Assessment of *Pseudomonas aeruginosa* epidemiology and the wider microbial diversity within the bronchiectatic lung

Submitted by Philip Mitchelmore
to the University of Exeter Medical School
as a thesis for the degree of
Doctor of Philosophy
In Medical Studies
March 2018

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: ……..

Abstract

The bronchiectatic lung is a diseased state in which the airways are chronically damaged and dilated. This state is found in the clinical entities of cystic fibrosis and non-cystic fibrosis bronchiectasis. These are two highly relevant chronic suppurative lung diseases in which an understanding of the microbiology of these patients is considered key to appropriate management. This has traditionally been via the use of traditional culture techniques. However, with the development of molecular methodologies, the previously perceived wisdom is being challenged.

In both cystic fibrosis and non-cystic fibrosis bronchiectasis, *Pseudomonas aeruginosa* is considered the most significant pathogen. In CF there has been considerable concern surrounding the risk of transmission of *Pseudomonas aeruginosa* between patients on the basis of a significant quantity of research into this matter. In contrast, there has been very little research performed into the equivalent risk in non-cystic fibrosis bronchiectasis. In this thesis we describe an extensive single-centre epidemiological review of *Pseudomonas aeruginosa* spanning both these diseases. Via this we have shown evidence of cross-infection within a non-cystic fibrosis bronchiectasis cohort. This epidemiological review has included multiple genotyping methods including multilocus sequence typing and whole genome sequencing, As an extension of the epidemiological review, we have performed an *in silico* prediction of hypermutator status from the whole genome sequencing data to provide greater understanding of the likelihood of cross-infection, and have also demonstrated a culture-independent adaption of multilocus sequence typing for potential screening for cross-infection.

In addition to *Pseudomonas aeruginosa*, we have also looked at the wider bacterial community in the lungs of patients with these two conditions via culture-independent techniques. We have shown that whilst *Pseudomonas aeruginosa* is often an important component, these are clearly complex communities. We have primarily investigated the cohort with non-cystic fibrosis bronchiectasis, but we have demonstrated associations between clinically-relevant markers and complexity of the bacterial communities within the lungs of both these cohorts of patients. Whilst we have used the gold-standard technique of 16S rRNA sequencing, we have also shown the validity of a simple and potentially more feasible profiling technique for standard clinical care.

In summary, through the application of culture-dependent and independent molecular techniques, this research has shed light on the epidemiology of *Pseudomonas aeruginosa* within our respiratory cohorts, and the complexity and clinical relevance of the wider microbial communities within these patients. Such studies are essential if we are to advance our understanding of the bronchiectatic lung and optimise strategies for patient management.

Acknowledgements

In order for this thesis to have been submitted, a large number of people have selflessly given their time and efforts towards the work contained within, and I would like to take this opportunity to thank them.

For a thesis containing significant clinical and scientific components, I have been blessed with four fantastic supervisors. When I first started talking to Dr Nick Withers and Dr Chris Sheldon several years ago about performing research in the fields of bronchiectasis and cystic fibrosis, it seemed like a pipe-dream, however their faith and persistence have allowed it to be realised. They have also been major influences on my development as a clinician and the development of my interests. Dr Chris Scotton of the University of Exeter Medical School was probably never expecting to have a supervisory role for a post-graduate student interested in the bronchiectatic lung, but his enthusiasm, support, guidance and input has been greatly appreciated. The biggest step into the unknown was entering the research laboratory in Biosciences at the University of Exeter where a large amount of time has been spent over my period of study. There Dr Alan Brown bravely took on the laboratory novice and patiently guided me from basic pipetting through to interpreting the output of sequencing data. I could go on, but my supervisors are fully aware of all the efforts they have put in and hopefully all know how grateful I am.

Clearly none of this would have been possible without all the samples obtained. For the logistics of setting up the collections I am indebted to Dr Nick Withers, Dr Chris Sheldon, Dr Alan Brown, Dr Chris Scotton, Dr Michael Gibbons, Dr Gill Baker, Chloe Slade, Lidia Romanczuk, Kathryn Shears, Dr Jo Randall, Nigel Richardson and Dr Cressida Auckland. For the actual collection of the samples I had great assistance from Hilary Mortimer, Sarah Murray, Rachel Rogers, Liz Lane, Jayne Trott, Roseanne Lowless, Miriam Green, Kat Thomson and Kerrie Hansell. The Microbiology team at the RD&E then kindly sorted and stored samples before transfer to the University. Dr Jo Randall also repeatedly heat-treated a very large amount of sputum to allow me to transfer. Finally with regards to samples, I am very grateful to the patients who have taken the time to listen to what I have been up to, and then subsequently consented to involvement in these studies.

The subsequent work on these samples has again been crucially aided by the help of many including Prof Esh Mahenthiralingam and Dr Matt Bull (Cardiff University); Dr Konrad Paszkiewicz, Dr Karen Moore, Paul O'Neill and Audrey Farbos (Exeter Sequencing Service); Dr Matt Robinson and members of the 4th Floor in Geoffrey Pope (University of Exeter). I have also been extremely lucky with generous funding via private charitable funds, the Dennis and Mireille Gillings Foundation, and Small Grants awards from the Royal Devon & Exeter hospital. I am very grateful to the contributions of Dr Nick Withers, Dr Chris Sheldon, Prof Andrew Hattersley, Prof Angela Shore and the local R&D department for their parts in enabling this.

Finally, I have been lucky to have had the patience, love and support of my wife and family throughout, particularly with the backdrop of the adorable chaos that Charlie and Olivia have brought into our lives through this period. Thank you.

Table of Contents

List of Contents **Page Number**

PART A

PART B

List of Tables

List of Figures

Abbreviations

ABPA- Allergic Bronchopulmonary Aspergillosis

BCC- *Burkholderia cepacia* complex

BP- Base Pair

BSI- Bronchiectasis Severity Index

BTS- British Thoracic Society

CF- Cystic Fibrosis

CFTR- Cystic Fibrosis Transmembrane Conductance Regulator

CI- Confidence Interval

CLED- Cystine Lactose Electrolyte Deficient

COPD- Chronic Obstructive Pulmonary Disease

CT- Computerised Tomography

DGGE- Denaturing Gradient Gel Electrophoresis

DNA- Deoxyribonucleic Acid

ESS- Exeter Sequencing Service

FEV_1- Forced Expiratory Volume in One Second

GO- Oxidized Guanine

HI- *Haemophilus influenzae*

HIV- Human Immunodeficiency Virus

HM- Hypermutator

HRCT- High-resolution Computerised Tomography

IV- Intra-venous

LES- Liverpool Epidemic Strain

MALDI-TOF MS- Matrix-assisted Laser Desorption/Ionisation- Time of Flight Mass Spectrometry

Mbp(s) - Mega Base Pair(s)

MC- *Moraxella catarrhalis*

MES- Manchester Epidemic Strain

MHA- Mueller-Hinton Agar

MHA/Rif- MHA with 300 µg rifampicin per ml

MHB- Mueller-Hinton Broth

MLST- Multilocus Sequence Typing

MRC- Medical Research Council

MRD- Maximum Recover Diluent

MRS- Mismatch Repair System

MRSA- Methicillin Resistant *Staphylococcus aureus*

NCFB- Non-Cystic Fibrosis Bronchiectasis

NHS- National Health Service

NIHR- National Institute of Health Research

NM- Normomutator

NTM- Nontuberculous Mycobacteria

NTM-PD- Nontuberculous Mycobacterial Pulmonary Disease

OTU(s) - Operational Taxonomic Unit(s)

PA- *Pseudomonas aeruginosa*

PCR- Polymerase Chain Reaction

PFGE- Pulsed Field Gel Electrophoresis

PM- Philip Mitchelmore

RAPD- Random Amplified Polymorphic DNA

RCT(s)- Randomised Controlled Trial(s)

RD&E- Royal Devon & Exeter Hospital

RPM- Revolutions Per Minute

RISA- Ribosomal Intergenic Spacer analysis

SA- *Staphylococcus aureus*

SMalt- *Stenotrophomonas maltophilia*

SM- Strong Hypermutators

SNP(s) - Single Nucleotide Polymorphism(s)

SPn- *Streptococcus pneumoniae*

ST- Sequence Type

T-RFLP- Terminal Restriction Fragment Length Polymorphism

TB- *Mycobacterium tuberculosis*

UK- United Kingdom

USA- United States of America

WGS- Whole Genome Sequencing

WM- Weak Hypermutators

Publications and Presentations

- **Original Research:**

Mitchelmore PJ, Randall J, Bull MJ, Moore KA, O'Neill PA, Paszkiewicz K, Mahenthiralingam E, Scotton C, Sheldon C, Withers N, Brown A. Molecular epidemiology of Pseudomonas aeruginosa in an unsegregated bronchiectasis cohort sharing hospital facilities with a cystic fibrosis cohort. Thorax. pii: thoraxjnl-2016-209889. doi: 10.1136/thoraxjnl-2016-209889. [Epub ahead of print]

- **Review Article**

Mitchelmore P, Wilson C, Hettle D. Risk of bacterial transmission in bronchiectasis outpatient clinics. Current Pulmonology Reports (in press)

- **Conference Abstracts**

Mitchelmore P, Brown A, Sheldon C, Scotton C, Bull M, Mahenthiralingam E, Withers N. S113 An Epidemiological Review of Strains of Pseudomonas aeruginosa in a Non-Cystic Fibrosis Bronchiectasis Cohort. Thorax Dec 2015, 70 (Suppl 3) A65; DOI: 10.1136/thoraxjnl-2015-207770.119

Mitchelmore P, Bull M, Sheldon C, Mahenthiralingam E, Withers N, Brown A. Bacterial community fingerprinting in a Bronchiectasis cohort. 2nd World Bronchiectasis Conference

- **Presentations**

An Epidemiological Review of Strains of Pseudomonas aeruginosa in a Non-Cystic Fibrosis Bronchiectasis Cohort. British Thoracic Society Winter Meeting 2015

Epidemiological review of Pseudomonas aeruginosa in Royal Devon & Exeter Hospital's CF and non-CF cohorts. 5th UK Cystic Fibrosis Microbiology Consortium Meeting 2016

Characterisation and management of the lung microbiome in Bronchiectasis. Exeter Expert Series- Respiratory. 2017

“The rather impressive status of *P. aeruginosa* as a 'harbinger of death' for CF patients has been greatly reduced by medical and scientific advances, and there is every reason to be optimistic that further research will show a way for it to be relegated to a lower and more manageable league of microorganisms in CF.”[1]

T. Pitt, 1986

Part A

Preface

This thesis has been written for submission to the University of Exeter Medical School for the degree of Doctor of Philosophy in Medical Studies. This is the product of a three year period working in the Department of Biosciences at the University of Exeter and the Department of Respiratory Medicine at the Royal Devon and Exeter NHS Foundation Trust. Additional support has been provided by the Department of Microbiology at the Trust, and the University of Exeter Medical School.

The main focus of this work is the bacteria *Pseudomonas aeruginosa* in the setting of the bronchiectatic lung. The majority of the patients studied have the diagnoses of non-cystic fibrosis bronchiectasis, but the work extends to those with cystic fibrosis as well. I have initially investigated the epidemiology of *Pseudomonas aeruginosa* in our patients, before looking at its place within the wider microbial community by culture-independent methods. This work has been performed through both the use of gold-standard molecular techniques, and also alternative options with their utility assessed.

This thesis is divided into three main sections. *Part A* begins with an introductory chapter reviewing our relevant knowledge of the bronchiectatic lung, *Pseudomonas aeruginosa* and the wider bacterial community in this setting. The techniques used to address these issues are also discussed, as are the aims of this body of work. This section is completed by a “Methods and Materials” chapter. *Part B* contains the investigative work where the research performed is presented in the style of extended journal papers. *Part C* presents the conclusions of this work and additional material.

It is hoped that the following text will both interest the reader, and provoke further thought around the subjects covered.

Chapter 1- Introduction

1.1 The Bronchiectatic Lung

The human lung is a highly complex organ which facilitates the transfer of oxygen into the body, and the expulsion of carbon dioxide from it. This is achieved through the ongoing ventilation of approximately 10,000 litres of air per day along the respiratory tract.[2] The respiratory tract is divided into upper and lower components, with the lungs contained in the lower component. The path of inhaled air through the lower respiratory tract starts within the main trachea before separating into the main bronchi, subsequently dividing into other bronchi and bronchioles, and then finally reaching the alveoli where gas exchange occurs (see Figure 1.1). Gases pass in the opposite direction in the expiratory phase of respiration. This ventilatory process helps the human body to maintain physiological homeostasis. However, the transfer of gases is not the only movement along these airways which is crucial to the maintenance of lung health. Along the respiratory tract there is also the potential travel of particles, including microbes. Immigration of these microbes can occur via simple inhalation, micro-aspiration and dispersion along the respiratory tract.[3] For the maintenance of health the lung also has various mechanisms of elimination, which help prevent acute infection and chronic colonisation with pathogens. These include coughing, mucociliary clearance and the host immune system. In situations where the balance between immigration and elimination is lost, lung health may not be maintained. An example of this is the bronchiectatic lung.

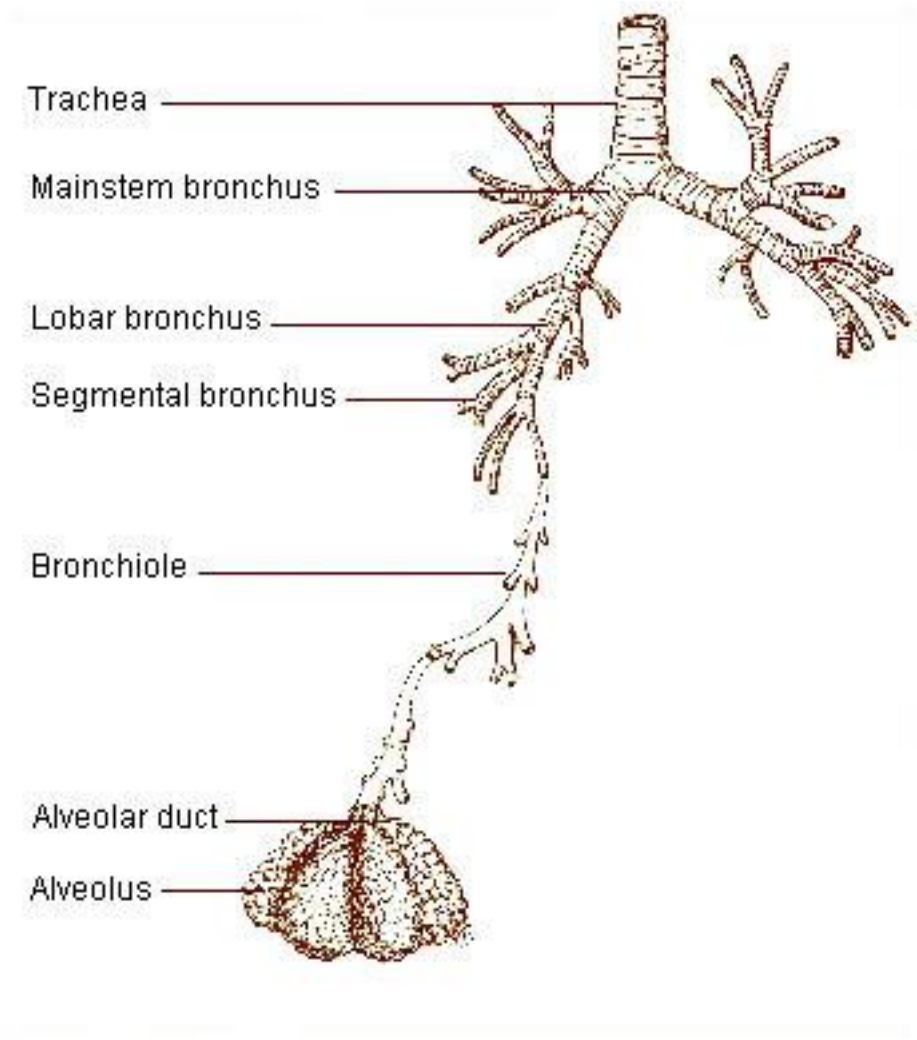

Figure 1.1 Anatomy of the lower respiratory tract. Image taken from NIH-National Cancer Institute

Rene Laënnec is generally credited with first describing bronchiectasis in the 19th Century.[4, 5] Bronchiectasis is a description of a pathological state where there is irreversible dilatation of part of the bronchi. In the 21st Century this is now defined by radiological assessment via high-resolution computerised tomography (HRCT) of the thorax.[6] Dilatation is defined by the internal diameter of a bronchus being larger than the accompanying vessel.[7] An alternative way of expressing this is an airway-to-vessel ratio of greater than one. Despite this apparently straight-forward criteria and the sophistication of HRCT, this remains a flawed process. Assessment of the images produced can be affected through the optical illusion of comparing hollow circles (as bronchi appear) and solid circles (as vessels appear), with the consequence being the over-diagnosis of bronchiectasis in some cases.[8] More worryingly for the integrity of the diagnosis is that the airway-to-vessel ratio may change with age. Potentially over 40% of

healthy subjects over the age of 65 meet the radiological criteria for bronchiectasis.[9] A further concern for the robustness of the radiological diagnosis is illustrated by work suggesting that an abnormal airway-to-vessel ratio can be driven by small vessels rather than enlarged airways.[10]

Rather than relying entirely on radiological findings, it is clearly important for a patient to have a relevant clinical presentation for a useful and relevant diagnosis to be made. The presentation of those considered to have clinical bronchiectasis is characterised by features such as a chronic cough, increased sputum production, haemoptysis, dyspnoea, fatigue, chest pain, and recurrent exacerbations of respiratory symptoms which are often suspected to be driven by infection. Clinicians will identify patients with these features who do not meet the radiological criteria for bronchiectasis.[11] This group of patients consequently do not have a diagnosis of bronchiectasis. They may as a result be poorly-served and miss out on specific trials and therapies. Young patients are at particular risk of this due to the radiological reasons described above. However, despite the issues raised above, there is clearly a significant group of patients who have the characteristic clinical syndrome alongside the radiological findings. This is the group of patients studied in this thesis.

When a patient is suspected and subsequently confirmed to have bronchiectasis, a key step in the management of the patient is the screening of potential aetiologies.[12] One cause which may be screened for is the life-limiting genetic condition of cystic fibrosis. This brings about the main split in the diagnostic labelling of patients with bronchiectasis: those described as having cystic fibrosis (which will be referred to as CF for the remainder of this thesis); and those with non-cystic fibrosis bronchiectasis (which will be referred to as NCFB hereafter). This distinction is a very important one to make, particularly for those with CF where the care and support structures are better established. Historically CF has received more research attention and as a result there have been many examples of treatments being used in NCFB on the basis of success in CF. Notably however, this has not always been successful, and indeed has on occasion resulted in harm.[13] CF research and drug development is now reaching the point where genomic-based therapeutics may transform the health-potential of many patients.[14] This may lead to further divergence in the management of the two categories of patients with bronchiectasis.

1.1.1 Non-Cystic Fibrosis Bronchiectasis

The majority of the work in this theses has involved patients with NCFB. Having previously been seen as an orphan disease of decreasing relevance, NCFB is now receiving significantly more attention.[15] I shall now describe this disease in more detail including its importance, its impact on patients, and the microbiological issues associated with it.

Epidemiology

The epidemiology of NCFB is surprisingly difficult to express clearly. This is for multiple reasons. There have been a variety of quoted rates of NCFB around the world. A frequently quoted prevalence had been 52/100,000 based on data from the United States of America (USA).[16] This dataset suggested prevalence ranging from 4.2/100,000 in 18-34 year olds to 272/100,000 in those over 75. It has subsequently been suggested that these figures are underestimates.[17] Australasian paediatric data revealed a prevalence rate of 1,470/100,000 in a Central Australian Indigenous population.[18] This was described as unacceptably common in this cohort. German data published as recently as 2015 reported a prevalence rate of 67/100,000 overall, and 228/100,000 when investigating men aged 75-84 years old.[19] A very large dataset from the United Kingdom (UK) was published in 2016 by Quint et al and reported a point prevalence of 566/100,000 in adult women and 486/100,000 in adult men.[20] In those over 80 years old, the prevalence rates for the population were 1,206/100,000. These UK figures are truly striking and significantly change our impression of how common NCFB is. Also of note from this report is the increasing prevalence rates between 2004 and 2013. What is clear from these various cohorts is that NCFB is a disease that is more prevalent in women, the middle-aged and the elderly.[21, 22] Other details are more subtle and require more explanation.

Studies where low rates had been described, may have been underestimates due in part to under-recognition of the condition. One study has suggested a striking 17 year delay in diagnosis in a cohort.[23] Due to the symptoms of NCFB being non-specific and shared by other respiratory diseases, if radiological investigations by HRCT is not performed, another diagnostic label may be applied. An example would be chronic obstructive pulmonary disease (COPD). A

patient presenting with chronic cough, recurrent exacerbations of respiratory symptoms, and abnormal lung function tests could easily be diagnosed with COPD, particularly if they had a history of smoking. If however, an HRCT was performed, the diagnosis may change to NCFB. A study which illustrates this was a prospective review of patients who had been seen by in primary care and treated in the community for an exacerbation of COPD.[24] The patients subsequently underwent an HRCT when they were stable, and almost a third showed evidence of bronchiectasis. A secondary care study found 69% of patients admitted for their first exacerbation of COPD had NCFB.[25] It is becoming clear that these diagnoses are not mutually exclusive and interest is developing in the concept of a COPD-NCFB overlap syndrome.[26, 27] As well as COPD, asthma may be another respiratory disease which may distract clinicians from a diagnosis of NCFB, particularly with the non-allergic/neutrophilic phenotype.[28] Another possible reason for lower reported rates may be intricate politico-economic reasons. It has been verbally suggested in international forums that the rates in Germany may have been falsely low due to differences in financial charges for patients when the diagnosis is COPD rather than NCFB.

Of the studies with the higher rates of NCFB the most striking is in the Australian indigenous population, particularly when it is taken into account that this is a paediatric cohort. This is used as an example of how poor access to healthcare and high rates of childhood infection increases the risk of developing NCFB. Of note, data from neighbouring New Zealand showed that the socioeconomically disadvantaged Maori and Pacific population had the highest rates of hospital admissions with NCFB.[29] That there is such a burden in high-income countries, suggests that there could be even greater problem in less developed nations. The other study with particularly high rates is the UK based study by Quint et al, and this is the most relevant in relation to this piece of work.[20] The diagnosis of NCFB was based on it being entered on a primary care system. As NCFB is a diagnosis confirmed in secondary care in the UK, even this study could still be an under-estimation.

On the face of it, the Quint study and others suggest that the prevalence of NCFB is rising.[20, 21] This is not necessarily true. Trying to interpret the changing prevalence of NCFB is complicated by the increasing use of HRCTs demonstrating less severe NCFB, and also revealing NCFB when a different

diagnosis was suspected.[30] The increase in imaging was illustrated by an estimation that the numbers of computerised tomography (CT) scans in 2014 in the USA was approximately 27 times greater than in 1980.[15] The increasing awareness of NCFB alongside the increasing use of imaging modalities may be increasing the diagnosis of NCFB, rather than the true prevalence rising. However, it would seem likely that even if we do not have a true rise in prevalence yet, that it is likely in the future. It would be expected that an ageing population would have a higher incidence of NCFB. In addition increasing antibiotic resistance, and the threat of multi-drug resistant pathogens such as *Mycobacterium tuberculosis* (TB) may lead to an increase in NCFB following significant infection.[31, 32] Consequently, NCFB is likely to remain an important and relevant condition to understand in the future.

Aetiology

Bronchiectasis is a pathological state which is reached in NCFB as a result of wide-ranging aetiologies. In the UK the national guidelines set out recommendations for investigations to be performed to establish the aetiology of bronchiectasis.[6] This is important because there are certain aetiologies where identification can influence treatment and prognosis (such as immune deficiencies, allergic bronchopulmonary aspergillosis (ABPA), nontuberculous mycobacterial pulmonary disease (NTM-PD) and CF). There have been several studies over the last two decades describing aetiology.[12, 20, 23, 33-35] When assessing these data, there are a couple of points to make. First of all, despite being conducted by researchers with a specialist interest, there are a large number of patients with a diagnosis of “idiopathic” or the vague “post-infective” label. Secondly, there is almost certainly issues with consistency between studies with regards to what are considered causative factors, and what are just associations.[15] When looking at specific studies, the high representation of COPD and Asthma in the study by Quint et al is likely to be due partly to the recording of disease in primary care, and is in part a reflection of association rather than aetiology.[20] The study by Lonni et al spans seven European centres, has large numbers and had aetiology assessed in a rigorous way.[33] Consequently, it is probably the most relevant study of aetiology. Despite the focused approach, “idiopathic” was the leading aetiology in this study. Interestingly COPD was associated with more severe disease. Idiopathic disease

was associated with lower severity as was “inflammatory-bowel disease” and “aspiration/oesophageal reflux”.

Like all idiopathic diagnoses in medicine, there is almost certainly a cause which has not yet been revealed. It is possible that some of the cases are due to genetic predispositions.[36] An interesting outlier with regards to aetiology is the paper by McShane et al.[34] In this single centre study in the USA, an aetiological cause was identified in over 93% of patients. “Immune dysregulation” was the predominant cause with a rate of 63%. This umbrella term included ABPA, autoimmune diseases, haematological malignancy, and immune deficiency. It may be that better understanding of genetics and immunology will lead to future reductions in the numbers labelled with the idiopathic aetiology.

Disease Burden

The impact of NCFB on patient health is variable. Work has looked into what patients find difficult with regards to their disease. The six issues which were found difficult by over 60% of patients were “Sputum”, “Exacerbations”, “Tiredness”, “Not feeling fit for daily activity”, “Shortness of breath”, and “Cough”.[37] Many patients will experience a baseline level of symptomatology punctured by episodes which are referred to as “exacerbations”. Exacerbations will be discussed at various points through this thesis so it is worth briefly introducing the concept now. Historically there has been no universal definition of what an exacerbation is, and indeed some studies include multiple criteria.[38] Essentially what is meant by the term is an acute worsening of respiratory symptoms relating to their chronic condition. Very recently a consensus statement has been published on what an exacerbation is, but it will take a long time before a significant amount of clinical research has used this.[39] The aetiology of these events is often unknown, though a bacterial driver is often blamed, and patients often receive antibiotics for the management of their exacerbations. Exacerbations can crudely be described as severe if hospital admission and/or intra-venous (IV) antibiotics are administered. However, sometimes symptoms may not be severe but a lack of oral options mean IV therapy is required. Despite its flaws, the concerns of exacerbations are of great importance to both patients and experts.[37] This is exemplified by data showing that a rate of more than three exacerbations per year is significantly associated with mortality.[40]

Some appreciation of exacerbation rates can be gleaned from recent European multi-centre data. In a cohort of 1,310 patients, an exacerbation frequency of 1.8-3 per year and a hospital admission rate over a two year period of 26.6-31.4% was described.[40] It should be noted that exclusion criteria for this dataset included patients on long-term antibiotic therapy, and patients with active malignancy, NTM-PD or Human Immunodeficiency Virus (HIV). It would be expected that these patients would have a higher exacerbation rate, and consequently the "real-world" exacerbation rates are likely to be higher.

As well as a symptom burden, there is also a significant finance burden to NCFB. Recent estimates of the mean healthcare costs per year in Spain were over 4,500Euros per patient per year.[41] Australian data suggested a higher healthcare burden for individual NCFB patients than those with heart failure or diabetes.[42]

In addition to significant morbidity, NCFB has also been associated with a reduced life-expectancy compared to the general population. A UK study following a cohort for 13 years reported a mortality rate of almost 30%.[43] In this study the survival rate at 4 years was 91% which approximately fits with findings from Belgium where survival rates were 89.4% over 41 months.[44] Despite this group of patients tending to be elderly with other co-morbidities, respiratory failure tends to be a prominent cause of death. Various cohorts have suggested that between 50-70% of deaths of patients with NCFB are due to respiratory complications.[43, 45, 46] NCFB has a clear burden on patient's health and requires proper directed therapies and research.

Management

The principles of management of NCFB are relatively straightforward and Peter Cole's "Vicious Cycle" hypothesis is a good starting point (see Figure 1.2).[47] Strategies and treatments should be put in place to break the cycle. By achieving this, patients should have a reduction in symptoms and exacerbations, reduction in sputum volume and purulence, and improvement in quality of life. An example of the consequences of breaking this cycle is that the reduction in exacerbations is likely to have an impact on the rate of decline of lung function, and reduced lung function is associated with an increased risk of exacerbation.[40, 48]

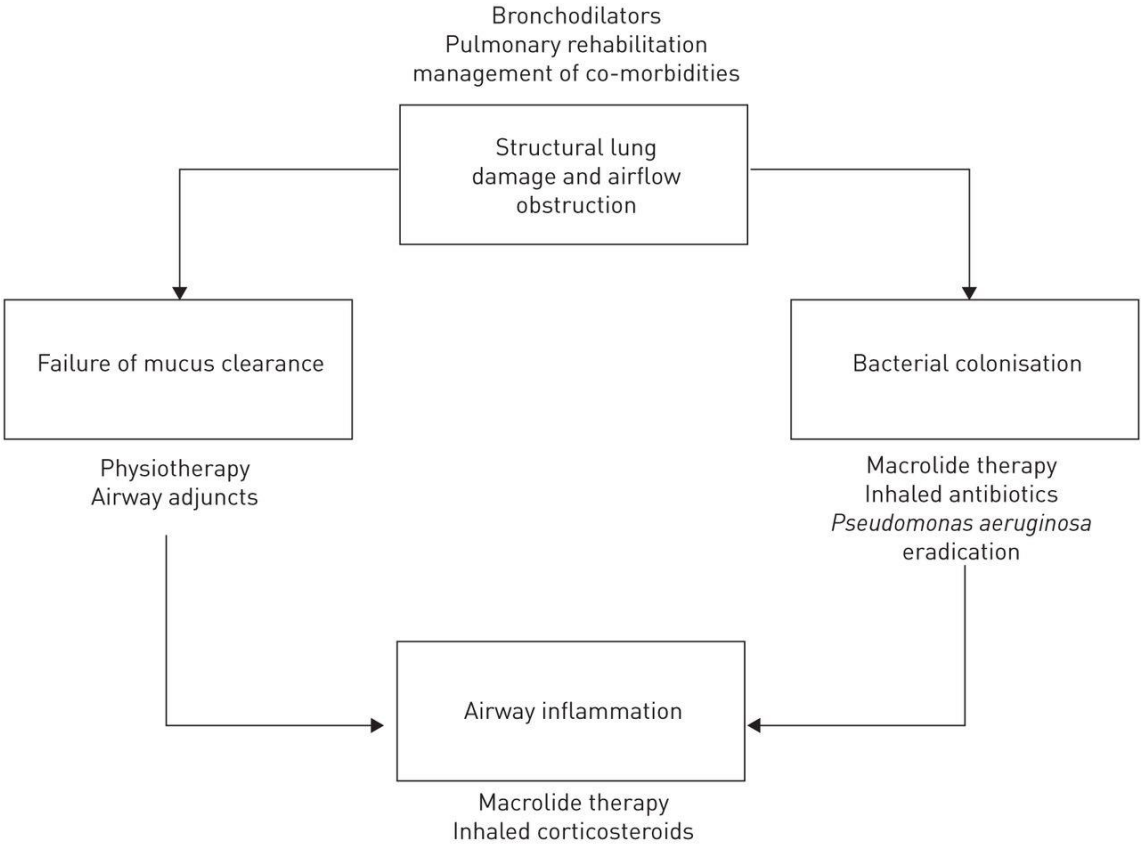

Figure 1.2 Cole's vicious cycle hypothesis Image taken from Chalmers et al. [49] Current therapies are listed alongside each component of the "cycle".

As would be expected of a chronic suppurative lung disease, the use of antibiotics are key to the management of NCFB. Their usage tends to fall into four main categories, namely acute oral courses, acute IV courses, prophylactic oral antibiotics, and prophylactic nebulised antibiotics. The results of previous microbiological investigation will commonly impact the choice of antimicrobial agent used. Acute courses for the treatment of exacerbations are often given over a longer period than in other lung conditions. The evidence for when to use IV antibiotics is limited.[50] The British Thoracic Society (BTS) guidelines suggest considering IV usage if a patient is unwell and requires hospital admission, or if resistant organisms are not responsive to oral therapies, or if the patient has not responded to oral therapy.[6]

Prophylactic antibiotics have been an area of increasing interest and research. The aim of this therapy is to reduce exacerbation rates and consequently reduce further damage to the lung's structure and function. With regards to oral therapies, the macrolides have been prominent with three randomised controlled trials (RCTs) in recent years.[38, 51, 52] All three RCTs demonstrated a reduction in exacerbation rates, though questions remain about long-term consequences and side effects as none of these studies lasted beyond a year. These therapies are thought to have anti-inflammatory and immunomodulatory effects in addition to antibacterial properties. Promotion of gastric emptying may also have an impact if reflux is indeed clinically significant to the lung pathology. However, of particular interest and concern, there appears to be an increasing potential for antibiotic resistance.[53]

Several antibiotics have been used in inhaled form such as gentamicin, tobramycin, ceftazidime and colistin. Whilst the concept of inhaled antibiotics is very appealing, tolerability has often been an issue.[17] Of late, ciprofloxacin has been trialled in the inhaled form and hopes remain with regards to its potential.[54, 55]

All patients should be engaged in the practice of chest physiotherapy.[6, 42] The aim is to compensate for impaired mucociliary clearance and consequently reduce the retention of respiratory secretions and enable the elimination from the lung of particles for the maintenance of health. Mechanical devices can be included to help the process.[56] A Cochrane review only contained 51 patients but did suggest that airway clearance techniques improved expectoration, reduced hyperinflation, and improved quality of life scores.[57] A recent small RCT has been published since and suggests benefits with regards to exacerbations and quality of life measures.[58] Whilst the evidence-base for these techniques is weak, they continue to be considered an important part of management. Additional therapies have been trialled as adjuncts such as mucoactive agents and inhaled saline solutions.[59-62] One highly significant study of a mucoactive agent trialled DNase.[13] Despite being an effective therapy in CF, a RCT suggested it may be harmful in NCFB. Another aspect of care linked to physiotherapy is the undertaking of a pulmonary rehabilitation course. This should be offered to patients who have breathlessness affecting their daily lives.[6]

Others therapies that are used in NCFB are bronchodilators, inhaled corticosteroids and anti-inflammatory agents.[63-70] The evidence for these therapies is variable and taken from small studies. Surgical interventions, such as lung resection in localised disease, are occasionally considered though they are not without their risks.[71, 72] Additional basic standards of care include pneumococcal and influenza vaccinations, and smoking cessation services being offered. Finally, with a significant proportion of NCFB cohorts being elderly, co-morbidities are to be expected and therefore should be sought and assessed.[73]

There is clearly scope for improving the armoury available to the managing physician. Unsurprisingly, antimicrobial therapies will continue to be at the forefront of interest. This will include attempts to utilise old therapies in a more effective way (such as inhaled ciprofloxacin), but also the investigation of novel agents.[74, 75] Another potential target could be the inflammatory process leading to the destructive changes found in NCFB. A single-centre study has looked at the use of statins in NCFB with the hope of utilising the anti-inflammatory effects of this class of drug.[70] Some improvements were noted and further studies are awaited. With neutrophilic inflammation so prominent in NCFB, this is also viewed as a potential therapeutic target. CXCR2 is a neutrophil trafficking receptor and it is hoped that antagonism of this may reduce, but not prevent, neutrophil recruitment.[76] The danger of immune system manipulation is uncontrolled infection, and consequently achieving the right balance will be crucial.

1.1.2 Cystic Fibrosis

The other group of patients studied in this work are those with CF. Historically this has been a far more studied disease. In our local hospital, they are a smaller cohort than those with NCFB and in this piece of work they are also in the minority. I shall now describe the condition in order to illustrate its differences from NCFB.

Epidemiology

Whilst the epidemiology of NCFB is muddled for the host of reasons discussed previously, it is a clearer picture in CF, particularly in settings with well-maintained registries.[77, 78] As would be expected with a genetic disease, different populations can have very different rates. The continents with the highest

prevalence are North America, Europe and Australasia, with the occurrence seeming to be in those of European descent. Within Europe, there are higher rates in the UK and Ireland than in Eastern European and Baltic nations.[79]

It is commonly quoted that approximately 1 in 25 Caucasians carry a disease causing mutation and that 1 in 2,500 live births have CF. Data for the Cystic Fibrosis Trust in the UK from 2014 reported 10,583 patients and in the USA approximately 28,000 have the condition.[80, 81] My work has taken place in Devon, a county in the South-West of England. Approximately 2.3% of the UK CF population are living in Devon, and the prevalence in the county is approximately 21 per 100,000. Despite having greater public awareness, CF is a far rarer disease than NCFB in the UK. Another noticeable difference is the lack of female predominance seen in NCFB.

As would be expected of a life-limiting genetic disease, the age profile of CF patients is very different to those with NCFB. New-born screening is now common-practice in the UK, and the median age of diagnosis of paediatric patients was 26 days in 2015.[82] When the UK population is looked at as a whole, the current median age of a CF patient is 19 years of age.

Aetiology

Unlike NCFB, the aetiology of CF is relatively well understood. CF is an autosomal recessive condition. The disease was 1st named “Cystic Fibrosis” in the 1930s by the American pathologist Dorothy Andersen who went on to note its autosomal recessive nature in the 1940s.[83, 84] However, it took until the 1980s for advances in genomics to identify the relevant gene on the long arm of chromosome 7.[85] This is now known as the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) gene. The gene codes for a protein called CFTR which enables the transport of chloride ions across the cell membrane, as well as other regulatory roles including sodium transport.[86] The dysfunction of CFTR may have numerous consequences which contribute to a pathological state including hydration status, mucociliary clearance, increased inflammation and impaired immunity.[87] CFTR is expressed in many organs including the lungs, the liver, the pancreas and the gastro-intestinal tract.

Since the discovery of the CFTR gene, approximately 2,000 mutations have been identified, however only a small minority are confidently considered to be disease-

causing.[88] CFTR mutations are grouped into 6 classes based on the resulting mechanistic abnormality. Classes I, II and III are associated with a more severe phenotype and no CFTR function. By far the most commonly identified abnormality is the class 2 mutation known as F508del. In the UK 90% have this mutation at least once, and 50% are homozygous for it.[82] The next most common mutation in the UK is known as G551D and is a Class III, or "gating" mutation. As will be explained later on, this genetic information is becoming increasingly important in the management of CF.

Disease Burden

The clinical impact of CF is variable but often devastating and the disease is considered the most severe autosomal recessive disease in Caucasians. Whilst it is a multi-organ disease the respiratory system endures the greatest burden in patients. Patients will tend to suffer from progressively declining lung function, with persistent symptoms of cough, sputum production, shortness of breath and chest pain. Periods of relative stability will be interrupted by episodic exacerbations.[86] Exacerbations are associated with worse quality of life, worsening of lung function, and mortality.[89-91] This may lead to multiple courses of oral and IV antibiotics for patients. Registry data suggest that the median number days of IV antibiotic usage in the UK for all CF patients is 28 days.[82] As well as the burden of symptoms and intermittent requirements for acute antibiotic courses, patients will often have a long and complex list of time-consuming daily therapies.

The UK registry data also gives a clear picture of the non-respiratory based morbidities in CF patients. In those over the age of 16 approximately a third have CF-related diabetes. Liver disease and gastro-oesophageal reflux are present in approximately a fifth of patients. Other notable complications include intestinal obstruction, osteopenia/osteoporosis, renal dysfunction, infertility and depression.

The natural disease process leads to a premature death either with or without organ transplantation. In the UK the current median predicted survival is 47 years.[82] Whilst this appears a shockingly young age, it is a remarkable improvement on where life expectancy used to be (see Figure 1.3). CF used to essentially be a paediatric disease, however in developed areas we are now

seeing more adults with the condition than children.[81, 92] The reasons for this improvement are multi-factorial but include pancreatic enzyme replacement, nutritional input, mucoactive therapies and a better understanding of treatment of infection.[92-94] The expectation should be for this trend to continue though there is still a long way to go for the life-expectancy to be comparable to those with NCFB.

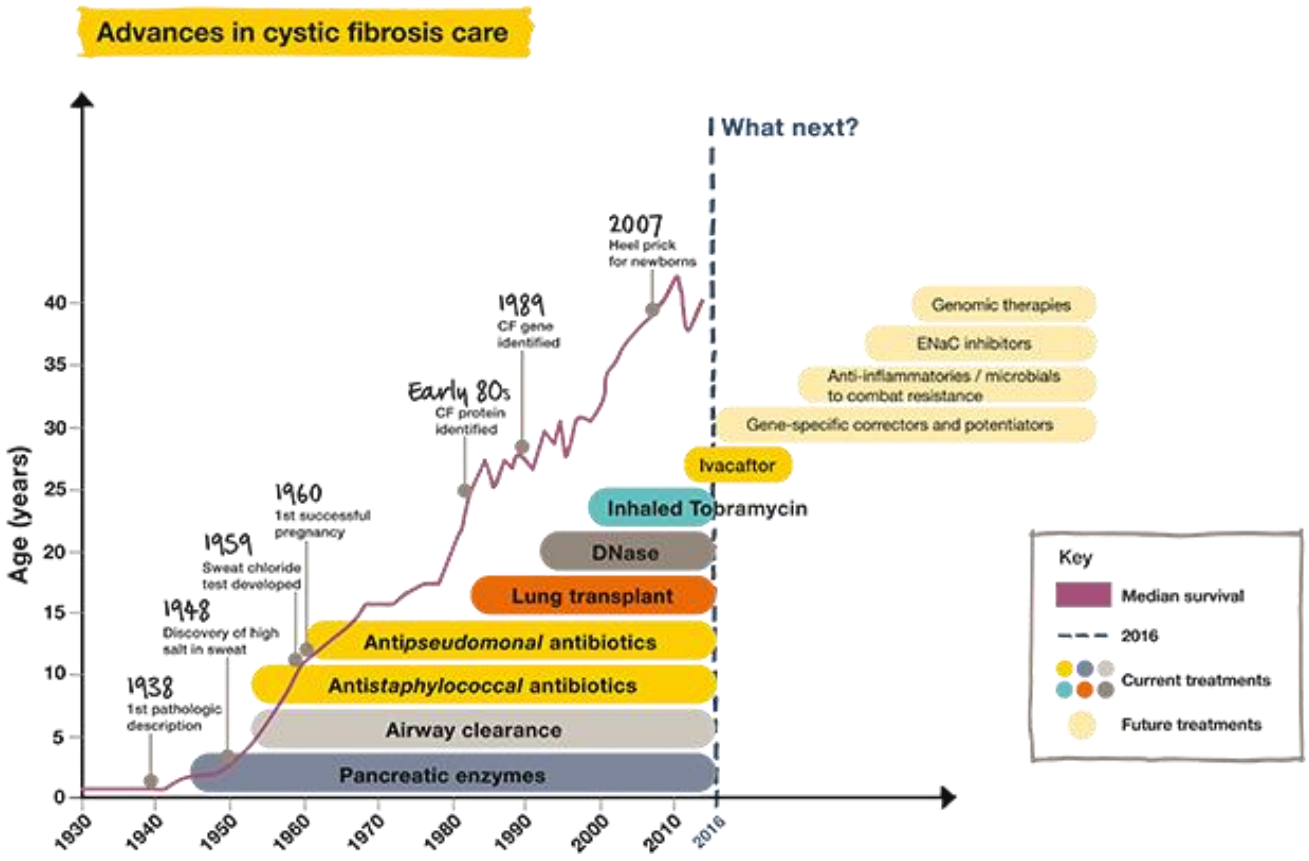

Figure 1.3 Life expectancy in cystic fibrosis Image taken from CF Trust Website. Relevant advances in management are included in the graph of improving life expectancy, as are potential therapies of the future.

Management

There is clearly some cross-over in the management of NCFB and CF, particularly with regards to the management of the pulmonary system. In both conditions there is an aim to preserve lung function through physiotherapy based

airway clearance techniques, and the management of acute exacerbations and chronic airway infection.

As in NCFB, even though there is a paucity of high-quality evidence, physiotherapy techniques to facilitate airway clearance are considered of great importance in prominent guidelines.[95] Unlike NCFB, there is good evidence for the effectiveness and safety of DNase as an adjunct therapy.[96] Further adjuncts with better evidence in CF are hypertonic saline and Mannitol.[97, 98] In the UK currently, over half of patients use DNase, approximately a quarter use Hypertonic Saline and less than 5% use Mannitol.[82]

Antimicrobial usage (both antibiotics and antifungals) is the cornerstone of CF management. Like NCFB, oral, IV and inhaled antibiotics are used in a variety of regimes for the management of exacerbations, chronic prophylaxis and attempted eradication of pathogens. Antibiotic administration is highly dependent on the pathogens isolated by microbiological techniques. The microbiology of these conditions will be discussed in more detail shortly.

The non-respiratory management of CF is dependent on the disease-related complications experienced by the patient. This may include insulin for CF-related diabetes, supplemental pancreatic enzymes and lipid soluble vitamins, fertility techniques and psychological support.

Perhaps the most exciting development has been the interest in therapies which actual address the CFTR defect. There has long been interest in the actual delivery of a cloned CFTR gene to the lung via nebulisation. Whilst hugely appealing, optimum delivery and efficacy are yet to be achieved.[99] The most striking example so far of addressing the CFTR defect is Ivacaftor, an orally administered small molecular therapy. Pre-clinical studies demonstrated that Ivacaftor corrected Chloride transport in most class III mutations.[100, 101] Subsequent clinical trials have shown impressive effects, particularly in patients with the G551D mutation.[102-104]. For the greatest impact on the CF community as a whole, therapy which improves CFTR function in those with F508del would be the most desirable. Ivacaftor has been used in this group in combination with another small molecule therapy (lumacaftor), and whilst benefit has been shown, it has not been as impressive.[105, 106] The future of this line of therapy is very exciting and has come about through deep understanding of the underlying

pathology in CF. Unsurprisingly, in NCFB where there is far greater heterogeneity of pathogenesis, there are no equivalents to this route of therapy.

1.1.3 The Microbiology of the Bronchiectatic Lung

In the diseases where bronchiectasis is present, and patients suffer recurrent symptoms which are believed to be driven by pathogens, an understanding of the microbiology would seem to be essential. Consequently sampling of respiratory secretions has become common practice in the management of both NCFB and CF, and the results of these investigations often dictate subsequent antimicrobial management.

Culture-based techniques have been used since the 1880s to investigate the presence of potential pathogens. In these techniques, samples are inoculated onto selective media to enable the growth and subsequent identification of bacteria and fungi. In the last few years culture-independent techniques have been employed in research facilities to investigate the microbiology of the bronchiectatic lung, but this has not crossed over to standard clinical practice yet. I shall now describe the current knowledge of bacterial pathogens in these conditions as revealed in the everyday practice of culture. Insight gleaned from culture-independent techniques will be addressed later.

Non-Cystic Fibrosis Bronchiectasis

There are now many cohorts whose microbiological status has been described from a variety of different settings and countries (see Table 1.1). Some of this data comes from cohorts enrolled for drug trials, others from severity score derivation studies, and some from unselected observational cohorts.[38, 40, 107]. Despite the potential inherent biases to some of these studies, when looking at the question of prevalence, it is undeniably clear that *Pseudomonas aeruginosa* (PA) and *Haemophilus influenzae* (HI) are the predominant pathogens. Many other pathogens are cultivated from sampling of NCFB patients and include *Moraxella catarrhalis* (MC), *Streptococcus pneumoniae* (SPn), *Staphylococcus aureus* (SA), *Aspergillus* species and Non-tuberculous mycobacteria (NTM).

Reference	Year	Origin	N	PA	HI	SA	SP	MC	Myco	NP
[40][a]	2014	UK	608	12	29	8	6	10	ND	28
[44][a]	2012	BEL	479	30	31	23	20	15	NR	NR
[107][b]	2014	UK	155	49	57	23	33	25	3	6
[38][c]	2012	NZ	141	12	28	3	3	4	NR	NR
[108][d]	1995	USA	123	31	30	7	11	2	23	23
[109][e]	2002	THA	50	20	14	NR	6	4	6	36

Table 1.1 Pathogens of non-cystic fibrosis bronchiectasis

The identification rates of each pathogen are expressed as a percentage. The abbreviations for the pathogens are as follows: PA- *Pseudomonas aeruginosa*, HI- *Haemophilus influenzae*, SA- *Staphylococcus aureus*, SP- *Streptococcus pneumoniae*, MC- *Moraxella catarrhalis*, Myco- Mycobacterium. NP- no pathogen. ND- not done and NR- Not reported. The studies are from a variety of countries. UK- United Kingdom, BEL- Belgium, NZ- New Zealand, USA- United States of America and THA-Thailand. The data has been recorded in different ways. [a]Colonisation status. [b]Pathogen isolated during median follow-up of 46 months. [c]Clinical trial cohort. Cross-sectional baseline data. [d]Pathogens isolated during retrospective review. [e]Cross-sectional first visit data

Even though HI is frequently reported as the more prevalent pathogen in NCFB, PA has received more attention. As shown in Table 1.1, the rates of pathogen prevalence vary, and in part depend on whether they are referring to isolation in a cross-sectional study, or colonisation. As an aside, the definition of colonisation is not consistent, as demonstrated by a meta-analysis identifying 8 different definitions between 21 studies.[110] A recent study in the North-East of England reported that approximately half of their patients had at least one sample positive for PA in a cohort who had a median follow-up of 46 months.[107] However of these, only 62% met the defined criteria for colonisation. Whilst it has long been acknowledged that there is an association between PA isolation and extensive disease in NCFB, it has been debated whether the isolation of PA is a marker or

a cause of disease severity. Some studies have suggested it is either not clear, or that PA is a marker of disease severity.[111, 112] The cohort from the North-East of England did show isolation of PA from patients with minimal airflow obstruction, though at a much lower rate than in those with more severe limitation.[107] Various studies have shown PA's association with morbidity. These have included worse quality of life, worse radiological appearance, worse lung function, increased inflammation in the airways, more frequent exacerbations and a greater rate of decline in lung function.[44, 48, 111, 113-115]. A mortality review of a historical cohort suggested that PA had an independent effect on mortality and was therefore likely to not just be a marker of disease severity.[43] A recent meta-analysis showed a very strong association between PA colonisation and mortality with an odds ratio of almost 3.[110] There are no studies showing such worrying trends with HI. There is however a suggestion that colonisation with recognised pathogens other than PA, is associated with increased disease severity.[40]

It is this author's belief that PA colonisation is a marker of disease severity, but also a contributor to it. Its contribution to disease severity scores clearly illustrate it status as a marker of disease. PA has been extensively studied in the *in vitro* setting with many virulence factors, toxin production pathways and mechanisms for pathogenicity discovered.[116-119] We also see clinical responses in patients who reliably use antipseudomonal therapies.[120] In light of these findings, it seems very unlikely that PA is just a marker of disease severity, and not also an active aggressor against the lung.

Cystic Fibrosis

A good insight into the microbiology of CF can be derived from registry data.[81, 82] In CF the prevalence of various pathogens varies with age. In the initial years HI and SA tend to be the prominent pathogens (see Figure 1.4). In the UK the rates of SA are possibly lower than other countries such as the USA (see Figure 1.5). This difference may well be due to the aggressive anti-staphylococcal regimes used in UK paediatric patients.[95] What is consistent across the UK and USA data is the rising prevalence of PA with age. In adulthood, PA becomes the most prevalent pathogen as assessed by culture techniques.

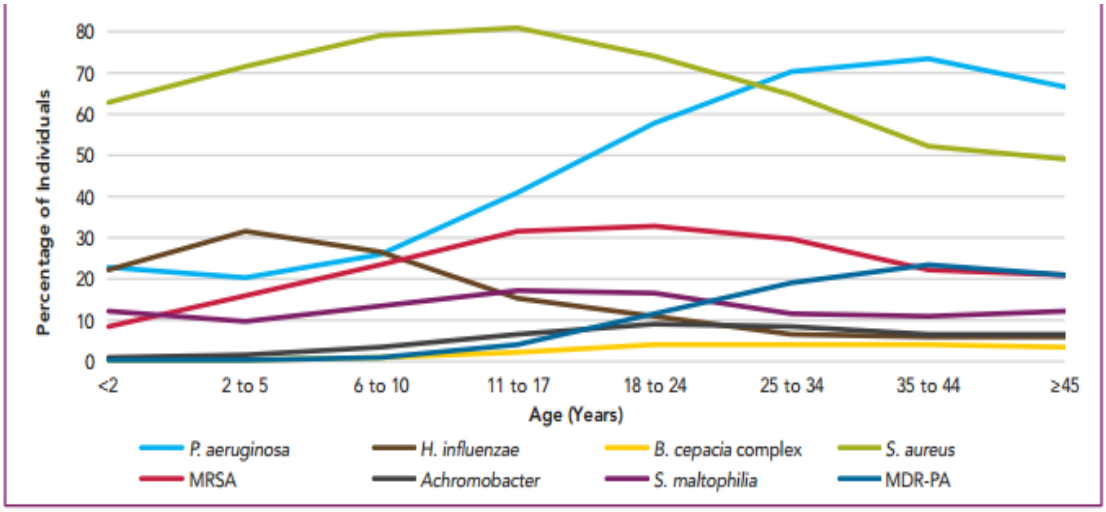

Figure 1.4 CF pathogens in the USA by age. Image taken from the CF Foundation Patient Registry Annual Data Report 2015. B. cepacia complex- *Burkholderia cepacia* complex, S. aureus- *Staphylococcus aureus*, MRSA- Methecillin-Resistant *Staphylococcus aureus*, S. maltophilia- *Stenotrophomonas maltophilia*, MDR-PA- Multi drug resistant *Pseudomonas aeruginosa*

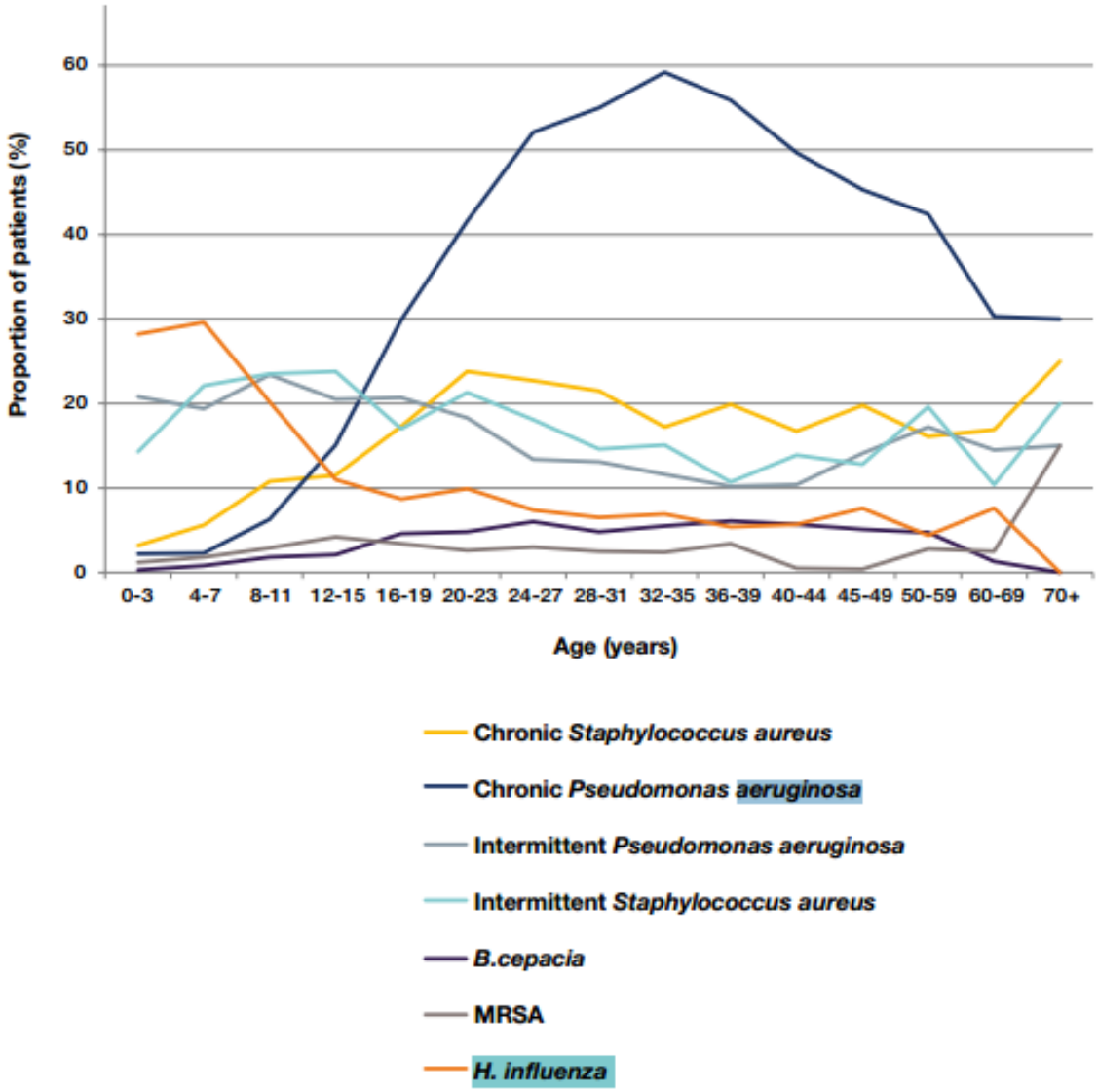

Figure 1.5 CF pathogens in the UK by age. Image taken from the CF Trust's UK CF Registry 2015 Annual Report.

As in NCFB, PA is not just prevalent but also of high clinical importance. Various studies have illustrated its impact on lung function, morbidity and survival.[121-128]. Whilst its prevalence changes with age, its impact affects all ages. Registry data showed that the presence of PA was the greatest predictor of mortality in those aged 1-5 years old.[125] Other data showed that the majority of those under the age of 7 who were chronically infected had a marked reduction in lung function.[122] Another study showed 72.5% of patients had isolated PA within their first 3 years of life.[129] A commonly quoted figure for prevalence of PA in

adult CF patients is 80%.[130, 131] In the UK this figure is not currently accurate. The latest annual report by the CF Trust suggested that just under half of all patients over the age of 16 are chronically colonised, whilst a further 15% intermittently isolate PA.[82]

Of possible encouragement, the rates of chronic infection with PA in the UK appear to be falling.[82] There are however plenty of other pathogens for the CF community to be aware of. There are lower rates in the UK than the USA of the NTMs and Methicillin-Resistant *Staphylococcus aureus* (MRSA), but they still provide a degree of disease burden. The *Burkholderia cepacia* complex (BCC) is rarely of clinical significance outside the CF community and is not common within it. When present however, it can be associated with poor outcomes, particularly in the case of the so called Cepacia syndrome, when a rapidly progressing pneumonitis is accompanied by bacteraemia. Other pathogens which are receiving increasing interest are the “emerging” non-fermenting gram-negative bacilli, which include *Stenotrophomonas maltophilia*, *Achromobacter xylosoxidans*, *Pandorea* species and *Ralstonia* species.[132, 133]

Potential pathogens in CF are not limited to just bacteria. Viruses are noted as a driver of some exacerbations.[134] An example of a potential mechanisms is rhinovirus disrupting PA biofilms resulting in freed motile bacteria.[135] Fungi such as *Aspergillus* species have a clear association with a variety of clinical syndromes.[136-138] For others such as *Exophiala*, the significance is less clear.[139]

The prevalence of pathogens will likely evolve as life expectancy changes and new therapies develop. Our description and understanding will advance through culture-independent sequencing techniques.

1.2 Pseudomonas aeruginosa

As has been illustrated above, PA is both highly prominent and clinically significant in the bronchiectatic lung, and clearly warrants research attention. PA is classified in the Proteobacteria phylum alongside other significant human pathogens such as *Escherichia coli* and *Salmonella*.[140]. It is a gram-negative rod and is part of the Pseudomonadaceae family. It has an adaptable genome of 5.2-7.1 Mega Base Pairs (Mbps) and is made up of a core genome of

approximately 4,000 genes and an accessory genome of approximately 10,000 genes.[141, 142].

1.2.1 Niches

With its high level of adaptability and complexity, PA is present in a large number of settings. I shall briefly describe its presence in the environment and the human lung. Whilst it is associated with pathology in other parts of the human body, such as the urinary tract and the soft tissues of burns victims, this will not be focused on here.

PA in the Environment

PA is often described as ubiquitous. It is commonly found in soil and water, and on the surfaces of hosts ranging from worms to mammals.[143] Further striking examples of its adaptability including its survival in fluids ranging from distilled water through to jet fuel.[1]

PA has many characteristics which facilitate its resilience to a wide-variety of environments. Whilst its metabolism is respiratory, it can grow in the absence of oxygen. It can be found in biofilms on surfaces or as a unicellular organism. In its unicellular state it is seen as a vigorous swimmer propelled by a polar flagellum. It can survive with minimum nutrition as demonstrated by its presence in distilled water. However, it can also use more than 75 organic compounds for growth.[144] Whilst it will ideally grow at a normo-thermic human temperature, it can grow at temperatures as high as 42°C.[144] As well a temperature tolerance, PA can also deal with high concentrations of salts, dyes and antibiotics.

PA's resistance to antibiotics is due to a variety of reasons. The permeability barrier of it gram-negative outer membrane causes natural resistance to some antibiotics. In addition, its resistance has evolved due to its natural environment in soil being host to bacilli, Actinomycetes and moulds and their naturally occurring antibiotics.[144]

Through these characterises and others, PA has become a highly successful environmental bacteria.[145]

Adaptation to the human lung

The most important PA niche within the limits of this work is the human lung. Whilst PA can be present in many lung conditions, it is at its most significant in CF and NCFB.

The human lung inhales a very large quantity of air each day, estimated to be of the magnitude of 10,000 litres per person.[2] This results in frequent and repetitive inhalation of micro-organisms including ubiquitous bacteria such as PA. Mechanisms are however in place to prevent acute infection and chronic colonisation of the lung in health. These include effective mucociliary clearance, as well as epithelial cells producing antimicrobial peptides, and macrophages in the airways.[146] However these mechanisms can fail if there are structural, functional or immunological deficits. This can therefore allow inhaled pathogens to persist and multiply, with the outcome of lung damage and chronic infection.

Experimental models have suggested that depressed mucociliary clearance is associated with PA infection.[147] Mucociliary clearance is impaired in CF and NCFB. With clearance reduced, PA is likely to initiate a foothold in the lung through adhesins such as pili and flagellae. *In vitro* work suggests that PA preferentially adheres to mucus and damaged epithelium, as well as creating more local damage.[148-150] PA may then inhibit mucociliary clearance further through pyocyanin production.[151] Once PA is established in the lung, multiple adaptations are employed to aid long-term colonisation.

A particularly well-documented adaptation is the formation of a biofilm. This protects the bacteria from attack from phagocytes, humoral responses, and antibiotics. Despite being under this parapet, PA can still drive significant antigen-antibody responses that lead to neutrophil recruitment. It may even be the case that isolates which do not produce biofilms *in vitro* can be incorporated in the biofilms of those who do produce.[152]

Interestingly, PA does not seem to maintain colonisation through increased virulence, but in fact the opposite.[153, 154] This includes a switch to a non-flagellated, non-motile phenotype and downregulation of quorum sensing (an inter-organism signalling system).[155-157] This tactic may be of benefit due to a reduced energy cost and reduced detection by the host.

Another important adaptation to highlight is a change to a hypermutable state.[158] Hypermutators occur when there is a defect in DNA repair leading to increased mutation, of which some may have a benefit to the bacteria. There are suggestions that the chronic state of a condition like CF drives hypermutation more than an acute environment.[159, 160] Hypermutators are considered an important factor in antimicrobial resistance and are common in CF with possible rates of approximately 40%.[159, 161]

When considering adaptation to the human lung, it should be remembered that more research has been done in CF patients than NCFB patients. It is possible that there are specific factors in the CF airway which enhance PA colonisation. This could include dehydration of the airway surface and possible receptors of PA on the membranes of epithelial cells.[162, 163]. Others factors may be less unique to CF and include hypoxic niches in mucus plaques.[164]

The complex process of adaptation is highly successful in those with NCFB and CF as demonstrated by high prevalence rates in these conditions. This is despite CF and NCFB guidelines advocating aggressive attempts at eradication of early PA.[6, 95] Regimes vary but can include prolonged usage of a combination of oral, IV and inhaled antibiotics.[165-167] If this fails, patients will often be put onto long-term anti-pseudomonal therapy. This is with the aspiration of managing chronic infection, and no longer of eradicating it. This approach will often involve inhaled antibiotics such as tobramycin or colistin.[120, 168-170]

1.2.2 Acquisition

With these disease processes being so impacted by PA, it is logical for there to be a desire to understand how it is acquired. With its status as a ubiquitous pathogen it was long assumed that the environment was where patients acquired their strain from. As investigation into this has advanced, it has been suggested that patient-to-patient transmission is also a possible route.

I will now discuss these two sources of acquisition.

Environmental

PA's ubiquity clearly make the environment a highly plausible source of acquisition. This is further backed up by data which suggest that the local environment influences acquisition. The majority of the relevant research has

been into the CF population. Interestingly it has seemed rare for patients to initially acquire PA from their own home.[171] However, it is very likely that a patient can colonise the home environment including everyday high-use objects such as a toothbrush.[172] This type of on-going reservoir is then a potential threat to successful eradication. The environment outside the home seems to be significant to acquisition risk, and the presence of water is a regular theme. An Australian study suggested an increased risk of first PA acquisition in non-city dwellers, with sprinkler use seemingly being associated with acquisition.[173] In a Belgium study, the proximity that a patient lived to a large body of water was associated with PA infection in a case-control study.[174] Water's capacity for containing a huge collection of strains has been illustrated by the presence of almost the entire global diversity of PA in a Belgium river.[175] Other associations which have been made with PA acquisition and the environment include warmer ambient temperature, dew point, longitude, latitude and elevation.[176, 177]

Patient-to-Patient Transmission

Whilst it is would seem difficult to prevent acquisition from the environment, it may be an expectation for cross-infection between patients not to be a significant issue. It has however become clear that cross-infection occurs and can be of great clinical significance.

In general it is felt that human-to-human transfer of respiratory pathogens takes place by contact transmission (direct and indirect), droplet transmission and airborne transmission.[178, 179] A pertinent point of note with regards to contact transmission is that PA can survive in dried sputum for at least a week.[180] An unusual example of indirect contact which was recently described was via aroma-therapy oils used by different patients on an Intensive Care Unit.[181] Droplet transmission generally refers to particles greater than 5 µm in size which do not remain airborne, while droplet nuclei are smaller than 5 µm and can be inhaled.[182]

Cross-infection via the airborne route with droplet nuclei is of particular concern. Knibbs et al demonstrated that cough aerosols containing viable PA could travel at least 4 metres and be detected at a 45 minute time period.[178] This 45 minute time period was consistent with a previous aerobiological model of viable PA.[183] An older study found evidence of a transmissible strain via air sampling

1-3 hours after patients left their ward room. This study also found the same strain when sampling ward corridors, as well as spirometry tubing and chairs after use by colonised patients.[184] It is possible that survival in aerosols may be enhanced in strains expressing a mucoid phenotype- a common finding in chronic CF strains.[183]

It used to be common practice for patients to mix at CF holiday camps. Whilst not a universal finding, several studies suggested that the sharing of pathogens, including PA, was occurring at these camps.[185-188] There were studies advocating the continuation of holiday camps as they felt the benefits outweighed the risks, however the overall body of work led to the cessation of these camps.[189] Person-to-person transmission was also suspected between siblings and also the participants of CF fundraising events.[185, 190, 191]

It had been hoped that cross-infection was limited to settings of particularly high personal contact, and that the hospital setting would not be of concern. Unfortunately, CF units were also appearing to be a site of transmission.[192, 193] Not only were patients seemingly acquiring the pathogen from other patients, but it appeared in many cases to be associated with clinically significant deterioration, increased resistance and virulence, and increased use of health-care resources.[194-199] However, it should also be noted that not all transmissible strains have clinically significant impacts.[200] A further counterpoint to note is that the risk of cross-infection can be low, particularly in small centres.[201-203]

The above findings have contributed to the varied practice of segregation in CF cohorts.[95, 179] These policies may include complete segregation or specific clinics for patients with specific strains or pathogens. Before these practices there was a strong argument that shared strains were merely a consequence of common source acquisition rather than cross-infection. However, changes in hospital policy with regards to hygiene and segregation have shown significant effects, and hence strengthened the case for cross-infection over common-source acquisition. The demonstrated benefits of these policies have included delayed chronic infection with PA, stopping the spread of a common strain, and reduced prevalence of a transmissible strain.[200, 204-206]. It appears that just implementing measures against contact and droplet transmission does not halt the spread of transmissible strains without strict segregation.[207, 208] This adds

weight to the concerns of transmission with droplet nuclei being a major factor. Beyond PA, there is also evidence of epidemic strains of BCC, MRSA, *Mycobacterium abscessus*, *Stenotrophomonas maltophilia*, *Achromobacter xylosoxidans*, *Ralstonia*, *Cupriavidus* and *Pandoraea*.[179]

There has been less research into NCFB patients, but they are not protected from transmissible strains. A case in Australia demonstrated the transmission of a multi-drug resistant strain of PA to a 14 year old after sharing facilities with CF patients over several in-patient admissions.[209] Work looking specifically at NCFB-to-NCFB transmission is very limited. Pujana et al looked at 16 patients and concluded that cross-infection or common-source acquisition did not appear to have occurred.[210] The largest piece of work looking at this took place in the North-East of England on a site where CF patients did not attend for health-care provision.[131] The study included 40 patients with NCFB, of whom 36 were seen in a specific bronchiectasis clinic. The authors felt that there was probably one case of cross-infection. They did question if they would have obtained the same results if they were on a site which also cared for CF patients. A recent whole genome sequencing (WGS) epidemiological study of PA in NCFB showed closely related isolates, which highlighted the possibility of cross-infection.[211] A publication by a large European Network listed the issue of cross-infection and segregation as a research priority in NCFB.[37]

Cross-infection will continue to be a concern in these two lung diseases. Whilst CF patients look set to continue with segregation, there is at present no apparent appetite to segregate patients with NCFB. The evidence in this cohort is limited and needs re-visiting in settings where both cohort groups share site facilities.

1.2.3 Epidemiology

The development of various molecular techniques has allowed more in-depth analysis rather than simply identifying a bacteria's species, or describing its phenotype and antibiogram. This has facilitated the identification of strains of PA. Techniques have suggested that on a global scale PA essentially has a non-clonal population structure but with the presence of a few highly successful clones.[212-215] This has not been an entirely universal finding and others have suggested that PA has a clonal population.[216] When trying to understand the true population structure it is difficult to know if the actual population has been

accurately represented in sampling. Understandably the investigation of clinical samples is likely to be disproportionately high due to interest and accessibility, but diversity is probably different in different niches. For example, there may be more diversity in the general environment and less in CF, COPD and acute infection.[217] A reasonable expectation would be that different clones have different capacities to infect humans.

The different strains of PA have become a major area of research interest. As discussed previously, this may be in relation to cross-infection but also with regards to virulence, pathogenesis, and resistance patterns. Whilst techniques such as Multilocus Sequencing Typing (MLST) have recorded over 2,000 strains, there is a narrow range of strains which have received the most attention.[213] I shall now review some of these.

Common Strains

As described above, there are some strains which appear to be more abundant than others. This is important when considering cross-infection. The mere presence of shared strains in a population may not mean that cross-infection has taken place, as a high prevalence of the strain in the local environment could suggest that independent acquisition has occurred. Two examples of common strains are Clone C and PA14, which are claimed to be the most common worldwide clones.[214, 217, 218] Clone C and PA14 have been described in a wide range of environments including salt and fresh water, domestic and wild animals, plants, and acute and chronic human infection.[145, 219, 220]

With recombination believed to be common in PA, it may be expected that strains could be relatively transient. Clone C was described over 20 years ago in a German study however, and is still frequently found.[218] In this initial study, isolates were investigated from a variety of sources including CF patients, non-CF clinical isolates, isolates from a CF clinic environment and from distant aquatic environments. Clone C was found across the various settings including approximately a quarter of a CF cohort, and a fifth of the local CF unit environment, and an unrelated aquatic environment 300-400KM away. The strain had also been reported in an ear infection in another German city. An interpretation of this study was that the high prevalence of this strain in clinical samples was due to its high background rate. The lead author of this initial work

published further epidemiological work, reporting Clone C in the UK from urine and peritoneal fluid.[221] It has also been reported in the literature from North America to Australasia.[145, 222]. In the UK estimates of its prevalence in the clinical setting are in the range of 2-6% based on CF, NCFB and all-source studies.[131, 223, 224]

Transmissible Strains

As discussed above, there are multiple studies suggesting cross-infection in CF populations. This has brought about the concept of the "transmissible strain". The list of widely acknowledged transmissible strains is limited. This may be in part due to a genuinely low number, but also difficulties in convincingly proving transmissibility. There will be a very large number of populations who have not undergone extensive epidemiological study, and even in those who have, the presence of shared strains may be due to common-source or independent acquisition of a highly-abundant background clone.

An early suggestion of a transmissible strain originated in Denmark.[193] At this point in time the techniques for strain identification were weak and the Danish study was based on antibiograms. They implemented an isolation policy, which was claimed to stop the epidemic strain. Subsequently the availability of molecular techniques allowed the more convincing reporting of other putative transmissible strains including the Liverpool Epidemic Strain (LES). In a single paediatric CF unit in Liverpool, 55 patients were shown to have the same strain as defined by Pulse Field Gel Electrophoresis (PFGE) and flagellin phenotyping.[192] Subsequent review of the adult unit revealed the presence of the strain in almost 80% of patients who were positive for PA.[225] Whilst there is no clear evidence of the presence or absence of LES in the environment, the significant rates in these populations and the lack of evidence for an alternative source, strongly supported LES as a truly transmissible strain. This was further supported by segregation policies having a marked effect on the prevalence of LES.[205] Whilst clearly present in Liverpool, LES has also been found in multiple other CF centres, both in the UK and internationally. A UK survey of CF isolates in 2004 found LES in almost half the centres involved in England and Wales and an estimated overall prevalence of 11%.[224] Canada is another country with a notable presence.[194]

Whilst the principle to avoid chronic colonisation of PA is a firmly held one, the importance of transmissible strains are in part dependent on their clinical significance and pathogenicity. LES's significance has been clearly demonstrated. LES has been shown to cause superinfection in those already colonised by other strains, to be associated with increased morbidity, to have a significant impact on 3 year survival; and to be associated with renal failure and empyema.[194, 196, 226-228] It also appears to have enhanced airborne survival.[184]

There has been evidence for other transmissible strains in the UK. In the previously mentioned 2004 survey, after LES, the second most common genotype was referred to as Midlands1. This was present in 10% of patients and 29% of centres. When considering its high rate it should be noted that the Midlands as a region was a very large contributor to this survey. One centre found rates of 30%, though despite its marked prevalence, it has not been as heavily researched or shown to be as pathologically significant as LES. Whilst likely to be a transmissible strain, the evidence is not nearly as convincing as for LES.

The other UK transmissible strain of note is the Manchester Epidemic Strain (MES). This was first reported in 2001 when 14% of patients were found to share a novel strain by PFGE.[229] The evidence for this being a transmissible strain is very convincing. The cohort who were studied contained three different patient groups who did not come into contact with each other: namely a group of patients with CF and PA, a group with CF and BCC and PA, and a group with PA but not CF. MES was not present in either of the last two groups. In addition to this, its spread was controlled by a strict segregation policy.[207] With regards to clinical impact, MES has been shown to be highly-resistant to anti-pseudomonal therapies and has been associated with a greater use of healthcare resources and antibiotics.[198, 229]

Outside of the UK other strains have been claimed to be transmissible including AES-1, AES-2 and AES-3 from Australia.[222] Of significance AES-1 has been associated with greater treatment requirements than unique strains; AES-1 and AES-2 with IV antibiotic usage; and AES-3 with exacerbations requiring hospitalisation.[222, 230, 231] In Europe claims have been made with regards to transmissible strains in Holland, Denmark and Norway.[232] North American work had shown LES to be present in Canada and also another potential

transmissible strain, termed Strain B.[194] A further putative transmissible strain described in Canada is the Prairie Epidemic Strain (PES).[233] Its association with increased morbidity and mortality has recently been shown.[234] Another suggested transmissible strain in North America was Houston-1.[235] This strain was associated with recent hospitalisation, and transmission was halted following adoption of infection control guidelines.

At this point in time, there has not been a clear explanation of what makes some strains transmissible and others not. It may be that there are many currently unidentified transmissible strains. There may be some which have been described as transmissible, but may not be as transmissible as others. Conceivably all strains could be transmissible some of the time with specific exposures and conditions, but that some are more transmissible than others. Either way, in some circumstances patient-to-patient transmission is clearly a risk to patient's long-term health and prognosis. Consequently, avoidance is highly desirable.

High Risk Clones

The concept of high risk clones has been promoted in recent years.[236-238] These are commonly referred to in the context of significant antibiotic resistance. Examples of these include clones producing metallo-beta-lactamases. High risk clones have been reported globally and include the MLST described Sequence Types (STs) 111, 175, 233, 235, 277, 357, 654, and 773.[239-242]

Epidemiology in the South West of England

The work described in the investigative chapters of this thesis was performed on isolates obtained from the Royal Devon & Exeter Hospital (RD&E) in Exeter. Exeter is the smaller of the two cities in the county of Devon and is situated in South West of England. Historical data of which strains are prominent in the local area is limited. The 2004 survey of PA in CF did however give a breakdown of the prevalence of LES, MES, Clone C and Midlands1 by hospital.[224] The hospitals were not described by name in the publication, but three centres in the South West were listed as contributing 57 isolates. The rates for LES were 18%, and 2% for Clone C. Midlands1 and MES were not found.

1.2.4 Genotyping Techniques

Techniques exist which define different examples of the same species. These various techniques are of great importance for the understanding of bacteria, their function, their impact and their epidemiology. In settings where cross-infection or the source of infection is deemed of interest and importance, these techniques are crucial.

A concept often described when considering these issues is that of a “strain”. Whilst it is a commonly used word, it is actually a little ambiguous. A strain can be defined as a different genetic variant of a species. Due to this, it suggests that knowledge and assessment of the genome is required to differentiate strains. However, early descriptions were based heavily on phenotypic traits.[180] As molecular microbiology has advanced, the sophistication of techniques to differentiate between different examples of the same species has markedly changed. It could be said the technique of WGS is the ultimate strain typing technique due to it unparalleled resolution, however differentiating strains on the basis of a small and potentially irrelevant part of the genome is of little practical use. It may be best to refer to the techniques as “typing techniques” or “genotyping” rather than “strain typing”. Another concept or description used is that of lineage, or clonal complexes. This is when a collection of isolates can be clustered together due to their alikeness with the inference of a recent common ancestor. The concept of “strains” in relation to real-world application is perhaps best summarised as those with a similar genetic core who behave, or are able to behave, in a predictable way to various environments and pressures.

There have been many techniques used to differentiate genotypes. Since the utilisation of extracted Deoxyribonucleic acid (DNA) for these purposes, these techniques can largely be divided into Molecular Fingerprinting and Sequencing techniques. In these techniques, DNA tends to be extracted following cultivation of the bacteria in the laboratory, and subsequent picking of a colony. One potential issue with this is that the results of a single colony may not be representative of the sample overall, and may lead to the under-reporting of important strains.

In this thesis three techniques have been used and they will now be focused on within their categories.

Molecular Fingerprinting

Molecular fingerprinting techniques involve the utilisation of template DNA to produce a visual pattern where different examples of a species can be differentiated. Examples of these techniques include PFGE, Random Amplified Polymorphic DNA (RAPD), and Polymerase Chain Reaction (PCR) based Open Read Frame Typing.[243-245] They often involved a PCR step followed by separation of products via gel electrophoresis. The resulting pattern can then be interpreted either by eye or with assistance of computer software. The fingerprinting technique used in this work has been RAPD.

Random Amplified Polymorphic DNA (RAPD)

RAPD has been used to genotype a variety of bacteria including PA for over 20 years.[246-248] It works by using short arbitrarily chosen primers for PCR with the aim of producing a "strain" specific pattern. Isolates with similar patterns are considered highly likely to be the same strain. RAPD has been a prominent method in PA genotyping studies and it has been a key technique in the study of CF cohorts, burns units, and comparisons between water supplies and patient genotypes.[189, 195, 203, 249, 250] It is appealing due to being able to cheaply and quickly discriminate between isolates.

As with any technique, initial comparisons with established techniques are required, and subsequently again with novel ones. Early on, RAPD illustrated it usefulness by showing greater discriminatory potential than traditional serotyping schemes.[251] PFGE has been considered a gold-standard of genotyping, and some early work suggested that RAPD could be as discriminatory.[244, 247] These were however in small and selected cohorts. The most recent useful study comparing molecular techniques which included RAPD was published by Waters et al in 2012.[252] The study was blinded and multi-centred. RAPD, MLST and PFGE were compared. A potentially significant design factor in the outcome of this study was that an isolate was described as clonal if there was agreement between at least two methods. As the study showed a greater congruence between PFGE and MLST (though overall this was relatively poor), there was a bias against RAPD. Even considering this bias, RAPD did appear to be inferior in all of the tested criteria in comparison to MLST and PFGE.

Additional limitations of RAPD should be highlighted. A major weakness of the technique is its lack of portability. Output between laboratories cannot reliably be

compared as very subtle differences can change the results. This includes which batch of enzymes are used and the source of the primers; and heat-transfer issues such as the thickness of the PCR tube, and the make and age of the thermocycler.[253] Even within the same laboratory there may be a lack of reproducibility by factors such as the ambient temperature that day. A further potential source of difference may be the hand that the pipette is in. It has been demonstrated that the duration the pipette tip is immersed can impact reproducibility due to errors when very small quantities are being used. Consequently an experienced technician may get a different output to a ponderous novice despite all other conditions being equal.[253]

Once RAPD has been performed there is still potential for differences in interpretation. With human visualisation there will clearly be a degree of variability in interpretation. As the number of isolates increases, so does the distance over which comparisons are made, and consequently the accuracy may wane. Other discrepancies occur when considering what a sufficiently different pattern is. In some cases the total number of bands will be taken into consideration when assessing “unique” patterns, or alternatively the presence or absence of a certain number of bands. This may vary depending on the question asked. For example, if you were looking to screen isolates to take forward for a confirmatory second technique, you may choose a different cut-off to if it is your only technique.

Despite these limitations, and the development of more sophisticated techniques, RAPD continues to have value in the correct setting and it is recognised by the CF Foundation as an appropriate genotyping method in CF cohorts.[179] We believe it has value as a cheap first-line screening technique when examining a large number of samples. This can then allow more targeted use of more sophisticated techniques. To minimise its limitations it ideally should be performed in a single run with the same vials of reagents and the same scientist. When the number of samples analysed creates such a distance between products on a gel that uncertainty occurs, computer analysis should be incorporated.

Sequencing

The use of sequencing technology to determine genetic differences within a species is a logical step in genotyping particularly as its cost decreases, and

quality and availability increase. There are a variety of ways that sequencing is used, and this ranges from the sequencing of specific regions or genes, through to sequencing the entire genome. In this thesis the two examples we have used are MLST and WGS, which I shall now describe.

Mulitlocus Sequence Typing (MLST)

MLST was first described for *Neisseria* at the end of the 20th Century and was an advancement on the concepts of Multilocus Enzyme Electrophoresis.[254] The scheme was subsequently adapted for PA and the most commonly used version was published in 2004.[213]

MLST contains two main steps: the amplification of housekeeper genes by PCR and subsequently the sequencing of the products. In the scheme for PA, seven loci were selected. The selection of these seven were based on biological role, size, location, suitability of primer design, and sequence diversity.[213] The sequencing data is used to assign novel alleles within each locus an arbitrary and unique number. The combination of the allocated numbers is then used to create an ST. For examples the allele combination of 1,1,1,1,1,1,1 is defined as ST-1. With this output a database has been created where scientists can submit and compare their “Sequence Types” (STs) with a global population. An alternative way to analyse MLST data is by a nucleotide based interpretation of concatenated data. Whilst this may be more accurate, it is less accessible and more complex to perform and report.

With other typing techniques preceding MLST, work was obviously required to compare this to established techniques. At the point of the introduction of MLST, PFGE was considered the gold-standard technique. Various studies have shown a variety of outcomes on this matter. There is some evidence that PFGE may be more discriminatory for detecting genetic differences, whilst MLST is better at detecting genetic relatedness.[252, 255, 256] It is worth noting that MLST exclusively interrogates well-preserved genes which is not the case in PFGE where a genome-wide interrogation takes place. With PFGE a single mutation can cause significant profile changes.[257] Therefore in strains with significant instability of the accessory genome, the discriminatory power of PFGE can fail to identify isolates with a common core genome. There have been occasional descriptions of a change in MLST type despite a consistent PFGE profile, though this is rare.[258, 259] It could be argued that PFGE is more suited to situations

where there is no time for mutations to influence output, such as a real-time outbreak, whilst MLST is more suited to a longer study duration. Issues with previous typing techniques including PFGE, revolved around the reproducibility between laboratories and the comparison of outputs globally. The unambiguous MLST scheme and an easy to negotiate database has helped to address this. However, it is worth remembering that the genome of PA is in the range of 5.2-7.1 Mbp, and the MLST scheme for PA only interrogates regions totalling 2882 base pairs (bp)[260]. Consequently there may be a gene of particular interest to an investigator, such as those responsible for specific virulence factors, which is not examined.

For MLST to be a strong longitudinal technique, the so-called housekeeper genes (*acsA*, *aroE*, *guaA*, *mutL*, *nuoD*, *ppsA* and *trpE*) need to be stable. Molecular clock speed calculations have suggested that this is the case, though there have been concerns in other studies about the stability of the *mutL* allele.[258, 261]. Other limitations of the scheme include its cost in certain environments. Others have suggested the time consuming nature of the technique to be a drawback, and alternatives and variations have been tried to combat this.[262-264]

Despite some of the weaknesses described above, the portability and lack of ambiguity of this technique has been valuable.

Whole Genome Sequencing (WGS)

Whilst techniques such as MLST have their undoubtable use, there can be circumstances requiring greater resolution. This may be when considering transmission pathways or interrogating the bacteria for genes associated with virulence or resistance. WGS provides this greater resolution by sequencing the entire genome. Whilst it is a relatively new tool with regards to cross-infection, it has already shown its real-world ability to enhance the understanding of outbreaks and the evolution of strains.[265-267]. It is a realistic expectation that WGS will be a very useful tool for both transmission studies and antibiotic susceptibility testing, with *Mycobacterium tuberculosis* a potential exemplar pathogen.[268]

Despite the obvious appeal of WGS, there are downsides which must be considered. The cost of WGS is continuing to decrease but is clearly more expensive than the preceding techniques. The output is vast and a large amount

of the information gleaned may be of no interest or even if it is, may well not be fully understood. When considering episodes of cross-infection, the technique can give ambiguity. In cases where there is essentially no difference between isolates, or a very large difference, then the chances of cross-infection can become much easier to assess. In-between these scenarios, it is less clear. Single Nucleotide Polymorphisms (SNPs) are often used as a value to compare samples. Approximations of mutation rates may allow interpretation of whether a difference of SNPs is likely to be significant over a period of time.[266, 267]. However, we know that clinical isolates of PA can acquire a hypermutator status and this can have a dramatic impact on mutation rates.[269] Consequently interpretation in the presence of hypermutators is far from straightforward, and has not been adequately addressed in cross-infection studies so far.

1.3 *Pseudomonas aeruginosa* and the Wider Bacterial Community

A large amount of the research into PA has been conducted in the *in vitro* environment. It is also often done with PA being the sole focus. Whilst there is a logic to this, we live in a polymicrobial world where bacteria are constantly coming into contact with each other. It would therefore seem sensible that when considering real-world scenarios, PA is not looked at merely in isolation, but in the context of the other bacteria which may be present in the lung. Examples of why this may be important include the seemingly antagonistic relationship between PA and HI; the potential of PA to thrive off *Streptococcus anginosus*; and the possible synergy between PA and *Stenotrophomonas maltophilia*.[270-273] Consequently, this thesis will contain work into the wider bacterial community in the lung, including PA's place within it.

1.3.1 Culture-independent Techniques

The microbiological assessments which has been described so far in this thesis have been based on culture-based techniques. Culture-based techniques have been used for over a century and remain the process used in everyday clinical care. These techniques have been refined and adapted to provide information on a specific clinical question: is there evidence of a pathogen which is a potential cause of the patient's current illness? Consequently, in different clinical scenarios, different protocols will be used to increase the chances of finding a pathogen which is believed to be clinically-relevant. Extended and adapted

culture protocols which can yield obligate anaerobes are not routinely performed, but are possible and an excellent example of how protocol can influence outcome.[274, 275] An inherent bias of this is that our knowledge of putative pathogens is based on what has previously been cultured and therefore bacteria which are difficult to cultivate in a laboratory are likely to be less well understood in a disease process. Despite this limitation, traditional culture has remained useful. What it has not been optimised for, and unsurprisingly is not very useful for, is describing a bacterial community. For understanding of this specific area, culture-independent techniques are required.

Culture-independent techniques typically use extracted DNA as a template to reveal community diversity and, potentially, composition. They are theoretically providing unbiased insight into what is actually there, rather than what grows in a biased artificial laboratory environment. However, culture-independent techniques do have some caveats, and these can be before the technique even begins.

In respiratory medicine, the samples which are used for microbiological investigation are predominantly sputum samples or samples obtained via bronchoscopy. In both these scenarios, between the sample originating in the lungs and it's containment in a sample pot, it must pass through the upper respiratory tract, either directly or via a bronchoscope. It is consequently susceptible to contamination by the upper respiratory tract. Despite this risk, various studies have suggested that this does not seem to impact results significantly.[276, 277] This may in part be due to aspiration from the oropharynx being a significant contributor to the bacterial community in the lung. A further issue with the sample itself is that it may only be representative of a small area of the lung, and a sample from a different region may give a different result.[278] These caveats are also factors when interpreting data from culture results.

Once the sample has been adequately stored, the next key step is the extraction of DNA. This stage can also impact the output of these techniques. Bacteria differ with the ease with which DNA can be extracted and different approaches may influence the success of DNA extraction. Examples of this include gram-positive bacteria such as SA, from which DNA may be difficult to extract.[279, 280] Steps such as mechanical disruption of the cell wall are included in some but not all protocols to try and mitigate for this.[281] A further concern, particularly with low

biomass samples, is that DNA extraction kits can be an unpredictable source of contamination.[282] A “sterile” kit will not contain viable bacteria but may have the presence of dead bacteria from which DNA can be extracted. This is a reminder of a further limitation: culture-independent techniques may not be reflecting which bacteria are actually living in the lung, but what has been there. This may still give a useful signal clinically. For example, the presence of a large amount of PA may reflect a moderate amount of viable PA and impaired airway clearance resulting in the retention of non-viable cells. Methods have been employed to differentiate DNA from viable and non-viable sources, however these are not routinely incorporated in clinical studies.[283]

At present culture-independent techniques are not used in everyday healthcare systems like the National Health Service (NHS). Whilst there are technical and financial reasons for this at the moment, there is also a deficit in understanding how to utilise these techniques and what their outputs actually mean for patient-care. When considering these techniques, like the previously described genotyping techniques, they can be divided into community fingerprinting and sequencing techniques. I shall now describe these in more detail with specific attention to the one example of each that I have used.

Community Fingerprinting

These are the original techniques which provided some insight into bacterial communities. They tend to involve a PCR step which amplifies a discriminatory genomic region, and then visualising the products via electrophoresis. Examples of community fingerprinting techniques include Denaturing Gradient Gel Electrophoresis (DGGE), Terminal Restriction Fragment Length Polymorphism (T-RFLP) and Ribosomal Intergenic Spacer analysis (RISA).[133, 284, 285] These techniques have the universal appeal of being relative quick, cheap and simple in comparison to sequencing techniques. These advantages predictably come with limitations. The techniques give a visual overview of the community structure but do not provide actual taxonomic identification. Taxonomy can be speculated based on the output, which could direct subsequent investigation.[133] They also have reduced sensitivity to low abundance species.[285, 286] Ecological communities can however be dominated by relatively few bacteria, and hence despite this underestimation of richness, these techniques can still differentiate communities.[287]

In this body of work, one community fingerprinting technique has been used- RISA. I shall now describe this in more detail.

Ribosomal Intergenic Spacer Analysis (RISA)

The RISA technique involves the interrogation of a specific region between two genes: the 16S rRNA gene and the 23S rRNA gene. The principle of the technique is that this intergenic region is variable in length between different species. Following a PCR step to amplify this region, the PCR products from different bacteria are separated on an electrophoresis gel providing a complex band pattern for visualisation. As with the RAPD technique previously described, computer software can be used to analyse the PCR products more accurately and consistently.[133]

Its background has historically been in complex environmental samples and some evidence suggests the technique has greater accuracy than T-RFLP.[288-291] It has also been used in mammalian work with the suggestion of reasonable concordance with more sophisticated sequencing data.[292] RISA has not been used much in Respiratory research and when it has been, it has exclusively been in CF cohorts.[133, 270, 293, 294] All of these studies have suggested potential usefulness in clinical practice.

Like other fingerprinting techniques, there appears to be a reduced sensitivity to all species. A further potential issue is that the bands of different species may be close together and inaccuracies in observing the output may impact the observations. At present there have been no investigations into the utility of RISA in NCFB.

Sequencing

As described already, sequencing techniques are potentially very powerful tools. As well as being of use when investigating a single species, they can also be applied to complex microbial communities. A metagenomic approach can be taken with shotgun sequencing where all the genes are sequenced to give an appreciation of the capability of a community. An alternative, and currently a more frequent approach, is the cataloguing of species present by 16S rRNA sequencing. This is the approach used in this body of work and will be described in more detail below.

These techniques have the same problems as the fingerprinting techniques with regards to the template obtained to work with. They are also subject to sequencing errors and considerably more cost that the fingerprinting techniques despite recent reductions.[270] Following the sequencing, considerable bioinformatic input is then required.

16S rRNA Sequencing

16S ribosomal RNA is a part of the 30S subunit. The gene which codes for it, the 16S rRNA gene, has been interrogated since the late 1970s for the purposes of identification.[295, 296] Initially it was met with significant scepticism, but is now widely used and accepted.[297] The reason that this technique works is that the 16S rRNA gene is conserved across bacterial species; is not present in non-bacterial species; and has nine hypervariable regions. Consequently it allows both the detection of bacteria and phylogenetic identity. It has an impressive ability to detect very low levels of abundance. For example it may pick up abundance levels of 0.01%, whilst in contrast fingerprinting techniques may just pick up 1%.[286]

However, the technique is not perfect. The usual grouping of bacteria is based on a certain level of sequence identity (normally greater than 97%). The groupings are referred to as Operational Taxonomic Units (OTUs). OTUs have variable resolution to species or genus level, and consequently can group non-pathogenic and pathogenic species in the same group. Different variable regions have been sequenced for the purposes of identification, and sequencing different regions may give you a different result.[298] Also, using a different sequencing platform may also change your result.[299]

Despite these limitations, this technique has taken our understanding of bacterial communities in the lung to a new level.

1.3.2 Bacterial Communities in the Lung

Interest in the interaction between bacteria and health has been on the rise in the last decade, and has been typified by the Human Microbiome Project.[300] The long-held dogma had been that the lung was essentially a sterile site in health, and it was not included in this initial project. Subsequently it has been shown that this dogma is incorrect.[301]

In retrospect it does not seem a surprising finding that the lungs are not sterile. It has long been acknowledged that we breathe in a huge number of organisms and that micro-aspiration occurs from the bacteria laden upper airways.[302]

Once it has been acknowledged that the lung is not sterile, it is worth considering how the community in the lung is formed and develops. The ecological modelling description as presented by Dickson et al is a clear way of conceptualising this.[303] Like any population, a community is reflective of the effects of immigration, emigration and intra-site reproduction. Immigration is essentially via inhalation, micro-aspiration and direct mucosal dispersion. Elimination in the healthy state will be by cough, mucociliary clearance and the host immune system. In a steady state you would expect a balance between immigration and elimination, and subsequently very little intra-site residence and reproduction. With this in mind, it becomes conceptually easy to see how dysbiosis could occur through changes in immigration and/or elimination (see Table 1.2). In an environment where a steady state is not maintained, long-term residents will exist in this situation, and local factors will bias some community members over others. These factors could include pH, oxygen tension, nutrient availability, local microbial competition and local inflammatory response.

Factor affecting Immigration	**Factors affecting Elimination**
Oro-pharyngeal microbial burden	Mucociliary clearance
Minute ventilation	Cough
Gastro-oesophageal reflux	Microbial growth conditions
Medications	Innate Immune Response
Gross aspiration	Adaptive Immune Response
Laryngeal Dysfunction	Medications

Table 1.2 Factors influencing immigration and elimination of bacteria into and out from the lung

These concepts would point to certain disease states having specific community profiles. This has been investigated in both health and disease over the last few years and it has indeed been the case that different diseases appear to have different profiles. Whilst this has been of great interest, there remain key

questions and issues. These include: how much of a pathogenic entity is the microbiome or is it more a consequence of the process; can knowledge of the microbiome positively influence therapy; and if so how can this information be delivered in the real-world in a prompt and affordable manner? The microbiome may influence the inflammatory process on epithelial surfaces, and the potential of microbiome manipulation is intriguing.[304] It is likely that important components are the fungal and viral communities in the lung (the "Mycobiome" and "Virome"), but that is beyond the scope of this work which will focus on bacteria.[305]

I shall now describe the main findings with regards to bacterial communities in the healthy lung and in CF and NCFB.

The Healthy Lung

The first main piece of work which included healthy controls was a study looking at Asthma and COPD patients.[301] The key finding from this work was that the healthy lung appeared to have a bacterial community similar to the upper airway and that those with chronic respiratory disease had distinctly different communities. There has been subsequent work which has created an appreciation of the composition of a "normal" microbiome.[302] It appears that the prominent phyla are Bacteroidetes and Firmicutes, whilst *Prevotella*, *Veillonella* and *Streptococcus* are the common genera. The healthy lung has also been shown to contain *Staphylococcus*, *Pseudomonas*, *Fusobacteria*, *Acinetobacter* and others.[306, 307] There are varying degrees of similarity between the mouth and the lung, and this may be a reflection on factors such as oral hygiene and difference in reflux and aspiration. Aspiration is common during deep sleep in healthy subjects and there is data to suggest that the GI tract may have a significant impact on the lung microbiome.[308-310] The undamaged airways of the healthy lung will have less regional variation in growth conditions, resulting in the communities being relatively uniform throughout the lung.[311]

The CF Lung

The CF lung was unsurprisingly an early target for culture-independent investigation and, via fingerprinting techniques, this gave the first hints of the wealth of detail not revealed by culture.[284] As more 16S rRNA sequencing has been performed, more specific observations have been made. There appears to

be a reduction in community diversity with increasing age and disease severity.[286, 312-314] In end-stage disease, there can be evidence of just a single pathogen which has complete domination of the site. The loss of diversity appears most tightly correlated with cumulative antibiotic usage.[314, 315] The true driver of reduced diversity is however very difficult to identify as cumulative antibiotic usage will increase with both age and disease severity. However, there does seem to be an association with greater rate of decline of lung function in less complex communities, which may suggest a pathogenic aspect to certain bacterial communities.[316] The intriguing link between the microbiota in the gut and the lung has shown reduced gut diversity in those with severe lung disease.[317] Again the impact of cumulative antibiotics may be a significant contributor to this observation.

Whilst traditional culture results effectively give a binary outcome for a specific pathogen (i.e. either cultivated or not), 16S rRNA sequencing reports both the presence and the abundance of a pathogen. As previously described, PA and SA are recognised as the predominant pathogens by culture. When considering mere presence or absence though, *Streptococcus* and *Prevotella* may be more common. *Pseudomonas* may be the next most common OTU, with *Veillonella*, *Rothia*, *Granulicatella*, *Gemella* and *Fusobacterium*, all more frequently present than *Staphylococcus*.[318] However, there may be an anomaly with *Staphylococcus* as it appears to be more frequently revealed by culture than culture-independent techniques. It is unclear if this is due to difficulty extracting the DNA from SA, or poor amplification of the 16S gene in a mixed sample.[319] When abundance rather than mere presence is looked at, there are notable pathogens which when present appear to be in high abundance and dominate the community. These include the OTUs of *Pseudomonas*, *Staphylococcus*, *Stenotrophomonas*, *Burkholderia* and *Achromobacter*.[318] With the exception of SA, all the main pathogens in CF are potentially under-reported by culture. Consequently, important clinical information may be being missed.

The NCFB lung

As in CF, significant weight is placed on culture-dependant results when clinical decisions are being made. Often to the frustration of the treating physician, microbiological investigation of sputum can fail to give useful guidance. One study suggested that 40% of purulent samples did not isolate a recognised

pathogen.[320] Culture-independent investigation is therefore clearly of interest in this disease cohort as well.

So far there has been a small collection of interesting studies into NCFB and the associations between microbiomes and severity. A study by Rogers et al looked at samples from 41 patients and concluded that bacterial diversity positively correlated to the forced expiratory volume in one second (FEV_1).[321] This study identified “core” bacteria in NCFB and unsurprisingly included *Pseudomonas* and *Haemophilus*, as well as others including *Streptococcus*, *Veillonella* and *Prevotella*. Further work by Rogers also suggested that richness was negatively correlated with sputum weight, age and inflammatory markers; that communities dominated by *Pseudomonas* or *Haemophilus* had reduced richness; and that those dominated by *Pseudomonas* were most likely to suffer an exacerbation followed by those dominated by *Veillonella*.[322, 323] Of possible note, they did not find an association between diversity and antibiotic use.

Not all work in NCFB has found associations between the bacterial community and disease severity.[271, 319, 324] There are a few possible explanations for these conflicts. It is possible that in fact there is not a strong association between community profiles and disease severity, and that in some cohorts an association will not be revealed. This may particularly be the case in studies with limited numbers. It could be that geographical location makes an impact. The work of Rogers et al is often Australian-based whereas most of the other main papers are UK based. There are some interesting points to make from the UK papers which may give some hints about the relationships of the bacterial communities and NCFB patients. In the NCFB cohort from the North-East of England, there was greater richness in those who were “culture-negative” and less in those who cultured PA.[271] The study based in London found reduced diversity in those on long-term antibiotics, those with a positive culture result, and if mucoid PA was isolated.[319] The studies from London and Belfast showed impressive longitudinal stability of profiles even in the face of exacerbations and antibiotics.[319, 324] These studies and others are a reminder of our lack of understanding of exacerbations in the bronchiectatic lung and what antibiotic therapy actually does.

A further reason why we are not entirely clear on the association between disease severity and community profile may in part be due to the markers of severity

studied. For example, FEV_1 certainly has limitations as a disease marker.[283] There are now two severity indexes for NCFB.[40, 325] While they both contain FEV_1, additional factors include age, BMI, culture results and radiological changes, and they predict morbidity and mortality. These severity scores may be a useful tool when comparing bacterial communities to severity.

1.4 Key Points of Chapter 1

In the above sections key points have been raised which are of particular note when considering the investigative work that has been performed for this thesis. They can be summarised in the following main points which can be divided between clinical and scientific issues.

1.4.1 **Key Points- Clinical**

- The bronchiectatic lung is a pathological state found in patients with NCFB and CF.
- NCFB and CF are important diseases associated with significant morbidity and mortality rates.
- PA is a highly prominent and clinically-important pathogen in both these diseases.
- PA has been shown to be transmissible between patients in CF, leading to changes in practice. The equivalent risk is unknown in NCFB and patients are potentially in danger of significant cross-infection.
- The lungs are a polymicrobial communities. The relationship between PA, the wider microbial community, and disease severity is unclear in NCFB.

1.4.2 **Key Points- Scientific**

- Multiple genotyping techniques are available with different advantages and disadvantages.
- Whilst WGS may be the future of these techniques, resources may limit its use in some settings. Also, determining cross-infection with WGS is not clear-cut, particularly in circumstances where hypermutator strains are involved.
- Most genotyping technique involve the cultivation of a pathogen in the first instance. This is time-consuming, not always possible, and may lead to the typing of an unrepresentative colony.

- Culture-independent techniques such as 16S rRNA sequencing have revealed polymicrobial communities in the lung, but issues with complexity, time and cost may limit clinical utility.

1.5 Aims of the Thesis

With these key points in mind, this thesis had set aims which have been addressed and described in Part B. This has been achieved with the methods and materials which are described in Chapter 2. The aims were as follows

1) To perform an epidemiological review of PA in NCFB utilising a variety of genotyping techniques to investigate the likelihood of cross-infection with PA.
2) To gain insight into the prevalence of hypermutable PA in NCFB and its impact on WGS-based studies addressing cross-infection.
3) To test the utility of a novel culture-independent genotyping technique for PA.
4) To investigate the relationship between disease severity and the bacterial community composition in NCFB.
5) To test the clinical utility of RISA as a cheaper and quicker alternative to 16S rRNA sequencing.

Chapter 2- Methods and Materials

In this chapter this I will describe the main methods and materials used through this body of work.

2.1 Sample Acquisition

All the samples used in this work have come through clinical practice at the RD&E hospital. All respiratory samples were expectorated sputum rather than via bronchoscopy and were all collected for microbiological investigation as part of the patients routine care. All respiratory patients were recruited prospectively and consented either as agreed with the local ethics committee, or for the donation of the sputum sample to the local National Institute of Health Research (NIHR) tissue bank. The recruitment process included the collection of clinically-relevant information. The only other clinical samples obtained and studied were isolates of PA from non-respiratory origins. For these samples no patient details were known beyond type of sample (e.g. urine) and health-care setting of origin (e.g. hospital ward). Consequently, no consent was obtained from any non-respiratory patients.

The sputum samples all initially went to the Exeter Clinical Laboratory which contains the accredited microbiology services used by the RD&E. Following receipt of a sputum sample, an equal volume of Mucolyse (dithiothreitol) was added before vigorous shaking for 10 seconds and being left at ambient temperature for 15 minutes. After a further 10 seconds of shaking, up to 4mls was aliquoted into 2 ml vials for the samples identified for culture-independent investigation (see Chapters 5, 6 and 7). The samples in these vials were subsequently heat-treated for de-activation for safe transfer. The heat-treatment was performed at 95°C for 20 minutes. This protocol was chosen on the basis of advice from a National Mycobacteria laboratory, as the principle reason for heat-treatment was the potential of samples to harbour *Mycobacterium tuberculosis*. Before commencing our studies, DNA extraction had been performed on samples in the Exeter Clinical Laboratory both before and after heat-treatment to look for possible impact on downstream uses of the DNA. Heat-treatment was found to have no significant impact on DNA yield or PCR success.

The aliquoted sputum samples were transferred on dry-ice to a University of Exeter laboratory for storage at -80°C. From the sample remaining in the Exeter Clinical Laboratory, a 10 µl loop of digested sputum was added to 5ml of Maximum Recover Diluent (MRD). Plates were then inoculated immediately after dilution. The plates used were based on the clinical diagnosis. For NCFB patients, Blood and Chocolate Agar plates were incubated in Carbon Dioxide; and Cystine Lactose Electrolyte Deficient (CLED), Mannitol and Sabouraud Agar plates were incubated in air. For samples from CF patients, *Burkholderia cepacia* agar was also used.

For the isolates from which the epidemiological review of PA was done (see Chapter 3), 10 representative colonies were picked and subsequently stored on Microbank™ microbial storage beads (Pro-Lab Diagnostics). These were transferred on dry-ice to a University of Exeter laboratory for storage at -80°C. PA was identified by technician observation, an oxidase test and confirmation with the VITEK®MS MS Matrix-assisted Laser Desorption/Ionisation- Time Of Flight Mass Spectrometry (MALDI-TOF MS) (Biomerieux).

2.2 DNA Extraction

DNA extraction has been a key step in much of the investigative work described in this thesis. There are many different techniques and commercial kits available. For different stages of this work, three different approaches were required. These were for the extraction of DNA from PA colonies; the extraction of DNA from whole sputum to investigate PA; and the extraction of DNA from whole sputum to investigate the wider bacterial community. The reason for two different protocols when working with sputum is that some gram-positive bacteria are harder to extract DNA from. An example of this is the difficulty of extracting the important respiratory pathogen SA.[280] To improve extraction of DNA from SA, a mechanical disruption step is required, however this is not required for the gram-negative bacteria PA.

I shall now describe the different methods used.

2.2.1 DNA Extraction of PA

This process was performed for the extraction of PA from stored isolates obtained from sputum from NCFB and CF patients, and from stored isolates from non-respiratory samples.

PA was re-grown from microbial storage beads on plates of LB agar and incubated at 37°C. Re-plating and ongoing incubation took place if required to allow single colony picking.

Single colonies were inoculated into 5mls of LB broth for overnight incubation (37°C, 200RPM). 2mls from the overnight culture was then used for the DNA extraction protocol. The process was performed with the GeneJET Genomic DNA Purification Kit (ThermoFisher Scientific). A modified version of the Gram-Negative Bacteria Genomic DNA Purification Protocol was performed. The only adaption was a reduced eluting volume (100 µl) in order to obtain an adequate concentration for downstream applications.

The DNA was checked for quantity and quality by a NanoDrop™ 1000 Spectrophotometer (ThermoFisher Scientific).

2.2.2 DNA Extraction From Whole Sputum for PA MLST

In order to perform DNA extraction to apply the MLST scheme on whole sputum a different kit was used, namely the QIAamp DNA mini kit (Qiagen). The “DNA purification from Blood or Body Fluids (Spin Protocol)” was followed using 400µl of sputum and modifications as described previously.[326]

2.2.3 DNA Extraction From Whole Sputum for the Assessment of the Bacterial Community

A different approach to DNA extraction was performed for the purposes of assessing the bacterial community of the lung. This is due to the need to extract DNA from both gram-positive and gram-negative bacteria. Concerns persist about gram-positive bacteria being under-represented in microbial community profiling studies due to a tougher cell wall and the need for mechanical disruption to overcome this.[280, 281] Several kits and protocols were trialled including: the protocol used in 2.2.2 with and without bead-beating, the UltraClean Microbial DNA Isolation Kit (Mo Bio), and the FastDNA™ Spin Kit for Soil (MP Biomedicals). As RISA was the principle downstream application, several samples were tested

via these different kits and visualised by gel electrophoresis. There was no discernible difference between the samples based on RISA output. Due to no evidence of inferiority or failure to produce DNA which could be amplified, it was decided that a protocol that included mechanical disruption should be used to avoid the concerns in the literature surrounding extraction from gram-positive bacteria. Due to personal recommendations and prominent use in the literature, the FastDNA™ Spin Kit for Soil was used.[319, 327-330]

DNA extraction was performed as described in the manufacturer instructions with minor adaptations. An updated protocol was released by the company after the first few samples were processed, and these were consequently re-done so that all underwent the same protocol. The manufacturer's protocol used 500mg of soil as the starting product and we used 500µl of sputum. In addition, instead of the FastPrep Instrument, the Precellys 24 (Bertin Instruments) was used for homogenisation. This was performed three times at 6,440 RPM for 30 seconds with cooling on ice in-between cycles. The other adaptation was that the SpinTM Filter underwent its 5 minutes of air drying at 60°C in order to aid removal of residual ethanol. This step was adapted following personal correspondence with the company.

2.3 Random Amplified Polymorphic DNA (RAPD)

In this work, RAPD was used as an initial screening method so that we could work with a large number of isolates in the first instance, before narrowing down to a lower number to enable targeted higher-resolution investigation. The process involved an initial PCR reaction on template DNA with negative controls to ensure an absence of contamination. The subsequent PCR products were then analysed either by human visualisation or via microfluidic amplicon separation analysis. PCR products were only compared if they were part of the same experiment- i.e. with the same batch of reagents and were amplified in the same run.

For assessment by human visualisation, the RAPD products were separated by electrophoresis using 1.5% TAE-agarose gel supplemented with Midori Green Advanced DNA stain (5µl/100mls). 10µl of PCR product was mixed with 2µl of 6X DNA loading dye (ThermoFisher Scientific) before placement into the wells. The gel was run with a GeneRuler 1KB DNA ladder (ThermoFisher Scientific). The gel was then visualised under ultra-violet light. Bands were considered either

“major” or “minor” on the basis of visual interpretation of band intensity within each sample. Samples with a difference of one major band or two minor bands were considered different. All visual assessment were conducted by Philip Mitchelmore (PM).

For assessment via microfluidic amplicon separation analysis, 1µl of each PCR product was run on an Agilent 2100 Bioanalyser using a DNA 7500 chip (Agilent). The resultant fingerprints were subsequently clustered using GelCompar II software (Applied Maths). Pearson’s method was chosen with 2% optimisation.

In order to gain optimal performance, significant experimental modifications were trialled to previously described protocols before the isolates were formally tested.[244, 331] The initial PCR was performed in a 25µl reaction volume containing 1x reaction buffer, 40 pmol of primer, 0.25mM of dNTP, 1 unit of Taq DNA polymerase (New England Biolabs) and 40-80 ng genomic DNA. Primer 272 (AGCGGGCCAA) (Eurofins Scientific) was primarily used, with 208 (ACGGCCGACC) (Eurofins Scientific) also used to confirm appropriate output. PCR was performed on a TProfessionalTRIO Thermocycler (Biometra). The mixture was cycled 4 times through 94°C for 5 minutes; 36°C for 5 minutes and 68°C for 5 minutes then 30 cycles at 94°C for 1 minute; 36°C for 1 minute and 68°C for 2 minutes. A final extension of 68°C for 10 minutes completed the process. This initial process did not lead to any visible bands on electrophoresis. A variety of concentrations of Magnesium Chloride ($MgCl_2$) were trialled with the first evidence of visible bands on electrophoresis from reactions containing 1.5mM $MgCl_2$, in the 208 Primer group. Subsequently the protocol of Hafiane et al was trialled which included the use of Invitrogen reagents instead of New England Biolab.[331] This led to faint bands so a modified cycle programme was set up consisting of 94°C for 2 minutes followed by 5 cycles of 94°C for 1 min; 35°C for 1 min and 72°C for 5 min before 40 cycles of 94°C for 1 min; 35°C for 1 min and 72°C for 2 minutes before a further extension at 72°C for 10 minutes. The modifications included an extended DNA chain extension phase in the first 5 cycles, followed by an additional 5 cycles ontop of what was used in the original protocol. This lead to inconsistently visible bands when using both primers.

After discussions with a collaborator (Prof E. Mahenthiralingam, Cardiff University), a different manufacturers products were used (Qiagen), and an extended cycle. This included the addition of Q Solution, which was part of the

manufacturer's reagent package. The PCR mixture was performed in 25µl containing 1x reaction Buffer, 40 pmol of primer, 0.25 mM of dTNPs, 1.5mM of $MgCl_2$, 1x Q Solution, 1 unit of Taq DNA Polymerase, and 40ng of genomic DNA. The reaction took place under the following cycling conditions: 3 minutes at 94°C; 35 cycles of 94°C for 1 minute, 36°C for 1 minute, and 72°C for 2 minutes; before a final 10 minutes at 72°C. This produced visible bands on electrophoresis for both primers with no evidence of contamination in the negative control lane. Consequently, this protocol was adopted for all further studies.

2.4 Multilocus Sequence Typing (MLST)

The protocol was performed according to the *Pseudomonas aeruginosa* MLST website (http://pubmlst.org/paeruginosa/).[332] The initial phase was amplification of the seven housekeeper genes (see Table 2.1). The amplification mixture contained a total volume of 50µl. This was made up of 10-40ng of DNA; 20 pmol of both forward and reverse primers; 1x PCR Buffer; dNTP solution (10mM each); and 1.25 units of Taq DNA Polymerase (Qiagen). The reaction conditions were as follows: initial denaturation at 96°C for 1 min, followed by 30 cycles of denaturation at 96°C for 1 minute, primer annealing at 55°C for 1 min and extension at 72°C for 1 min. A final extension period of 10 minutes at 72°C then completed the process. To look for evidence of successful amplification, 10µl of the products were then put through gel electrophoresis.

When MLST was performed on DNA extracted directly from sputum there was only intermittent success with the standard protocol for the *nuoD* locus. This persisted even with increasing the number of cycles during PCR and titrating the primer annealing temperature. Consistent PCR product was achieved after using some of the modifications suggested by Van Mansfeld et al, including the addition of Q buffer (Qiagen) to the PCR mixture and using the primers described in their work for *nuoD* (forward primer: GGGACATGTACGGCATCACCT and reverse primer: GCGCAGGATGCTGTTCTTCA).[333]

Subsequent sequencing of the products was performed commercially by Eurofins Scientific with the sequencing primers listed in Table 2.1. This was initially performed on PCR products purified by Eurofins Scientific as part of the commercial service. However, due to a disappointing return of data, we subsequently conducted purification of PCR products ourselves before sending

those purified PCR products for sequencing. This was performed with the GeneJET Gel Extraction kit (ThermoFisher Scientific), using the manufacturer's protocol for simple PCR clean-up. The DNA was checked for quantity and quality by a NanoDrop^{TM} 1000 Spectrophotometer (ThermoFisher Scientific).

Following the return of sequencing data, the forward and reverse reads were aligned using Vector NTI software (ThermoFisher Scientific). The consensus sequence was then interrogated via the MLST database website and the allele number recorded. This was repeated for all 7 loci and our sequences were used to search the database to identify corresponding loci and ST numbers.

Locus	Gene Product	Amplification Primers		Sequencing Primers	
acsA	Acetyl coenzyme A synthetase	Forward	ACCTGGTGTACGCCTCGCTGAC	Forward	GCCACACCTACATCGTCTAT
		Reverse	GACATAGATGCCCTGCCCCTTGAT	Reverse	AGGTTGCCGAGGTTGTCCAC
aroE	Shikimate dehydrogenase	Forward	TGGGGCTATGACTGGAAACC	Forward	ATGTCACCGTGCCGTTCAAG
		Reverse	TAACCCGGTTTTGTGATTCCTACA	Reverse	TGAAGGCAGTCGGTTCCTTG
guaA	GMP synthase	Forward	CGGCCTCGACGTGTGGATGA	Forward	AGGTCGGTTCCTCCAAGGTC
		Reverse	GAACGCCTGGCTGGTCTTGTGGTA	Reverse	GACGTTGTGGTGCGACTTGA
mutL	DNA mismatch repair protein	Forward	CCAGATCGCCGCCGGTGAGGTG	Forward	AGAAGACCGAGTTCGACCAT
		Reverse	CAGGGTGCCATAGAGGAAGTC	Reverse	GGTGCCATAGAGGAAGTCAT
nuoD	NADH dehydrogenase I chain C, D	Forward	ACCGCCACCCGTACTG	Forward	ACGGCGAGAACGAGGACTAC
		Reverse	TCTCGCCCATCTTGACCA	Reverse	TGGCGGTCGGTGAAGGTGAA
ppsA	Phosphoenolpyruvate synthase	Forward	GGTCGCTCGGTCAAGGTAGTGG	Forward	GGTGACGACGGCAAGCTGTA
		Reverse	GGGTTCTCTTCTTCCGGCTCGTAG	Reverse	GTATCGCCTTCGGCACAGGA
trpE	Anthralite synthetase component I	Forward	GCGGCCCAGGGTCGTGAG	Forward	TTCAACTTCGGCGACTTCCA
		Reverse	CCCGGCGCTTGTTGATGGTT	Reverse	GGTGTCCATGTTGCCGTTCC

Table 2.1 Primers used for the MLST scheme

.

2.5 Whole Genome Sequencing (WGS)

The WGS included in this thesis was performed by the Exeter Sequencing Service (ESS) (including all the bioinformatics analysis) with DNA extracted by PM. For this, the data was generated from the Illumina MiSeq system using 2 x 300bp read lengths. The subsequent analysis was carried out using the MRC CLIMB infrastructure with a Virtual Machine pre-installed with the Nullarbor package (https://github.com/tseemann/nullarbor).[334] This package then generated core genome variants, and determined MLST profiles, resistomes, SNP distance matrices and a pan-genome report. It also trimmed low quality and adaptors present in reads using Trimmomatic, used Kraken to assign reads to taxonomic groups and SPAdes to assemble genomes (using the --accurate option).[335-337] Annotation was then performed using Prokka and variants called with Snippy (https://github.com/tseemann/snippy).[338]

Publicly-available genomes were also included in analysis by the Nullarbor package. Many of these were not created via Illumina sequencing and in order for these to be incorporated into the Nullarbor package, the genomes were processed using the wgsim package (v 0.3.2) to generate simulated short paired-end reads for re-assembly. The following parameters were used to generate 2x300bp reads with zero changes with respect to the reference genome: -e 0.000000 -d 600 -1 300 -2 300 -r 0.000000 –R 0.0000 -X 0.0000 -h -s 0 -N 1100000 -A 0.000.

Additional assessment of WGS data was performed for *in silico* prediction of hypermutators. Reads were quality and adapter trimmed using fastq-mcf using parameters -q 20 with skew settings switched off. Reads were aligned using bwa mem to the reference genome (*Pseudomonas aeruginosa* PAO1; NC_002516) and sorted and converted to BAM using samtools. The mpileup component of samtools was used to call variants and perform local re-alignment of sequences. The BCF formatted files were converted to VCF format using bcftools and filtered to exclude sites with coverage < 10 or variant quality < 60. Data was analysed using the Zeus computational infrastructure at the University of Exeter. From the list of SNPs and Indels generated by this method, we identified those occurring within seven genes known to be implicated in proofreading and/or DNA repair,

namely *mutS* (PA3620), *mutL* (PA4946), *mutY* (PA0357), *mutM* (PA5147), *dnaQ* (PA1816), *mutT* (PA4400) and *uvrD* (PA5443). SNAP2 and PROVEAN were then used to predict whether the observed SNPs and Indels would be neutral or deleterious with regards to protein function.[339, 340] If the prediction was for a deleterious effect in one of these genes, then the isolate was predicted to be a hypermutator.

2.6 Hypermutator Assay for *Pseudomonas aeruginosa*

Hypermutator assays were performed on PA isolates as previously described by assessing spontaneous resistance to rifampicin.[158] Stored isolates of PA were re-cultured on Mueller-Hinton agar (MHA) plates at 37°C. Overnight cultures were then set up in triplicates, each inoculated with 3-5 colonies in 20mls of Mueller-Hinton broth (MHB) (37°C, 200 RPM). The cells from these overnight cultures were then pelleted by centrifugation at 3,000 RPM for 5 minutes and re-suspended in 1ml of MHB. Serial dilutions with Phosphate Buffered Saline were made from this neat stock ranging from 10^{-1} to 10^{-8}. Pre-prepared plates of MHA, and MHA with 300 µg rifampicin per ml (MHA/Rif), were subsequently inoculated. The neat, 10^{-1} and 10^{-2} solutions were plated onto the MHA/Rif plates, and the 10^{-6}, 10^{-7} and 10^{-8} solutions were plated onto MHA plates. The plates were incubated for 36 hours at 37°C and the colonies subsequently counted to obtain a mutant frequency. The counts were obtained from the most appropriate MHA and MHA/Rif plates as defined by the plate with the highest number of countable colonies. A mean was calculated from the triplicates for each sample. All batches included isolates from the same PAO1 stock for inter-batch comparison. Examples of mutators were streaked onto rifampicin plates in order to check stability of their resistance to rifampicin.

2.7 Ribosomal Intergenic Spacer Analysis (RISA)

RISA was performed as previously described.[133] The initial step involved a PCR step to amplify the intergenic spacer regions. The 25µl reaction mixture contained 1µl of template DNA (approximately 20 to 40 ng), 10pmol of each primer, 1x PCR buffer, 0.25mM of each dNTP, 1x Q solution, 1 unit of Taq DNA polymerase (Qiagen) and an additional 1.5mM of $MgCl_2$.The primers were 1406F (TGYACACACCGCCCGT) and 23SR (GGGTTBCCCCATTCRG), both from Eurofins Scientific. The reaction conditions were as follows: 95^{o}C for 5 min; 34

cycles of 95°C for 30 sec, 55°C for 30 sec and 72°C for 60 sec followed by 72°C for 5 min.

Following the PCR reaction, products were visualised by gel electrophoresis in the same manner that RAPD PCR products had been (see Section 2.3). After confirmation of PCR product, further analysis was performed with micro-fluidic amplicon separation on the Agilent Bioanalyser as had previously been done for RAPD PCR products except with 2 µl of product rather than just 1 µl. Analysis of this data was again performed with the use of GelCompar II software (Applied Maths). The analysis included a dendrogram created by Pearson's correlation co-efficient, an estimate of richness based on counting bands, and estimations of Shannon's diversity and evenness based on the band concentrations as a surrogate of abundance.

2.8 16S rRNA Sequencing

The 16S rRNA sequencing process was conducted by the ESS on DNA extracted by PM according to the protocol in section 2.2.3. Negative controls from each DNA extraction kit and DES used to pre-dilute samples were also submitted for sequencing in order to detect suspected contaminants. This was conducted as per the 16S rRNA gene amplification Illumina MiSeq protocol with primers used to amplify the V1-V2 region. Subsequent bioinformatics analysis was performed by the ESS via the Dada2 software package.[341] This is a specific software package for the modelling and correction of Illumina sequencing errors. It also gives higher resolution to the standard OTU output. Removal of chimeras occur after trimming and denoising of data.

Part B

Chapter 3- Genotyping of *Pseudomonas aeruginosa* in an unsegregated bronchiectasis cohort sharing hospital facilities with a cystic fibrosis cohort

3.1 Abstract

Introduction: Whilst PA cross-infection is well-documented amongst CF patients, the equivalent risk amongst NCFB patients is unclear. There is some evidence of rare cross-infection with PA in a NCFB cohort managed in a centre without CF services. There is no evidence with regards to the risk of cross-infection in a centre where both NCFB and CF patients are managed

Methods: We performed an epidemiological analysis of PA within a single centre that manages an unsegregated NCFB cohort alongside a segregated CF cohort. This was performed through the prospective collection of sputum samples from patients who had previously isolated PA. Ten isolates were analysed per patient by RAPD and unique profiles were further analysed by MLST. Shared strains as defined by MLST underwent WGS with comparison to publicly-available unconnected genomes of the same strain. Further comparison with non-respiratory clinical isolates of PA was performed by MLST.

Results: Positive isolates of PA were obtained from 46 NCFB patients and 22 CF patients. A further 76 isolates were obtained from non-respiratory clinical isolates. We found no evidence of cross-infection between the two cohorts or within the segregated CF cohort with the exception of a pair of siblings. However, within the unsegregated NCFB cohort, evidence of cross-infection was found between three patients. Multiple other shared strains were found, but these were believed to be due to environmental acquisition.

Conclusion: This epidemiological review has shown evidence of uncommon cross-infection between NCFB patients. At present this level of evidence does not suggest that the implementation of segregation policies would be appropriate, though longitudinal surveillance should be performed.

3.2 Introduction

PA is a highly significant pathogen in chronic suppurative lung diseases. Its significance in CF is extensively described, where it is the most frequently cultured pathogen in adults with CF, and is a predictor of mortality.[127, 342] It is less well studied in other suppurative states such as NCFB, but is again associated with significant morbidity and mortality.[43, 44] The Bronchiectasis Severity Index (BSI) illustrates its significance by giving the same score for colonisation with PA as an FEV_1<30%, or a Medical Research Council (MRC) dyspnoea score of 5.[40]

It is well established that PA cross-infection occurs between CF patients having been observed in a variety of settings including siblings, holiday camps and CF units.[187, 190, 192] Whilst not universal, some transmissible strains appear to have greater clinical impact than environmentally acquired strains.[196, 197] Studies highlighting cross-infection with pathogens such as PA have led to the practice of segregation of CF patients.[205] NCFB and other cohorts affected by PA tend not to be segregated. The need to review this has been highlighted as a research priority [37].

NCFB was considered an “orphan” disease and unsurprisingly there has been little research into cross-infection risk.[343] One UK study concluded that cross-infection was rare.[131] However, in that study, the NCFB patients were managed in a different hospital to the local CF cohort. Similarly, a recent multi-centre study highlighted the potential for PA cross-infection through identifying closely related isolates by WGS in the same clinical centre.[211] This was again exclusively focused on NCFB cohorts. In many hospitals, including ours, CF and NCFB patients share the same facilities and healthcare professionals. Therefore, in the study described herein, we have conducted an epidemiological review of strains of PA in our unsegregated NCFB cohort and our segregated CF cohort to assess cross-infection within and between cohorts. In addition, parallel analysis of local non-respiratory PA isolates has allowed us to compare CF and NCFB PA isolates with those present in the wider hospital population. This work incorporates three different genotyping techniques (namely RAPD, MLST and WGS) and highlights the strengths and weaknesses of these in this setting. The true definition of a “strain” is debatable. For the purposes of this work, it refers to an ST-type as

described by the MLST scheme. The ST-type can be determined either by MLST or by WGS.

3.3 Methods

3.3.1 Study Population

All patients were recruited at the Royal Devon & Exeter Hospital, a teaching hospital in the South-West of England with a catchment area of approximately 400,000. Isolates of PA were obtained from three cohorts. The first cohort were patients with a documented diagnosis of NCFB ("NCFB Cohort"), followed by collections from patients with CF ("CF Cohort"), and from clinical samples not from a respiratory source ("Non-Resp Cohort").

3.3.2 NCFB Cohort and CF Cohort

Patients in these cohorts were opportunistically recruited from adult services only. For inclusion, patients required the confirmed diagnosis of NCFB or CF, and to have previously grown PA. Maintenance anti-pseudomonal therapy was not an exclusion for recruitment. Patient data was collected including details on previous hospital and departmental attendance. If PA was not identified from the sample, then a future sample could be included if the patient submitted a further sample for clinical assessment within the recruitment window.

3.3.3 Sputum Processing

After consent, spontaneous sputum samples were submitted to the hospital's microbiology services and cultured in the usual manner (as described in Chapter 2). When PA was identified, 10 representative colonies were picked per sputum sample (based on colony morphologies) and stored on Microbank™ microbial storage beads (Pro-Lab Diagnostics) at -80°C. Initial identification of PA was based on colony morphology and a positive oxidase test, and was subsequently confirmed by VITEK MS MALDI-TOF MS (Biomerieux).

3.3.4 Non-Resp Cohort

Isolates of PA were obtained prospectively from both community and hospital samples from different patients. These were samples submitted to the local microbiology department as part of standard practice and confirmation of PA identity was performed as described above. The only detail obtained was the

origin of the sample (for example: urine sample from General Practice). One colony was picked per patient and stored on beads at -80°C.

3.3.5 DNA Extraction

Following re-growth of isolates on Luria broth (LB) agar, single colonies were inoculated into LB broth and incubated overnight at 37°C with shaking. Subsequently, DNA was extracted from bacterial cell pellets using the GeneJET Genomic DNA Purification kit (ThermoFisher Scientific) as per the relevant manufacturer's instructions. The extracted DNA was checked for quantity and quality by the Nanodrop^{TM} 1000 Spectrophotometer (ThermoFisher Scientific) and stored at -20°C.

3.3.6 Random Amplified Polymorphic DNA (RAPD)

RAPD was performed on each isolate from the NCFB Cohort and the CF cohort, using a modified version of the previously described protocol.[244] In brief, PCR was carried out in a 25µl volume containing 1x PCR Buffer, 1U Taq Polymerase (Qiagen), 0.25mM (each) deoxynucleoside triphosphate, 1x Q-Solution, 1.5 mM $MgCl_2$, 40 pmol primer, and 40 ng of genomic DNA. Initially, RAPD was performed on all isolates using primer 272, and re-checked using primer 208. Reaction mixtures were run on a TProfessionalTRIO Thermocycler (Biometra) as follows: 3 minutes at 94°C; 35 cycles of 94°C for 1 minute, 36°C for 1 minute and 72°C for 2 minutes; followed by 72°C for 10 minutes.

In the first instance, RAPD products were subjected to conventional agarose gel electrophoresis (1.5% (w/v) agarose) and visualisation via a Gel Doc XR (Bio-Rad) to enable identification of unique profiles. Batches were performed on one patient at a time to facilitate the visualisation of all unique RAPD profiles within the 10 isolates per patient. The criteria for visually defining a unique RAPD profile was (a) the presence/absence of a major band, (b) the presence/absence of two minor bands, or (c) a difference in one band being major/minor and the presence/absence of another band. From each NCFB patient, a representative of each unique profile (based on visual inspection of the two runs) was taken forward for repeat RAPD analysis within a single batch encompassing representatives from all patients. For the repeat RAPD analysis, RAPD profiles were analysed via microfluidic amplicon separation (Agilent 2100 Bioanalyser),

with cluster analysis subsequently performed using Gelcompar II software (Applied Maths), with Pearson's Method and 2% optimisation.

3.3.7 Multilocus Sequence Typing (MLST)

MLST was performed on the same NCFB isolates that underwent repeat RAPD analysis via the Bioanalyser, and on unique profiles from the CF isolates as defined by visual inspection. This was to provide an epidemiological analysis using a distinct method and to compare our isolates to a global database. This was performed as previously described, with allele sequences being compared between patients to identify shared strains, and also to the MLST database (www.pubmlst.org/paeruginosa) to identify matches to previously documented strains.[213] A modified "cascade" version of MLST was then performed on isolates from the Non-Resp Cohort to look for the shared strains identified within the respiratory cohorts. Analysis of the MLST alleles was performed sequentially, starting with the most discriminatory alleles in the original description of the scheme.[213] The cascade process continued until either all seven alleles were analysed or until the isolate could not match at least six of the seven alleles of a shared strain from the respiratory cohorts.

3.3.8 Whole Genome Sequencing (WGS)

WGS was performed on the shared strains from the respiratory cohorts. One isolate was chosen per patient with the exception of the most abundant strain for which three were chosen per patient. DNA was quantified by Qubit (Thermo Fisher Scientific) before submission to the ESS who undertook the process which included the running of the Bioinformatic pipeline. WGS was performed on the Illumina MiSeq platform. Bioinformatic analysis was performed via the Nullarbor pipeline. Further details of this process are described in Chapter 2.

In addition to our samples, relevant reference genomes obtained from the *Pseudomonas* Genome Database were included in our Nullarbor pipeline analysis to enable isolates to be compared to clonally-unconnected representatives of the same sequence type.[344] Three genomes were selected for each Sequence Type (ST) that had undergone WGS. This was not possible for ST564 for which no publicly-available genomes were available. Relevant genomes were selected on the basis of them being either complete or with a high Contig N50 (Table 3.1).

ST type	Isolate identifier	Accession Number	Origin	Country
17	BL02	SAMN02360715	CLIN (Vitreous fluid)	USA
	C20	SAMN02360744	ENV	Unknown
	C23	SAMN02360745	ENV	Unknown
27	AZPAE14980	SAMN03105677	CLIN (Intra-abdominal)	USA
	BWHPSA011	SAMN02360683	CLIN (Tissue middle turbinate)	USA
	BWHPSA022	SAMN02360694	CLIN (Sputum)	USA
146	LES431	SAMN02641592	CLIN (Parent of CF patient)	UK
	LESB58	SAMEA1705916	CLIN (CF isolate)	UK
	AZPAE13757	SAMN03105416	CLIN (Respiratory Tract)	Canada
235	BTP032	SAMN03787333	CLIN	USA
	JJ692	SAMN02360667	CLIN (UTI)	USA
	NCGM2.S1	SAMD00061003	CLIN (UTI)	Japan
252	AZPAE12420	SAMN03105411	CLIN (CF isolate)	USA
	AZPAE15012	SAMN03105709	CLIN (Intra-abdominal)	Germany
	BWHPSA028	SAMN02360700	CLIN (Sputum)	USA
253	BL16	SAMN02360729	CLIN (Corneal scraping)	USA
	BWH058	SAMN02402442	CLIN	Unknown
	UCBPP-PA14	SAMN02603591	CLIN (Burn wound)	Unknown
274	AZPAE14926	SAMN03105624	CLIN (UTI)	Brazil
	AZPAE14981	SAMN03105678	CLIN (UTI)	France
	BWHPSA040	SAMN02360704	CLIN (Sputum)	USA
395	3581	SAMN02584694	CLIN	Unknown
	BWH059	SAMN02402443	CLIN	Unknown
	BWHPSA045	SAMN02360709	CLIN (Sputum)	USA

Table 3.1 Publicly-available genomes used for comparison with shared strains from the NCFB and CF cohorts Where available, relevant information is provided on the origin of isolates. CLIN, Clinical; ENV, Environmental; UTI, Urinary Tract Infection. Based on the available information, there is no evidence that any of the isolates above are directly linked to our patient cohorts.

3.3.9 *In silico* Prediction of Hypermutators

An *in silico* prediction model was performed on the WGS data as described in Chapter 2. SNPs and insertion-deletion events (Indels) within seven genes implicated in proofreading and/or DNA repair were identified. These were *mutS* (PA3620), *mutL* (PA4946), *mutY* (PA0357), *mutM* (PA5147), *dnaQ* (PA1816), *mutT* (PA4400) and *uvrD* (PA5443).[158, 345] Two software tools (SNAP2 and PROVEAN), were used to predict the impact of SNPs and Indels on protein function.[339, 340]

3.4 Results

3.4.1 Cohort Demographics

In a twelve month recruitment window 63 NCFB patients were recruited of whom 46 produced samples positive for PA. Of those with positive samples, 89% were under the care of the two Respiratory consultants who care for all the adult CF patients in the RD&E Hospital. In a five month recruitment period, 32 CF patients were recruited. The demographics of these patients are shown in Table 3.2

	NCFB	**CF**
Recruitment Period (Month/Year-Month/Year)	07/14-06/15	09/15-01/16
Subjects	46	22
Median Age (in years)	69	27.5
IQR for age	65.8-76	24.5-40.3
Male	10 (21.7%)	10 (45.5%)
Time since 1st PA isolate		
- **Less than 1 year**	12 (26.1%)	0
- **1-5 years**	11 (23.9%)	9 (40.9%)
- **5-9 years**	11 (23.9%)	4 (18.2%)
- **More than 9 years**	12 (26.1%)	9 (40.9%)
Co-pathogens		
- ***Staphylococcus aureus***	1 (2.2%)	9 (40.9%)
- ***Aspergillus fumigatus***	0	1 (4.5%)
- ***Exophiala* species**	0	2 (9.1%)
- **Other**	7 (15.2%)	3 (13.6%)
Antibiotic therapy		
- **Current azithromycin use**	18 (39.1%)	12 (54.5%)
- **Current inhaled anti-pseudomonal use**	16 (34.8%)	18 (81.8%)
- **Neither azithromycin nor inhaled anti-pseudomonal use**	17 (37%)	1 (4.5%)

Table 3.2. Patient demographics for PA-positive NCFB and CF cohorts

Between September and November 2015, 76 isolates of PA were collected for our Non-Resp Cohort. Their origins are displayed in Table 3.3

Source	Community	Hospital[a]	Total
Genito/Urinary	25	16	41
Wound Site/Skin	1	15	16
ENT	1	8	9
Faeces	4	5	9
Unspecified Pus	0	1	1
Total	31	45	76

Table 3.3. Origin of samples for the non-respiratory cohort

[a]Samples were considered as "Hospital" origin if they were collected from a patient attending out-patient services or during a ward admission.

3.4.2 RAPD/MLST Genotyping Pipeline- NCFB

From the 46 samples which grew PA, 10 representative colonies were stored for all but one sample, for which 9 were stored. This resulted in 459 isolates being available for interrogation. Following initial RAPD analysis by the visualisation of products by gel electrophoresis, 59 isolates were identified as representatives of unique RAPD profiles. These were taken forward for repeat RAPD analysis via microfluidic amplicon separation on a Bioanalyser. This panel of 59 isolates included a minimum of one isolate per patient, and multiple isolates from 8 patients.

The Bioanalyser analysis of the RAPD output revealed the presence of 7 clusters containing samples from more than one patient with a cut-off of 90% for clonality, suggesting that shared strains were present in the cohort (Figure 3.1). Of the 8 patients who were identified as having multiple unique RAPD profiles by visual inspection, 5 (63%) did not appear to have multiple RAPD profiles as defined by a 90% cut-off.

The panel of 59 isolates subsequently underwent MLST analysis. By this method, 35 different concatenated sequences were identified with 5 of these being shared between patients, and 25 had exact matches on the MLST database. Two

patients displayed evidence of multiple strains. The most prevalent ST type was ST17 which predominated in the largest cluster on RAPD analysis. This is better known as Clone C. By MLST, this was found in 8/46 patients (17%). The other shared ST types were found in either two or three different patients (Table 3.4)

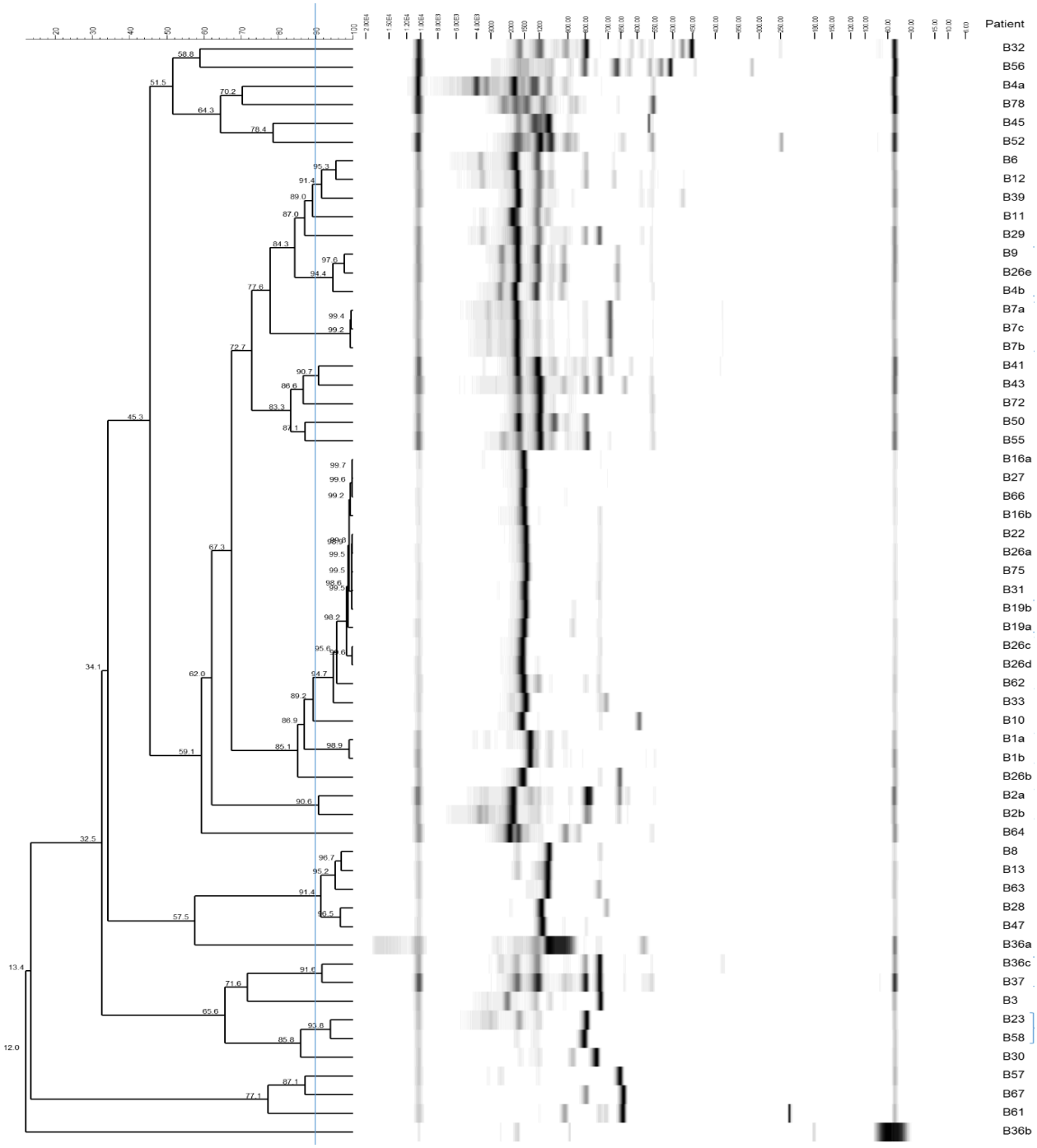

Fig 3.1. Dendrogram of RAPD profiles of NCFB PA isolates, derived using microfluidic amplicon separation (Agilent 2100 Bioanalyser). The dendrogram was produced using Gelcompar II (Applied Maths) with Pearson's method and 2% optimisation. A 90% cut-off (denoted by the vertical blue line) was used to identify clusters.

3.4.3 RAPD/MLST Genotyping Pipeline- CF

From the 22 samples from which PA was grown, 10 representative colonies were stored for all but one sample, for which 9 were stored. This resulted in 221 isolates being available for interrogation. Following initial RAPD analysis by the visualisation of products by gel electrophoresis, 26 isolates were identified as unique profiles to be taken forward for further investigation by MLST. This panel of 26 isolates included a minimum of one isolate per patient, and multiple isolates from 3 patients.

By MLST, 20 ST types were identified of which two were shared between multiple patients. ST146 (the Liverpool Epidemic Strain, (LES)) was shared between a pair of twins, and ST27 was shared by three patients. One patient had evidence of multiple strains. When reviewed alongside the NCFB data, five ST types were seen in both the NCFB and CF cohorts (ST27, ST235, ST274 and ST395).

3.4.4 Cascade MLST for Non-resp Isolates

The cascade method of MLST resulted in 28 of the 76 samples being completely assessed. Of the 9 shared strains evident within the NCFB and/or CF cohorts, 5 were also found in the non-respiratory samples (Table 3.4). Of these, the most prominent in the non-respiratory samples were ST253 (better known as PA14) and ST17 (Clone C).

MLST type	Alias	NCFB (n=46)	CF (n=22)	Non-resp (n=76)	Total (n=144)
ST17	Clone C	8 (17%)		7 (9%)	15 (10%)
ST27		1 (2%)	3 (14%)	5 (7%)	9 (6%)
ST146	LES		2 (9%)		2 (1%)
ST235		1 (2%)	1 (5%)		2 (1%)
ST252		3 (7%)	1 (5%)		4 (3%)
ST253	PA14	2 (4%)		10 (13%)	12 (8%)
ST274		1 (2%)	1 (5%)	2 (3%)	4 (3%)
ST395		3 (7%)	1 (5%)	3 (4%)	7 (5%)
ST564		3 (7%)			3 (2%)

Table 3.4 Shared strains of PA identified within the respiratory (CF and NCFB) and non-respiratory cohorts, as defined by MLST One NCFB patient was co-infected with ST17 and ST564. Isolates in the non-respiratory cohort originated from genitourinary, wound, ENT and faecal samples from community and hospital investigation.

3.4.5 Whole Genome Sequencing of Shared Strains

WGS was focused on the ST types which were shared among the NCFB and/or CF patients. WGS was performed on three isolates of ST17 from each of the 8 patients who harboured it, and one isolate per patient of the other 8 shared strains. The primary measure was the pairwise comparison of SNP difference between isolates. This ranged from 4-3925. The analysis also included three publicly-available genomes per ST type, with the exception of ST564 for which there is not one. This was performed to enable comparison to unconnected representatives of the same ST type. These additional pairwise comparisons are

illustrated in Figure 3.2 and demonstrate that the CF and NCFB isolates belonging to ST17, ST27, ST235, ST252, ST253, ST274 and ST395 are as divergent from each other (Figure 3.2; circles) as they are from unconnected representatives of the same sequence type (Figure 3.2; grey crosses). This was not the case with ST146 where the pairwise comparison between two of our isolates showed far less divergence from each other than the unconnected genomes. These two isolates came from siblings with CF who have long been colonised with PA and were previously known to harbour ST146 (more commonly known as LES). The analysis also showed that the ST17 isolates revealed significantly greater ST17 diversity between patients than within patients. This gives further confidence in the interpretation that this does not represent likely cross-infection with this strain.

Apart from ST146, the significant outlier in this analysis is ST564. The isolates with this ST type were near-identical and differed by only 4-12 SNPs. ST564 was found in three bronchiectasis patients and not in either the CF cohort or among the non-respiratory isolates. Two of the three patients were co-infected with another ST type. Via a review of hospital attendance data, it was noted that two of the three patients shared a waiting area and lung function room approximately 17 months prior to recruitment. This potential cross-infection event did not coincide with a clear change in PA culture status as one of the patients intermittently isolated PA before and after this event, while the other patient had evidence of multiple PA strains (and therefore superinfection may have occurred). We were unable to identify a potential cross-infection event involving the third patient, but a difference of only 4 SNPs strongly supports cross-infection.

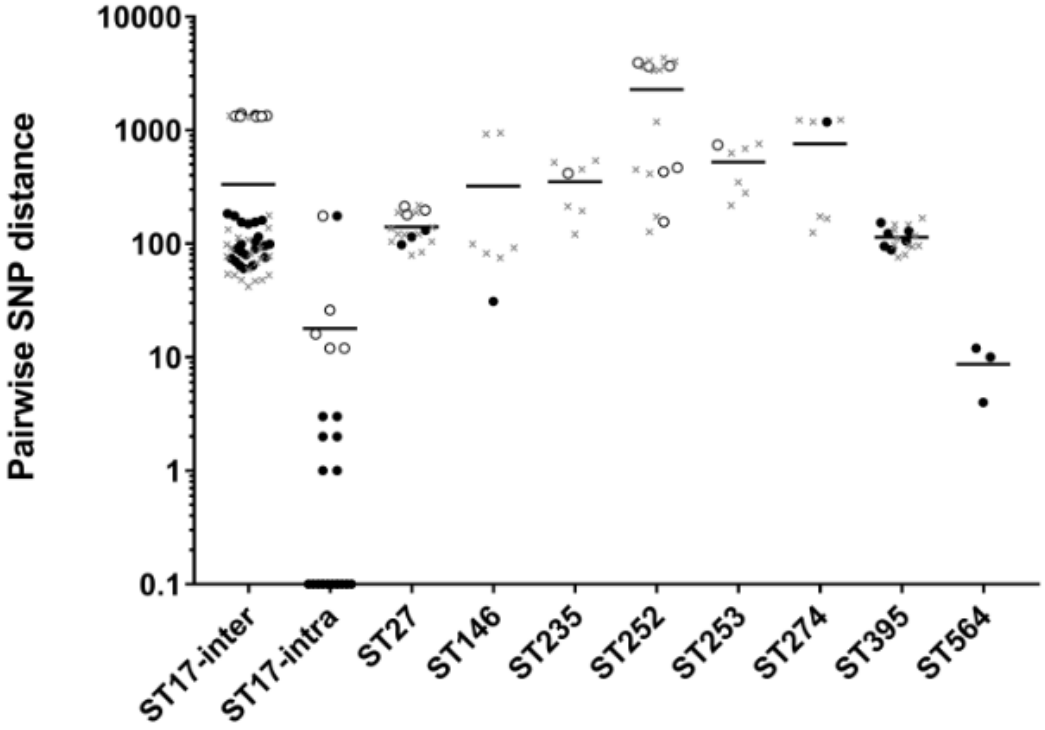

Figure 3.2. Genetic diversity within PA isolates, as defined by whole genome sequencing. The number of single nucleotide polymorphisms (SNPs) was calculated across the core genome of all sequenced isolates. Each data point represents a pairwise comparison within each ST, with the bar representing the mean. Circles represent pairwise comparisons that are exclusively between PA isolates from our own respiratory cohort (CF or NCFB), with the open circles representing those comparisons in which at least one isolate is a predicted hypermutator. The grey crosses represent pairwise comparisons in which one isolate is from our respiratory cohort and the other is an unconnected representative of the same sequence type (using publicly-available genomes). For ST17, SNP numbers are shown that reflect the diversity observed between patients (ST17-inter) and within individual patients (ST17-intra; based on sequencing of three isolates per patient).

3.4.6 Hypermutators by *in silico* Prediction

The *in silico* prediction model identified nine patients in whom isolates had putative hypermutator status (see Table 3.5). No deleterious mutations were found in *uvrD* or *mutT*. The pairwise comparisons which included at least one putative hypermutator are shown in Figure 3.2 by open circles, and these

comparison have a significantly greater difference than those which do not include at least one putative hypermutator (see Figure 3.3).

		Gene and nature of mutation				
Isolate[a]	**ST[b]**	***mutS***	***mutL***	***mutY***	***mutM***	***dnaQ***
PIB16	ST17	L52P	Q52X		L342P	
PIB26	ST17	Frameshift				
PIB45	ST27	V264E				
PIB01	ST235	ΔL541-S544				
PIB23	ST252					R33H
PIB58	ST252		H469R			R33H
PIB67	ST252					R33H
PIC30	ST252		Frameshift			R33H
PIB63	ST253	Frameshift		H72R		

Table 3.5 Prediction of hypermutators based on the identification of deleterious mutations in genes conferring DNA proof-reading and mismatch repair functions All of the indicated amino acid substitutions are predicted to be deleterious by both SNAP2 and PROVEAN, whilst the frameshift mutations each cause premature truncation of the gene product. [a] PIB isolates are from NCFB patients, whilst PIC isolates are from CF patients. [b] Sequence Type, as defined by Multilocus Sequence Typing (MLST).

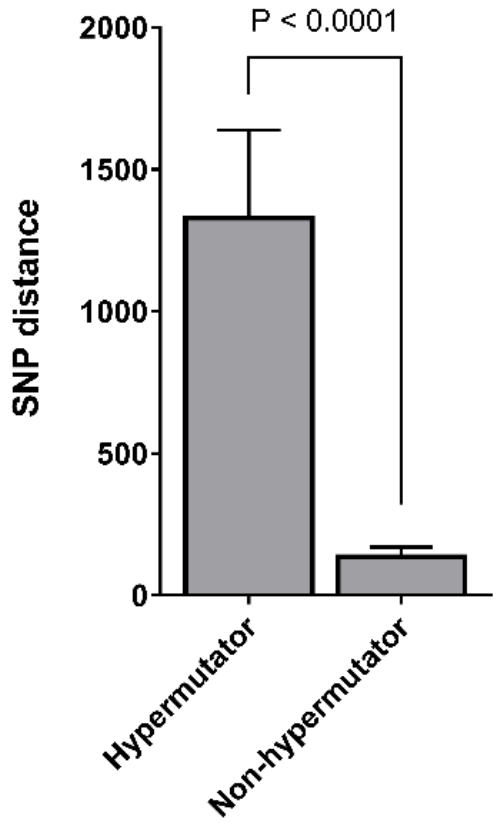

Figure 3.3 SNPs distances in pairwise comparison with and without predicted hypermutators Predicted hypermutable PA isolates exhibited significantly elevated levels of genetic divergence (SNP distance) relative to predicted non-hypermutable PA.

3.5 Discussion

This study shows that while cross-infection with PA is highly likely to have occurred in our NCFB cohorts, it does not appear to be common. The possibility of cross-infection was raised by the findings of shared strains by three molecular techniques. The declaration of a transmissible strain however requires more than just the confirmation of shared strains and deserves particular consideration and judgement.

Current technology does not allow certainty with regards to cross-infection. In addition, the assessment of sputum may not be representative of all regions of the lungs. Consequently, a transmissible strain may be missed. Despite these issues, investigators can accumulate evidence to comment on the likelihood of

cross-infection. This may include: the proportion of shared strains in a cohort; how genetically related isolates are; the plausibility of transmission episodes occurring; and a knowledge of abundant strains and interventions affecting new acquisition.

Cross-infection is plausible in NCFB. Growing cohorts of patients with NCFB, the establishment of specialist clinics for NCFB, and appropriate referral to pulmonary rehabilitation courses enhance this risk.[6, 42] The LES, a known transmissible strain in CF cohorts, has been identified in the air where patients have been present several hours before, and other work has shown the potential for viable PA to remain airborne for at least 45 minutes.[178, 184] Consequently, aerosolisation is a plausible route and therefore patients may not even have had contemporaneous presence for transmission to occur. Whilst basic infection control measures should ideally prevent the risk of direct and indirect contact transmission being an issue, it would be naïve to assume these are always followed. In addition, hospitals contain non-clinical areas such as shops and canteens where infection control guidelines cannot be enforced.

Previously we have had significant evidence for cross-infection in CF but not in NCFB, and the reasons why need to be considered. It should be noted that there have been multiple studies where cross-infection has not appeared to be a significant issue in CF cohorts and some evidence for cross-infection has come from episodes of high exposure including holiday camps and between siblings.[187, 202, 203, 346] CF cohorts are also more likely to be seen at more frequent intervals in disease-specific clinics, and the duration of appointments may be longer due to multidisciplinary reviews. Consequently the period of time that patients may be at risk is likely to be greater. In addition, the burden of the pathogen is likely to be higher in CF cohorts in whom a considerably higher proportion are chronically colonised with PA.[107, 130, 347] NCFB patients have therefore traditionally not been exposed to as much of a bacterial burden. It may also be that the relative lack of investigation of NCFB cohorts for shared strains has not revealed significant cross-infection. However, there have been two recent publications suggesting the possibility of the occurrence of cross-infection.[131, 211]

The strong evidence of cross-infection in this study comes from isolates from three patients which were identified as ST564 by both WGS and MLST, and

clustered together by RAPD. Between the isolates from these three patients there was considerably less SNPs difference by pairwise comparison than observed among isolates from other patients. This strain was also absent from the other cohorts. ST564 only has 3 entries on the global database but these submissions are from three separate countries and two continents. It is not a well-described strain in the literature. There was an identifiable occasion when two of these patients (B9 and B26) attended the same clinic and would have used the same waiting and lung function rooms. This potential cross-infection event did not coincide with a clear change in PA culture status as one of the patients intermittently isolated PA before and after this event whilst the other patient had evidence of multiple PA strains (and therefore super-infection may have occurred). Whilst we were unable to identify potential cross-infection event(s) involving the third ST564-infected patient (who also carried multiple strains), a difference of only 4 SNPs strongly supports cross-infection. Interactions may have occurred in or outside the hospital that are not apparent via the review of clinical notes. Furthermore, we believe ST564 acquisition from a common environmental source is highly unlikely due to its absence from other cohorts, and it being so sparsely described previously.

Whilst the lack of SNPs difference creates a strong argument, interpretation of other ST types is not clear cut. There has been some longitudinal work showing mutation rates of 1-5.5 SNPs/year which may help to interpret cross-infection risk.[154, 348-350] Our example of two patients sharing facilities and within 18 months having isolates only 10 SNPs different, does therefore seem plausible. When considering the other strains in this study, it must be noted that other work has demonstrated that hypermutators may generate rates of change in the region of 100 SNPs/year.[351] We have provided evidence that hypermutator status is a real issue in interpretation of this kind of data, by showing significantly greater genetic diversity by pairwise comparison of isolates when predicted hypermutators are involved. Also of note, paired samples from the same patient may have significant differences in SNPs.[352] Consequently, large numbers of SNPs difference cannot provide absolute certainty that cross-infection has not occurred and a strict cut-off value cannot be used to confirm cross-infection. Therefore, there may have been episodes when cross-infection has occurred but WGS analysis does not make it as apparent. In order to gain greater insight into

the expected level of diversity of unconnected isolates of the same strain, we incorporated publicly-available genomes of the relevant ST types. With these points in mind, it is important to consider the other shared strains in detail.

The most abundant strain in the NCFB cohort was Clone C (ST17). This is a well-described abundant strain which has high rates in clinical and environmental settings.[218] Multiple studies in both CF and mixed cohorts have consistently demonstrated its presence.[221, 223, 224] Strikingly, we did not identify Clone C within our CF cohort. Among the NCFB patients who have this strain, there were time points when cross-infection could have occurred. The WGS data however suggested greater between-patient diversity than within-patient diversity and large SNP differences. In those in whom hypermutator status was predicted, the pairwise comparisons showed particularly marked genetic diversity (see Figure 3.2). Importantly, the comparison with the unconnected genomes which were incorporated showed similar diversity. Also of note, Clone C was shown to be present in non-respiratory samples from both the hospital and community setting. With this accumulation of evidence, we believe that Clone C is likely to be a prominent strain in the local environment and that cross-infection with this strain has probably not taken place. The only other strain which was shared among NCFB patients but not CF patients was ST253, which is also known as PA14. This was shared between two patients and had a large number of SNPs difference (741). Whilst hypermutator status was predicted in one of the two, the lack of an identifiable cross-infection event, the isolates having greater diversity between them than some of the unconnected genomes, the common-occurrence of this strain in our non-respiratory cohorts, and ST253 being recognised as a major globally abundant strain, suggest that this is not a case of transmission between patients.

There were 2 shared strains in our CF cohort including one strain (LES/ST146) shared between a pair of twins. The twins had been infected with PA for many years, and before their care began in our unit. Their high levels of personal contact and relatively few SNP differences, point strongly to cross-infection. These two isolates were noticeably closer to each other genetically than they were to the unconnected genomes which were incorporated in our analysis. The WGS analysis of the other shared strain in our CF cohort (ST27) does not support cross-infection. ST27 is a common environmental strain and our isolates had

similar diversity when compared to unconnected genomes. Owing to our strict segregation policies for our CF cohort, these patients should not have come into contact with each other. In the out-patient clinics, patients go directly to their own consultation room, avoiding a communal waiting area. No other patient in the study uses that room during the clinic. During in-patient admissions, patients are managed in a single room and do not use shared facilities. The exceptions to these procedures are siblings. Consequently we do not have evidence of our segregation policy failing in its purpose.

In addition to the strains shared exclusively within the respiratory cohorts, there were 5 strains which were shared across the cohorts including the previously described high risk clones ST175 and ST235.[238] Importantly these isolates have shown as high levels of diversity when compared to the unconnected genomes as they did to each other, as illustrated in Figure 3.2. When looking at this Figure, it should be noted that with ST252 a pairwise comparison between two clinical samples was clearly less than the others in our cohorts. Whilst similar to some unconnected genomes, its predicted hypermutator status may raise concerns of this being an example of cross-infection which is not immediately apparent by simple SNP comparison. ST252 was a strain found in 3 NCFB patients, 1 CF patient, and has been found in distant locations such as Australia and Brazil on the MLST database.[332] The pairwise comparison in question had a difference of 156 SNPs and was between two NCFB patients. Review of clinical attendance could not identify a potential cross-infection events. In light of this, the degree of genetic difference and the demonstration of some unconnected strains having similar diversity, it is felt that this is unlikely to have been an example of cross-infection. The large genetic difference between the CF isolate of this strain and the others (greater than 3600 SNPs), does not suggest cross-infection between the CF cohort and the NCFB even when taking into account the hypermutator status of all the ST252 isolates. None of the other shared strains raised concerns of further cross-infection events either within or across cohorts. ST395 showed a lower level of diversity than the other strains which were shared across cohorts, however none of the isolates were predicted to be hypermutators and the pairwise comparisons showed very similar diversity amongst themselves as to when they were compared to the unconnected genomes.

This study has multiple strengths. It gives a comprehensive epidemiological review of PA in an unsegregated NCFB cohort who share both healthcare professionals and facilities with a CF cohort. The collection took place over a relatively short period of time and hence provides a “snap-shot view”. Examining multiple isolates reduces concerns that strains were being missed through investigating just single colonies. The use of three molecular techniques adds to the robustness of this study but their limitations must be noted, and warrant highlighting.

RAPD is an effective discriminator of strains, but its output is not reliably interpretable between cycles and gel electrophoresis of RAPD products introduces unacceptable variability.[244] Our study design circumvented these issues by ensuring that all isolates with unique RAPD profiles were re-analysed within a single batch, and employing Bioanalyser analysis to enable a robust and reproducible comparison of RAPD profiles. It could be argued that a large number of isolates did not undergo full analysis due to their exclusion following the initial visual inspection stage, and that this is a limitation. Whilst this is true, there are multiple examples of samples being considered “different” by visual inspection but then appeared to be clonal by Bioanalyser analysis and the same ST-type by MLST. With the design of the study retaining those felt to be different by visual inspection, this would suggest that more isolates could have been excluded for subsequent further analysis. This may have reduced costs but also increased the risk of missing other shared strains. As a caveat to this, there is also evidence of strains appearing clonal by RAPD analysis, which then had different ST types by MLST, hence suggesting there remained a small chance of additional shared strains being missed during the RAPD process.

The next technique, MLST, is considered a “Gold-Standard” strain typing technique.[353] It provides unambiguous data which can be compared to a substantial global database. The scheme’s principle involves the sequencing of highly conserved housekeeper genes, but therefore only a very small amount of the genome. Consequently, like RAPD, it does not provide sufficient resolution to discriminate closely-related strains. MLST is also considerably more expensive than RAPD. It is worth noting that there was complete concordance between the ST type as defined by MLST and that defined by WGS in those isolates which

underwent both techniques, suggesting that the highlighting of shared strains for WGS by MLST was an effective pipeline.

The final technique we have used is WGS. This has the obvious appeal of a comprehensive review of the genome. The calculation of SNP difference reflects the genetic diversity between bacterial genomes, but its interpretation has the issues described above. We have looked to reduce these weaknesses through the prediction of hypermutators and the inclusion of unconnected publicly-available genomes. Whilst these approaches do not completely eradicate these issues, they add a level of analysis which has not been applied to similar work, and we believe is a major strength to this study. Whilst we were not able to include unconnected genomes for ST564, this is not a major issue as our results did not show an intermediate amount of genetic difference- the range that causes the greatest uncertainty in this type of analysis. WGS's additional weaknesses as a technique for this sort of work include its lack of universal accessibility and that it requires more complexity and cost than RAPD or MLST.

This work has obtained a cross-sectional review of PA epidemiology in a NCFB cohort with the added context of a parallel CF cohort and non-respiratory isolates from the same healthcare and geographical setting. Longitudinal analysis would be a valuable extension of this study and would give an insight into the chronicity and persistence of individual strains. It would also be important to assess for evidence of ST564 becoming more prevalent across the NCFB cohort. This, however, should not be performed too soon. The hospital attendance of patients with NCFB tends to be less frequent than those with CF and it may take a considerable amount of time for a strain to show increasing prevalence in the cohort. The absence of environmental PA data in this study may be seen as a limitation but the absence of a truly dominant strain means that it would be very unlikely to find a relevant hospital reservoir. In addition, the marked differences in Clone C prevalence across the respiratory cohorts, despite their sharing of facilities, argues against an environmental reservoir in the hospital. Our non-respiratory samples do provide a comparator group to our respiratory cohorts, and the prominence of Clone C and PA14 does not point to a PA population structure out of keeping with previous data.[220]

A further limitation is the identification of potential cross-infection events. Our review of this information could only include documented activity within the RD&E

hospital. Therefore there is the potential for cross-infection events to occur which would not be identified by this approach. This could include visits to community healthcare, attendance with a relative, picking up a prescription from the in-hospital chemist, spending time in hospital amenities, or the sharing of public spaces. There may also have been further shared strains that were not identified within the patient cohort. This could have been for a multitude of reasons including, not all patients attending the department at a point when recruitment was possible, the patient not being able to produce a sample, the pathogen not being grown in the laboratory, and the strain being excluded by RAPD. Consequently there is a small possibility that the reported findings are an underrepresentation of the level of cross-infection occurring.

Our findings in this study are consistent with previous work. The only similar NCFB UK study was performed in Newcastle where cross-infection may have taken place, but was believed to be rare.[131] Our study is an important addition to this work, not just by adding to the evidence-base, but also due to the difference in the set-up between sites. In the Newcastle study, patients with NCFB were managed on a separate site to the CF cohort, whilst our cohorts are managed on the same site and within the same department. A recent study performed a cross-sectional genetic review of isolates of PA obtained through a multi-centre drug trial.[211] Due to the methodology and the multi-centre approach, this isn't as relevant a publication for the specific question of cross-infection within NCFB cohorts. It did however show evidence of shared lineages within single centres and that isolates from the same sample could show greater diversity than the comparison of isolates from different patients. The inference of this is that cross-infection could be occurring, which is again in keeping with our findings.

It could be argued that if cross-infection is occurring then patients should be segregated. This may, however, not be the most pragmatic approach. The segregation of our CF cohort has logistical implications in terms of clinic capacity, time and space. If these were applied to NCFB clinics, the number of patients who could be seen per clinic would have to be reduced significantly. This would impact follow-up at the interval desired. Physiotherapy clinics and pulmonary rehabilitation courses would also be affected. From a patients perspective this may be highly frustrating, and also provide a negative psychological impact via segregation and the implication of being an infectious threat to others. Overall,

we believe that the negative impacts may outweigh the low risk of cross-infection. Some may argue that cohorts could be segregated by colonisation status, but this would not stop super-infection and patients will be infected with PA for a period of time before clinical investigation has detected it. We do not advocate a change in policy based on these results. Longitudinal data or evidence of cross-infection with other pathogens may however change opinion on this. Without a segregation policy it is clearly important that basic infection control measures should not be ignored.

3.6 Conclusions

Our study has shown evidence of likely but uncommon cross-infection between NCFB patients. At present, we believe the negative impacts that would be associated with implementing a segregated NCFB cohort (including reduced patients per clinic and reduced access to pulmonary rehabilitation courses) outweigh the low risk of cross-infection. However, with growing NCFB cohorts nationwide and cross-infection possible, ongoing longitudinal surveillance is clearly warranted.[20, 354, 355]

Chapter 4- Comparison of an *in silico* prediction model of hypermutability in *Pseudomonas aeruginosa* to a traditional hypermutator assay

4.1 Abstract

Introduction: The hypermutator phenotype of PA is well recognised in chronic lung disease. It results in a higher spontaneous mutation rate and is associated with antibiotic resistance. Interpretation of cross-infection studies with WGS require an appreciation of this phenotype when investigating beyond a short-term outbreak. In Chapter 3, we used a novel *in silico* prediction model to identify hypermutators. This was to aid our interpretation of the epidemiology of PA in cohorts with chronic infection and evidence of shared strains. Here we compare this *in silico* prediction model to the traditional culture-based method of identifying the hypermutator phenotype.

Methods: Isolates which were predicted to be hypermutators by the *in silico* prediction model underwent traditional phenotypic assessment of hypermutability. Where available, examples of the same strains which were not predicted to be hypermutators were also included, alongside PAO1 as a control.

Results: Nine isolates predicted to be hypermutators and 5 isolates predicted not to be hypermutators (normomutators), underwent traditional phenotypic assessment of hypermutability. Isolates which were predicted to be hypermutators had a significantly higher actual mutation frequency when compared to both the predicted normomutators ($p<0.00001$) and PAO1 ($p=0.005$). Seven out of 9 predicted hypermutators displayed phenotypic hypermutator behaviour. In contrast, none of the predicted normomutable isolates displayed hypermutator status.

Conclusion: Our *in silico* prediction model displays compatibility with a traditional culture-based phenotypic assay for highlighting potential hypermutable isolates of PA. This highlights its usefulness in WGS cross-infection studies for chronic infection.

4.2 Introduction

As already described in this thesis, PA has demonstrated its adaptability to a wide range of environments, including the airways of diseased lungs. An adaptive phenotype seen in this setting is that of the hypermutator where microorganisms display a higher spontaneous mutation rate. In PA this phenotype is rare in the environment and acute infection, but is more prevalent in chronic infection in the lungs. Previous data has suggested that 36-55% of CF patients may harbour this phenotype.[159, 345] This may even be an underestimation due to the significant PA divergence found in the lung and the observation that hypermutators can occur alongside normomutators.[356] A single sputum sample may not fully reveal this. Whilst studied in less detail, hypermutators have also been found in NCFB patients.[161] The hypermutator phenotype is thought to be due mainly to inactivation of the mismatch repair system (MRS) through loss of function of antimutator genes like *mutS* and *mutL*, but also through defects in the GO system which deals with error avoidance.[357, 358]

Whilst a high mutation rate could be considered a disadvantage due to an increased risk in deleterious mutations, it may also have advantages in an airway environment, such as an association with biofilm growth.[359] A further association, and a major reason for the ongoing interest in this phenotype, is antibiotic resistance.[360-362] Patients with the chronically diseased airways of CF or NCFB are likely to have been subjected to a high historical usage of antibiotics. However, antibiotic resistance is thought to not solely be due to antibiotic use, and may instead arise as a consequence of hypermutability.[363] Furthermore, some studies have reported an association between clinically-relevant markers of disease and a hypermutable state, though conflicting results exist, and confounders may explain these associations.[361, 364, 365] At present the investigation of hypermutator status has no place in standard clinical practice due to the uncertainties of its significance.[366]

As WGS continues to become more accessible, hypermutator status may become important for a different reason. Among the many potential benefits of WGS, the use of the technique for cross-infection studies has been shown to be invaluable, both through our work described in Chapter 3, and others.[266, 367] In short-term outbreak studies, the mutation rate of microorganisms may not have much of an impact on interpretation of results. However in longer-term studies,

an appreciation of the mutation rate becomes highly relevant, particularly in cases where there is a reasonable amount of genetic difference but a prolonged period for that divergence to have occurred. Significantly different mutation rates have been reported for PA in chronic lung disease. The normal rate has been reported by some as 1-5.5 SNPs per year, whilst hypermutators have been reported with a rate of 100 SNPs per year.[154, 348-351] Consequently the potential range of difference in SNPs between two isolates involved in a cross-infection event 2 years previous, may be very large. In cases where hypermutators are involved, the differences may be similar to those seen between seemingly unconnected strains. In Chapter 3 we used a novel *in silico* prediction model of hypermutator status to enhance our understanding of potential cross-infection in an epidemiological study of PA. This model predicted hypermutable isolates in 8 out of the 21 (38%) NCFB patients interrogated by WGS and 1 out of the 9 (11%) CF patients interrogated in this way. The pairwise comparisons of SNPs in these isolates were significantly greater than in those not predicted as having hypermutator status, suggesting an accuracy and relevancy to this output (see Figure 3.2 and Figure 3.3).

The investigation for the presence of the hypermutator phenotype has traditionally been performed through the interrogation of PA growth on plates of culture media with and without rifampicin.[159] The frequency of mutants found on the rifampicin plates compared to the number of isolates found on plates without rifampicin allows the calculation of a quantifiable measurement known as the mutation frequency. The subsequent interpretation of this value has been variable, but often includes comparisons to the mutation frequency of PAO1 or a known hypermutator laboratory strain.[158, 159, 363, 368] Whilst straightforward, the technique is not rapid due to the requirement of time for growth, and the plating and subsequent counting of bacteria onto a large number of plates as triplicates are required to ensure robustness. As the novel *in silico* prediction model was integral to the investigation performed in Chapter 3, this chapter will compare the output from this model to the traditional culture-dependant technique.

4.3 Methods

A sub-set of isolates were chosen from the collection investigated in Chapter 3 to undergo further hypermutator assessment. The isolates which by *in silico*

prediction were described as hypermutators were all included as were examples of isolates of the same ST type which were not predicted as hypermutators by the model. Consequently, all isolates reviewed had undergone genotypic assessment by RAPD, MLST and WGS. Furthermore, all but one of these isolates had undergone further phenotypic assessment as part of an undergraduate student project which had included antibiotic resistance to ciprofloxacin, gentamicin and ceftazidime. These antibiotics are all clinically-relevant for this cohort of patients.

The hypermutator assay was performed as previously described (see Chapter 2).[158] All batches included PAO1 to provide a batch-to-batch comparison.

The primary assessment of hypermutability was the frequency in comparison to PAO1 within that batch. Those that were 20-fold greater or more were defined as strong hypermutators (SM); those 10-20 fold greater as weak hypermutators (WM), and others as normomutators (NM). This was an arbitrary definition for this study having taking into account alternative criteria and is referred to from here forth as M-status. Additional assessment of hypermutator status was also performed based on alternative criteria.[158, 159, 368] Clinical details for these patients had been obtained previously as part of the work involved for Chapter 3. The antibiotic resistance patterns from the undergraduate project were also reviewed.

4.4 Results

4.4.1 Samples Investigated

In total 14 isolates underwent the Hypermutator assay, of which 9 were predicted to be hypermutators by the *in silico* prediction model. The 5 isolates which were not predicted to be hypermutators provided matches for ST17, ST27, ST235 and ST253. All ST252 isolates that had previously undergone WGS were predicted to be hypermutators and consequently no comparators were available for this strain. Of the 14 isolates, 11 originated from NCFB patients and 3 from CF patients. The duration of chronic infection in these patients are shown in Table 4.1. A variety of mutations were seen which were predicted to have deleterious effects in MRS or GO systems, and these have previously been illustrated in Table 3.5.

	HM[a]	NM[a]
Subjects	9	5
NCFB	8	3
Time since 1st PA isolate		
- **Less than 1 year**	0	0
- **1-5 years**	5 (55.6%)	1 (20%)
- **5-9 years**	1 (11.1%)	0
- **More than 9 years**	3 (33.3%)	4 (80%)

Table 4.1. Time since 1st documented PA-positive sputum culture [a]HM (Hypermutator) and NM (Normomutator) status are based on the *in silico* prediction.

4.4.2 Variation in PAO1 Mutation Frequency

Each batch of hypermutator assays performed contained no more than 4 isolates in order to maintain accuracy, and each batch contained a PAO1 control from the same master-strain stock. Consequently the same PAO1 master-strain was measured in 5 separate triplicates and provides insight into the consistency of the assay. The mean mutation frequency was 3.48×10^{-8} (95 % Confidence interval (CI) 2.7×10^{-8}- 4.28×10^{-8}) with a standard deviation of 1.55×10^{-8}. The within triplicate standard deviation ranged from 6.22×10^{-9} to 1.34×10^{-8} (mean within triplicate standard deviation of 1.04×10^{-8}). As expected there was less variance within triplicates than across all samples and therefore the mutation frequency is most robustly assessed by comparison to the PAO1 triplicate result within the same batch.

4.4.3 Mutation Frequencies

Mutation frequencies calculated from the mean of the triplicates were obtained from all samples. One sample required prolonged incubation in order to grow sufficiently to record. The mean mutation frequency among all experiments on clinical samples was 1.87×10^{-6} (95% (CI) 1.85×10^{-6}- 1.88×10^{-6}). The predicted hypermutators had significantly greater actual mutation frequencies as well as mutation frequencies compared to PAO1 (see Table 4.2).

	HM[a]	NM[a]	P value[b]
xPAO1[c]	126.49	1.41	0.005
Mutation Frequency	2.88×10^{-6}	4.16×10^{-8}	<0.00001

Table 4.2 Comparison of mutation frequencies between predicted hypermutators and predicted normomutators [a]HM and NM status are based on the *in silico* prediction. [b]The p values are calculated by Mann-Whitney Test. Values are displayed as means. [c]The xPAO1 is calculated by the raw mutation frequency divided by the mutation frequency of PAO1 in that batch

By our *a priori* definition of the hypermutator status phenotype, 7 out of 9 of those predicted to be hypermutators reached the criteria (5 were SMs and 2 WMs). None of those predicted to be normomutators displayed the hypermutator phenotype (sensitivity 100%; specificity 78%) (see Table 4.3 and Table 4.4). All predicted hypermutators had a higher frequency than PAO1. Two isolates did not meet our definition of a hypermutator for this assay. Both of these only had an R33H mutation in *dnaQ* and no other identified relevant mutations. Mutation frequencies were also compared to alternative criteria for defining the strength of the hypermutator phenotype (see Table 4.5). The greatest correlation appeared to be with the definition used by Kenna et al.[158]

4.4.4 Antibiotic Resistance

Antibiotic testing had been performed for 8 out of the 9 predicted hypermutators and all the predicted normomutators in the previous undergraduate study. No duplication of this testing has taken place. There was significant resistance to clinically-relevant antibiotics reported in this work (see Tables 4.6 and 4.7). No isolate was reported as being sensitive to all three antibiotics. The most marked observation was with ceftazidime, where 1 out of 5 of the predicted normomutators were resistant, whilst seven out of eight of the predicted hypermutators were (p= 0.015). The only predicted hypermutator not to be resistant contained the R33H mutation in *dnaQ*. A similar finding was found when defining the isolates by the hypermutator assay result. All 6 of the hypermutators which had sensitivity data were resistant to ceftazidime whilst only 2 out of the 7 normomutators had were resistant (p=0.008).

Sample (ST)	Mutation	Mean Mutation Frequency	xPAO1	M-Status	Status by Kenna et al [158]	Status by Oliver et al [159]	Status by Lutz et al [368]
PIB26 (ST17)	*mutS*- Frameshift	5.78 x 10^{-7}	14.5	WM	WM	NM	SM
PIB45 (ST27)	*mutS*-V264E	5.59 x 10^{-6}	140.25	SM	SM	HM	SM
PIB16 (ST17)	*mutS*- L52P *mutL*- Q52X *mutM*- L342P	5.09 x 10^{-7}	36.82	SM	WM	HM	SM
PIB63 (ST253)	*mutS*- Frameshift *mutY*- H72R	9.21 x10^{-6}	666.48	SM	SM	HM	SM
PIB58 (ST252)	*mutL*- H469R *dnaQ*-R33H	7.92 x 10^{-6}	226.92	SM	SM	HM	SM
PIB67 (ST252)	*dnaQ*- R33H	2.46 x 10^{-7}	6.29	NM	WM	NM	SM
PIC30 (ST252)	*mutL*- Frameshift *dnaQ*- R33H	4.61 x 10^{-7}	11.81	WM	WM	NM	SM
PIB01 (ST235)	*mutS*- ΔL541-S544	1.27 x 10^{-6}	32.57	SM	SM	HM	SM
PIB23 (ST252)	*dnaQ*- R33H	1.3 x 10^{-7}	2.78	NM	NM	NM	WM

Table 4.3 Mutation frequencies and hypermutator status by various criteria for predicted hypermutators

Sample	Mean Mutation Frequency	xPAO1	M-Status	Status by Kenna et al [158]	Status by Oliver et al [159]	Status by Lutz et al [368]
PIB10.9	1.91 x 10^{-8}	0.48	NM	NM	NM	NM
PIB22.6	4.2 x 10^{-8}	3.04	NM	NM	NM	WM
PIB78.6	5.19 x 10^{-8}	1.49	NM	NM	NM	WM
PIC07.9	3.24 x 10^{-8}	0.69	NM	NM	NM	WM
PIC19.1	6.27 x 10^{-8}	1.34	NM	NM	NM	WM

Table 4.4 Mutation frequencies and hypermutator status by various criteria for predicted normomutators

Criteria	Specificity	Sensitivity
M-Status	77.78%	100%
Kenna et al [158]	88.89%	100%
Oliver et al [159]	55.56%	100%
Lutz et al [368]	100%	20%

Table 4.5 Sensitivity and specificity for *in silico* predictions by various criteria

	HM	NM
Isolates	8	5
Ciprofloxacin		
- **Sensitive**	5 (62.5%)	2 (40%)
- **Intermediate**	1 (12.5%)	1 (20%)
- **Resistant**	2 (25%)	2 (40%)
Ceftazidime		
- **Sensitive**	1 (12.5%)	4 (80%)
- **Resistant**	7 (87.5%)	1 (20%)
Gentamicin		
- **Sensitive**	1 (12.5%)	0
- **Resistant**	7 (87.5%)	5 (100%)

Table 4.6 Antibiotic resistance profiles for predicted hypermutators and predicted normomutators Results reported from previous undergraduate project

	HM	NM
Isolates	6	7
Ciprofloxacin		
- **Sensitive**	1 (16.7%)	3 (42.9%)
- **Intermediate**	1 (16.7%)	1 (14.3%)
- **Resistant**	4 (66.7%)	2 (28.6%)
Ceftazidime		
- **Sensitive**	0	5 (71.4%)
- **Resistant**	6 (100%)	2 (28.6%)
Gentamicin		
- **Sensitive**	1 (16.7%)	0
- **Resistant**	5 (83.3%)	7 (100%)

Table 4.7 Antibiotic resistance profiles for assay defined hypermutators and normomutators Results reported from previous undergraduate project

4.5 Discussion

In this chapter we have shown that those isolates predicted to be hypermutators by our *in silico* prediction model do indeed have a greater mutation frequency as determined by the traditional testing for the hypermutator phenotype. There was however large variability in the mutation frequencies observed among the predicted hypermutators, and by adopting different criteria some of these isolates would not be defined as phenotypic hypermutators. All isolates had originated

from patients who had been infected with PA for at least a year, and therefore the isolates were unlikely to represent acute infection.

Our primary assessment was the comparison of mutation frequency to PAO1. A wide range of frequencies were seen, varying from a frequency of approximately half that of PAO1, though to a 666-fold increase. Whilst there are only small numbers available for comparison, there was no suggestion that hypermutator status correlated with ST type. There were four isolates belonging to ST252 which were predicted to be hypermutators and whilst two of them did not possess a mutation frequency greater than 10-fold that of PAO1, one ST252 isolate had a frequency which was over 200-fold the frequency of PAO1. What may be expected to be more relevant than ST type is the mutations noted during the *in silico* prediction. From the early literature into this topic, it has been made clear that *mutS* is a highly relevant gene.[357] Mutations were seen in *mutS* in 5 of the 9 isolates and all but one were phenotypic hypermutators by all criteria detailed in Table 4.3 and all 5 were at least weak hypermutators by our criteria. Three isolates had mutations in *mutL* and just single isolates had mutations in either of the oxidized guanine (GO) system genes *mutY* and *mutM*. The isolate with the *mutY* mutation had the greatest mutation frequency, though it should be noted that there was also a *mutL* frameshift in this isolate. It is possible that a *mutY* mutation on its own may not have a large effect, but in combination with another mutation, may have led to a more dramatic change.[369] The other gene in which mutations were identified was *dnaQ* which is involved in proofreading activity and is also known as *mutD*. The same mutation (R33H) was found in all four of the ST252 isolates and no others. In two samples it was the only identified mutation of note, and these were the only two isolates which were not even weak hypermutators by our criteria (mutation frequencies PAO1 x 2.78 and PAO1 x 6.29). This mutation on its own appears to have minimal mutator effect as defined by the hypermutator assay. The results from those that we did not predict to be hypermutators were as we would expect. None of the isolates reached criteria for hypermutator status by our criteria, or those described by Kenna et al or Oliver et al.[158, 159] This was not the case with the criteria by Lutz et al (Table 4.4), highlighting the discrepancies that exist between different hypermutator definitions.[368]

These discrepancies in definitions are worth considering. A 20-fold frequency in comparison to PAO1 was used to define hypermutator status in several studies.[159, 161] In the work by Kenna et al, whose methodology we have replicated, the definition of normomutable strains was based on the modal point of the distribution of mutation frequencies.[158] Three other categories were defined (hypomutators, weak and strong mutators), and those with a frequency 20-fold greater were considered strong hypermutators. Consequently there were some which were considered weak hypermutators that had a frequency less than 20-fold that of PAO1. To take this into account, we defined our categories with 20-fold greater than PAO1 being noted as strong hypermutators and also noted those with frequencies 10-fold greater as weak hypermutators. Others have discriminated between strong and weak hypermutators with specific mutation frequency cut-offs.[368] In theory, it would be hoped that as the studies tend to have very similar protocols, including the same media and the same concentration of rifampicin in the media, that the mutation frequencies would be comparable. We have shown that even with the same master-strain and the assays being performed in the same laboratory by the same researcher, that variability is displayed. This is not an issue unique to ourselves, as this point was also made by Kenna et al.[158] It therefore seems logical that batch-to-batch variability should be considered in assessment of hypermutator status, rather than have specific mutation frequency cut-off points. As described above, a common option is an arbitrary threshold in relation to PAO1, however it is not clear what this should be.

It is worth considering these results in the context of our pairwise comparison of SNPs in Chapter 3. Two ST17 isolates were included in this chapter (PIB16 and PIB26). The isolate PIB16 was more divergent from other ST17 isolates than PIB26, and had mutations in three genes, whilst PIB26 just had a frameshift in *mutS*. It may be expected therefore that PIB16 would have a higher mutation frequency, and indeed it did (PAO1 x 36.82 v PAO1 x 14.5). ST252 was the only other strain to have more than one predicted hypermutator. As mentioned above, 2 of the 4 ST252 isolates did not meet our criteria for hypermutator status based on phenotype. These two had the least mean difference from other isolates of ST252. With regards to the two that were hypermutators, PIB58 had a far greater mutation frequency, though was less divergent from other isolates than PIC30.

This doesn't not necessarily contradict our findings, as the origins of PIC30 may have been very different from the other ST252 isolates, and importantly isolates can switch their hypermutability. In a cross-sectional study it is impossible to know if an isolate has only recently obtained a hypermutable state. Duration of infection cannot be used as a surrogate of this as we can clearly see in our predicted normomutators that long-term infection does not necessary equate to a hypermutator phenotype.

There was a large amount of antibiotic resistance noted in our isolates for ciprofloxacin, gentamicin and ceftazidime. Antibiotic resistance would be expected to some degree in hypermutators, but there was not a discernible difference from the resistance levels observed in normomutators for ciprofloxacin and gentamicin, albeit with very small numbers. There was however a statistically significant difference in rates of ceftazidime resistance in both predicted, and assay confirmed, hypermutators. All those which were resistant to ceftazidime were also resistant to gentamicin. This is of note as ceftazidime and gentamicin is a combination used for courses of IV antibiotics in patients with a history of PA, and is noted as common practice in national guidelines.[6]

Taking things forward, a key question is why do we want to identify a microorganism as being a hypermutator? It has previously helped our understanding of adaptation to an environment, but as discussed above, at present there is no indication in clinical practice to investigate for this. If impacts on antibiotic resistance are of interest, then more direct measures would surely be of more relevance. However, identification of hypermutators should be of direct relevance to WGS-based cross-infection studies involving chronic infection. It is anticipated that the WGS method will become a gold-standard technique for this type of investigation. In a situation when WGS has already occurred, it would seem logical to then use information contained within the sequences to identify isolates that may accumulate SNPs at a higher rate- a potential modern way of defining a "Hypermutator". This approach could evolve as WGS becomes more common and more longitudinal studies are performed. More genes may be revealed, and certain mutations may be identified as more or less important. The work in this chapter presents a starting point which even with small numbers, displays predictable differences between isolates.

There are clearly some weaknesses which should be noted. First of all, this study contains small numbers, though the hypermutator group did include every patient in whom we had predicted hypermutator status. Potential extensions of this work could include enlarging the investigation of those not predicted to be hypermutators. If we found isolates which displayed a hypermutator phenotype, then the WGS data could be re-examined for other mutations not previously identified. Further extensions could also include looking at multiple isolates from the same patient and investigating the heterogeneity of the potential hypermutator status and phenotype. With more resources, WGS on our large collection of isolates may provide greater insight into the accuracy of our *in silico* prediction model and the apparent significance of certain mutations. However, even with larger numbers the caveat of not knowing how long a microorganism has or has not been a hypermutator would prevent complete accuracy of interpretation. With this in mind, a study with the sequencing of longitudinal isolates with interrogation by this *in silico* prediction model and the hypermutator assay would improve understanding. A further weakness of this study is the hypermutator assay itself as it displays variability between batches. This was illustrated by the interrogation of PAO1 from the same stock across 5 batches. Our definition of hypermutator phenotype being based on comparison to PAO1 within the same batch, should reduce this limitation. The marked difference in ceftazidime resistance was interesting to note but this was in the situation of low numbers and without replication performed of antibiotic sensitivity testing, which is a limitation of this finding. Significant resistance to this antibiotic in hypermutators is certainly consistent with recent work.[363] The fact that the one isolate which did not display resistance had the mutation with the weakest effect on hypermutator status, adds to the suggestion that further work would reveal mutations of less significance with regards to potentially useful markers such as antibiotic resistance and mutation frequency.

4.6 Conclusions

Interrogation of a sub-set of PA isolates which had undergone WGS, has supported the *in silico* prediction of hypermutator status which was used in Chapter 3. Despite small numbers, highly significant differences were seen between those predicted to be hypermutators and those which were not. Whilst the clinical application of testing for hypermutator status is unclear, this *in silico*

prediction model may be very useful for interpretation of cross-infection studies in the WGS-era. As more isolates of PA are analysed by WGS, the accuracy and relevance of this model can evolve.

Chapter 5- Culture-independent genotyping of *Pseudomonas aeruginosa* for cross-infection screening

5.1 Abstract

Introduction: Evidence has shown that cross-infection with PA occurs in CF and may occur in NCFB. Consequently, intermittent screening of cohorts is important. Current methods rely on cultivation of PA before employing a molecular technique such as MLST. This approach may be affected by false-negatives and bias towards faster growing strains when multiple strains are present. Culture-independent approaches are developing in molecular microbiology, and we have compared the output from the traditional culture-dependent approach of MLST with a novel culture-independent approach in a CF cohort.

Methods: Patients with CF who had previously isolated PA were prospectively recruited and submitted a sputum sample for investigation as part of our investigations described in Chapter 3. DNA was extracted from all sputum samples and genotyping by MLST was attempted. If PA had been cultivated, then DNA was extracted from isolates and genotyped by MLST as reported in Chapter 3.

Results: Thirty-two patients were recruited and sample processing by both approaches was performed on 31. Of these 31 sputum samples, PA was cultured from 22. Genotyping by MLST was achieved in 23 patients by the culture-independent approach and 22 by the culture-dependent approach. In three instances genotyping was possible when culture did not isolate PA, and in another two instances genotyping was only successful via the culture-dependent approach. The culture-independent approach showed evidence of four strains being found in multiple patients (ST146- the Liverpool Epidemic Strain, ST217- the Manchester Epidemic Strain, ST27 and ST395) of which two were not revealed by the culture-dependent approach. From the 20 cases where genotyping results were obtained by both methods, there were exact matches for

130 of the 140 loci interrogated (92.86%). In 19 of these 20 patients, strains from the same clonal complex were revealed by both approaches.

Conclusion: Genotyping for PA via a culture-independent technique is feasible and can reveal shared strains missed by culture-dependent techniques. There is strong agreement between the two approaches and the culture-independent approach offers results unbiased by culture. This approach has potential for cross-infection screening, including in those who only intermittently isolate PA. However, vigilance is required towards the impact of multiple strains within individual samples.

5.2 Introduction

As previously described in this thesis, cross-infection with PA is an important issue in CF. We have demonstrated highly likely cross-infection in a NCFB cohort and this has also been either reported or highlighted as a possibility by others.[131, 211, 355] Practical genotyping techniques are therefore important for centres to be able to screen for cross-infection in CF and NCFB, and potentially other respiratory diseases.

As described in Chapter 3, we undertook an epidemiological review of PA in our NCFB and CF cohorts using three molecular techniques- namely RAPD, MLST and WGS. All three of these required the cultivation of PA on culture media from sputum samples. This first step provides a first potential limitation of this process as it is well-recognised that patients can intermittently isolate PA and at times this may be due to the technique not isolating PA, rather than PA being truly absent. In our study all patients recruited had previously isolated PA. We recruited 63 NCFB patients but only obtained PA from 46. Of the 17 from whom we failed to obtain isolates, 12 isolated PA again after their recruitment. In our CF cohort we failed to obtain PA from 9 out of 31 and 5 re-isolated PA after their recruitment.

If PA is isolated, a second issue is faced - of all the colonies on the plate how many and which ones will you investigate? It is recognised that multiple strains of PA can be found within a single sputum sample and therefore if just one colony is examined the presence of multiple strains will always be missed.[261] In addition, the isolate that you pick may not be the predominant, or the most important strain in the lung. In our work, ten isolates which were morphologically representative of those seen on the plate were picked for investigation. The

numbers of isolates picked in previous studies is highly variable but clearly the more isolates picked, the more accurate the representation you will have of the predominant strains.[131, 261] Also of note, PA can become visible on plates at different rates. Consequently there may be a bias towards fast-growing strains if there are multiple strains present.

In recent years there have been many studies looking into the bacterial communities in the lungs and this has required the extraction of DNA directly from sputum samples to reveal the presence of bacteria which have not been cultured by standard techniques.[302] As would be expected, these studies have shown that PA can be detected by culture-independent techniques in samples of patients with chronic lung disease. With growing interest in culture-independent microbiological investigation, it is likely that the extraction of DNA will become a common starting point in standard diagnostic testing. This allows the potential for genotyping directly from sputum samples rather than requiring the initial step of culture. There has been a precedent for this with MLST having been applied directly to DNA from sputum for BCC genotyping,[370] however it has not been described in PA. If a similar approach was feasible for PA then the benefits would be multiple. The process would be quicker, less biased by culture and the colony picking process, and potentially allow the revealing of multiple strains. This could allow rapid and regular monitoring of a cohort for evidence of cross-infection. It has been shown that PA can be detected by molecular techniques, even in culture-negative samples.[318] Therefore this approach may allow strain typing of culture-negative samples. This may allow greater understanding of patients who intermittently isolate PA and also with regards to the cycle of seeming “eradication” followed by “re-infection”.

Due to the above described issues, we have performed a study to look into the feasibility of applying the MLST scheme for PA directly from sputum samples. This has been performed alongside culture-dependent MLST to provide insight into the concordance of the output from these techniques. DNA from samples known to harbour different strains have also been artificially mixed in order to assess the impact of multiple strains on the techniques output. For the purposes of this thesis, the application of the MLST scheme for PA directly from sputum will be described as culture-independent MLST, whilst the traditional methods, as used in Chapter 3, will be described as culture-dependent MLST.

5.3 Methods

For the CF patients investigated in Chapter 3, the sputum samples from which isolates originated were stored for investigation alongside the isolates themselves. Sputum samples were also stored from those who were recruited but whose investigation by culture methods did not reveal PA. The isolates were investigated as described in Chapter 3. The sputum samples subsequently underwent DNA extraction with the GeneJET Genomic DNA Purification Kit (ThermoFisher Scientific) as described in Chapter 2.2.3. MLST was performed on all sputum samples as described in Chapter 2.4. As an initial screen, PCR amplification was first performed for two loci (*acsA* and *guaA*), and checked for product by gel electrophoresis. If this was positive for both then the scheme was performed for the 5 remaining loci, if it was negative for both then the sample was identified as "MLST negative" for PA, and if one did not produce a band on electrophoresis then the process was repeated on the allele that did not produce a band. In no circumstances did the repeat PCR fail to produce product if one loci had been positive in the first round.

To enable appreciation of the impact of multiple strains within individual sputum samples, the DNA extracted from two samples known to harbour different strains were mixed into five aliquots with DNA based ratios of 9:1; 7:3; 1:1, 3:7 and 1:9 for each sample. These mixtures then underwent the amplification and sequencing process for the *trpE* gene. This was also done with DNA extracted from bacterial culture for the same previously identified alleles. The two chosen were based on two samples with a large number of nucleotide differences within the *trpE* locus as revealed by the investigations described in Chapter 3 from culture, and from DNA extracted from sputum in this chapter.

5.4 Results

5.4.1 Culture-independent MLST Results

Sputum samples were collected from 31 CF patients who had previously grown PA. One other patient was recruited, however the sample was not collected. In 22 instances PA was isolated from the sample and 10 representative isolates from that sample were stored. The results of these isolates have been described in Chapter 3. DNA was successfully extracted from all of the sputum samples, with DNA yields ranging from 1.1-76 µg. MLST was attempted on all 31 samples

and successfully provided a result in 23 (74.19%). This compares to 22 by culture-dependent MLST (71%). Of the eight in whom culture-independent MLST did not reveal PA, 6 did not isolate PA on culture. Culture-independent MLST allowed genotyping in three instances where the sputum sample had been reported as culture-negative by the clinical laboratory (see Table 5.1).

From the 23 sputum samples from which results were obtained by culture-independent MLST, only 17 distinct concatenated sequences were obtained. In total there were 4 concatenated sequences which were found in multiple patients. These shared genotypes were ST27 and ST146 from 3 patients, and ST217 and ST395 from 2 patients.

5.4.2 Concordance Between Culture-dependent and Culture-independent MLST

Overall there were 20 samples with paired MLST results from both the culture-dependent and the culture-independent technique. Therefore the sequences of 140 alleles could be compared as the MLST scheme sequences 7 housekeeper genes. For the purposes of this comparison, the ST-type from the sample representative of the more abundant RAPD profile was chosen (see methods section in Chapter 3). From the comparison of these 140 alleles, there was an exact sequence match in 130 cases (92.86%). When assessing from a sample perspective, 16 out of the 20 had an identical concatenated sequence by both techniques and a further 3 were identified as the same clonal complex (defined as having exact matches for 6 out of the 7 alleles). In two of these three cases there was just one nucleotide difference between the samples, whilst in the other there were 6 nucleotides different. In the one case where the results did not identify the same clonal complex, the culture-dependent technique identified ST195 and the culture-independent technique identified ST217. These two ST types have differences in all 7 alleles and 53 nucleotide differences across their concatenated sequence (see Table 5.1). The investigations in Chapter 3 had not shown any indication of multiple strains in this sample. There was no evidence of ST195 in any other samples however, as previously reported, ST217 was present in a sample from another patient.

By the culture-dependent method there had previously been evidence of shared strains with ST146 and ST27. However, by the culture-independent method there

was also evidence of ST217 and ST395 in multiple patients. Potentially the most significant finding from the discrepancies between the two techniques is the revealing of previous missed known transmissible strain (see Table 5.1). MES was reported in two patients by this technique whilst the LES was found in a patient who was not related to the siblings described in Chapter 3 who harboured this strain.

Patient	Culture-dependent	Culture-independent	Relevant notes
PIC09	ST235	No exact match	Difference of 1 nucleotide. Same clonal complex.
PIC13	No exact match	ST217/MES	Difference of 1 nucleotide. Same clonal complex.
PIC16	Culture-negative	No exact match	
PIC17	ST379	Negative	
PIC25	No exact match	ST1717	Difference of 6 nucleotides. Same clonal complex.
PIC27	Culture-negative	ST146/LES	Shared stain not revealed by culture-dependent MLST
PIC28	No exact match	Negative	
PIC29	Culture-negative	ST395	Shared stain not revealed by culture-dependent MLST
PIC32	ST195	ST217/MES	Difference of 53 nucleotides. Shared stain not revealed by culture-dependent MLST

Table 5.1 Discrepancies between culture-dependent and culture-independent MLST Samples listed as “no exact match” did not have an exact match on the MLST database.

5.4.3 Mixed Samples and the Presence of Multiple Strains

As described in Chapter 3, one patient showed evidence of multiple strains by RAPD and MLST. With regards to the RAPD profiles, 9 out of the 10 had been the same and the MLST result for a representative of this prevalent profile was ST146/LES. When the other RAPD profile was assessed by MLST, it did not have an exact match but was closely related to ST146/LES. Unsurprisingly when the matched sputum sample was interrogated by culture-independent MLST, the result was ST146/LES. At the *ppsA* allele, where the isolates obtained from the sputum sample differed by 3 nucleotides, the trace from the culture-independent MLST was examined at these positions of difference. No evidence of the second strain was seen by this examination.

The extracted DNA from two sputum samples was combined into 5 different mixes. The combinations in these mixes was based on total DNA in ratios of 9:1, 7:3, 1:1, 3:7 and 1:9. The sequences for the two samples at the *trpE* allele were previously identified as *trpE* 7 and *trpE* 47, which differ by 13 nucleotides. In the 5 mixtures, those with at least 50% of the mixture being from the sample identified as *trpE 7*, there was no bi-directional disagreement and the consensus sequence was *trpE 7*. In neither of the other two mixtures was the allele identified as *trpE* 47 and bidirectional disagreement was seen. The sequence with the most bidirectional disagreement was the mixture with 30% from the *trpE* 7 sample, where the resulting consensus sequence harboured a mix of nucleotides from both *trpE* alleles. Examining the traces at these points of disagreement suggests the presence of multiple strains (as displayed in Figure 5.1).

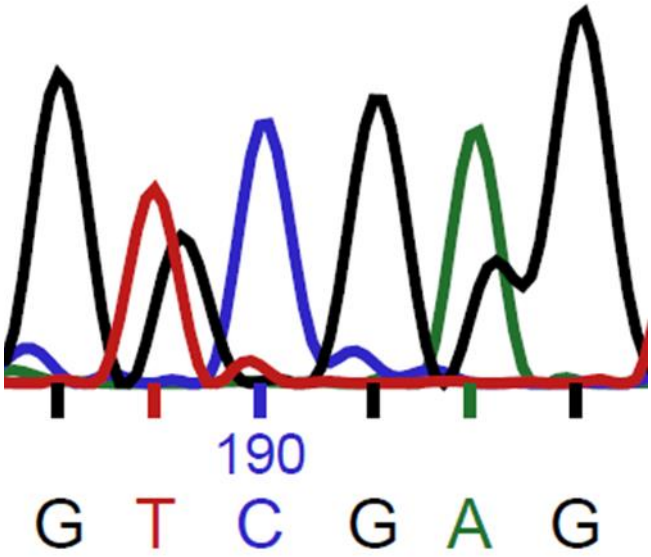

Figure 5.1 Multiple traces on sequencing data from DNA mixed from two sputum samples Traces taken from the Forward sequence of a mixture containing 30% of DNA from a sample identified as *trpE* 7 and 70% from a sample identified as *trpE* 47. Bi-directional disagreement occurred at position 189 and 192. For *trpE* 7 the sequence would be G**G**CG**G**G whilst for *trpE* 47 it is GTCGAG

The same experiment was performed on DNA extracted from culture for the same alleles so that the sequences should have been pure from each strain. In this instance, the consensus sequences were reported as *trpE* 7 when its proportion in the mix was 90%, 70%, 50% and 30%. For the mixture with 10% DNA from a *trpE* 7 sample and 90% DNA from a *trpE* 47 sample, there were multiple instances of bidirectional disagreement. Again, on examining the trace, there was evidence of multiple traces compatible with the known constituent sequences (see Figure 5.2).

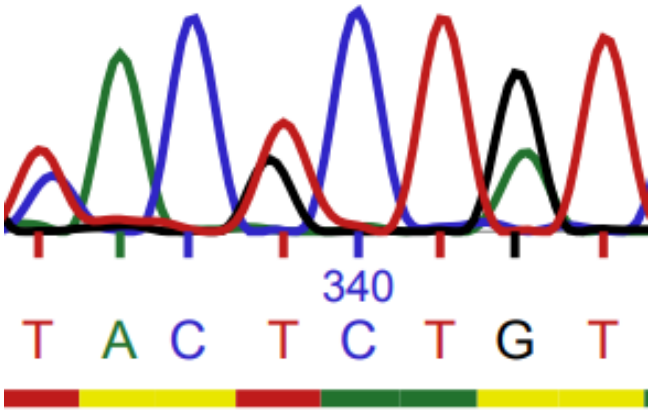

Figure 5.2 Multiple traces on sequencing data from DNA mixed from culture
The trace displays the suggestion of multiple traces at positions 336, 339 and 342, from a mixed sample of 10% from *trpE* 7 sample and 90% *trpE* 47 sample. Bidirectional disagreement occurred at positions 336 and 342, and the sequences for the allele also differ at position 339. The sequence for *trpE* 47 would have been TAC**G**CTGT, whilst for *trpE* 7 is would have been **C**ACTCT**A**T

One patient showed evidence of a completely different strain by culture-dependent and culture-independent MLST with reported results of ST195 and ST217 respectively. When looking at the sequences derived from culture-independent MLST, there was occasional evidence of disagreement between forward and reverse sequences and assessment of traces did show evidence of multiple traces (see Figure 5.3) suggesting possible multiple strains.

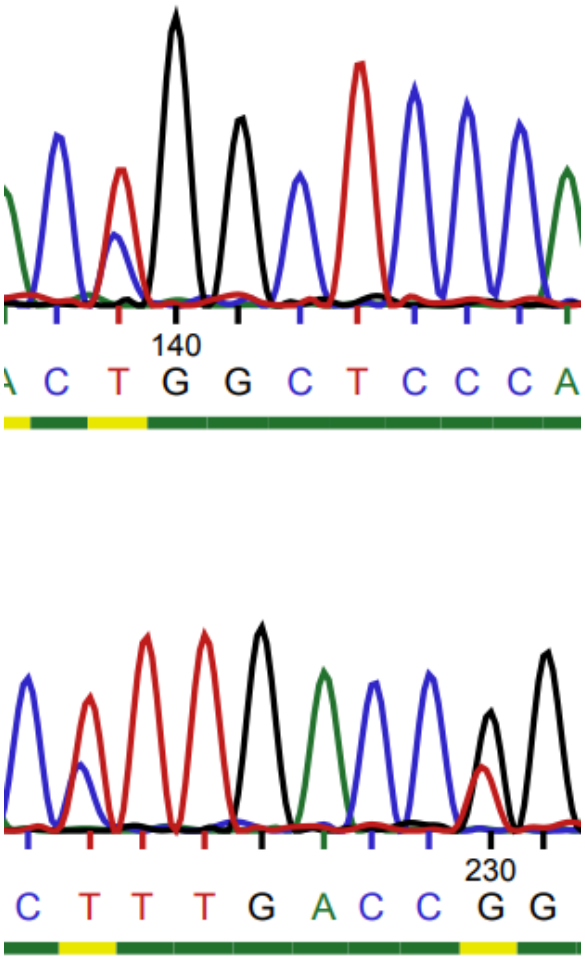

Figure 5.3 Multiple traces from sample with conflicting results Multiple traces seen at various points including positions 139, 223 and 230

5.5 Discussion

In this chapter we have applied culture-independent MLST to a cohort of CF patients who have previously isolated PA. This has demonstrated the presence of renowned transmissible strains which were not reported by traditional MLST. It has also highlighted the potential presence and issues of multiple strains when investigating by these techniques.

One of our aims in this study was to investigate the feasibility of this novel technique of culture-independent genotyping of PA. This was with consideration of this being a technique that could be used for epidemiological studies of PA in clinical cohorts. We have demonstrated that this is possible and can provide genotyping data in the globally recognisable format of MLST. In this particular cohort it has shown evidence of shared strains between patients. When this cohort was studied by culture-dependent methods two shared strains were found-

namely ST146/LES and ST27. By the culture-independent method described here, these shared strains were again found in the same patients, with the addition of ST146/LES being found in one other. This therefore could represent missed epidemiological data of key importance and be evidence of breaches in infection control. In this instance review of clinical notes suggests that this is unlikely to be the case. The patient was already colonised with PA before they transferred from a CF unit in another part of the country to our care. In addition, before the patients care was transferred, their PA had been genotyped and reported as LES.

In addition to ST27 and ST146/LES, two other shared strains were revealed by this method - ST217 and ST395. ST217 is also known as the Manchester Epidemic Strain (MES) and has been reported as a transmissible strain associated with increased healthcare burden.[198, 229] MES was not reported in Chapter 3, as one of the patients who has displayed this had a sequence with a slight difference by the culture-dependent technique (differing at a single locus) whilst the other patient had a completely different strain reported by the culture-dependent approach. ST395 was reported in Chapter 3 in one CF patient and 3 NCFB patients. It is not currently recognised as a transmissible strain. By the culture-independent approach, it was again found in the same CF patient, but also in one other despite being culture-negative. This ability to generate genotyping data from culture-negative samples is potentially very useful. This is because the intermittent isolation of PA in culture can mean that false-negative culture results may result in missed shared strains. In addition, being able to perform repeated genotyping on consecutive samples could give insight into the effects of attempted eradication regimes and whether later re-isolation is with a new strain or not. It should however be noted, that there were two episodes when PA was grown on a plate but not found by this technique. A possible explanation may be that the amount of DNA in the sample which was from PA may be very small. Previous 16S rRNA sequencing data has shown that samples can be culture-positive for PA despite having an exceedingly low relative abundance.[133]

With culture-independent MLST for PA not previously described, and the recognised potential for the presence of multiple strains in CF, it was important to perform a pairwise comparison of the output to the traditional approach. With

the traditional approach, MLST is performed on DNA originating from a single colony and therefore should be a pure sample. If the presence of multiple strains in samples was having a major impact on the consensus readings from the sequencing data, then it may be expected that there would be poor concordance between the two results. However, this has not been the case and concordance was very high between the samples, even when the traces may suggest multiple strains. In all but one case where data was available by both approaches, there was an exact match for at least 6 of out 7 of the loci sequenced. In one sample (PIC32) there was no match for any of the loci. However on examining the traces, there did appear to be the presence of multiple traces. Consequently a plausible explanation is that multiple strains were present in the same sample and that the faster growing strain was picked from plates for the culture-dependent approach, but on the basis of total DNA a different strain predominated. This then begs the question, which is the better approach? We picked multiple isolates from our patients in order to reduce multiple strains going unobserved, however in patient PIC32 we appear to have missed multiple strains. However via the culture-independent approach whilst we got just one result, there was the observation of the likely presence of multiple strains. To give further perspective on this issue we mixed samples to provide insight into the impact this can have on results.

In Chapter 3 we found evidence of multiple strains in one patient, however by the culture-independent approach we just picked up what we believed to be the most abundant strain (based on RAPD profiles and targeted MLST). When we mixed the DNA from two sputum samples that were known to harbour different strains, we observed inconsistencies in the MLST output as the ratios changed. A potential flaw of this experiment was that whilst a total DNA content was known, it was conceivable that one sample had a high relative abundance of PA, whilst the other had a low relative abundance, consequently distorting the ratios. We therefore repeated this process using a mixture of DNA extracted from pure cultures rather than sputum. Again inconsistent outputs were observed as the ratios of the DNA samples changed. A few important points need to be taken from these experiments. Firstly, the use of a different alignment sequence can bias towards a certain result if there are disagreements between the bidirectional sequencing. This was noticeable in some of the mixed experiments. Secondly, if the consensus results contain sections from more than one strain, then the

process may generate a hybrid sequence that is in not actually present within the sample. If this was a common issue then we would not see the high level of agreement that we saw between the two approaches. It does however highlight an occasional limitation of the technique. The other limitation of culture-independent MLST is the inability to identify specific strains when the presence of multiple strains is suspected. We can collect evidence of the presence of multiple strains via this culture-independent approach by viewing traces and noticing bidirectional disagreement in sequences but this does not allow identification of what the other strains are. If there are two traces or more, then it is unclear which alleles the differences at different positions come from, except in the circumstance where there is just a single nucleotide different throughout the entire concatenated sequence. Even then doubts remain, as the commercial sequencing company reports that the multiple traces seen can just be noise rather than definite multiple sequences (this is from personal correspondence with the sequencing company). This limitation must be put into context however. We picked multiple colonies in order to get an appreciation of the multiple strains present but this culture-dependent approach failed to detect multiple strains in samples such as PIC32.

On the basis of the above findings, it could be argued that a dual approach should be taken for extensive epidemiological reviews of PA in respiratory cohorts. By the dual approach you may pick up evidence of multiple strains as in PIC32, including the faster growers and the most predominant strain by total DNA. Due to the occasional issues of multiple strains giving seemingly false MLST results, if confirmation of shared strains was required, the culture-dependent approach would be essential and it may take multiple sputum samples and picking of a significant number of colonies to prove this. The use of two typing techniques is common in epidemiological reviews to ensure robustness of findings, and indeed in our work we have used three techniques. Consequently it should not be seen as a failing to use the two techniques. If rather than an extensive epidemiology review, a screening process was wanted, then culture-independent MLST may have its place as a first line investigation. By this approach two loci could be screened and if there is a match between samples for one of these, then a cascade approach could be taken with sequencing the other loci until all 7 loci are sequenced or differences at greater than one loci is seen. Culture-dependent

MLST could then be performed in cases where shared strains needed to be proved and where WGS may be wanted, or in cases where there was the suggestion of multiple strains by either sequence or trace examination. A further place for culture-independent MLST would be for monitoring and understanding the process of attempting to eradicate PA. As previously mentioned, patients can isolate PA intermittently and can seemingly "self-clear". Relapse after attempts of eradication is also recognised with some questioning if true eradication ever truly occurred. By performing intermittent culture-independent genotyping, assessment of strain persistence could be better understood, even in culture-negative samples. It should however be noted that as culture-independent methods do not measure viability, a lag period may exist when PA is still detected despite no longer being viable.

Whilst culture-independent identification of strains has been performed before, as far as we are aware this represents the first time it has been reported for PA.[370, 371] This is clearly a real strength of this study, as is the pairwise comparison with the culture-dependent approach. A limitation is the size of the collection. If a larger collection was examined, more instances of multiple strains giving conflicting results may have occurred. It is also of potential significance that this was performed exclusively in CF patients. It is possible that in other diseases, the relative abundance of PA may be lower and consequently only a small amount of PA DNA is present following extraction. The methods described here could be used in other diseased populations to investigate this further.

5.6 Conclusion

We have demonstrated the feasibility of genotyping of PA on DNA extracted directly from sputum. This has shown high concordance with culture-dependent approaches and has also revealed transmissible strains in culture-negative samples which were missed by the traditional approach. The approach can highlight potential situations of multiple strains in a specimen but it should be noted that this can on occasion negatively affect the result by creating a hybrid sequence that fails to match any individual alleles present. This novel approach offers potential for rapid screening of a cohort, as part of an epidemiological review, and the investigation of intermittent infection.

Chapter 6- Utilising RISA for interrogating bacterial communities of the bronchiectatic lung

6.1 Abstract

Introduction: A cornerstone of the management of the chronic suppurative lung diseases is the understanding and treatment of acute exacerbations and chronic infection. These decisions are often based on pathogens identified by traditional culture-based methods. Culture-independent techniques have however revealed that the bacterial communities of the lungs are complex and have far greater diversity than that reported by culture-based methods. These methods have suggested associations between ecological measures of a bacterial community and disease severity in multiple diseases, however there are inconsistencies in conditions such as NCFB. Whilst 16S rRNA sequencing has been the gold-standard technique for examining these communities, it is not accessible in everyday clinical care. In this chapter, we investigate the utility of an alternative cheaper microbial community profiling technique (RISA), and examine the relationship between the bacterial community and patient characteristics in a NCFB cohort, and a comparator CF cohort.

Methods: Sputum samples were prospectively collected from patients with NCFB or CF when they were submitting sputum samples as part of their standard clinical management, as well as relevant clinical details and patient characteristics. Following DNA extraction, RISA was performed on all submitted samples. Investigation into the output from RISA was performed by cluster analysis and the assessment of ecological measures of the bacterial communities in relation to patient characteristics.

Results: 208 sputum samples were submitted from 99 patients with NCFB and 51 patients with CF. RISA analysis was successfully performed on all samples and revealed a wide range of community complexity. Comparison between the two diseases suggested that our CF patients had more complex bacterial communities. In NCFB, cluster analysis revealed groups with potentially significant characteristics, such as reduced diversity and evenness. Increased

community complexity was associated with a low exacerbation phenotype in NCFB, and a positive correlation was seen between community richness and lung function. A relationship was not found between disease severity (as measured by the Bronchiectasis Severity Index) and ecological measures of the bacterial community. In the CF cohort, a positive correlation was seen between lung function and both community richness and diversity.

Conclusion: RISA offers a potential alternative to 16S rRNA sequencing, and as an adjunct to traditional culture, for characterising of bacterial communities in chronic suppurative lung disease. In our NCFB and CF cohorts it has revealed associations with clinically-relevant markers, such as lung function and exacerbation phenotype. It now requires validation of its output by 16S rRNA sequencing.

6.2 Introduction

CF and NCFB are chronic suppurative lung diseases whose clinical courses are believed to be heavily influenced by the management of both their chronic infections and their acute exacerbations. The cornerstone of managing chronic infection and exacerbations is antimicrobial therapy. In the acute setting this is often with 2-week courses of antibiotics via the oral or intra-venous (IV) route. Chronic infection is often managed with prophylactic oral or inhaled antibiotics. In current practice, traditional culture techniques are applied to sputum and other respiratory samples to identify potential pathogens. This information is then used to guide decision-making with regards to antimicrobial therapy. This process tends to lead to PA and SA being common targets in CF, and PA and HI in NCFB. Other pathogens which are observed and targeted, though less frequently, include *Streptococcus pneumoniae* (SPn), *Moraxella catarrhalis* (MC), *Stenotrophomonas maltophilia* (SMalt), species of the BCC, and NTMs.

Over the last decade, research using culture-independent molecular techniques has illustrated that there are far more complex bacterial communities in diseased lungs than is revealed by culture.[302] It has also been established that the healthy lung is not sterile.[301] Investigation of different disease states has shown disease-specific patterns. CF has understandably been subject to considerable interest, and it appears that there is an association between bacterial communities lacking diversity and disease severity.[133] In end-stage disease

the lung may be dominated by a single pathogen such as PA.[372] This concept of reduced diversity and disease severity is not limited to CF, or even the lung, with evidence of reduced diversity in other pathological conditions such as inflammatory bowel disease.[373] There have also been studies into NCFB cohorts. Studies by Rogers et al on a RCT cohort have suggested associations between diversity and lung function, and richness and several disease markers.[321, 323] However, this has not been a universal finding and other studies on unselected cohorts have not found such a correlation.[271, 319] Of note, these studies contained samples collected before the trio of RCTs which have led to chronic macrolide therapy being standard care for those regularly suffering from exacerbations.[38, 51, 52] Also the studies were performed without comparison with a validated severity index (the Bronchiectasis Severity Index (BSI)), which is now available and gives a score between 0-21.[40] The relationship between bacterial community composition and disease severity remains unresolved.

With clinicians basing management on traditional culture methods, a microbiological technique that is limited in the detail it reveals, it would seem likely that a technique providing more detail of the bacterial community would be desirable. The main culture-independent technique currently used in the research community is 16S rRNA sequencing. Despite reducing costs of sequencing, it clearly remains far more expensive and time consuming than traditional culture. In addition, it requires significant bioinformatic processing. A further reason that it is not used in standard NHS practice is that we do not know what the results actually mean for the clinical management of a patient. It would consequently be appealing for a technique to be available that could quickly and cheaply provide information beyond what is provided by culture, which could be of benefit in understanding the microbial communities of the lung, and could subsequently guide management strategies. Such a method would allow large clinical studies into bacterial communities and the impacts of therapies at a more affordable cost. This would then also be potentially usable in daily clinical practice to help personalise treatment in an evidence-based manner.

Ribosomal Intergenic Spacer Analysis (RISA) is a simple and cheap microbial profiling technique which has been used predominantly in environmental sampling.[288, 290, 291, 374] It has been suggested that RISA can discriminate

at the species level in most cases and has shown a degree of consistency with pyrosequencing in environmental work.[287, 291, 375] The technique involves PCR amplification of the intergenic spacer region between the 16S and 23S subunits. It has recently been applied to CF cohorts with the suggestion of potential benefit in the management of CF and other conditions including NCFB.[133, 270] However, its use has not been described in a NCFB cohort at present. Other community fingerprinting techniques exist and have been used for studies of the diseased lung such as T-RFLP, however RISA has demonstrated greater sensitivity and resolution in environmental work.[284, 291] Consequently, in this study we have collected sputum samples and clinical data from NCFB patients to examine the utility of RISA as a technique to be used in clinical practice, and the relationships between microbial profiles and disease severity. In parallel, we have applied RISA to samples from our CF cohort as a comparator group. In this chapter we will describe our cohorts, the output obtained by RISA interrogation, and insights into bacterial communities alongside disease and patient characteristics. In the subsequent chapter we will compare its output to 16S rRNA sequencing in a subset of NCFB patients.

6.3 Methods

Patients were recruited prospectively if they were submitting a sputum sample as part of their routine care. Patients were eligible for recruitment if they had a confirmed diagnosis of CF, or radiologically-defined bronchiectasis and a consistent clinical presentation in the absence of CF. Up to three samples per patient could be included and recruitment could take place irrelevant of exacerbation status, though this was noted at recruitment. The “exacerbation” state was defined by a patient currently taking an acute course of antibiotics due to respiratory symptoms, or if antibiotics were prescribed to them at that clinical consultation due to a change in respiratory symptoms. If the sample was submitted within 2 weeks of an exacerbation it was defined as “post-exacerbation”, whilst beyond 2 weeks was classified as “stable”. At the point of recruitment, clinical details were recorded and a sample sent to the local microbiology laboratory for standard processing. Samples were only included in this study if they were available at the consultation. This was to prevent the situation where samples produced at home are then submitted at the local health-centre before transfer to the hospital laboratory. In this scenario, samples can be

awaiting courier pick up for variable periods of time and potentially stored at a variety of temperatures. By just interrogating samples submitted immediately we could guarantee minimal delay in transfer to the laboratory.

In the NHS laboratory an aliquot was taken for our investigation, with the remainder used for standard microbiological testing. Consequently all samples had a culture result reported by the laboratory. Once the aliquots were transferred to the University laboratory, they underwent DNA extraction and subsequent RISA analysis as described in Chapter 2. In addition, aliquots of Mucolyse were also obtained from the laboratory and 500µl were subjected to the DNA extraction process with each of the 5 batches of the DNA extraction kit and reagents used.

The ecological measures of diversity, evenness and richness were estimated by using the concentration of the bands and numbers of band on the bioanalyser output as a surrogate for species abundance and presence.[291] The estimates of these ecological measures by RISA are referred to in this thesis as e-Diversity, e-Evenness and e-Richness. This is to distinguish the results from the subsequent output from 16S rRNA sequencing described later in this thesis. Statistical testing was performed in GraphPad Prism 7.03. For the purposes of statistical testing, all variables were assessed for normality by histogram observation. Of note, e-Diversity and e-Evenness were not consider to have a normal distribution for the purposes of analysis.

6.4 Results

6.4.1 The NCFB Cohort

As would be expected by the epidemiology of the conditions, the NCFB cohort was the larger of the two. In total 102 patients without CF were recruited, however three were subsequently excluded. One patient on re-review of radiological investigation did not have definite NCFB, one patient's sputum sample was lost, and from one patient the laboratory were unable to process the sample due to insurmountable adherence to the sputum pot. From the remaining 99 patients, 123 samples were submitted, of which 98 were taken in the stable state. A wide range of disease severity was seen, with BSI scores ranging from 1-18. With regards to categories, 21.2% were "Mild" (BSI = 0-4,), 44.4% were "Moderate" (BSI = 5-8) and the remaining 34.3% were "Severe" (BSI greater than 8). The main clinical details at recruitment are displayed in Table 6.1. All the BSI

components either showed significant differences between the BSI categories or a trend towards, with the exception of radiological severity (Chi-square= 1.603, p= 0.4487). There was no significant difference in the use of prophylactic antibiotics between the different BSI categories.

	Mild (n=21)	Moderate (n=44)	Severe (n=34)	Overall (n=99)
Age, years (Range)	56.86 (31-69)	66.23 (17-85)	69.82 (35-88)	65.47 (17-88)
BSI (Range)	2.71 (1-4)	6.5 (5-8)	11.29 (9-18)	7.34 (1-18)
%Female	66.67%	72.73%	52.94%	64.65%
Aetiology				
- **Idiopathic**	14.28%	40.91%	38.24%	34.34%
- **Asthma**	38.09%	13.64%	11.76%	18.18%
- **Post-infective**	14.28%	18.18%	14.71%	16.16%
- **Other**	33.33%	27.27%	35.29%	31.31%
FEV_1 %predicted (Range)	84.86% (33-116%)	72.05% (30-108%)	55.09% (23-96%)	69.94% (23-116%)
Colonised at recruitment[a]				
- **Any**	9.52%	31.82%	70.58%	40.40%
- **PA**	0	13.64%	41.18%	20.20%
- **HI**	9.52%	13.64%	26.47%	17.17%
BMI	25.68 (4.22)	25.97 (6.35)	23.54 (4.44)	25.07 (5.45)
Exacerbations in last 12 months[b]	1.71 (1.93)	3.55 (2.57)	4.32 (3.69)	3.42 (0.97)
MRC Dyspnoea score	2 (0.69)	2.25 (0.88)	2.85 (1.06)	2.4 (0.97)
In-patient hospital days in last 2 years	0	0.43 (2.2)	4.44 (8.9)	1.59 (5.66)
3 or more lobes affected/cystic changes on CT	57.14%	45.45%	58.82%	52.5%
Antibiotic Prophylaxis				
- **Any**	61.9%	47.72%	47.06%	50.5%

- **Oral**	42.86%	43.18%	35.29%	40.4%
- **Inhaled**	19.05%	18.18%	20.59%	19.2%
Any IV antibiotics in last 12 months	0	4.55%	26.47%	11.11%
<10 pack year smoking history[c]	90.48%	72.73%	64.71%	73.7%

Table 6.1 Baseline demographics of NCFB cohort at recruitment. The data in this table are presented as mean values or percentages of that group. In those continuous measures where range is not of specific interest, standard deviations are reported alongside the mean value. [a]Colonisation was defined by the isolation of a pathogen twice in a twelve month period and at least three months apart. [b]Exacerbations were by patient recall of a change in respiratory symptoms which lead to a course of antibiotic treatment. [c]A pack year is the smoking of 20 cigarettes a day for a year. The figure was recorded as reported by the patient.

6.4.2 The CF Cohort

In the CF cohort, 51 patients were recruited, of whom 28 handed in multiple samples. Overall 85 samples were submitted and 53 of these were taken from a patient in a stable state. A wide range of ages (17-65 years) and lung function (26-113% predicted) were seen. The main baseline demographics are shown in Table 6.2.

Age, years	32.80
(Range)	(17-65)
%Female	50.98%
Mutations:	
- **F508del homozygous**	39.22%
- **F508del heterozygous**	52.94%
- **Other**	7.84%
FEV_1 %predicted	61.86%
(Range)	(26-113%)
BMI	23.62
	(3.84)
MRC Dyspnoea score	1.9
	(0.95)
Antibiotic Prophylaxis	
- **Any**	90.2%
- **Oral**	80.39%
- **Inhaled**	66.67%
Any IV Antibiotics in last 12 months	52.94%

Table 6.2 Baseline demographics of CF cohort at recruitment Values are reported as means or as a percentage of the cohort. For continuous data, if the range is not considered to be of specific interest, then a standard deviation is reported. FEV_1 %predicted and BMI values were available for 50 out of the 51 patients.

6.4.3 Culture Results

All sputum samples underwent standard microbiological assessment at our NHS laboratory. Samples are not routinely tested for acid-fast bacilli in order to detect TB or NTM but are requested intermittently for a variety of reasons including clinical suspicion, part of a CF annual review, or before consideration of commencing a prophylactic antibiotic such as azithromycin. Consequently the

lack of samples positive for an NTM should not be compared directly to other pathogens such as PA.

In the NCFB cohort the most frequently isolated pathogen was HI, both overall and when looking at just one result per patient (as assessed by the first sample). As would be expected, PA had a very similar prevalence, and these two pathogens accounted for approximately three-quarters of cultured pathogens (see Table 6.3). Of note a large number of samples did not reveal bacterial pathogens (43.09%).

Bacteria	**All samples (n=123)**	**Baseline samples (n=99)**	**Stable (n=98)**	**Post-exacerbation (n=4)**	**Exacerbation (n=21)**
PA	25	19	19	1	5
HI	26	22	23	1	2
SA	7	5	6		1
Coliform	3	3	2		1
SPn	3	3	3		
Klebsiella pneumoniae	2	2	1		1
MC	2	2	2		
Serratia marcescens	2	1	2		
SMalt	2	1	2		
Burkholderia multivorans	1	1	1		
Proteus sp	1	1		1	
Culture-Negative	53	43	42	1	10

Table 6.3 Culture results for NCFB cohort From some samples multiple pathogens were isolated. A sample was called “Culture-negative” if a potential bacterial pathogen was not isolated from the sample.

Traditionally PA has been seen as a more significant pathogen than HI and colonisation with it results in 2 more points in the BSI than other potentially pathogenic organisms. Consequently, it may be expected that those isolating PA

at baseline would have a higher BSI than those isolating HI. This was the case, however this difference did not reach statistical significance (though it did if colonisation status was considered). The mean difference in the BSI score between those who isolated PA at baseline and those who isolated HI was only 2.01 (see Table 6.4). Of interest, in those who isolated HI rather than PA, a worse exacerbator phenotype was seen (see Table 6.4). The patients who isolated HI had significantly less prophylactic therapy or IV antibiotics in the preceding 12 months. When patients were assessed on the basis of colonisation status rather than baseline culture result, the difference in exacerbator phenotype between HI and PA no longer reached statistical significance and the difference in oral prophylaxis no longer persisted (see Table A6.1).

	PA (n=19)	HI (n=22)	Other (n=15)	Culture-negative (n=43)	P Value (PA vs. HI)
Age[a]	65.79 (11.51)	62.68 (18.11)	63 (13.06)	67.63 (11.81)	0.5335
BSI[a]	9.37 (4.5)	7.36 (2.76)	5.87 (3.38)	6.95 (3.28)	0.0967
FEV_1 %predicted[a]	66.58 (24.54)	61.77 (21.77)	73.27 (24.1)	72.14 (20.72)	0.5208
Exacerbator phenotype[b,c]					**0.0220[d]**
- High	10.53%	50%	40%	46.51%	**0.0068**
- Medium	47.37%	31.82%	40%	30.23%	0.3087
- Low	42.11%	18.18%	20%	23.26%	0.0931
Oral antibiotic prophylaxis[b]	63.16%	27.27%	26.67%	41.86%	**0.0296**
Inhaled antibiotic prophylaxis[b]	31.58%	0	13.33%	25.58%	**0.0060**
Received IV antibiotics in last 12 months[b]	31.58%	0	13.33%	9.3%	**0.0060**

Table 6.4 Clinical measures for different baseline culture results in the NCFB cohort Data is presented with mean values and standard deviations, or percentage of the group. Statistical testing was by [a]t-test, or [b]Chi-squared test. [c]Exacerbator phenotype was described as “High”- greater than 3 exacerbations in preceding 12 months; “Medium”- 2 or 3 exacerbations in preceding 12 months; “Low”- 0 or 1 exacerbations in preceding 12 months. [d]Chi-squared test from 3x2 contingency table.

In the CF cohort the culture results were again dominated by just two pathogens, but in this group it was PA and SA (see Table 6.5). There was a noticeable

amount of co-infection with both of these pathogens which needs to be taken into account when reviewing results alongside clinical detail (see Table 6.6). Statistical analysis may be further limited due to small sub-groups. However, in the absence of SA, there is a trend to suggest PA may be seen in the more severely diseased patient and those with PA and SA appear to be in-between those just isolating PA or SA with regards to lung function and IV antibiotic requirement (see Table 6.6). The above reservations with regards to small groups and the impact of co-infection should be remembered though.

Bacteria	All samples (n=85)	Baseline samples (n=51)	Stable (n=53)	Post-exacerbation (n=3)	Exacerbation (n=29)
PA	45	24	28	1	16
SA	32	21	22		10
Achromobacter sp	4	3	1		3
Burkholderia vietamensis	4	1	4		
Burkholderia multivorans	3	2	2	1	
Burkholderia gladioli	2	1			2
HI	2	2	2		
Proteus mirabilis	2	1	1		1
SMalt	2	2			2
Achromobacter xylosoxidans	1				1
Acinetobacter baumanni	1	1	1		
Acinetobacter haemolyticus	1	1			1
Coliform	1	1	1		
Not processed	1				1
Culture-negative	14	11	11	1	2

Table 6.5 Culture results from CF patients From some sputum samples multiple pathogens were isolated. A sample was called “Culture-negative” if a potential bacterial pathogen was not isolated from the sample.

	PA (n=12)	SA (n=9)	PA+SA (n=12)	Others (n=7)	Culture-negative (n=11)	P Value
Age, years[a]	35.25 (11.45)	35.44 (15.12)	30.42 (10.52)	29.43 (7.58)	32.73 (9.31)	0.7250
%female[b]	50%	66.67%	33.33%	57.14%	54.55%	0.3776
FEV_1 %predicted[a]	52.36 (10.46)	69.22% (18.20)	64.42% (20.58)	50.71% (18.89)	67.27% (21.18)	0.1371
Oral prophylaxis[b]	75%	77.78%	91.97%	85.71%	72.73%	0.7791
Inhaled prophylaxis[b]	66.67%	44.44%	100%	71.43%	63.64%	0.0722
IV days in last 12 months[c]	11.5 (0-29.75)	0 (0-12)	0.5 (0-25.75)	28 (0-64)	0 (0-14)	0.2467

Table 6.6 Baseline culture results and clinical characteristics of CF cohort
Data is present either as a mean value with standard deviation, median with a range, or as a percentage of the group. Statistical analysis performed by [a]t-test, [b]Chi-squared test, or [c]Mann-Whitney.

6.4.4 RISA Profiles

Following DNA extraction from all 208 samples and 5 negative controls, RISA was performed with PCR amplification of all samples. We were able to obtain observable DNA fragments which were seen on both standard gel electrophoresis and via the bioanalyser for all the clinical samples. In total, 1258 bands were observed. No PCR product was detected for any of the negative controls. Over 95% of bands had a length between 400-1200 bp with many samples yielding bands in the region of 530 to 570 bp (see Figure 6.1). Cluster analysis with the creation of a dendrogram for all samples displayed significant diversity among the samples studied (see Figure 6.2).From this dendrogram potentially significant bands were identifiable as being of high intensity with values around 530, 560 and 750 bp. The bioanalyser output quantified the band intensity by concentration. When the bands with the highest concentration within each sample were reviewed, the prevalent bands were around 530-570 bp and

750-770 bp (see Figure 6.3). Previously performed unpublished *in silico* predictions by collaborators, suggests that the prominent bands in the 530-570 bp region may be due to *Streptococcus* and that the 750-770 bp region may be dominated by PA and HI.

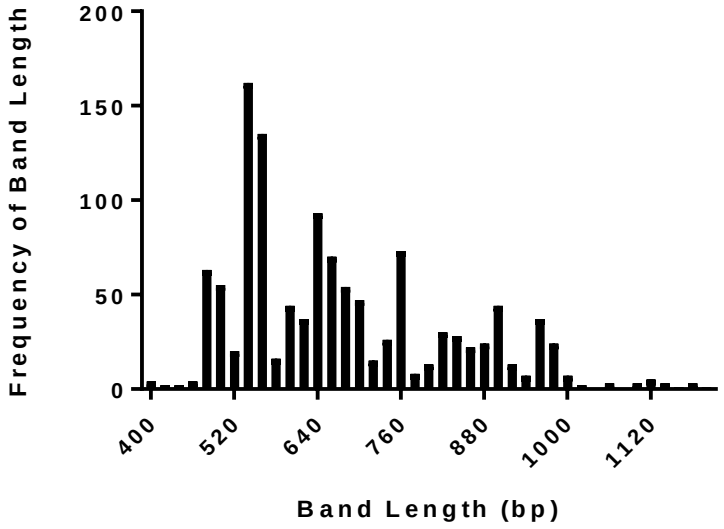

Fig 6.1 Frequency distribution data of all bands between 400-1200 bp

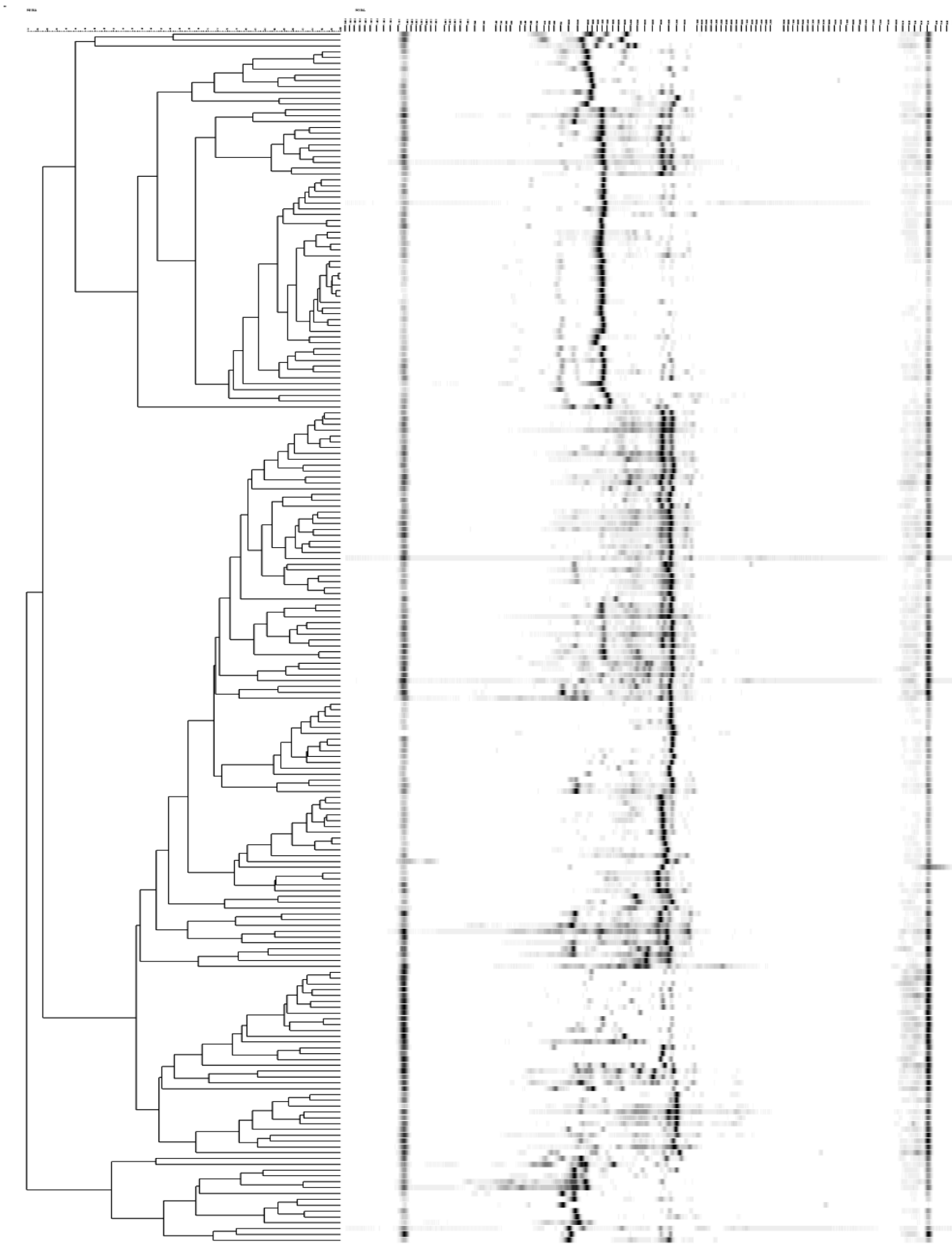

Fig 6.2 Dendrogram of RISA fragments from all clinical samples

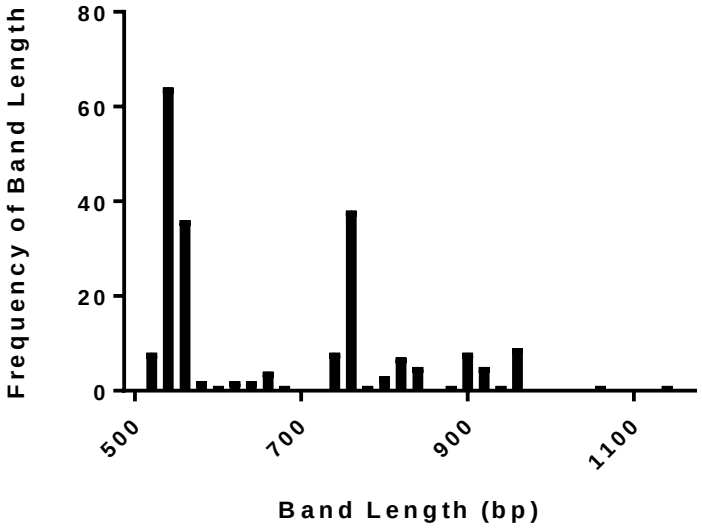

Figure 6.3 Frequency distribution data of the dominant band in each sample

6.4.5 Ecological Measures from RISA Profiles Between Cohorts

The ecological measures of diversity, evenness and richness were all estimated from the bioanalyser data output. Comparisons were made between the two patient cohorts with the baseline sample included for analysis. By all measures the NCFB cohort appeared less complex (see Figures 6.4, 6.5 and 6.6). This analysis included samples in various clinical states. We have also performed comparison of these ecological measures by just looking at the first available sample from a patient in the stable state. This reduced the numbers available in the analysis but still showed statistically a significantly less complex community in the NCFB cohort (see Table 6.7). Comparisons of samples from the same patient in both the stable and exacerbation state was only available for 18 patients. However, from these small numbers there was no clear difference in the ecological measures between these states (see Table 6.8).

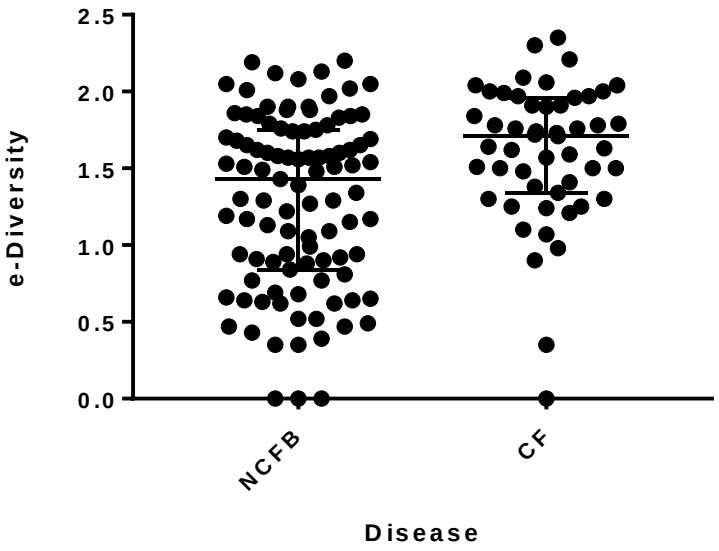

Figure 6.4 Comparison of e-Diversity between NCFB and CF Median values were 1.43 for NCFB and 1.71 for CF. The p value as calculated by Mann-Whitney test was 0.0009. The bars represent the median and the interquartile range.

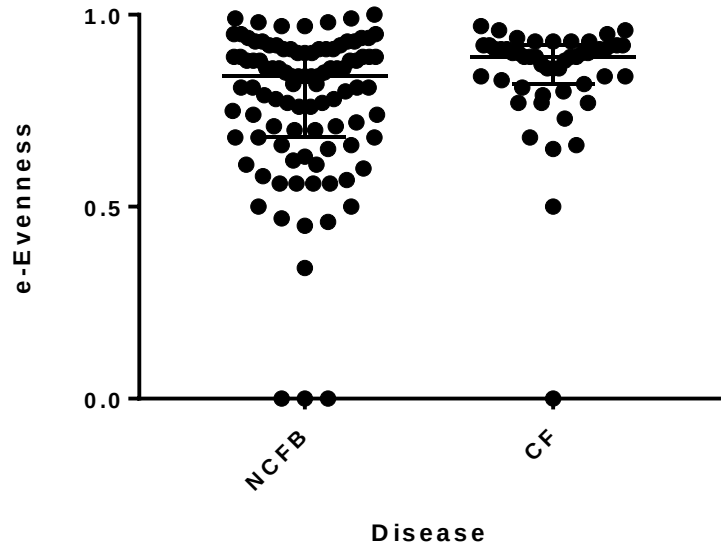

Figure 6.5 Comparison of e-Evenness between NCFB and CF Median values were 0.84 for NCFB and 0.89 for CF. The p value as calculated by Mann-Whitney test was 0.0096. The bars represent the median and the interquartile range.

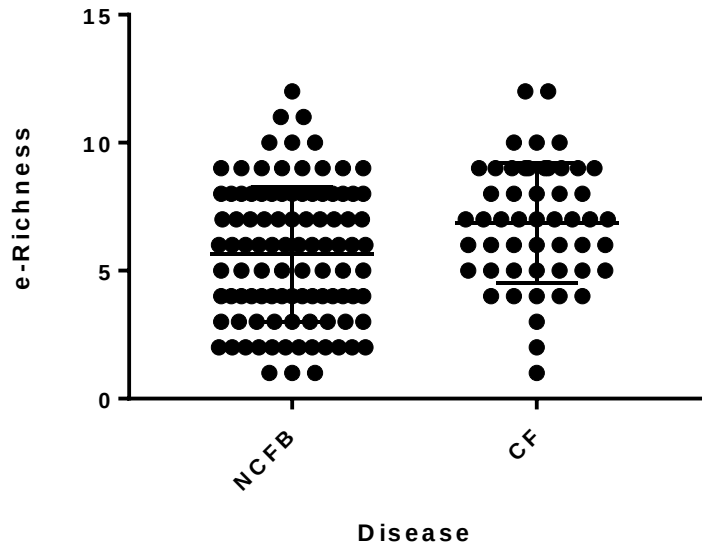

Figure 6.6 Comparison of e-Richness between NCFB and CF Mean values were 5.65 for NCFB and 6.83 for CF. The p value as calculated by unpaired t-test was 0.0063. The bars represent the mean and the standard deviation.

	NCFB (n=85)	**CF (n=39)**	**P Value**
e-Diversity	1.51	1.71	**0.0064**
e-Evenness	0.84	0.91	**0.0015**
e-Richness	5.79	7.05	**0.0136**

Table 6.7 Comparison of ecological measures in stable NCFB and CF patients Includes 1st stable sample per patients available. Comparison was performed by Mann-Whitney test for e-Diversity and e-Evenness with values reported as medians. e-Richness is reported as a median and statistical comparison was performed by the unpaired t-test.

	Stable	Exacerbation	P Value
e-Diversity	1.56	1.57	0.8317
e-Evenness	0.87	0.88	0.3982
e-Richness	6.5	5.78	0.3231

Table 6.8 Comparison of ecological measures in the stable and exacerbation state from both cohorts Wilcox matched pairs signed rank test performed for comparison of e-Diversity and e-Evenness with medians reported. Paired t-test was performed for comparisons of e-Richness and mean values reported

6.4.6 RISA in the NCFB Cohort

When cluster analysis was performed on the baseline samples from the NCFB cohort, the dendrogram divided into three main clusters (Groups A-C) and a small disparate group (Group D) (see Figure 6.7). When compared to culture results, Group A had a high proportion of culture-negative samples (16/22) and Group C had a high proportion of *H. influenzae* culture-positive samples (14/25). Reduced e-Diversity and e-Evenness were present in Group C. Group B featured two dominant bands around 530 and 560bps, which based on previous *in silico* predictions may be representative of *Streptococcus*. Two samples were culture-positive for MC and these two clustered right next to each other in Group C with a highly similar dominant band. Details of the comparison of the Groups created through cluster analysis are presented in Table 6.9.

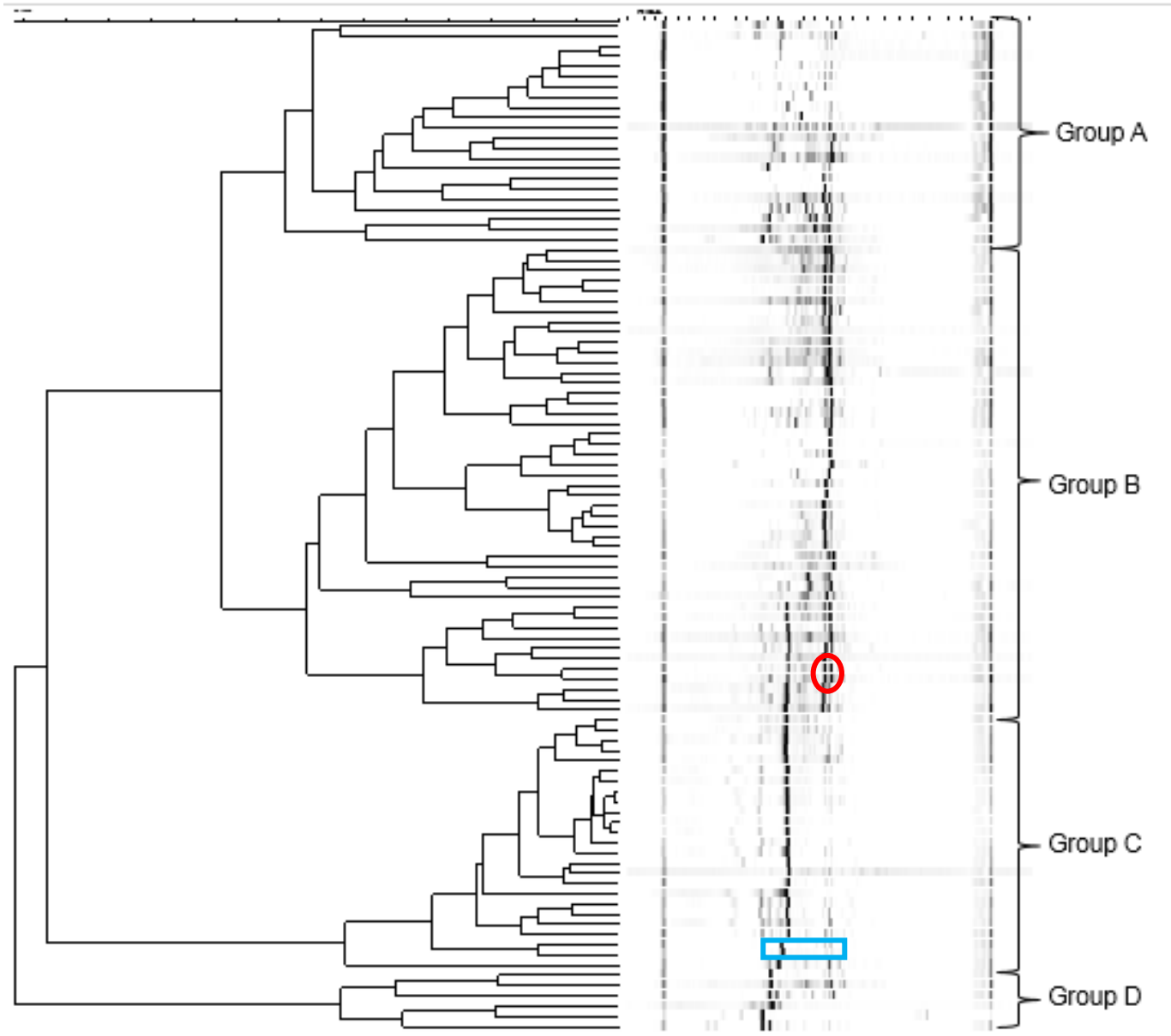

Figure 6.7 Dendrogram of baseline samples from NCFB cohort The red circle indicates typical bands around 530 and 560 bps. The blue box highlights the two patterns of the two samples positive for *Moraxella catarrhalis* clustering next to each other.

	Group A (n=22)	Group B (n=46)	Group C (n=25)	Group D (n=6)	P<0.05
Age[a]	65.73 (13.91)	69.11 (10.06)	60.28 (15.95)	58.33 (17.75)	**0.0374 B v C**
BSI[a]	7.59 (4.15)	7.24 (3.21)	7.2 (3.69)	7.83 (4.22)	
FEV_1 %predicted[a]	69.91 (24.5)	71.54 (21.92)	65.52 (22.18)	59.67 (19.48)	
Oral Prophylaxis[b]	36.36%	41.3%	40%	50%	
Inhaled prophylaxis[b]	31.82%	13.04%	24%	0%	
PA-positive[b]	9.09%	23.91%	12%	50%	
HI-positive[b]	0	15.22%	56%	16.67%	**<0.0001**
Culture-negative[b]	72.73%	47.83%	16%	16.67%	**0.0006**
High exacerbators[b] (>3)	36.36%	41.3%	40%	33.33%	
Medium exacerbators[b] (2-3)	40.91%	32.61%	36%	33.33%	
Low exacarbators[b] (0-1)	22.73%	26.09%	24%	33.33%	
e-Diversity[c]	1.26 (0.64-1.9)	1.58 (1.13-1.83)	0.94 (0.63-1.54)	1.33 (0.9-1.85)	**0.0171 B v C**
e-Evenness[c]	0.93 (0.89-0.97)	0.85 (0.76-0.89)	0.62 (0.56-0.78)	0.76 (0.65-0.87)	**<0.0001 A v B A v C B v C**
e-Richness[a]	4.82 (3.04)	6.24 (2.51)	5.12 (2.12)	6.33 (2.36)	

Table 6.9 Comparison of cluster analysis groups in the NCFB cohort Statistical analysis by [a]ANOVA or [b]Chi-squared test or [c]Krusal-Wallis. Values are quoted as means with standard deviations, or percentages of their group, or medians with IQR. The "P<0.05" column highlights statistically significant results, and if by ANOVA or Krusal-Wallis, then which groups differ significantly.

As discussed above, the NCFB cohort appears to have less complex bacterial communities than those of the CF cohort. We have interrogated potential factors for this to seek an indication of association. Of note there was a positive correlation between age and e-Diversity and e-Evenness, and between FEV_1 and e-Richness, and a negative correlation between exacerbations and e-Diversity (see Table A6.2). When looking at the culture results, there is a suggestion that HI-positive samples are less complex that those of PA, with a statistically significant difference in e-Evenness (p=0.0064) and trending towards it for e-Diversity (p=0.1645) (see Table A6.3). This is in keeping with Group C on the dendrogram. In the samples where a pathogen was isolated from the laboratory, a statistically significant reduced e-Evenness score was seen (p=0.0262) (see Table A6.4). When assessing categorical factors such as antibiotic burden, gender or home environment (urban and rural setting and the ownership of animals), no significant relationships were seen with ecological measures except urban living having higher e-Diversity and e-Richness scores than rural living (p values of 0.0294 and 0.0098 respectively (see Table A6.5)).

Of particular interest, and relevance to patients, there were significant differences in exacerbation phenotype and ecological measures. The patients classified as having a low exacerbation phenotype (as defined by an either 0 or 1 exacerbations in the last 12 months) had higher e-Diversity and e-Richness and a trend towards more e-Evenness (see Table 6.10).

	Exacerbator Phenotype			
	High (n=39)	**Medium (n=35)**	**Low (n=25)**	**P Value**
e-Diversity	1.29 (0.68-1.62)	1.22 (0.69-1.68)	1.65 (1.15-1.96)	**0.0201**
e-Evenness	0.77 (0.68-0.88)	0.84 (0.65-0.91)	0.88 (0.8-0.91)	0.1826
e-Richness	5.54 (2.63)	4.97 (2.54)	6.76 (2.35)	**0.0313**

Table 6.10 Comparison of exacerbation phenotypes Statistical analysis is performed by Krusal- Wallis test when e-Diversity or e-Evenness is involved with median values reported, or by ANOVA for e-Richness with mean values reported

6.4.7 RISA in the CF Cohort

Cluster analysis was also performed for the CF cohort (see Figure 6.8). Of note, it did not fall into as tight clusters as the NCFB cohort and consequently a useful comparison of clusters could not be performed. Ecological measures were again calculated from the bioanalyser output and compared to patient characteristics and clinical data. With regards to continuous data, no significant association was found between age and BMI, and any ecological measures (see Table A6.6). There was however a positive correlation found between lung function and e-Diversity and e-Richness (see Figure 6.9, Figure 6.10 and Table A6.6).

The ecological measures were also compared to different culture results with significant findings with regards to all measures. Those who isolated PA or did not isolate a pathogen appeared to have the least complex communities (see Table 6.11). The samples which did not isolate any pathogens had lower e-Diversity (p=0.0182) and lower e-Richness (p=0.0350), than those who isolated pathogens (see Table A6.7). As with the NCFB cohort, other factors such as gender, antibiotic burden and environment were assessed, but no associations were noted (see TableA6.8). As well as antibiotic prophylaxis, this also applied to IV antibiotics in the last 12 months (see Table 6.12).

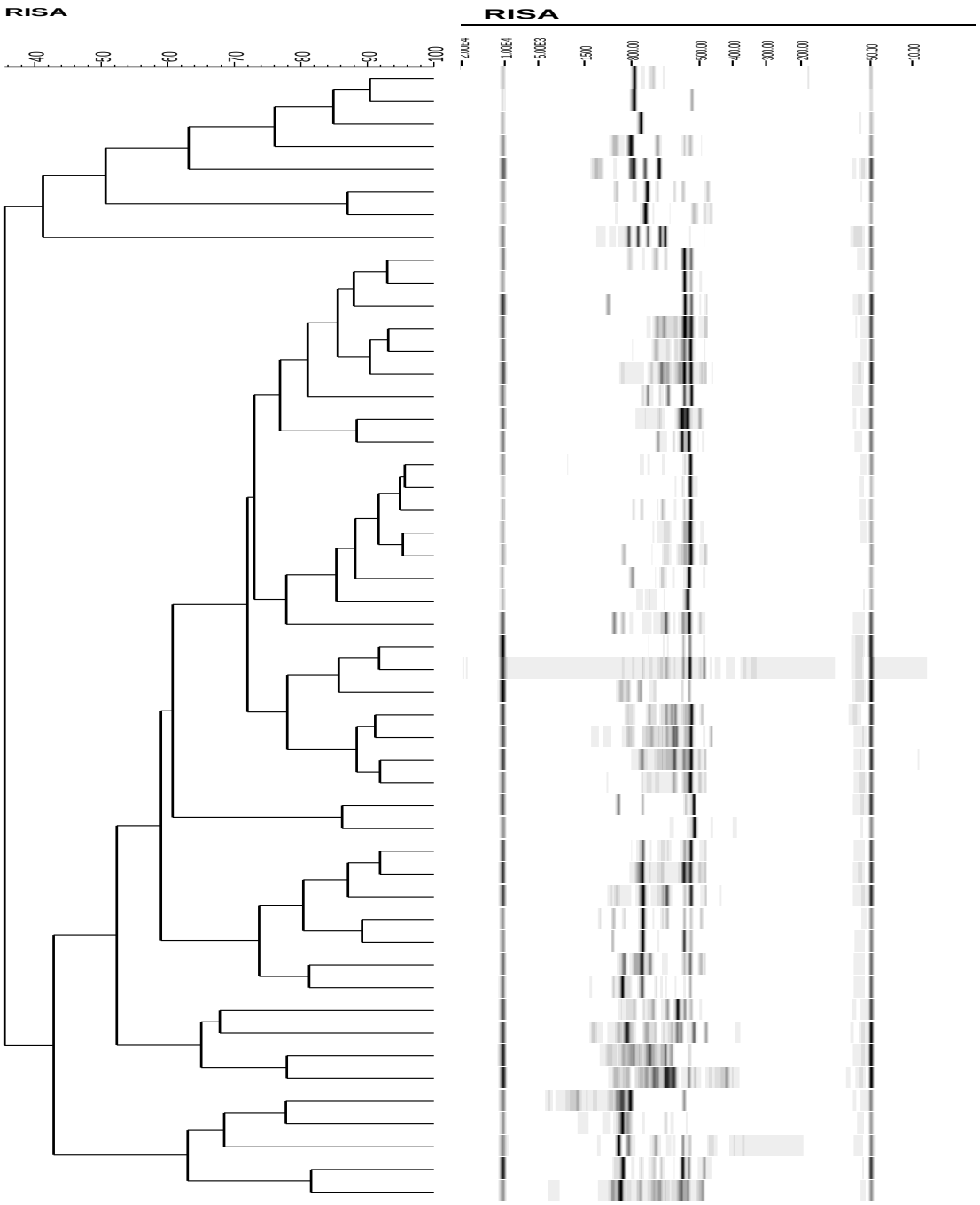

Figure 6.8 Dendrogram of baseline samples from CF cohort

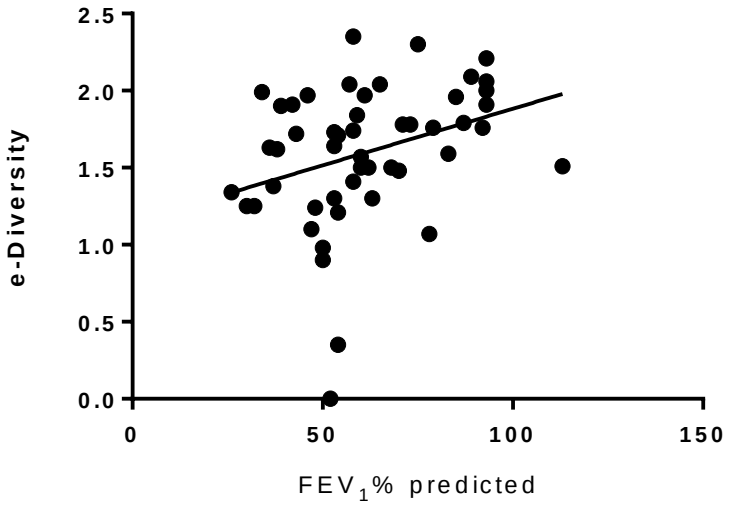

Figure 6.9 Comparison of e-Diversity and FEV_1 %predicted in CF cohort

Assessed by linear regression. P=0.0216 and R^2=0.1052

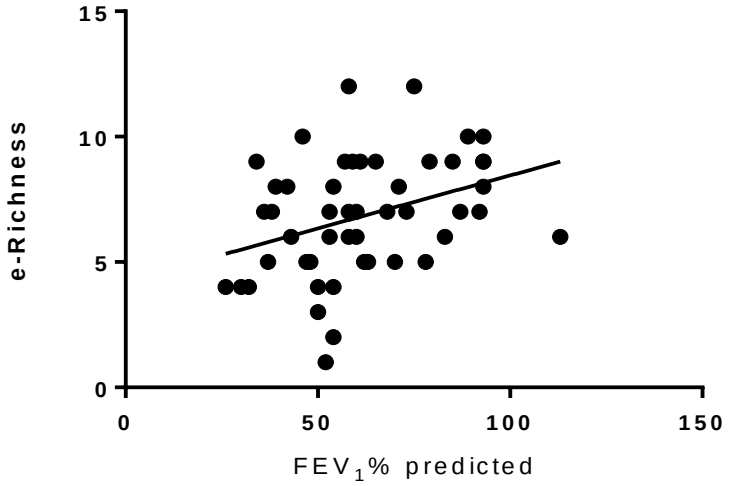

Figure 6.10 Comparison of e-Richness and FEV_1% predicted in CF cohort

Assessed by linear regression. P=0.0111 and R^2=0.1269

	PA (n=12)	SA (n=9)	PA+SA (n=12)	Other (n=7)	Culture-negative (n=11)	P Value
e-Diversity	1.48 (1.14-1.73)	1.91 (1.63-2.09)	1.9 (1.65-2.04)	1.62 (1.41-1.99)	1.48 (1.07-1.74)	**0.0056**
e-Evenness	0.86 (0.71-0.9)	0.92 (0.85-0.94)	0.92 (0.85-0.92)	0.86 (0.79-0.91)	0.89 (0.73-0.9)	**0.0168**
e-Richness	5.5 (2.29)	8.11 (2.13)	8.08 (1.93)	7.57 (2.06)	5.55 (1.44)	**0.0043**

Table 6.11 Comparison of CF culture result and ecological measures Calculations involving e-Diversity and e-Evenness are performed by Krusal-Wallis test and values reported as medians with IQRs. Calculations involving e-Richness were performed by ANOVA with values reported as means with standard deviations. “Other” refers to culture positive samples which did not isolate PA or SA. Multiple comparisons for e-Diversity revealed a statistically significant difference in the comparison of PA+SA v Other. Multiple comparisons for e-Richness revealed a statistically significant difference in the comparison of PA v PA+SA and PA+SA v Culture-negative.

	>14days (n=13)	1-14days (n=14)	No IVs (n=24)	P Value
e-Diversity	1.62 (1.3-1.88)	1.62 (1.35-1.9)	1.76 (1.33-1.99)	0.6197
e-Evenness	0.87 (0.84-0.92)	0.9 (0.79-0.92)	0.89 (0.81-0.92)	0.9916
e-Richness	6.62 (1.94)	6.57 (2.58)	7.17 (2.34)	0.6940

Table 6.12 Comparison of IV antibiotic days in preceding 12 months and ecological measures in CF cohort Calculations involving e-Diversity and e-Evenness are performed by Krusal-Wallis test and values reported as medians with IQRs. Calculations involving e-Richness were performed by ANOVA with values reported as means with standard deviations

6.5 Discussion

In this chapter we have successfully performed microbial profiling with the RISA technique on NCFB and CF cohorts. This has demonstrated an output of potentially clinically-relevant information in relation to markers of disease severity such as lung function and the exacerbator phenotype, as well as clustering samples into potentially useful categories such as low diversity communities. It has also given insight into the comparison of bacterial communities in the lungs of those with NCFB and CF and the position of their traditional pathogens- namely PA, HI and SA, within their communities.

When considering our dataset it should be noted that the cohorts involved in this study are fairly typical of what may be expected in other centres around the UK. In the NCFB cohort there was a predominance of females and a typical mean age. As is often the case, a significant proportion of the cohort did not have a clear aetiology for their NCFB. Interestingly there appeared to be a higher rate of asthma as an aetiology in the mild category. Conceivably some of these patients only remain under out-patient follow-up because of their concurrent asthma, and otherwise they would have been discharged from follow-up with their mild NCFB. There was a wide range of severity, with the majority either having moderate or severe disease. Microbiological data was also typical, with PA and HI the most prominent pathogens by the traditional culture approach. However, there appeared to be a suggestion of HI-positive cultures being seen in patients with higher exacerbation rates than those with PA-positive cultures. PA is usually considered to be the more pathogenic of these two prominent microorganisms and is traditionally associated with more severe disease. Those with PA did have a higher severity score, but it should be considered that the difference of approximately 2 points between these sub-groups is the same as that awarded to patients colonised with PA. There could be a couple of explanations for this observations. It is plausible that certain strains of a generally less pathogenic organism, may have greater pathogenicity than strains of another more notorious microorganisms. A relatively benign local PA population may have less impact on exacerbations than a more pathogenic HI population. Another explanation maybe that those with PA receive better treatment. This may in part be due to greater vigilance with PA, but also available therapies may be more effective for PA. For example, macrolides have been shown to be more effective in those with PA, and

the use of inhaled colistin in those who are compliant is effective.[52, 120] In our cohort, those with PA received significantly more prophylactic antibiotic which may explain this finding.

In the CF cohort there was a predominance of the main mutation - F508del as would be expected. There was also a wide range of disease severity, varying from patients listed for transplant, through to those with supra-normal lung function. Traditional microbiological techniques revealed PA and SA to be the main pathogens grown on culture. Consequently we have studied two cohorts with characteristics and microbiological status which will be comparable to many others.

Following DNA extraction of sputum from these cohorts, RISA profiles were successfully obtained from all samples. Of note, no RISA profiles could be generated from accompanying negative controls. This is very important, as sensitive molecule techniques such as 16S rRNA sequencing, can amplify DNA when processing negative control samples. When samples were viewed together via cluster analysis there were clearly some dominant bands. RISA cannot be used to formally identify the presence of pathogens by bands but our collaborators (E. Mahenthiralingam and M. Bull, Cardiff University) have an unpublished database of ITS band sizes. This allows us to speculate that prominent microorganisms in our communities may be PA, HI, SA and various Streptococcal species. However, this database highlights why RISA cannot be used primarily as a pathogen identification tool, as the band size for HI and PA can be very close. The bioanalyser has an error margin of +/- 5% and consequently two bands which appear identical could be either PA or HI. Indeed in our dendrograms, examples can be highlighted where a PA-positive sample is between two very similar HI samples. Others have also noted issues with close band lengths.[292] In addition, our collaborators database and other published work has shown how the same species can have more than one band length.[376] These are major limitations, though for other pathogens, RISA may give an indication of a likely pathogen which may direct towards more targeted investigation for that microorganism. For example, the two samples positive for MC both had a dominant bands around 800 bp (797 bp and 805 bp). Three out of the four samples positive for SMalt had a dominant band around 820bp (819-821 bp) and the other sample's second most prominent band was 823 bp.

Whilst RISA has limitations as a pathogen identification tool, what it does provide is a visual and quantifiable assessment of the bacterial community. For example, the image of a single dense band suggests a community overwhelmingly dominated by a single pathogen. The bioanalyser output allows quantification of this through band counting and giving a concentration for the band. These outputs can then be used to calculate estimates of ecological measures such as diversity, evenness and richness. These are the ecological measures which have commonly been used in investigation of the bacterial communities of the lung and have shown associations with clinically significant markers and phenotypes.[133, 286, 321] Whilst these ecological measures are inter-linked in their calculations, they do provide different descriptions of the community. Richness is the most straightforward and is a count of the number of taxa (whether defined by species, genera, OTU or other criteria). Evenness takes into account how equal the abundance is of the taxa that make up that community. Diversity takes into account both the richness and the evenness of that community. Whilst these metrics may all point in the same direction when describing how complex a community is, this is not always the case. For example, a community with just two species with similar relative abundance will have a low richness but a high level of evenness. Another community with 10 species may have a greater richness, but if this community is dominated by one species then the evenness may be very low and the diversity less than other communities with similar richness.

Ecological estimates in our cohorts suggest that our NCFB bacterial communities are less complex than our CF bacterial communities by all measures. The statistical significance of all of these measures gives strong confidence in this observation, as does the persistence after removal of non-stable state samples. In previous studies the diseases have not been directly compared as they have here, and this finding is both interesting and novel, and well worth considering the reasons why. The expectation may have been for the opposite. CF is often seen as a more severe disease and the situation of the lung being dominated by a single pathogen in late-stage disease is well described.[286, 329, 372] However, with improvements in management and increasing life expectancy, patients are transferring from paediatric to adult care further away from the end-points of lung transplantation or death. In fact a recently published paper assessing three transitioning groups found the greatest diversity in the group who

had most recently transitioned.[377] Within our CF cohort there is a great range of disease severity and a significant proportion have well-maintained lung function. We have shown an association between lung function and estimates of diversity and richness, so whilst the lack of complexity of end-stage CF patients may hold true, this is not the picture for the whole cohort. Another reason why the CF cohort may have been expected to have less complex communities is the antibiotic burden that these patients endure. A higher proportion of CF patients are on prophylactic antibiotics and have had courses of IV antibiotics in the last 12 months. It would therefore be expected that this would have an impact on their communities. However, neither the usage of prophylactic antibiotics nor the burden of IV antibiotic therapy was associated with a difference in any ecological measures in either cohort. On the face of it, this may be counter-intuitive, but this may not be a great surprise. Some previous work has demonstrated remarkable resilience of the bacterial communities of the diseased lung in the face of antibiotics.[324] The technique may also play a part. A technique with a lower sensitivity than 16S rRNA sequencing might not show the impact on less prominent bacteria which also contribute to the community. The lack of bands generated by negative control samples may highlight the lower sensitivity of RISA, and this reduced sensitivity compared to 16S rRNA sequencing has been previously recognised.[292]

If we think about why NCFB lungs may have less complex communities, then age would be a potential factor. The NCFB cohort is clearly markedly older and if diversity reduced with age, then this may be the explanation. However our data suggests an increase in diversity with age in NCFB. This is surprising as work in CF cohorts has suggested a trend of decreasing diversity with age.[312] This may in some ways be a quirk of sampling an unselected everyday NHS cohort. For a younger patient to be seen in clinic regularly they are likely to have significant disease, whilst those who have exceeded life expectancy may have milder disease as suggested by their longevity. The answer to why our NCFB bacterial community appears less complex may be at least partially explained by looking directly at the bacteria rather than focusing on exogenous reasons. In the NCFB cohort those who isolated HI appeared to have less complex communities than those with PA. We saw no evidence of co-infection with PA and HI, and this is not a novel observation.[270, 271] In the CF lung however, instead of HI, SA is

the other main pathogen seen. Unlike HI, SA appears to be associated with more complex communities than patients with PA. In addition, co-infection with SA and PA was common and appeared to be associated with a more complex community. It would thus seem plausible that a contributor to our observation is that HI dominates a community more than SA, and HI is seen far more in NCFB whilst SA is seen far more in CF. It is worth noting that statistically significant differences in the ecological measures persisted when the numbers assessed were reduced by just looking at samples taken in the stable state. Whether or not studies of bacterial communities in the lungs should just be performed on stable state samples are unclear. Whilst it may seem that by doing so will give a fairer representation, some studies have shown that very little change occurs during exacerbations and acute antibiotic therapy, though this is not universal.[319, 324, 378] Combining our two cohorts allowed comparison of 18 patients from whom we had obtained samples in both the stable and the exacerbation state, and in these patients we did not see a significant change in any of the ecological measures.

One of our primary aims in this study was to investigate the microbial profiles generated by RISA and their associations with clinical characteristics and disease severity. The main focus has been on NCFB due to the size of the cohort and due to the disease having not previously been investigated by this technique. The use of cluster analysis on the NCFB cohort sorted samples into groups with certain characteristics. For example, Group C had less complex communities whilst Group B profiles were dominated by a couple of bands with lengths around 530 and 560 bp. By *in silico* prediction this may represent *Streptococcus*. Other work has highlighted that *Streptococcus* is often not reported by traditional culture.[314] Interestingly, there was no discernible difference between the groups with regards to disease severity as defined by the Bronchiectasis Severity Index. There was also no significant difference with others simple markers of severity such as lung function and exacerbation rate. It may be seen as a limitation that RISA cannot highlight disease severity by cluster analysis, but there should not be a need for this as clinical assessment should be able to do this. The fact that it does not distinguish disease severity may in fact be a significant advantage. It would seem plausible that two patients with similar disease severity but very different microbial communities may benefit differently from the same

prophylactic therapy. Consequently, for drug trials, RISA may provide simple sub-group analysis for little added cost in patients without discernible clinical differences. If certain sub-groups benefitted more, it would be a technique that could easily be incorporated into the planning of long-term therapies in standard clinical care.

As well as the creation of dendrograms, the output from RISA has allowed us to describe communities with estimates of diversity, evenness and richness. These descriptions could be an adjunct to traditional culture to help provide a more detailed overview of the patient's bacterial community. For example, rather than a sample simply being reported as HI-positive, a report of HI-positive within a community of high richness, low evenness and moderate diversity, may be more helpful. As our understanding of the importance of the microbiota increases, this simple addition to clinical reports may aid the managing clinician and therefore the patient.

For RISA, or an alternative technique, to be really useful the ecological measures need to mean something with regards to the impact of the disease on the patient. We have not demonstrated an association with disease severity as defined by the BSI. However, there does need to be some caution not to over-interpret this result. The BSI is a scoring system which provides clinicians an overview of disease severity and researchers a quantifiable measure. It does award a significant number of points for being colonised by PA and being elderly. To illustrate this impact, if you are 80 years old and have isolated PA in sputum twice in the last 12 months, you would be categorised as severe even if you had no other markers of disease. There is a suggestion in our data that PA may not be behaving more pathogenically than HI in our local cohort, and Devon does have a large geriatric population. The BSI also awards points for hospital admissions. This often occurs when a patient requires IV antibiotics. We provide a home IV service which tends to be used more by younger patients who self-administer their own drugs. Consequently a patient may have had courses of IV antibiotics without being admitted to hospital. This may reduce the score in some of our younger patients and hence distort their reported severity. For many patients, rather than a score, what matters to them is their frequent symptoms including exacerbations and breathlessness. In our cohort, patients with low exacerbation rates had more diverse and richer communities, whilst lung function was

associated with a richer community. These findings clearly do not show causation, though the concept of a more complex community being less susceptible to exacerbations is appealing and greater investigation is warranted.

Beyond lung function and exacerbation rates, other results of note in the NCFB cohort included the observation of reduced evenness in culture-positive samples and less complex communities in rural dwellers. The reduced evenness in culture-positive samples is logical and implies that the more prominent a pathogen is in a community, the more likely it will reveal itself when that community is used to inoculate agar plates. The reduced complexity of those in rural areas (defined as not living in a town or city) is interesting. It is not mechanistically clear why this is the case, as the opposite may have been expected. We have not recorded duration of living in that setting or where the patient was brought up as a child, and consequently this result may not be a reflection on the rural environment.

The CF cohort was also analysed on its own, however its smaller size needs to be considered when assessing results. The cluster analysis did not display multiple distinct clusters of significant size like in the NCFB cohort. This may suggest some greater heterogeneity of CF bacterial communities which should be expected in the cohort with the greater complexity of its communities. Of note, reduced diversity and richness was associated with worse lung function, whilst evenness was not. This is very similar to the results seen in our NCFB cohort and compatible with previous work when FEV_1 was used as a primary marker of disease severity.[133] Clearly causation cannot be claimed, but it would seem plausible that airflow obstruction could impact on the clearance of respiratory secretions and accompanying microbiota, and that certain communities with reduced diversity and richness may cause further airway damaged and worsening airflow. There has been work which raises the possibility of differential airway remodelling according to the bacteria present in the airway via matrix metalloproteinases.[379] When looking at culture results and ecological measures, less complex communities were seen for PA-positive samples and culture-negative samples, and reduced diversity in PA-positive samples has been observed before.[270] The culture-negative group is also of interest and can pose significant challenges in clinical decision-making due to the lack of a pathogen to target. The ecological pattern seen in our CF cohort is the opposite of our NCFB

cohort and may represent a different threat in CF where difficult to culture pathogens may dominate communities.[133] Another consideration in these patients may be the impact of fungal disease, which is not investigated by the RISA technique. The final point to highlight in the CF cohort is the lack of impact of antibiotic burden on community measures. The literature is mixed on the impact of antibiotics and ecological measures. Significant resilience in the microbial community can be seen, whilst there is also the suggestion that antibiotics drive loss of diversity in CF.[314, 315, 319, 329, 377] These concepts do challenge previously held views on what we are achieving, or looking to achieve, with antimicrobial therapy and the management of chronic lung disease. The concept of germ theory appears too simplistic in chronic infection of the lung. Consequently the concept of looking to kill one pathogen may also be flawed.

There is no previous work looking at bacterial communities in NCFB by RISA, however if RISA is a useful technique, then we would expect some similar findings to those seen in studies using the gold-standard technique of 16S rRNA sequencing. The previous findings in NCFB are mixed, and consequently it is unsurprising that some of our results are in keeping with some of the literature. Work by Rogers et al on a cohort from an RCT assessing long-term macrolides, reported a positive correlation between diversity and FEV_1.[321] Whilst we did not quite reach statistically significance for this, we did see a trend towards, and we did find a positive correlation between FEV_1 and richness which was also found by Rogers et al. However, reviews of unselected cohorts of NCFB at centres in London and the North-East of England failed to show an association between FEV_1 and diversity.[271, 319] Previous studies had not compared ecological measures to the BSI. In the CF cohort the findings of FEV_1 correlating to ecological measures has been consistently found.[133, 377] However, the seeming lack of impact of antibiotic prophylaxis on ecological measures in both cohorts is at odds with some previous work in both diseases.[314, 319]. As discussed previously, this may be due to a lack of sensitivity in the technique, the stability of the community or a combination of both.

With regards to the CF cohort itself, it is difficult to make direct comparisons with the most relevant work of Flight et al that also used RISA.[133] Many of their conclusions were based on the accompanying 16S rRNA sequencing data which we do not have. Their cluster analysis contained 9 more patients than ours and

divided into two large clusters. Our cluster analysis was performed in the same way and with collaboration with the same team, and yet if we divide our cluster analysis at the same level of similarity, we have 6 different clusters with only one of them containing more than 5 samples. The only clear difference between the process used by Flight et al and ours is the technique used for extraction of DNA. As we used mechanical disruption as part of our protocol, this may have changed the type of species extracted and hence lead to less tight clustering. The other difference of note was in the pathogens cultured. For example, we saw more SA and they found more BCC. This would very much be expected to influence the output.

There are several strengths to highlight in both this study and the technique described. In the current literature there is no current description of bacterial communities in NCFB in a cohort of this size. In addition, none of the preceding work takes into account the BSI or provides direct comparison to a CF cohort. This is also the first study to describe RISA profiles in a NCFB cohort. The technique has shown itself to be easy to perform and analyse. It has also been performed at a fraction of the cost and bioinformatics processing of 16S rRNA sequencing. This makes it a realistic technique in day-to-day clinical practice. Its limitations however must be noted. The 5% error margins of the bioanalyser and the small difference in band length among different potential pathogens means this cannot be used for diagnostic identification. In addition, there is no previous data to validate its estimates of ecological measures alongside the gold-standard technique, however we have performed this comparison and describe it in the next chapter. The likely reduced sensitivity compared to 16S rRNA sequencing increases the chances of samples having only a single band seen. A single value results in an e-Diversity and e-Evenness scores of 0. Whilst these samples are likely to have low diversity and evenness, using other more sensitive techniques, the values may well not be that low. For the purposes of this study, the non-parametric distribution means that the statistical tests are unlikely to be affected, and if the question asked is whether or not the sample has low diversity or evenness, then again this is not a significant issue. Limitations in this study beyond the technique include the sample size and the lack of longitudinal samples and clinical follow-up. However, as an initial step this study demonstrates what RISA is capable of and the feasibility of its use in larger

longitudinal trials. Additional limitations include the focus exclusively on bacteria. The impacts of fungi and viruses are beyond the scope of this work and neither RISA nor 16S rRNA sequencing can provide insight beyond bacteria. It may be that for better understanding of the microbiome that parallel investigations are required which take into account the virome and the mycobiome.

When considering the role of RISA in the future, it is important to remind ourselves where our understanding is at the moment. Research so far has highlighted associations between communities and patient phenotypes in chronic suppurative lung disease, other lung pathologies and in other organs. This has not however led beyond hypotheses of the importance of these observations. To move the field forward, the inclusion of microbial community assessment in large longitudinal clinical trials is essential. This may then give understanding into the real-world significance of these findings and which therapies should be administered to patients with different bacterial communities. For this to be done on a large scale in clinical trials, and then be used in subsequent clinical practice, the technique should ideally be simple, quick and cheap. RISA offers these attributes and therefore whilst it has clear limitations, it should be considered as a technique to take the field forward.

6.6 Conclusion

In this chapter we have described and shown the utility of RISA in a NCFB cohort and an accompanying CF cohort. The technique has allowed estimates of ecological measures of the bacterial community and has suggested a greater community complexity in the CF lung than in NCFB. It has also highlighted associations between community measures and clinically-important markers such as exacerbator phenotype in NCFB and lung function in NCFB and CF. RISA holds potential as an adjunct to traditional culture and stratifying patients for future trials and therapies. However, it needs validation of its output by 16S rRNA sequencing and this is the focus of the next chapter.

Chapter 7- Microbiota characterisation of a bronchiectasis cohort by 16S rRNA sequencing with comparison to RISA analysis

7.1 Abstract

Introduction: It has become clear that human lungs contain far more complex bacterial communities than previously appreciated. However, the clinical significance of these communities in NCFB is unclear, as is the best approach to practically measure them. In this chapter we look into the bacterial communities and patient characteristics, and compare community measures by a gold-standard technique (16S rRNA sequencing) to a cheaper and easier technique (RISA).

Methods: A subgroup of samples from the NCFB cohort described in Chapter 6 underwent interrogation by 16S rRNA sequencing. This technique was used to identify the genera making up these communities and to investigate for associations with patient characteristics. Comparisons were made with the ecological measures estimated by RISA and the sample characteristics proposed via cluster analysis.

Results: Seventy-two samples from 68 patients underwent interrogation by 16S rRNA sequencing. The most common genera by mean relative abundance were *Streptococcus*, *Haemophilus* and *Pseudomonas*, and these genera were the most prevalent in over 80% of samples. Those with *Streptococcus* as the most prevalent genera had greater community complexity than those with *Haemophilus* or *Pseudomonas* as the most prevalent. Whilst community measures did not reveal an association with disease severity as such (as defined by the BSI), there was a negative correlation between exacerbation frequency and diversity and evenness. Similarly, there was a positive correlation between exacerbation frequency and *Haemophilus* abundance. This analysis demonstrated a correlation between communities profiles as described by 16S rRNA sequencing and RISA.

Conclusion: In our cohort 16S rRNA sequencing has revealed complex communities which may have an impact on exacerbation frequency in NCFB. We have shown that RISA provides useful insight into the ecological characteristics of a bacterial community and may have utility in widespread routine use as a surrogate for 16S rRNA sequencing.

7.2 Introduction

In this thesis we have described and shown how molecular techniques reveal far greater detail and insight into the microbiology of the bronchiectatic lung found in CF and NCFB patients. An important concept in this field has been the realisation that the healthy lung is not sterile and in fact contains complex polymicrobial communities. A principle technique used in recent years has been 16S rRNA sequencing. This essentially allows a listing of bacteria and their relative abundance. From this, calculations of ecological measures are made such as diversity, evenness and richness. These measures appear to be associated with clinically significant markers in diseases both in and outside the lung. In asthma, bronchial hyper-responsiveness has been shown to be associated with bacterial diversity, and increased bacterial diversity is possibly protective against asthma development.[380, 381] In COPD, airflow obstruction has been associated with reduced diversity.[382] In CF, a loss of diversity has been associated with disease severity and less complex communities with quicker decline in lung function.[314, 316] In the gut, reduced diversity has been identified in inflammatory bowel disease.[373]

In NCFB the message is muddier. One of the earlier prominent works suggested a stable microbiome despite antimicrobial treatment.[324] Subsequent work by Rogers et al then suggested that bacterial diversity correlated with lung function and that patients' risk for exacerbation could be stratified by their bacterial community.[321, 323] However, there has been other work which has not linked clinical characteristics and disease severity with the microbial community measures.[271, 319, 324] Importantly, none of these studies have looked at using the Bronchiectasis Severity Index (BSI) to compare the community to the disease severity. Whilst it is unclear, it would seem likely that different bacterial communities are likely to behave differently, and also respond differently to both the host immune response and therapies like antibiotics. Consequently better

understanding of the association between disease, therapy and the bacteria communities in the lung environment is likely to be useful.

If appreciation of the bacterial community's ecological structure does indeed turn out to be useful with regard to the management of NCFB or other conditions, then a rapid, cheap and accessible test would have large appeal for everyday healthcare. Despite the falling costs of sequencing, 16S rRNA sequencing still has a significant cost and complexity, and requires significant bioinformatics processing before even considering its clinical significance. In Chapter 6 we have described RISA, a cheap and accessible profiling technique that allowed us to estimate the ecological measures of diversity, evenness and richness.

In the context of the above issues, we have conducted a study investigating the microbiota of NCFB patients in relation to potential influencing factors such as their disease severity, their previous therapy and their environment. The molecular investigation has been performed on samples which had previously undergone RISA, and a further aim of this chapter is to compare the information gleaned from that technique with the view of validating it as an appropriate method.

7.3 Methods

A subset of samples from those investigated in Chapter 6 were sent for 16S rRNA amplicon sequencing with the ESS, as described in Chapter 2. The samples were chosen exclusively from the NCFB cohort. The first stable sample available from each patient was submitted if available, resulting in 85 samples being submitted. Five negative controls were also submitted, generated by performing the DNA extraction process on Mucolyse from the NHS microbiology laboratory, which had not been mixed with sputum (this process was done once from each of the five batches of the DNA extraction kit used). In addition, DNA was extracted from one aliquot of the DES used to dilute samples for submission. Finally, a further five culture-negative samples were submitted in order to fill a 96-well plate for sequencing. Following sequencing, any sequence which had at least 10% of its reads in the negative controls was excluded from downstream bioinformatics analysis by the Sequencing Service. In addition, one sample was missed off the sequencing run and seventeen were removed from final analysis due to low concentrations of DNA after initial amplification, and hence preventing confidence

that all samples had been treated equally. This left 68 stable samples and a further 4 others.

7.4 Results

7.4.1 Sequencing Output and Read Removal

In total 90 clinical samples and 6 negative controls were submitted for 16S rRNA sequencing with the ESS. One sample was left off (due to laboratory error) and the remaining 95 underwent sequencing. From these, 8447 different sequences were obtained - of which 141 were only found in negative controls and 4 were found in all negative controls. These 4 sequences were identified at genus level as *Delftia*, *Achromobacter*, *Ralstonia*, and Propionibacterium. Sequences were removed when greater than 10% of the total reads came from the 6 negative controls. This left 8177 sequences. For 17 samples there was un-recordable quantity of DNA at a point in the process, and these samples had a higher proportion of contaminant reads. To ensure confidence in our results, these were removed from further analysis leaving 72 samples available for analysis. From these 72 samples, there were 7289 different sequences and 482 when collapsed to genus level for the purposes of taxonomic binning.

The bioinformatics pipeline DADA2, gave reported resolution to sequence level rather than OTUs, with the next level of differentiation at genus level. It was debated which taxonomic level would be used for the primary analysis of the ecological measures of diversity, evenness and richness. These measures were calculated by both methods and showed very high concordance (see Figures 7.1, 7.2 and 7.3). Due to such high concordance, it was felt that either could be used. We chose the taxa of genus due to issues with accurate species-level identification for many sequences. For example, one sequence identified as *Burkholderia ambifaria / anthina / arboris / cenocepacia / cepacia / contaminans / diffusa / glumae / lata / metallica / multivorans / plantarii / pyrrocinia / sp / stabilis / vietnamiensis*.

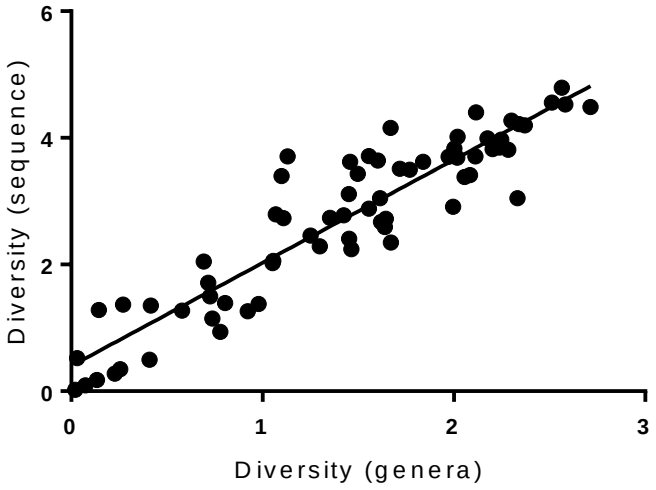

Figure 7.1 Diversity calculated at sequence and genera level Analysis performed by linear regression. P<0.001 and R^2= 0.8539.

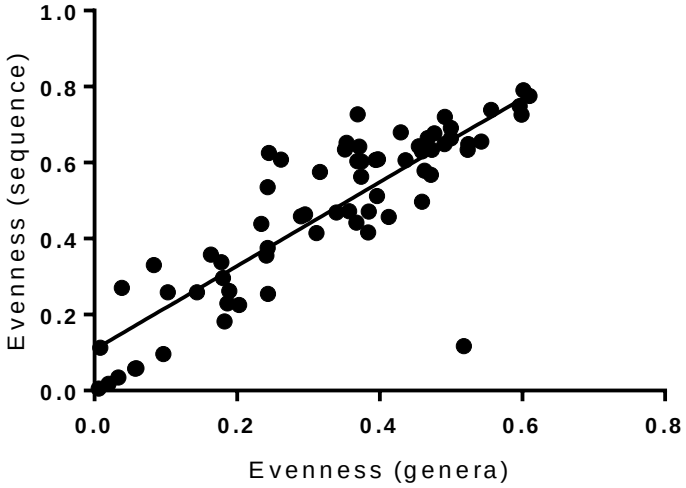

Figure 7.2 Evenness calculated at sequence and genera level Analysis performed by linear regression. P<0.0001 and R^2= 0.7255.

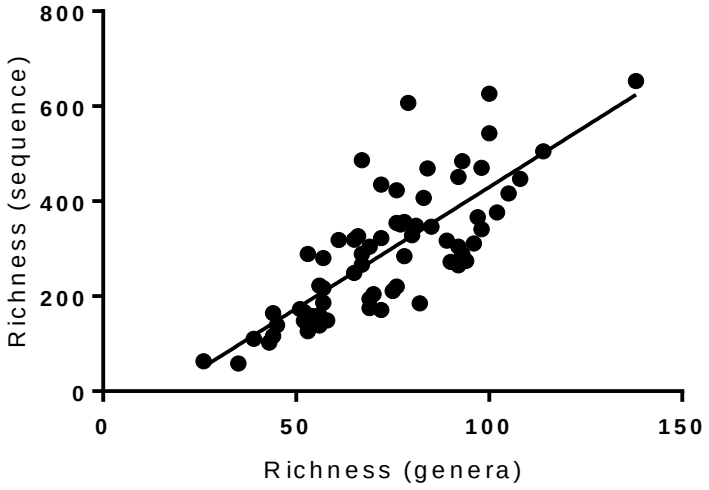

Figure 7.3 Richness calculated at sequence and genera level Analysis performed by linear regression. $P<0.0001$ and $R^2= 0.6178$.

7.4.2 Comparison of 16S rRNA Sequencing Subgroup to Original Cohort

The samples were selected for 16S rRNA sequencing by the first stable sample available for each patient. A further 4 additional samples were also investigated and came from patients who were already included. These additional samples are only included for comparison between techniques (e.g. comparing ecological measures generated by RISA and 16S rRNA sequencing). For all other analysis, the 68 stable samples with just one per patient are used. There were no significant differences found between the overall cohort which were described in Chapter 6 and the subset assessed here (see Table 7.1).

	Total Cohort	Subset	P Value
No	99	68	
Age, yrs.[a] (Range)	65.5 (17-88)	66 (17-88)	0.8198
BSI[a] (Range)	7.3 (1-18)	7.3 (2-18)	0.9714
%Female[b]	64.65	67.65	0.6878
FEV_1 %predicted[a] (Range)	68.94 (23-116)	68.87 (25-116)	0.9840
Colonised at Recruitment[b]			
- **Any**	41.41%	39.71%	0.8253
- **PA**	20.20%	20.59%	0.9514
- **HI**	17.17%	17.58%	0.9365
BMI[a]	25.07 (5.45)	25.58 (5.87)	0.5700
Exacerbations in last 12 months[c]	3.42 (3.06)	3.34 (3.05)	0.9235
MRC Dyspnoea score[a]	2.404 (0.97)	2.426 (0.98)	0.8808
Antibiotic Prophylaxis[b]			
- **Any**	50.51%	44.12%	0.4349
- **Oral**	40.40%	36.76%	0.7469
- **Inhaled**	19.19%	16.18%	0.6850
<10 pack year smoking history[b]	73.74%	72.06%	0.8600

Table 7.1 Comparison of demographics of NCFB cohort initially recruited and the subgroup that underwent 16S rRNA sequencing The data in this table is presented as mean values or percentages of that group. In those continuous measures where range is not of specific interest, standard deviations are reported

alongside the mean value. Statistical comparison was performed by [a]unpaired t-test, [b]Chi-squared or [c]Mann-Whitney test.

7.4.3 Relative Abundances of Genera and Ecological Measures

From the total reads within each sample, relative abundances were calculated. When the mean relative abundance was calculated from the 68 samples at genera level, only three were over 10%, namely *Streptococcus* (27.49%), *Haemophilus* (21.88%) and *Pseudomonas* (10.94%). The 15 most abundant general are shown in Figure 7.4.

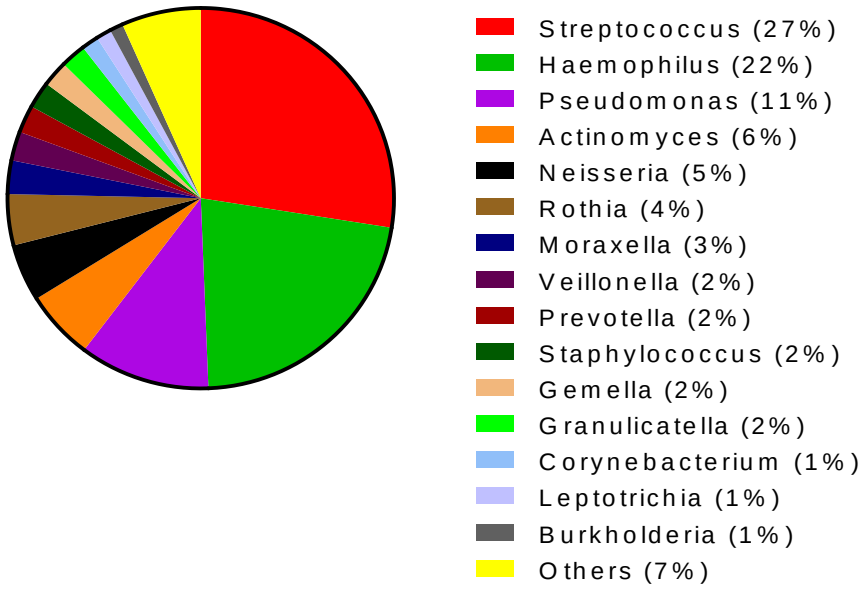

Figure 7.4 Mean relative abundance by genera in NCFB cohort

There was highly variable abundances of some genera across the samples as highlighted by *Haemophilus* and *Pseudomonas* having noticeably lower median values (4.22% and less than 0.01% respectively) than their mean value. *Streptococcus* had a similar median value (24.81%) to its mean value, demonstrating it's more consistent presence across communities. When assessing genera by frequency of being the most prominent genera within a given sample, these same three genera were the most prominent, and between them were the most prominent in 82.35% of all samples (see Fig 7.5).

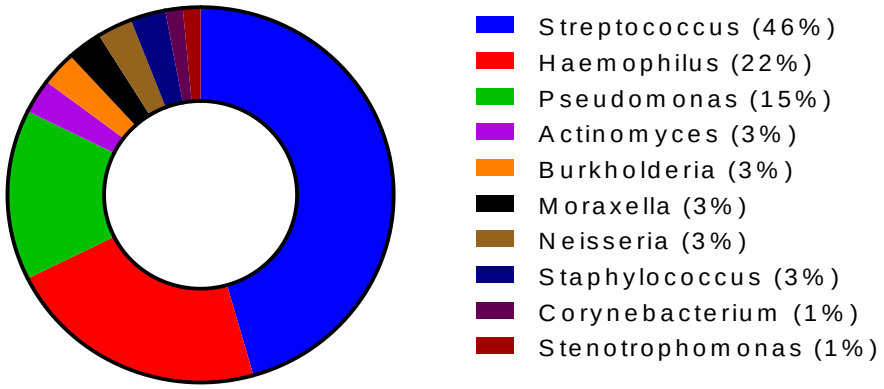

Figure 7.5 Proportion of communities dominated by each genera

The ecological measures of diversity, evenness and richness were compared between samples dominated by each of the three main genera and the others. A genera was considered to dominate a sample if it was the most prevalent. ANOVA assessment revealed a significant difference in diversity ($p < 0.0001$). This was due to differences between the *Streptococcus*-dominated and the *Haemophilus*-dominated samples, and between the *Streptococcus*-dominated and the *Pseudomonas*-dominated samples. In both instances *Streptococcus*-dominated samples had a higher diversity score (see Fig 7.6). The same pattern was also seen with regards to evenness, with *Streptococcus*-dominated samples being more even than *Haemophilus* - ($p<0.0001$) and *Pseudomonas* - ($p=0.0151$) dominated samples (see Fig 7.7). With regards to richness, there was a slightly different pattern, though again significant differences where seen by ANOVA. These were due to differences between *Haemophilus*-dominated samples and samples dominated by *Streptococcus* ($p=0.0004$), and between *Haemophilus*-dominated and those dominated by a bacteria other than *Streptococcus* or *Pseudomonas* ($p=0.0002$) (see Figure 7.8).

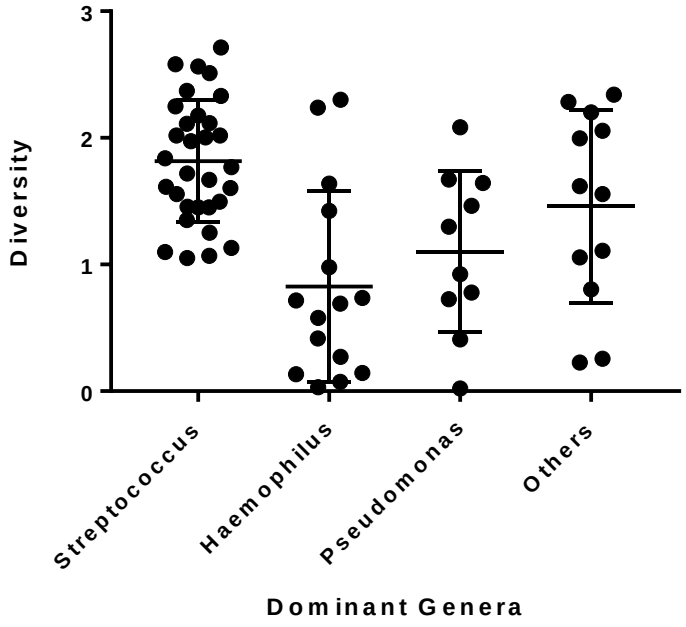

Figure 7.6 Diversity depending on the dominant genera Bars represent the mean with standard deviations.

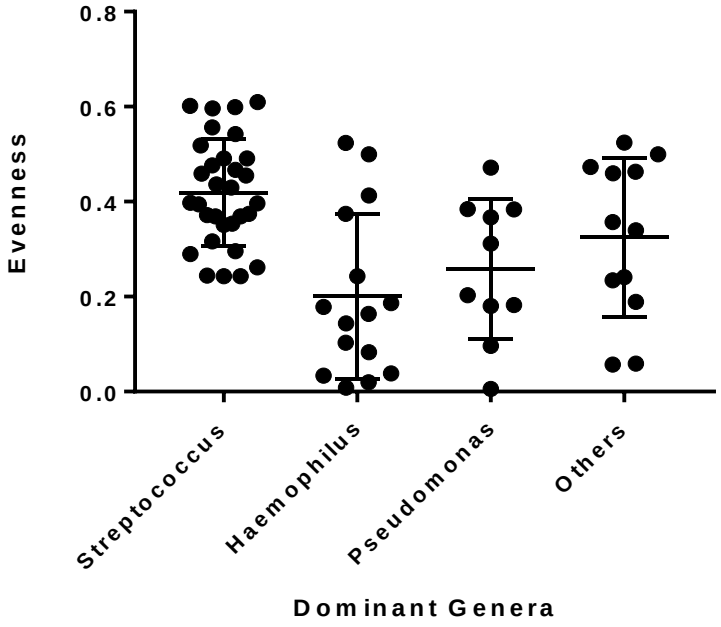

Figure 7.7 Evenness depending on the dominant genera Bars represent the mean with standard deviations.

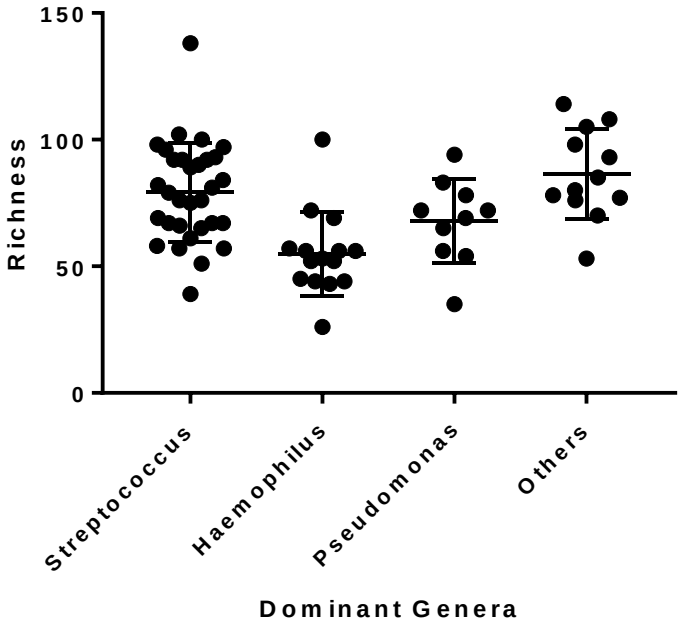

Figure 7.8 Richness depending on the dominant genera Bars represent the mean with standard deviations.

The above measures suggest that when either *Haemophilus* or *Pseudomonas* is the most prominent genus, the communities appear less complex than those dominated by *Streptococcus*. In order to investigate further, correlations were assessed between the presence of these three genera and the presence of other common genera found. As would be expected of *Haemophilus* and *Pseudomonas* (given their domination of less complex communities), no positive correlations were seen between the presence of either of these genera and the presence of other genera. Indeed, multiple negative correlations were seen. For *Haemophilus*, these negative correlation were seen with *Pseudomonas*, *Actinomyces*, *Rothia*, *Veillonella*, *Prevotella* and *Streptococcus*, whilst for *Pseudomonas*, the negative correlations were seen with *Veillonella*, *Streptococcus* and *Haemophilus*. On the other hand, *Streptoccocus* actually had positive correlation with *Rothia* and *Veillonella*. Other positive correlations were observed between other genera, as listed in Table 7.2.

Genera	Positive Correlates
Streptococcus	*Rothia* *Veillonella*
Actinomyces	*Veillonella* *Prevotella,*
Rothia	*Streptococcus*
Veillonella	*Prevotella* *Streptococcus* *Actinomyces*
Prevotella	*Actinomyces* *Veillonella*

Table 7.2 Positive correlations found between the relative abundances of the common genera

7.4.4 Comparison of Traditional Culture and 16S rRNA Sequencing Results

Comparisons were made between the culture results and 16S rRNA sequencing results from the 72 samples that had undergone both techniques. Of these 72 samples, 39 isolated potential pathogens and are referred to as culture-positive samples, whilst 33 did not and are referred to as culture-negative samples. When looking at results at this initial sub-division, there were clear differences in community complexity and composition. Culture-negative samples were significantly more complex when assessed by the measures of diversity ($p=0.0026$), evenness ($p=0.0046$) and richness ($p=0.0082$) (see Table 7.3). When looking at samples which were culture-positive for HI or PA, there was a trend towards lower diversity and evenness in the HI-positive samples, without quite reaching statistical significance. When reviewing composition with regards to mean relative abundance, there appears to be a higher abundance of *Haemophilus* and *Pseudomonas* in culture-positive samples (see Figure 7.9), although these do not reach statistical significance. There were however

significantly higher relative abundance of *Streptococcus*, *Actinomyces*, *Neisseria*, *Rothia* and *Prevotella* in the culture-negative samples.

	Culture-positive	Culture-negative	P=
Diversity	1.20	1.70	0.0026
Evenness	0.28	0.39	0.0046
Richness	67.54	80.85	0.0082

Table 7.3 Ecological measures of culture-positive and culture-negative samples Comparison of measures was performed by the unpaired t-test.

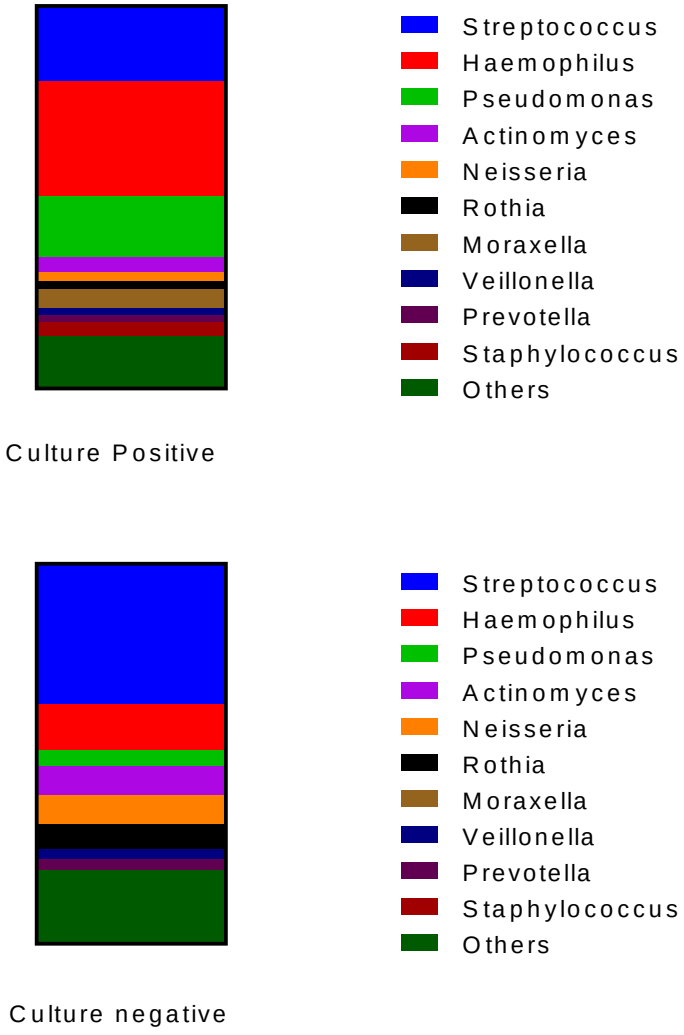

Figure 7.9 Mean relative abundance of genera in culture-positive and culture-negative samples

The genus with the highest mean relative abundance in culture-negative samples was *Streptococcus*, as opposed to *Haemophilus* in the culture-positive samples. When assessing the most prominent genus in culture-negative samples, *Streptococcus* was the most prominent in 22 of the 33 samples (66.67%). As well as these 22 samples, a further 10 culture-positive samples had *Streptococcus* as the most prominent genus. However, in these 10 samples, *Streptococcus* was not isolated at all by traditional culture. Instead, *Haemophilus influenzae*, *Staphylococcus aureus*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, and *Stenotrophomonas maltophilia* were isolated from these 10 samples. By 16S rRNA sequencing, *Haemophilus* was the most common genera in 15.15% of the culture-negative samples and 30.2% of the culture-positive samples. Of the 12 culture-positive samples in which *Haemophilus* was deemed the most abundant genus, 11 isolated HI and one isolated *Streptococcus pneumoniae*. *Pseudomonas* was the most prominent genera in 3.03% of culture-negative samples and 23.08% of the culture-positive samples. In these culture-positive samples, all of them isolate PA and one also isolated SA. The pathogens isolated by culture when a certain genera is deemed the most prevalent by 16S rRNA sequencing, are listed in Table 7.4.

Dominant Genera/Culture Result	BMV	HI	KP	MC	PA	SA	SerM	Smalt	Spn	Culture-Negative
Streptococcus		4	1		3	1		1		22
Haemophilus		11							1	5
Pseudomonas					9	1				1
Actinomyces										2
Burkholderia	1	1								
Moraxella				2						
Neisseria										2
Staphyloccocus		1				2				
Corynebacterium							1			
Enterococcus										1
Stenotrophomonas								1		
Total	1	17	1	2	12	4	1	2	1	33

Table 7.4 Culture results when a certain genera is most prevalent BMV- *Burkholderia multivorans*, HI- *Haemophilus influenza*, KP- *Klebsiella pneumoniae*, MC- *Moraxella catarrhalis*, PA- *Pseudomonas aeruginosa*, SA- *Staphylococcus aureus*, SerM- *Serratia marcescens*, Smalt- *Stenotrophomonas maltophilia*, Spn- *Streptococcus pneumoniae*.

Whilst the bioinformatic pipeline attempted to provide species level identification, such results tended to be ambiguous (only 12% of sequences detected had an unambiguous identification). *Klebsiella pneumoniae* was the only pathogen isolated by culture that was not identifiable by the accompanying 16S rRNA sequencing data. There was wide range of relative abundances seen for the genera of the pathogens identified by culture (Table 7.5). Of note, a sample had 1% abundance for *Staphylococcus* but was reported by culture as positive for SA and Candida. The only other genera to have less than 10% relative abundance but to be isolated by culture was *Pseudomonas* (see Table 7.5). When looking at culture-negative samples, there were examples of high relative abundance of the important genera of *Pseudomonas* and *Haemophilus*. One sample was 93.33%

Pseudomonas (of which at least 99.99% of the *Pseudomonas* reads were identified as PA), yet did not isolate PA. Twelve samples had a relative abundance of Haemophilus above 10% without isolating a *Haemophilus* species, including a sample with 67% relative abundance of *Haemophilus* (of which at least 66.34%% of the *Haemophilus* reads were identified as *Haemophilus influenzae*). Other noteworthy findings included the near complete dominance of *Moraxella* in the two samples which isolated it (both greater than 96% relative abundance), whilst it's relative abundance was never above 0.38% in any other sample. Also of note, the technique was able to report a relative abundance of *Staphylococcus* of 74.71% in a sample. This is noteworthy due to concerns with 16S rRNA sequencing not picking up SA.[280, 319]

Culture	N=	Lowest abundance (genera)	Highest abundance (genera)
BMV	1	41.89%	41.89%
HI	17	23.66%	99.66%
KP	1	0	0
MC	2	96.12%	96.32%
PA	12	9.47%	99.78%
SA	4	1%	74.71%
SerM	1	19.9%	19.9%
SMalt	2	24.55%	40.56%
SPn	1	18.46%	18.46%

Table 7.5 Range of relative abundance of relevant genera in culture-positive samples BMV- *Burkholderia multivorans*, HI- *Haemophilus influenza*, KP- *Klebsiella pneumoniae*, MC- *Moraxella catarrhalis*, PA- *Pseudomonas aeruginosa*, SA- *Staphylococcus aureus*, SerM- *Serratia marcescens*, Smalt- *Stenotrophomonas maltophilia*, Spn- *Streptococcus pneumoniae.*

7.4.5 Patient Characteristics and Ecological Measures

Patient characteristics were assessed alongside ecological measures as determined by 16S rRNA sequencing. For each characteristic, comparisons were made with diversity, evenness and richness. No significant correlations were found for multiple factors including age, BMI and FEV_1, and urban or rural living.

Disease severity as defined by the Bronchiectasis Severity Index (BSI) was assessed in two ways: (1) comparison between the BSI total score and the ecological measures, and (2) comparison between the categories generated by the BSI (i.e. mild, moderate and severe) and the ecological measures. No significant associations were found by either approach.

Patients were grouped by exacerbator phenotype (low, medium or high, as described in Chapter 6) based on their exacerbation frequency of the last 12 months. Whilst RISA-based ecological measures correlated with exacerbator phenotype (Chapter 6), equivalent 16S rRNA-based ecological measures did not. However, if the actual number of patient-reported exacerbations was compared to 16S rRNA-based diversity and evenness, then a significant correlation was seen ($p= 0.0085$; $R^2= 0.1003$ and $p= 0.0075$; $R^2= 0.1035$) (see Figures A7.1 and A7.2). There was one significant outlier with regards to reported exacerbations, and the correlation by evenness remained even after removing this patient.

The relative abundance of the 10 most common genera were compared against multiple clinical and patient characteristics. There was a positive correlation with age with regards to the *Streptococcus* genera ($p= 0.0202$; $R^2= 0.07901$), but a negative correlation with *Staphylococcus* ($p=0.0137$; $R^2= 0.08851$). As would be expected there was a positive correlation between *Pseudomonas* abundance and disease severity as defined by the BSI (see Figures A7.3). There was also a positive correlation between exacerbation frequency and abundance of *Haemophilus*, but not *Pseudomonas*.

Additional comparisons were made between patient therapies and ecological characteristics. There was a trend towards greater diversity and evenness in those not on inhaled corticosteroid, but this did not quite meet statistical significance. No significant associations were made between PPI usage, oral antibiotic prophylaxis or nebulised prophylaxis.

7.4.6 Comparison of RISA Output to 16S rRNA Sequencing Data

In chapter 6, cluster analyses were performed on the output provided by RISA. The baseline samples for the NCFB cohort divided into four clusters of varying size. Of these samples, 65 underwent 16S rRNA sequencing and this allowed insight into the characteristics of these previously described groups.

One of the RISA groups (Group C), appeared to have reduced diversity and evenness on the basis of the RISA analysis. This finding was supported by 16S rRNA sequencing data (see Figs 7.10 and 7.11). In addition, when using culture results, Group C had a higher proportion of *Haemophilus influenzae*. This was also found to be consistent with 16S rRNA sequencing data as 9 out of the 17 samples were dominated by *Haemophilus*. This compared to only 6 samples being dominated by *Haemophilus* out of the remaining 48 samples clustered into other groups. When assessing *Haemophilus* with a relative abundance criteria rather than a community dominance view point, there was a significant difference between the RISA groups due to the high abundance in Group C and the low abundance in Group B. Whilst the levels of abundance were also low in the other two RISA groups, they did not reach significance (likely due to low numbers).

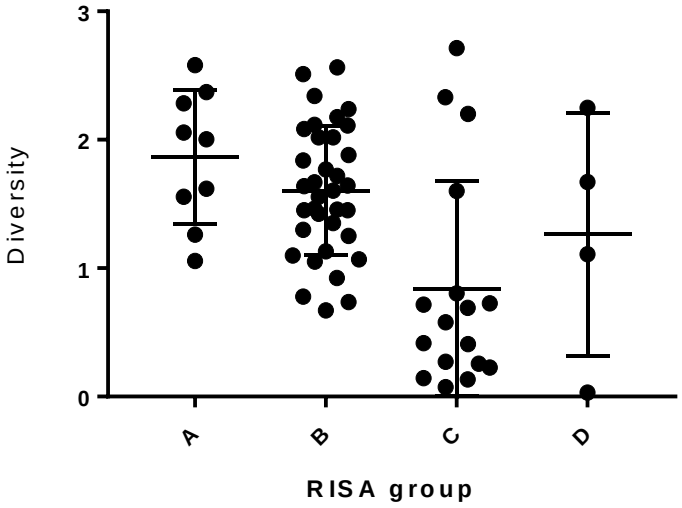

Figure 7.10 Comparison of 16S rRNA-based diversity of each RISA group Comparison were performed by ANOVA. Significant difference between A-C (p=0.0013) and between B-C (p=0.0008). Bars represent the mean value and standards deviation.

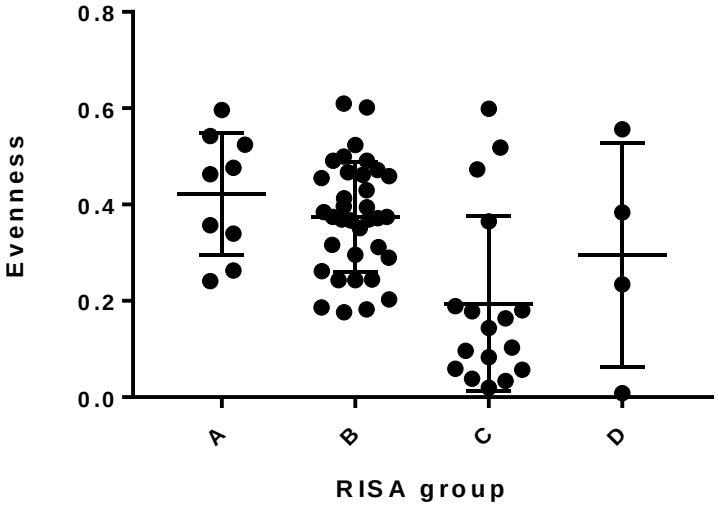

Figure 7.11 Comparison of 16S rRNA-based evenness of each RISA group Comparison were performed by ANOVA. Significant difference between A-C (p=0.0016) and between B-C (p=0.0005). Bars represent the mean value and standards deviation.

The largest group in the RISA analysis was Group B. This group had a large proportion of culture-negative samples and prediction based on RISA fragment lengths had suggested that this group may be dominated by *Streptococcus*. This again appeared to be consistent with the 16S rRNA analysis. Out of the 35 samples from Group B which underwent 16S sequencing, 23 were dominated by *Streptococcus*. Only 6 samples out a possible 30 were dominated by *Streptococcus* and clustered into the other RISA groups.

The estimated ecological markers of diversity, evenness and richness obtained from RISA were compared to the values generated by 16S rRNA sequencing. These comparisons were done using the values obtained by sequence, genera, and also looking at only the genera with a relative abundance over 1% ("1% genera"). Significant association was seen for all comparisons except richness by genera (see Table 7.6).

	Diversity (genera)	Diversity (sequence)	Diversity (1% genera)
p=	<0.0001	<0.0001	<0.0001
R^2=	0.32	0.3539	0.3282
	Evenness (genera)	**Evenness (sequence)**	**Evenness (1% genera)**
p=	<0.0001	<0.0001	<0.0001
R^2=	0.3418	0.3627	0.3767
	Richness (genera)	**Richness (sequence)**	**Richness (1% genera)**
p=	0.1815	0.0012	<0.0001
R^2=	0.02536	0.1404	0.1928

Table 7.6 Association between ecological measures estimated by RISA and calculated by 16S rRNA sequencing. The values are generated by linear regression.

7.5 Discussion

In this chapter I have examined the microbiota in a cohort of bronchiectasis patients with the use of 16S rRNA sequencing technology. This has provided insight into the bacteria present in a NCFB cohort from a region of the UK not previously investigated by culture-independent techniques. This has allowed comparison between bacterial communities and patient characteristics - an area of some conflicting evidence. It has also allowed comparison with a simpler and cheaper technique described in Chapter 6, namely RISA.

As part of the work in Chapter 6, samples were collected from both NCFB and CF patients with multiple samples from some patients. Whilst all samples were able to undergo RISA, it was not feasible for all samples to be interrogated by 16S rRNA sequencing. Due to capacity on a single run being 96 (and this being a feasible number), this number of samples were included. In order to obtain negative controls from all 5 batches of kit used for DNA extraction, and from the diluent used to standardise the samples, 6 of the 96 wells were used for these negative controls. It was decided to only use NCFB samples to maximise the opportunity for clinical correlates in this condition, whilst still enabling comparison

to RISA. There are suggestions from the literature that the microbial communities in several lung diseases can be relatively stable despite exacerbations and antimicrobial therapy.[324, 378] However for consistency we aimed to target samples obtained only in the stable state. We had samples from 85 bronchiectasis patients which met this requirement. A further 5 samples were submitted which were taken from a non-stable state. This was to utilise space on the sequencing run and provide added data for the comparison with RISA. For a proportion of samples, insufficient DNA was measurable during the sequencing process to allow confidence in a consistent process across all samples. Consequently these samples were excluded from final analysis. These samples had more overlap with the communities seen in the negative controls and this added to our justification for omitting these samples. Comparisons were made of the patient demographics in the subgroup of 68 patients undergoing 16S rRNA sequencing analysis, and the 99 patient who underwent RISA interrogation. No significant differences were found, providing confidence that our subgroup was an accurate representation of our overall NCFB cohort.

In previous microbiota analysis, the taxonomic identification is often expressed in OTUs. This is based on similarity of sequence and equates roughly to genus level identification. The DADA2 pipeline used herein offers the potential for species levels identification instead of OTUs as well as genera level identification, and can report ecological measures at a sequence or genera level.[341] In order to calculate ecological markers, it had to be decided what level of identification we would work at. Comparisons were made between the measures calculated at genera and sequence level identification. These showed very close correlation, and hence we concluded them to be good surrogates of each other. Whilst species-level identification was possible for some samples, in other samples the genera was reported with multiple possibilities at species level. Consequently it was not possible to perform useful ecological analysis at species level. Due to genera having recognisable names to work with, this was the measure chosen instead of sequence level.

Of the 482 different genera identified, there were 3 genera which combined to make up over half the mean relative abundance - namely *Streptococcus*, *Haemophilus* and *Pseudomonas*. More strikingly, over 80% of samples had one of these three as the most abundant genus. These findings are not out of keeping

with other UK NCFB cohorts. Data from a tertiary centre in London had these three as the most abundant genera, and Belfast data highlighted them as being the dominant bacteria in samples from NCFB patients.[319, 324] Work from the North East of England identified three main families within their NCFB cohort- Pseudomonadaceae (majority likely to be PA), Pasteurellaceae (majority likely to be HI), and Streptococcaceae.[271]. This does highlight that despite the inherent bias of traditional culture, the same common potential pathogens are suggested by both techniques. The results do however appear quite different from a recently published group described by Byun et al.[383]. The most striking difference is their significantly lower abundance of *Streptococcus*. There is also a noticeably larger abundance of *Haemophilus*, *Pseudomonas*, *Moraxella* and *Prevotella*, and less *Neisseria*, *Rothia* and *Veillonella*. When considering these differences it is worth noting that samples were obtained by bronchoscopy, there was an exacerbation and stable group, and only 14 patient were included. The patients were also from South Korea and consequently conceivably surrounded by different bacterial communities and treatment strategies. In addition, a different DNA extraction kit was used to those used in the UK, and it is conceivable different extraction processes may have yielded different results.

An observation from our data when considering the three main genera is the variable abundance. This is neatly illustrated by the median abundance for *Haemophilus* and particularly *Pseudomonas* being so much lower than the mean values. This is driven by the large dominance that these genera display in some patients. This feeds into the hypothesis that these bacteria are “aggressors” that dominate the microbial community in the lung at the exclusion of others. It would also be consistent with the concept of HI and PA being significant pathogens in this disease process. No positive correlations were found between *Haemophilus* or *Pseudomonas* and other genera, whilst there was a positive correlation between the abundance of *Streptococcus* and *Rothia*, and *Veillonella*. The concept of positive and negative relationships in the bacterial community is not new.[271, 321, 322] When we looked at the ecological markers of diversity, evenness and richness, significant differences were seen. These again illustrated the domination over a community that *Haemophilus* and *Pseudomonas* could have as opposed to *Streptococcus*. The comparison of culture-negative and culture-positive samples, revealed greater complexity in culture-negative

samples. This finding is likely to be interconnected with the samples dominated by *Streptococcus* being more complex and culture-negative samples often having *Streptococcus* as the prevalent genus. This data as a whole could point to a concept of *Streptococcus* as a community “stabiliser”, and its near absence in the South Korean cohort may explain the predominance of certain pathogens.

A concern of culture techniques is their seeming inability to identify a pathogen when a patient appears to be suffering from symptoms which are believed to be of bacterial aetiology. It is consequently of interest to examine the correlation of culture-positive results to 16S rRNA sequencing data, and to assess what is present in culture-negative samples. Whilst the overall picture suggests that *Pseudomonas* and *Haemophilus* are deemed prevalent by both culture-dependent and culture-independent methods, sequencing data from culture-negative samples shows that PA and HI may not be isolated on plates even when sequencing data suggests otherwise. One sample was almost completely dominated by PA when the sequencing data was analysed to species level, and yet was culture-negative. Five samples were dominated by *Haemophilus* by sequencing and yet were culture-negative in the NHS laboratory. In one instance the sequencing data reported a relative abundance of 67% (of which at least 66% was HI) without a pathogen being found. Beyond these examples, the main finding in culture-negative samples was the general abundance of *Streptococcus*. In two-thirds of these samples, *Streptococcus* was the dominant genus. Other genera which dominated a small number of samples were *Actinomyces*, *Neisseria* and *Enterococcus*.

In this cohort there were 9 different pathogens revealed by culture. Unremarkable findings were seen for samples positive for *Burkholderia multivorans*, *Stenotrophomonas maltophilia*, *Serratia marcescens* and *Streptococcus pneumoniae*. In these cases, relative abundances of the relevant genera ranged from 18.46% to 41.89%. For one pathogen, *Klebsiella pneumoniae*, no evidence of presence was found on sequencing data despite a positive culture. There are two plausible explanations here - either the culture result is a false-positive, or the culture-independent process has failed to either successfully extract the DNA from this organism, or to identify the DNA sequence. One pathogen which historically has been felt to be over-represented by culture or under-represented by 16S rRNA sequencing is SA. In one SA-positive sample, we found a relative

abundance of 74.71%. This is reassuring as it illustrates that our methods were able to successfully extract and sequence *Staphylococcus*. One sample which was positive for SA by culture revealed only 1% relative abundance by 16S rRNA sequencing. These results may suggest that laboratory culture techniques are very effective for isolating SA from sputum, even when it is present in low abundance. After *Staphylococcus*, *Pseudomonas* had the next lowest relative abundance despite a positive culture result with under 10% in one instance. At the other end of the scale, *Pseudomonas* and *Haemophilus* had examples of relative abundance over 99% in culture-positive samples. *Moraxella catarrhalis* is the other pathogen that was revealed by culture. It is notable that in the two samples which were positive for this pathogen, the relative abundance was above 96% in both. Although based on only two samples, it suggests that *Moraxella* may be able to dominate a community like other more prevalent pathogens such as *Haemophilus* and *Pseudomonas*. *Moraxella* has shown capacity to reduce community diversity in other disease processes such as COPD.[378]

When comparing the culture results to sequencing data, it is important to remember what the results actually mean. Culture results indicate whether or not a bacterial species which is considered a putative pathogen in a specific disease process, has been isolated and identified in a deliberately biased environment. Sequencing data, when reported as relative abundance, just reports the presence of DNA of that taxa as a percentage of the total bacterial DNA found. It does not identify either total amount (bacterial load) or actual viable bacteria. A degree of caution must therefore be applied. For instance, two samples may have a relative abundance of *Haemophilus* of 20% and one may be culture-positive for HI whilst the other does not isolate a pathogen. This may not represent inconsistency of the culture technique, as the culture-negative sample may have little viable HI (for example, due to a very recent course of antibiotics) and a low bacterial load. In chronic suppurative lung diseases such as NCFB and CF, issues with sputum clearance from diseased airways is a common problem. These patient are also commonly on prophylactic antibiotics, which may include inhaled antibiotics delivered directly to the bacterial communities. It could be hypothesized that in patients with significant disease burden (illustrated by poor airway clearance and high antimicrobial intake), the contribution of non-viable bacteria to the relative abundance could be high. As previously mentioned, attempts have been made

to just measure viable cells, however this is not common practice.[283] With regards to issues with the use of relative abundance, it is possible to perform quantitative PCR to give you a total abundance. Whilst this is likely to be more useful when comparing a sputum culture result to sequencing data from DNA extracted from the same samples, limitations remain. A sputum sample is only going to be a small component of what is present in the lung. Denser sputum may be harder to clear from the lung for sampling, and may contain a different bacterial load. Therefore even if qPCR were to be employed, it may not be a true representation of the total abundance of bacteria in the lung.

A key interest with regards to the data arising from 16S rRNA sequencing is what its clinical relevance is, and what factors may influence the communities. As previously discussed in this thesis, the preceding work in this field has been partly inconsistent.[271, 319, 321, 323, 324] Also, whilst previous work has attempted to seek associations between bacterial communities and clinical severity, none previously had use of the BSI as a measure of disease severity. One of our aims was to take the opportunity to use the BSI alongside ecological measures of the bacterial communities. In light of the previous work in NCFB and other disease processes, the influence PA colonisation has on the BSI score, and also the impact *Pseudomonas* seems to have on community complexity, it may have been expected that an association would be seen between disease severity (as quantified by the BSI) and the community measures. However, this was not the case. Similarly, community measures did not correlate with FEV_1, which has previously been used as a surrogate for disease severity. This was out of keeping with some but not all previous work.[319, 321]

Exacerbations are events that patients are acutely aware of, and the understanding of them and subsequent management is very important for patients.[37] In this subgroup, there appeared to be an association between patient-reported exacerbations and reduced diversity and evenness. There was a significant outlier with regards to patient-reported exacerbations. If this patient was removed then the association with evenness only remained and when grouping patient by exacerbations frequency (low, medium and high), no association was seen. This differed from our RISA analysis. This may in part be due to smaller numbers, but also that the greater sensitivity of16S rRNA sequencing data may result in measures being affected by irrelevant

contributions. It should also be noted that exacerbation frequency was based on the patient's recall and consequently potentially subject to bias, inaccuracy, or variability due to local prescribing. In addition, there are likely to be several different types of exacerbation phenotype and severity in bronchiectasis, and the community complexity may play a role. A recent COPD study suggested dysbiosis was associated with worse exacerbation severity and also with an eosinophilic process.[378] Other associations noted were a positive correlation between *Streptococcus* genus and age, a negative correlation between *Staphylococcus* and age, and a positive correlation between *Pseudomonas* abundance and BSI.

It is worth pausing to consider the significance of these findings. Whilst early work suggested a link between disease severity and communities, this has not been a consistent finding.[271, 319, 321, 323, 324] An appealing hypothesis is that a community dominated by a pathogen (and with a low diversity, evenness and richness) will be a pathology-causing entity leading to disease progression.[322, 384, 385] This could lead to the formulation of strategies aimed at altering communities in the lung to prevent disease progression. The success of faecal transplant in the bowel gives a precedent for potential benefit when the bacteria community is altered. Alternatively, a community of low complexity may appear in severe disease due to the toll of disease, recurrent therapies and time. However, this is not what we have found. The lack of correlation with markers of community complexity and the BSI is possibly surprising for many reasons. The BSI includes additional points for colonisation with PA, and unsurprisingly we observed a correlation between *Pseudomonas* abundance and BSI. We have seen less complex communities in those that are dominated by *Pseudomonas* compared to those dominated by *Streptococcus*. *Haemophilus* dominance was also associated with reduced community complexity, and *Haemophilus* abundance was also associated with exacerbation frequency (which is also a parameter that contributes to the BSI). Taking all this into account, an association may have been expected. However, a counterbalance to these factors is the contribution that greater age makes to the BSI. Interestingly, there was a positive correlation between age and *Streptococcus* abundance. There appears to be greater complexity in the *Streptococcus* samples and consequently, this may be having an impact on this interpretation. As discussed in Chapter 6, the

investigation of an unselected District General disease-specific cohort may be subject to certain quirks. For example, younger patients who are actually seen by the bronchiectasis service may have more aggressive disease which has led to them being referred at a younger age. Also, by performing a cross-sectional study, patients may be at different stages of the establishment of their maintenance therapy. For example some may have only been recently referred and may not have established airway clearance techniques, undergone pulmonary rehabilitation, become established on prophylactic therapy, undergone a full eradication course for PA, etc.

The observation of reduced complexity being associated with exacerbations allows for the continuation of the concept of dysbiosis causing susceptibility to these episodes. It could be the case that whilst the impact of exacerbations may have an impact on disease state, it may be over many years, and consequently a cross-sectional study may not bear out this impact. It may also be that the size of this study, and others, has not been large enough for significant associations to be clearly identified. It was interesting to note that *Haemophilus* was the only genus whose abundance was associated with exacerbation frequency. Traditionally, PA has been seen as the principle pathogen in this disease process. As mentioned elsewhere in this thesis, it is plausible that due to the attention given to PA, we may be better at treating patients with this pathogen. Outcomes may therefore be better, despite those patients harbouring a potentially more pathogenic organism.

When considering the importance of the polymicrobial community and NCFB, it should also be considered that the reason that there has not been a consistent finding of reduced community complexity and disease severity is that it does not exist. It may well be that the balance and structure of the community is not as significant as other considerations such as the behavioural capacity of the genes present. For example, the switching on of virulence factors may be more important than the evenness of a community and could explain why large changes in communities might not be seen during exacerbation.

Another reason for performing 16S rRNA sequencing on these samples was to allow comparison to our RISA data. The cluster analysis of RISA band patterns had created groups of which the characteristics had been speculated. These included lower complexity samples, high abundance of *Haemophilus*, and high

abundance of *Streptococcus*. These findings have been supported by our 16S rRNA sequencing data. However, it should be noted that the dominant characteristic of a sample did not ensure clustering to a specific group. For example, not all samples dominated by *Streptococcus* were in the same group. In Chapter 6 we also used RISA data to estimate our ecological measures. We compared these estimates to those recorded by 16S rRNA sequencing. For all three measures, positive correlations were seen, however the association was noticeably stronger for evenness and diversity than for richness. The lower concordance of richness is less of a concern as this is the least sophisticated of the measures, and probably least likely to be clinically relevant.

Potentially of more use practically is the ability of RISA to identify samples as either low or high diversity, evenness or richness. It is debatable what the criteria for “low” or “high” should be. However the correlation seen between RISA and 16S rRNA sequencing suggests that with larger data sets, useful parameters could be obtained. We hypothesised that as RISA is believed to be a less sensitive technique, that concordance may be better when discarding genera with less than 1% relative abundance from the 16S rRNA sequencing data. There was a suggestion from our results that this may be the case with diversity and richness but not evenness.

With 16S rRNA sequencing considered the gold-standard, our findings provide encouragement for the output from RISA. Whilst it does not provide a perfect correlation of ecological measures and does not allow for confident taxonomic identification, it clearly provides a meaningful measure of the community. A similar message also appeared between the two techniques - namely a suggestion of less exacerbations with more complex samples. There have been other studies comparing RISA and sequencing data, both in the environment and the lung.[133, 287] As evidence and use increases, RISA may develop utility in clinical practice.

There are various strengths and weaknesses of this work to comment on. This work has provided taxonomic identification and ecological measures of the communities found in an unselected bronchiectasis cohort. This has been done with highly detailed clinical correlates and hence adds to our understanding of the bacteria present in patients with bronchiectasis. The gold-standard technique has been used for this process, allowing comparison with RISA - an older

technique that has not been extensively used for interrogation of clinical samples. It has also used a pipeline not previously reported in lung studies, and this has allowed some greater identification to species level.

However, it should be noted that the gold-standard process is not perfect. Whilst we were able to process all the RISA samples, we were not confident enough with the consistency of some of our samples to report all of them by 16S rRNA sequencing. In addition there is the potential issue of contamination. As mentioned multiple times in this thesis, the molecular techniques I have used deal with DNA and not necessarily the DNA of viable bacteria. An environment may be sterile as the bacteria have been killed as part of a rigorous sterilising process, but this does not mean that the DNA of the dead bacteria may not be amplifiable and therefore detectable. There are multiple possible sources for this, including the DNA extraction kits.[282] To ensure rigor, I extracted DNA from negative controls from each of the 5 batches of DNA extraction kit used. These were sequenced, as was the DES used to dilute samples. This revealed that bacterial DNA was indeed detectable within these negative controls. This is an undeniable limitation, and is likely to be more of an issue with low biomass samples. Whilst a database of likely local contaminants and kit-specific contaminants can be collated, the fact that common contaminants such as *Ralstonia* have also been cultured from sputum in certain circumstances can provide uncertainty.[386] This highlights the potential importance of culture and how culture-dependent and culture-independent methods may complement each other. If a sample isolates a given species by culture that is also detectable by molecular analyses, then this is almost certainly accurate. A further issue is the reporting of relative abundance. It may be that bacterial load is more important than relative abundance. Whilst it is possible to do qPCR to give a better idea of this, you are still only getting a surrogate of the overall lung. It is very likely that the bacterial load in one sputum sample will be different to another. For example, the first cough of the day in a bronchiectasis patient may be very different to one later in the day. Hydration state may also influence the amount coughed up, as may the inhalation of therapies into the lung such as antibiotics or hypertonic saline. Also, sputum originating from one area of the lung may have a different biomass than another geographical region. Furthermore, our current use of antibiotic therapy does not take into account bacterial load as such, but merely the likely (or identified)

pathogen. Consequently, trying to measure total abundance rather than relative abundance, is not without significant limitations as well.

A further issue with 16S rRNA sequencing is the time and cost of the process. Whilst costs are undeniably falling, the process requires an intricate multi-step process likely to include multiple professionals. This may well not be feasible in day-to-day clinical practice.

It is worth commenting on the bronchiectasis cohort studied and the collection of data from them. The patients were recruited from routine clinical practice. They clearly needed to be able to produce a sputum sample. They are therefore representative of the patient from whom we use microbiological testing to drive management. Demographically they are also in keeping with other studied cohorts. As a cohort they are likely to be sicker than total population that have bronchiectasis and the likelihood of being recruited would be based on the need for healthcare interaction during the recruitment period. Overall, we can be confident that they are representative of those we treat in day-to-day practice. With regards to their clinical details, we can have high confidence of the accuracy of nearly all the data collected. The exacerbation rate however is possibly the measurement with the lowest accuracy. In this study it was based on patient reported exacerbation rate. This is obviously at risk of recall bias. There is also the issue of what actually constitutes an exacerbation. We considered the use of acute antibiotics for a change in respiratory symptoms to constitute an exacerbation. However, the threshold for antibiotic prescribing will vary from clinician to clinician. The only way around this is for a prospective study with protocol-defined exacerbations being recorded. This was not possible for this study design.

When assessing the outcomes of the chapter, it is clear that the bacterial communities of bronchiectasis patients are highly variable. It is also clear that culture can reveal the likely major pathogens, though they can be present in high relative abundance and yet not be isolated by culture. It seems likely that a single culture-negative sample cannot be taken as conclusive evidence that a previous pathogen is now eradicated. In a world where culture-independent techniques have not found their niche in everyday healthcare, culture still has an important place though. It remains unclear the exact significance of the bacterial community to the disease process, and mechanistic and prospective work is likely to be

required to answer this. It is also plausible that true understanding will be far more complex and require the additional understanding of the interactions of the bacterial community with: viruses and fungi; the host immune system; and the behaviour of the bacteria as displayed by virulence factors and resistance mechanisms. An understanding of such a complex environment currently seems a way off in the research world and consequently even further away in a resource-stretched healthcare setting such as the NHS.

This brings us along to the potential role and significance for RISA. In this chapter we have shown, that whilst not perfect, it provides reasonable correlation with 16S rRNA sequencing for ecological characteristics. By providing this at a fraction of the cost, complexity and time, RISA potentially represents a clinically-useful technique. Clearly for this to be the case, an appreciation of what it means for management is key. This will only be revealed by its prospective use in clinical trials. RISA is clearly more feasible than 16S rRNA sequencing to be used in a very large clinical trial of a therapy- for example prophylactic antibiotics, to identify subgroups who may respond to a therapy.

As our understanding of the microbiome and relevant techniques evolve, it may be the case that we see a hierarchy of investigations depending on the clinical situation. It is likely that culture will retain its place as a first line investigation and provide evidence of a viable pathogen. In clinical circumstances where an isolated bacteria is believed to be the important pathogen in a process, further techniques (including culture-independent techniques) may then be used to look at virulence factors, resistance genes etc. This is already being used in the management of TB with sequencing for rifampicin resistance.[268] An overview of the bacterial community may be relevant for more chronic management, such as prophylactic antibiotics or consideration of pathogen eradication regimes. This potentially could include RISA. In circumstances where culture has not revealed a significant pathogen, 16S rRNA sequencing may come into its own by also providing taxonomic identification alongside community characteristics. Metagenomics and metabolomics may be further levels of investigation in the future, however these may remain prohibitive from day-to-day clinical care for longer. These may eventually be reserved as an investigation for when patients are not responding to standard therapy as prescribed on the basis of clinical history, examination, and simpler microbiological tests.

7.6 Conclusion

In summary we have demonstrated complex bacterial communities in the bronchiectatic lung. The predominant genera are not out of keeping with traditional culture results, though high abundance of a particular genus does not always translate to culture-positive samples. Whilst we have not found an association between community characteristics and the BSI or FEV_1, there appears to be an association between exacerbation frequency and less complex communities. Finally, RISA has been shown to provide an appreciation of potentially relevant ecological markers, and may therefore have a role in the investigation and management of chronic respiratory disease.

Part C

Chapter 8- Summary and Discussion

In the thesis, I have used both culture-dependent and culture-independent techniques to investigate PA and the wider bacterial community within the bronchiectatic lung. This has been performed prospectively with the use of samples submitted by patients with NCFB and CF as part of their ongoing care at the Royal Devon & Exeter Hospital. In this final, Chapter I shall describe how I have addressed the aims which were set out in Chapter 1 of this thesis, highlight novel and important findings, and discuss potential areas for future research, before a final conclusion.

8.1 To Perform an Epidemiological Review of PA in NCFB Utilising a Variety of Genotyping Techniques to Investigate the Likelihood of Cross-infection With PA

As illustrated throughout this thesis, PA is both a common and highly important pathogen in NCFB.[40, 43, 44, 48, 107, 108, 110, 111, 113-115] Despite its status in NCFB, there is little known about the epidemiology of PA in this condition.[131, 210, 211] This is in contrast to CF where epidemiological studies of PA led to concerns of cross-infection with clinically significant strains.[185-188, 192, 194-199, 232] Cross-infection is taken very seriously in the CF community and this data has in part led to the widespread practice of segregation of patients.[95, 179] The risk of cross-infection with PA is unknown in NCFB, though there is some evidence suggesting occurrence.[131, 211]

In Chapter 3, I described the largest single-centre epidemiological review of PA in a NCFB cohort. This has been robustly performed due to a few key aspects of study design. Importantly 10 representative isolates per sample were used. This was done to reduce the risk of underestimating multiple strains within a sample. Previous work in CF and NCFB has illustrated the potential for multiple strains of a pathogen to be present within a single sample.[211, 261] The study was performed via the use of three genotyping techniques. These techniques included a rapid screening technique (RAPD), the highly portable MLST, and the higher resolution analysis of WGS. From the WGS data, *in silico* prediction of hypermutator status was performed to aid interpretation of differences between isolates. Interpretation of WGS data was also aided by the incorporation of publicly-available genomes into our analysis. In addition, to provide greater

context to the findings of the study, epidemiological analysis was also performed on both the local CF cohort and also on non-respiratory clinical samples. These are not aspects of any previous epidemiological study of PA in either CF or NCFB.

Due to its portability, MLST is the easiest method to describe the PA strains found in our NCFB cohort. Via MLST, 35 different sequence types were revealed and 5 were shared between patients. The most abundant strain was ST17, which is better known as Clone C. Clone C is a globally abundant strain which has been found in clinical and environmental sources over a prolonged period of time.[214, 217, 218, 220-223] It was also found in our non-respiratory samples and it should not be seen as surprising that it was prevalent in a NCFB cohort. Further investigation was performed on this strain with the interrogation of three isolates per NCFB patient by WGS. The demonstration of greater between-patient diversity than within-patient diversity, alongside the comparisons with unconnected publicly-available genomes and the background abundance, provided confidence that cross-infection was unlikely to have occurred with Clone C.

The epidemiological review revealed evidence of highly likely cross-infection with ST564, with three patients harbouring isolates that were genetically very similar. This is the most robust published data which suggests cross-infection in NCFB.[355] ST564 is a rarely reported strain and based on our clinical attendance records, may have super-infected a patient. Due to its small numbers, we are currently unable to comment on the clinical significance of this strain. It will be important to repeat this study in the future to see if we see evidence of increasing prevalence in the cohort, and if in large enough numbers, any suggestion of clinical impact. It is possible that there are other cases of cross-infection within this cohort, however we do not have sufficient evidence to state this. One factor that may well contribute to the uncertainty is that of hypermutable isolates. Unlike previous work, we have looked to address this issue, and have shown greater genetic diversity when predicted hypermutators are involved in pair-wise comparison. One of our other concerns was the potential for cross-infection between CF and NCFB patients. While these cohorts are seen in out-patients at separate times, they do use the same facilities and they do share healthcare professional such as physiotherapists, specialist nurses and doctors.

Pleasingly, there was no evidence of cross-infection between these cohorts or within the CF cohort, with the exception of two family members.

This chapter has achieved the aim of performing an epidemiological review of PA and is the largest of its kind in NCFB. It has provided the most robust evidence yet of cross-infection within a NCFB cohort, albeit rare. It has also highlighted the difficulties of performing a cross-infection study in a chronic disease cohort, even with the use of WGS. A principle issue is the interpretation of the genetic diversity between samples as seen in WGS. Despite this high level of robustness, cross-infection can only be considered as a likelihood and longitudinal data is required to further our understanding of the risk in this cohort.

8.2 To Gain Insight into the Prevalence of Hypermutable PA in NCFB and its Impact on WGS-based Studies Addressing Cross-infection

As highlighted above, WGS data in an epidemiological study can reveal wide-ranging levels of genetic difference. There are many reasons why two isolates may display a certain level of difference and one of these is hypermutation. Hypermutators are found in chronic lung disease and can indeed be common.[159, 161] In our epidemiological review we performed a previously undescribed *in silico* prediction of hypermutator status. We also performed a traditional hypermutator assay to ensure the reliability of this *in silico* prediction.

As part of the epidemiological review performed in Chapter 3, isolates from 31 patients underwent WGS. Of these, 22 were from NCFB patients and 9 from CF patients. By the *in silico* prediction 8 NCFB patients were predicted to have hypermutable PA and 1 CF isolate was predicted to be a hypermutator. This highlighted that hypermutable PA is not uncommon in NCFB. We also showed in Chapter 3 that those that were predicted to be hypermutators had significantly greater genetic difference when compared to other examples of the same strain. In addition, the results of the traditional hypermutator assay revealed a high frequency of the hypermutator phenotype amongst those isolates that were predicted to be hypermutators by the *in silico* approach, providing further confidence in this prediction model. Consequently, this prediction model appears to be highly relevant to the assessment of WGS output in the context of a cross-infection study.

We have clearly shown that hypermutators are not uncommon in PA found in NCFB patients. We have also demonstrated that *in silico* prediction of hypermutators from WGS data aids interpretation of results by facilitating the SNP differences to be put into relevant context and promoting the rational consideration of the likelihood of cross-infection. This common phenotype must now be taken into account in future cross-infection studies into chronic respiratory disease using WGS. Our testing of the model against the traditional technique has supported its use, however an interrogation of a larger collection of sequenced isolates may enable the refining of this model.

8.3 To Test the Utility of a Novel Culture-independent Genotyping Technique for PA

In recent times, the advances of molecular microbiology have resulted in the development of culture-independent techniques. These have revealed how traditional culture does not tell the full story and how samples may not grow certain pathogens even when culture-independent techniques reveal the presence of their DNA.[281] Consequently, when performing prospective cross-infection studies on samples, shared strains may potentially be missed if they do not grow on a plate. It is also recognised that not all pathogens will grow at the same rate on a plate, and that multiple strains can occur in a sample.[211, 261] When picking isolates off a plate, there may well be a bias towards those that are faster growing rather than the most important or abundant. With the increasing use of culture-independent techniques in research, DNA is increasingly extracted from sputum samples. Considering these factors, we have looked to see if it was possible to perform MLST on DNA extracted from sputum for PA, and if so, how it compared to the traditional method.

The results described in Chapter 5 show that it is indeed possible to perform MLST without the need for culturing the bacteria. We showed that genotyping data could be obtained from culture-negative samples, however there was the occasional example of not being able to obtain genotyping data despite a sample being culture-positive. By the culture-independent technique we found evidence of transmissible strains which were not revealed by the culture-dependent method. Whilst in our cohort this does not seem to have masked likely cross-infection, this does highlight the possibility of it being missed.

A potential concern with this technique is that the presence of multiple strains in a single sample may yield ambiguous sequence reads. This does not seem to be a common problem as there was very high concordance between the culture-dependent and culture-independent approaches. However, when artificially mixing DNA from different pure cultures in different ratios, outputs could be generated suggesting the presence of an MLST allele sequence which was not there. This limitation has to be acknowledged if using this technique and consequently we would recommend not using this as the only genotyping technique in a comprehensive epidemiological review. Despite this limitation, this technique could have a very useful role, either alongside a culture-dependent process or as a screening tool. As previously mentioned, longitudinal data in cross-infection studies is highly desirable, though often not performed. It is also unclear at what interval this should be done. By performing screening in a few loci in a cohort, a quick appreciation of changing epidemiology or the increasing presence of a particularly strain may be revealed, subsequently guiding the timing for an in-depth epidemiological review.

This study has achieved the aim of testing the utility of a culture-independent genotyping technique with the novel use of the MLST scheme for PA on sputum samples. As well as demonstrating the feasibility of this, we have also shown the potential benefits and suggested a place in clinical practice.

8.4 To Investigate the Relationship Between Disease Severity and the Bacterial Community Composition in NCFB

Our current understanding and management of NCFB has been based around the use of traditional culture methods. Via these techniques, a single pathogen is often reported. Data accumulated with the use of these processes has led to our view of PA as the most important pathogen in NCFB.[110] The development of culture-independent techniques has shown that these techniques do not reveal the significant complexity contained within the lung.

In Chapters 6 and 7 we looked into the communities in the bronchiectatic lung with the aim of investigating the relationship between community composition and disease severity. This was performed by two techniques- RISA and 16S rRNA sequencing. We are not aware of any previous work examining bacterial communities in NCFB by RISA, and via this technique we also examined our local CF community. Interestingly, the NCFB communities appeared less complex.

This is a novel finding, and unexpected. There has previously been an appreciation of PA dominating communities in both conditions, and as a result decreasing diversity and evenness. In CF the other main pathogen is SA, while in NCFB it is HI. Samples which were culture-positive for HI seemed to have lower complexity than those positive for SA. Consequently, in NCFB, both the main pathogens appear to have a large impact on community complexity, and hence may explain the lower overall complexity.

Previous work, both into NCFB and in other diseases, suggested that less complex communities may be associated with more severe disease.[133, 321] To our knowledge, we have undertaken the first study of whether the composition of these communities correlates with the BSI - a validated severity index for NCFB. Whilst we expected a relationship between the BSI and the communities, this was not found by either RISA or 16S rRNA sequencing. This could be for multiple reasons, including the possibility that there is not a link between disease severity and community complexity. Alternatively, it could be due to certain quirks of the index, and from collecting samples from an unselected cohort. Whilst there was no correlation with the BSI, we did observe associations between the communities and the clinical status of the patient. There was an inverse relationship between exacerbation frequency and complexity by both techniques, whilst RISA also showed a positive correlation between community richness and lung function.

Due to these findings, it remains plausible that there is a relationship between the complexity of the bacterial community and the clinical condition of the patient, even if this is not borne out by comparisons with the BSI. It is important to note that even if there is a relationship, this does not mean that a low complex community is necessarily more pathogenic, or makes patients more susceptible to exacerbations. Even if this is the case, it is likely to just be a component. Other components may include the viral and fungal communities, the collection of resistance genes in the community, and the specific strains of bacteria and their pathogenicity. This does not however mean that the community complexity may not be very important. In many disease processes, successful management includes targeting various components, and NCFB is likely to be no different.

When considering the pathogens dominating the community and disease severity, it was interesting to note the comparisons between HI and PA. By RISA,

samples which were positive for HI had less complexity, and by 16S rRNA sequencing there was a trend for samples where *Haemophilus* was the most prevalent genera to be less complex. Also of note there was an association between *Haemophilus* and exacerbations. This raises the possibility in our cohort, that HI may have the potential to dominate communities more than PA, and be associated with more exacerbations.

In these chapters we have achieved our aim of investigating relationships between bacterial communities and disease severity. The issue is not clear cut, and this may be due to there being far more to it than mere community structure. However, there is enough to suggest that it remains an approach of potential interest when considering our investigation and management of a patient with NCFB.

8.5 To Test the Clinical Utility of RISA as a Cheaper and Quicker Alternative to 16S rRNA Sequencing

A novel approach in the thesis has been to use RISA to investigate the wider bacterial community in the bronchiectatic lung. In Chapter 6 we used it to investigate both our NCFB and CF cohorts. In the following chapter we compared its output to the gold-standard technique of 16S rRNA sequencing in our NCFB cohort. As we now appreciate some of the limitations of traditional culture techniques, there is an obvious appeal to try and bring culture-independent techniques into everyday clinical practice. However, various considerations need to be made beyond what is simply the “best” test. Ideally a test will provide a clinician with a rapid and easily interpretable result with relevance to the treatment decisions. In a resource-limited environment like the NHS, the cost of the test will also be a consideration. While sequencing costs are falling and technologies evolving, the technique of 16S rRNA sequencing still comes at a significant cost in terms of finance, time and complexity. Consequently an alternative which could be incorporated into everyday clinical care would be appealing.

With this in mind, we have used RISA alongside 16S rRNA sequencing. It has been a cheaper, quicker and easier technique to perform. It has also required less external expertise to perform and produces outputs from. Furthermore, it has shown consistency with the gold-standard technique with regards to various measures, but also in showing trends with regards to relevant data. Consequently, it retains appeal as a potential tool. However, it has two clear

limitations - one shared with 16S rRNA sequencing (the current lack of understanding of the clinical relevance of the output), and the other not (a lack of capacity for pathogen identification).

As with 16S rRNA sequencing, at present we do not know how data derived from RISA can impact clinical decision-making. For this to change, we need prospective longitudinal clinical trials. These could give us insight into how a certain intervention in the face of different community characteristics can influence clinical outcome. These trials will clearly be difficult and expensive to run. RISA would both reduce the cost of the trial, but also allow subsequent integration into cash-strapped clinical settings to aid decision-making, in a way that 16S rRNA sequencing would not.

The clear difference between 16S rRNA sequencing and RISA is that RISA does not provide pathogen identification. This is an obvious limitation, however this does not need to be insurmountable when considering a place for RISA in clinical practice. There are putative band lengths which have been calculated and consequently candidate pathogens can be listed. For example, we were able to correctly predict the strong presence of *Streptococcus* in samples. In many cases, the use of concurrent culture techniques may provide evidence of a viable pathogen. If not, the extracted DNA could be used either for pathogen-specific PCR or (in cases of ongoing uncertainty) 16S rRNA sequencing.

In this thesis we have started to explore the utility of RISA and have demonstrated potential for its use in the clinical setting. It is conceivable that in the future, clinicians may have a variety of tests available to them of varying complexities and costs. For different clinical questions and problems, different tests may be appropriate, and RISA may sit within this suite.

8.6 Areas for Future Research

There is clearly scope for further research to be performed around the areas covered in this thesis. Recently the topic of cross-infection has gained attention in the NCFB community.[131, 211, 354, 355] At present there is no high quality longitudinal data and there is no evidence beyond PA. Consequently there is significant scope for future research into this area, and given the prominence we have displayed of *Haemophilus* and *Streptococcus*, these are logical targets for investigation.

Future studies into cross-infection are likely to include sequencing techniques. It is consequently critical to gain a better understanding of what constitutes a significant genetic difference between samples. While this is far more straight-forward in real-time outbreaks, this is very challenging in complex bacterial communities which have evolved over time. In this thesis we have attempted to gain an understanding of the impact of hypermutable strains on the genetic difference seen between samples by using an *in silico* prediction model. As more WGS data is generated from isolates, a better insight into the impact of certain mutations will be gained. In addition, longitudinal studies will give a greater understanding of the persistence or switching to a hypermutable state. Additional research will also aid a better understanding of the typical range of genetic divergence within a sputum sample. With these aspects better studied, greater context can be given to the differences seen between isolates.

Large scale cross-infections studies are time-consuming and costly. It is also unclear at what interval to perform them. There is consequently an appeal in performing a simpler screening test to give an indication of when to perform the follow-up study. We have shown the potential of MLST applied directly to DNA extracted from sputum. This has now been shown to be possible for both PA and BCC.[370] It would be expected that this would be possible for other important pathogens. Consequently, future research should look into the utility of rapid screening for evidence of putative transmissible strains in a variety of pathogens before embarking on large prospective studies.

It is very likely that there will be a large amount of culture-independent research into the lung microbiome. It would probably be most useful for this to be directed in two very different directions. There is the need for high quality studies with sophisticated techniques to further our understanding of the micro-organisms in the lung. However, there will come a time when culture-independent techniques will have to be integrated into clinical trials in a format which will be permissible to everyday clinical care. Without this, it is very difficult to see how the management of patients can move forward. A possible future research project would involve the use of RISA in a clinical trial involving an intervention and assess if its outputs can help personalised management.

8.7 Conclusion

In this thesis with have used molecular techniques to provide better insight into PA and the wider bacterial community within the bronchiectatic lung. We have found evidence of cross-infection in NCFB cohorts through an in-depth epidemiological study involving multiple genotyping techniques and with comparison to both CF and non-respiratory isolates. We have proposed an *in silico* prediction model for hypermutators to better assess WGS data and we have demonstrated a potential screening test for cross-infection in a cohort. For the wider bacterial community we have observed characteristics associated with clinically important markers and performed an alternative community assessment to the usual 16S rRNA sequencing approach. It is to be expected that culture-independent techniques will become increasing important, and the challenge now is how best to start utilising them for patient benefit.

Appendices

	PA Colonised (n=20)	HI Colonised (n=17)	Colonised by Other (n=4)	Not Colonised (n=58)	P Value PA v HI
Age[a]	66.3 (11.43)	60.82 (16.82)	69 (10.77)	66.31 (13.38)	0.2621
BSI[a]	10.6 (3.38)	7.53 (3.13)	7.5 (2.18)	6.16 (3)	**0.0154**
FEV_1 %predicted[a]	62.65 (22.03)	59.88 (21.1)	83.25 (26.45)	72.78 (21.61)	0.7080
Exacerbator Phenotype[b,c]					0.3089[d]
- high	20%	41.18%	100%	41.38%	
- medium	55%	47.06%	0	27.59%	
- low	25%	11.76%	0	31.03%	
Oral antibiotic prophylaxis[b]	55%	35.29%	25%	37.93%	0.3248
Inhaled antibiotic prophylaxis[b]	40%	0	50%	15.52%	**0.0039**
Received IV antibiotics in last 12 months[b]	30%	5.88%	0	6.9	0.0975

Table A6.1 Clinical measures for different baseline colonisation status- NCFB Data is presented with mean values with standard deviations or percentage of cohort. Statistical testing was by [a]t-test, or [b]Chi- squared test. [c]Exacerbator phenotype was described as “High”- greater than 3 exacarbations in preceding 12 months; “Medium”- 2 or 3 exacerbations in preceding 12 months; “Low”- 0 or 1 exacerbations in preceding 12 months. [d]Chi-squared test from 3x2 contingency table.

	P Value	R^2
Age		
- **e-Diversity**	**0.0402**	0.04269
- **e-Evenness**	**0.0243**	0.05119
- **e-Richness**	0.2064	0.01641
BSI		
- **e-Diversity**	0.8741	0.0002603
- **e-Evenness**	0.1312	0.02333
- **e-Richness**	0.9334	7.246e-005
FEV_1 %predicted		
- **e-Diversity**	0.0929	0.02884
- **e-Evenness**	0.4233	0.006622
- **e-Richness**	**0.0390**	0.04321
Exacerbations		
- **e-Diversity**	**0.0373**	0.04396
- **e-Evenness**	0.0538	0.03781
- **e-Richness**	0.0989	0.02782
BMI		
- **e-Diversity**	0.6603	0.002
- **e-Evenness**	0.5562	0.003582
- **e-Richness**	0.6618	0.001981

Table A6.2 Comparison of ecological measures and continuous data by linear regression for the NCFB cohort

	PA	HI	Others	P Value
e-Diversity	1.68	1.13	1.44	0.1645
e-Evenness	0.86	0.66	0.86	**0.0064**
e-Richness	6.32	5.68	5.5	0.4357

Table A6.3 Comparison of ecological measures and culture results in NCFB cohort Statistical analysis was performed by Krusal-Wallis test for e-Diversity and e-Evenness and by ANOVA for e-Richness.

	Culture-positive	Culture-negative	P Value
e-Diversity	1.41	1.48	0.5120
e-Evenness	0.81	0.86	**0.0262**
e-Richness	5.63	5.67	0.9269

Table A6.4 Comparison of ecological measures in NCFB cohort and whether sample is culture-positive or negative. P value generated by Mann-Whitney for e-Diversity and e-Evenness with values quoted as median. e-Richness is reported as a mean and analysed by the t-test.

	P Value
Gender	
- **e-Diversity**	0.2587
- **e-Evenness**	0.162
- **e-Richness**	0.6136
PA Colonisation	
- **e-Diversity**	0.4910
- **e-Evenness**	0.6202
- **e-Richness**	0.7011
Hi Colonisation	
- **e-Diversity**	0.8487
- **e-Evenness**	0.1082
- **e-Richness**	0.4207
Oral Prophylaxis	
- **e-Diversity**	0.3472
- **e-Evenness**	0.7641
- **e-Richness**	0.2151
Inhaled Prophylaxis	
- **e-Diversity**	0.3191
- **e-Evenness**	0.5981
- **e-Richness**	0.1678
IV antibiotic use in last 12 months	
- **e-Diversity**	0.2983
- **e-Evenness**	0.7688
- **e-Richness**	0.3909
Animal owner	
- **e-Diversity**	0.2448
- **e-Evenness**	0.4993
- **e-Richness**	0.3109
Urban living[a]	
- **e-Diversity**	**0.0294**
- **e-Evenness**	0.1046
- **e-Richness**	**0.0098**

Table A6.5 Comparison of ecological measures and potential influencing factors for NCFB cohort P values produced by the Mann-Whitney test when e-Diversity or e-Evenness was compared, or the t-test for e-Richness. [a]The scores for urban living were higher than rural living.

	P Value	R^2
Age		
- **e-Diversity**	0.7705	0.001753
- **e-Evenness**	0.6327	0.004698
- **e-Richness**	0.9247	0.00902
FEV_1 %predicted		
- **e-Diversity**	**0.0216**	0.1052
- **e-Evenness**	0.4268	0.01321
- **e-Richness**	**0.0111**	0.1269
BMI		
- **e-Diversity**	0.5694	0.006792
- **e-Evenness**	0.7736	0.001741
- **e-Richness**	0.5752	0.006588

Table A6.6 Comparison of ecological measures and continuous data by linear regression for CF patients Note one patient could not perform spirometry and BMI not performed on one

	Culture-positive	Culture-negative	P Value
e-Diversity	1.75 (1.43-1.99)	1.48 (1.07-1.74)	**0.0182**
e-Evenness	0.9 (0.83-0.92)	0.89 (0.73-0.9)	0.1956
e-Richness	7.23 (2.4)	5.55 (1.44)	**0.0350**

Table A6.7 Comparison of ecological measures and culture status in the CF cohort Statistical analysis by Mann Whitney test with medians reported for e-Diversity and e-Evenness. For e-Richness, mean values are reported an the t-test used

	P Value
Female Gender	
- **e-Diversity**	0.9888
- **e-Evenness**	0.9888
- **e-Richness**	0.8544
Oral prophylaxis	
- **e-Diversity**	0.9488
- **e-Evenness**	0.4077
- **e-Richness**	0.6189
Inhaled prophylaxis	
- **e-Diversity**	0.6312
- **e-Evenness**	0.5221
- **e-Richness**	0.7406
Urban Living	
- **e-Diversity**	0.3307
- **e-Evenness**	0.1876
- **e-Richness**	0.6622
Animal owner	
- **e-Diversity**	0.5941
- **e-Evenness**	0.6197
- **e-Richness**	0.6379

Table A6.8 Comparison of ecological measures and potential influencing factors for CF cohort P values produced by the Mann-Whitney test when e-Diversity or e-Evenness was compared, or the t-test for e-Richness

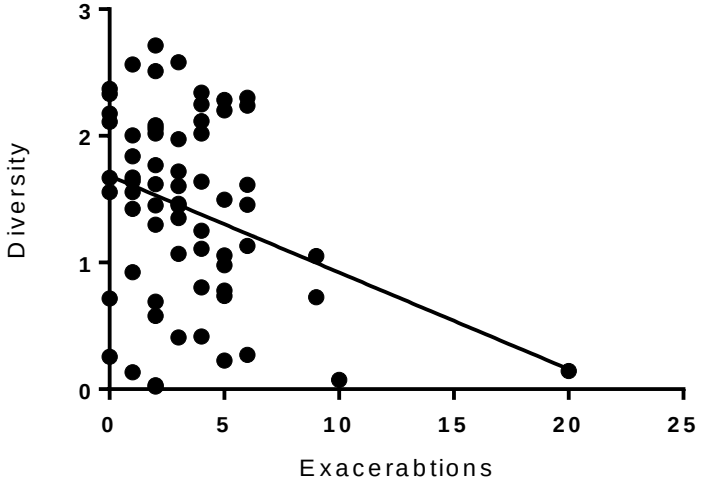

Figure A7.1 linear regression of Diversity v Exacerbations Exacerbations are patient reported over the preceding 12 months (p= 0.0085; R^2= 0.1003)

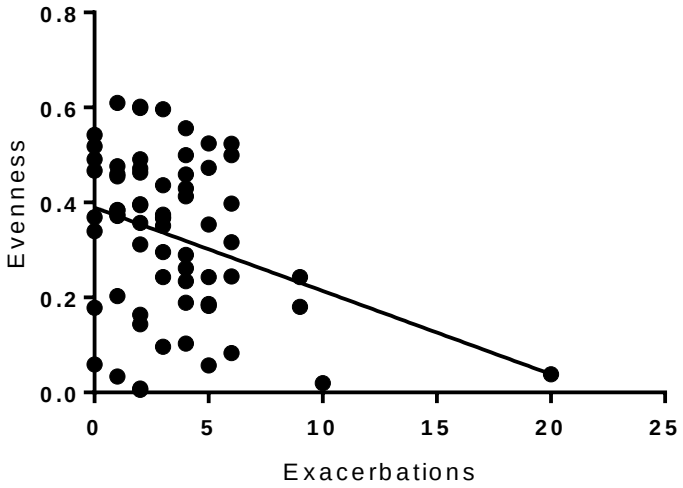

Figure A7.2 linear regression of Evenness v Exacerbations Exacerbations are patient reported over the preceding 12 months (p= 0.0075; R^2= 0.1035)

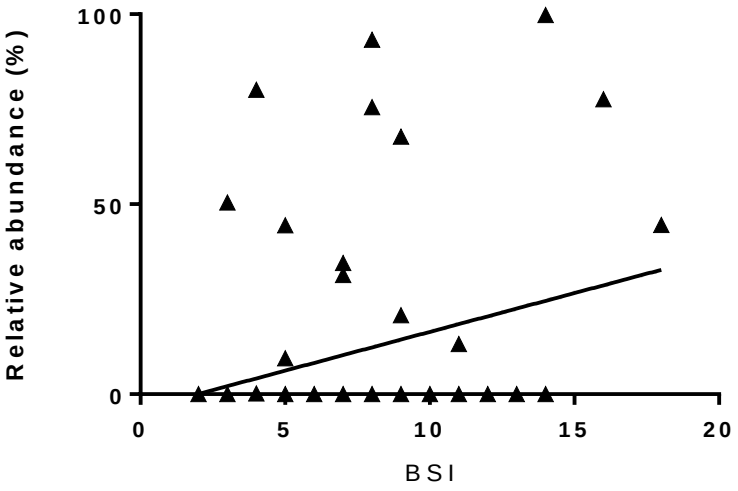

Figure A7.3 linear regression of relative abundance of *Pseudomonas* v BSI ($p= 0.0274$; $R^2= 0.07161$)

Bibliography

1. Pitt, T.L., *Biology of Pseudomonas aeruginosa in relation to pulmonary infection in cystic fibrosis.* J R Soc Med, 1986. **79 Suppl 12**: p. 13-8.
2. Cullen, L. and S. McClean, *Bacterial Adaptation during Chronic Respiratory Infections.* Pathogens, 2015. **4**(1): p. 66-89.
3. Dickson, R.P. and G.B. Huffnagle, *The Lung Microbiome: New Principles for Respiratory Bacteriology in Health and Disease.* PLoS Pathog, 2015. **11**(7): p. e1004923.
4. Roguin, A., *Rene Theophile Hyacinthe Laennec (1781-1826): the man behind the stethoscope.* Clin Med Res, 2006. **4**(3): p. 230-5.
5. Tomos, I., et al., *Celebrating Two Centuries since the Invention of the Stethoscope. Rene Theophile Hyacinthe Laennec (1781-1826).* Ann Am Thorac Soc, 2016. **13**(10): p. 1667-1670.
6. Pasteur, M.C., D. Bilton, and A.T. Hill, *British Thoracic Society guideline for non-CF bronchiectasis.* Thorax, 2010. **65 Suppl 1**: p. i1-58.
7. McShane, P.J., et al., *Non-cystic fibrosis bronchiectasis.* Am J Respir Crit Care Med, 2013. **188**(6): p. 647-56.
8. Amaral, R.H., et al., *Computed tomography in the diagnosis of bronchiectasis.* Eur Respir J, 2015. **46**(2): p. 576-7.
9. Matsuoka, S., et al., *Bronchoarterial ratio and bronchial wall thickness on high-resolution CT in asymptomatic subjects: correlation with age and smoking.* AJR Am J Roentgenol, 2003. **180**(2): p. 513-8.
10. Diaz, A.A., et al., *Bronchoarterial ratio in never-smokers adults: Implications for bronchial dilation definition.* Respirology, 2017. **22**(1): p. 108-113.
11. Metersky, M.L. and A. Mangardich, *Chronic suppurative lung disease in adults.* J Thorac Dis, 2016. **8**(9): p. E974-e978.
12. Pasteur, M.C., et al., *An investigation into causative factors in patients with bronchiectasis.* Am J Respir Crit Care Med, 2000. **162**(4 Pt 1): p. 1277-84.
13. O'Donnell, A.E., et al., *Treatment of idiopathic bronchiectasis with aerosolized recombinant human DNase I. rhDNase Study Group.* Chest, 1998. **113**(5): p. 1329-34.
14. Elborn, J.S., et al., *Current strategies for the long-term assessment, monitoring, and management of cystic fibrosis patients treated with CFTR modulator therapy.* J Cyst Fibros, 2017. **16**(1): p. 163-164.
15. Goeminne, P.C. and A. De Soyza, *Bronchiectasis: how to be an orphan with many parents?* Eur Respir J, 2016. **47**(1): p. 10-3.
16. Weycker, D., et al., *Prevalence and economic burden of bronchiectasis.* Clinical Pulmonary Medicine, 2005. **12**(4): p. 205-209.
17. Chalmers, J.D., S. Aliberti, and F. Blasi, *State of the art review: management of bronchiectasis in adults.* Eur Respir J, 2015.
18. Chang, A.B., et al., *Non-CF bronchiectasis: clinical and HRCT evaluation.* Pediatr Pulmonol, 2003. **35**(6): p. 477-83.
19. Ringshausen, F.C., et al., *Bronchiectasis in Germany: a population-based estimation of disease prevalence.* Eur Respir J, 2015. **46**(6): p. 1805-7.
20. Quint, J.K., et al., *Changes in the incidence, prevalence and mortality of bronchiectasis in the UK from 2004 to 2013: a population-based cohort study.* Eur Respir J, 2016. **47**(1): p. 186-93.
21. Seitz, A.E., et al., *Trends in bronchiectasis among medicare beneficiaries in the United States, 2000 to 2007.* Chest, 2012. **142**(2): p. 432-9.
22. O'Donnell, A.E., *Bronchiectasis.* Chest, 2008. **134**(4): p. 815-23.
23. Anwar, G.A., et al., *Phenotyping adults with non-cystic fibrosis bronchiectasis: a prospective observational cohort study.* Respir Med, 2013. **107**(7): p. 1001-7.

24. O'Brien, C., et al., *Physiological and radiological characterisation of patients diagnosed with chronic obstructive pulmonary disease in primary care.* Thorax, 2000. **55**(8): p. 635-42.
25. Gatheral, T., et al., *COPD-related bronchiectasis; independent impact on disease course and outcomes.* Copd, 2014. **11**(6): p. 605-14.
26. Hurst, J.R., J.S. Elborn, and A.D. Soyza, *COPD-bronchiectasis overlap syndrome.* Eur Respir J, 2015. **45**(2): p. 310-3.
27. Ni, Y., et al., *Clinical characteristics of patients with chronic obstructive pulmonary disease with comorbid bronchiectasis: a systemic review and meta-analysis.* Int J Chron Obstruct Pulmon Dis, 2015. **10**: p. 1465-75.
28. Paganin, F., et al., *Computed tomography of the lungs in asthma: influence of disease severity and etiology.* Am J Respir Crit Care Med, 1996. **153**(1): p. 110-4.
29. Bibby, S., R. Milne, and R. Beasley, *Hospital admissions for non-cystic fibrosis bronchiectasis in New Zealand.* N Z Med J, 2015. **128**(1421): p. 30-8.
30. Cohen, M. and S.A. Sahn, *Bronchiectasis in systemic diseases.* Chest, 1999. **116**(4): p. 1063-74.
31. King, P.T., et al., *Microbiologic follow-up study in adult bronchiectasis.* Respir Med, 2007. **101**(8): p. 1633-8.
32. Millard, J., C. Ugarte-Gil, and D.A. Moore, *Multidrug resistant tuberculosis.* Bmj, 2015. **350**: p. h882.
33. Lonni, S., et al., *Etiology of Non-Cystic Fibrosis Bronchiectasis in Adults and Its Correlation to Disease Severity.* Ann Am Thorac Soc, 2015. **12**(12): p. 1764-70.
34. McShane, P.J., E.T. Naureckas, and M.E. Strek, *Bronchiectasis in a diverse US population: effects of ethnicity on etiology and sputum culture.* Chest, 2012. **142**(1): p. 159-67.
35. Shoemark, A., L. Ozerovitch, and R. Wilson, *Aetiology in adult patients with bronchiectasis.* Respir Med, 2007. **101**(6): p. 1163-70.
36. Daniels, M.L., et al., *Enlarged Dural Sac in Idiopathic Bronchiectasis Implicates Heritable Connective Tissue Gene Variants.* Ann Am Thorac Soc, 2016. **13**(10): p. 1712-1720.
37. Aliberti, S., et al., *Research priorities in bronchiectasis: a consensus statement from the EMBARC Clinical Research Collaboration.* Eur Respir J, 2016.
38. Wong, C., et al., *Azithromycin for prevention of exacerbations in non-cystic fibrosis bronchiectasis (EMBRACE): a randomised, double-blind, placebo-controlled trial.* Lancet, 2012. **380**(9842): p. 660-7.
39. Hill, A.T., et al., *Pulmonary exacerbation in adults with bronchiectasis: a consensus definition for clinical research.* Eur Respir J, 2017. **49**(6).
40. Chalmers, J.D., et al., *The bronchiectasis severity index. An international derivation and validation study.* Am J Respir Crit Care Med, 2014. **189**(5): p. 576-85.
41. Sanchez-Munoz, G., et al., *Time Trends in Hospital Admissions for Bronchiectasis: Analysis of the Spanish National Hospital Discharge Data (2004 to 2013).* PLoS One, 2016. **11**(9): p. e0162282.
42. Chang, A.B., et al., *Chronic suppurative lung disease and bronchiectasis in children and adults in Australia and New Zealand.* Med J Aust, 2010. **193**(6): p. 356-65.
43. Loebinger, M.R., et al., *Mortality in bronchiectasis: a long-term study assessing the factors influencing survival.* Eur Respir J, 2009. **34**(4): p. 843-9.
44. Goeminne, P.C., et al., *Risk factors for morbidity and death in non-cystic fibrosis bronchiectasis: a retrospective cross-sectional analysis of CT diagnosed bronchiectatic patients.* Respir Res, 2012. **13**: p. 21.
45. Goeminne, P.C., et al., *Mortality in non-cystic fibrosis bronchiectasis: a prospective cohort analysis.* Respir Med, 2014. **108**(2): p. 287-96.
46. Chalmers, J.D., et al., *Vitamin-D deficiency is associated with chronic bacterial colonisation and disease severity in bronchiectasis.* Thorax, 2013. **68**(1): p. 39-47.

47. Cole, P.J., *Inflammation: a two-edged sword--the model of bronchiectasis.* Eur J Respir Dis Suppl, 1986. **147**: p. 6-15.
48. Martinez-Garcia, M.A., et al., *Factors associated with lung function decline in adult patients with stable non-cystic fibrosis bronchiectasis.* Chest, 2007. **132**(5): p. 1565-72.
49. Chalmers, J.D., S. Aliberti, and F. Blasi, *Management of bronchiectasis in adults.* Eur Respir J, 2015. **45**(5): p. 1446-62.
50. Sidhu, M.K., P. Mandal, and A.T. Hill, *Bronchiectasis: an update on current pharmacotherapy and future perspectives.* Expert Opin Pharmacother, 2014. **15**(4): p. 505-25.
51. Altenburg, J., et al., *Effect of azithromycin maintenance treatment on infectious exacerbations among patients with non-cystic fibrosis bronchiectasis: the BAT randomized controlled trial.* Jama, 2013. **309**(12): p. 1251-9.
52. Serisier, D.J., et al., *Effect of long-term, low-dose erythromycin on pulmonary exacerbations among patients with non-cystic fibrosis bronchiectasis: the BLESS randomized controlled trial.* Jama, 2013. **309**(12): p. 1260-7.
53. Hare, K.M., et al., *Nasopharyngeal carriage and macrolide resistance in Indigenous children with bronchiectasis randomized to long-term azithromycin or placebo.* Eur J Clin Microbiol Infect Dis, 2015. **34**(11): p. 2275-85.
54. Serisier, D.J., et al., *Inhaled, dual release liposomal ciprofloxacin in non-cystic fibrosis bronchiectasis (ORBIT-2): a randomised, double-blind, placebo-controlled trial.* Thorax, 2013. **68**(9): p. 812-7.
55. Wilson, R., et al., *Ciprofloxacin dry powder for inhalation in non-cystic fibrosis bronchiectasis: a phase II randomised study.* Eur Respir J, 2013. **41**(5): p. 1107-15.
56. Murray, M.P., J.L. Pentland, and A.T. Hill, *A randomised crossover trial of chest physiotherapy in non-cystic fibrosis bronchiectasis.* Eur Respir J, 2009. **34**(5): p. 1086-92.
57. Lee, A.L., A. Burge, and A.E. Holland, *Airway clearance techniques for bronchiectasis.* Cochrane Database Syst Rev, 2013. **5**: p. Cd008351.
58. Munoz, G., et al., *Long-term benefits of airway clearance in bronchiectasis: a randomised placebo-controlled trial.* Eur Respir J, 2018. **51**(1).
59. Kellett, F. and N.M. Robert, *Nebulised 7% hypertonic saline improves lung function and quality of life in bronchiectasis.* Respir Med, 2011. **105**(12): p. 1831-5.
60. Nicolson, C.H., et al., *The long term effect of inhaled hypertonic saline 6% in non-cystic fibrosis bronchiectasis.* Respir Med, 2012. **106**(5): p. 661-7.
61. Daviskas, E., et al., *Effect of increasing doses of mannitol on mucus clearance in patients with bronchiectasis.* Eur Respir J, 2008. **31**(4): p. 765-72.
62. Bilton, D., et al., *Phase 3 randomized study of the efficacy and safety of inhaled dry powder mannitol for the symptomatic treatment of non-cystic fibrosis bronchiectasis.* Chest, 2013. **144**(1): p. 215-25.
63. Hassan, J.A., et al., *Bronchodilator response to inhaled beta-2 agonist and anticholinergic drugs in patients with bronchiectasis.* Respirology, 1999. **4**(4): p. 423-6.
64. Tsang, K.W., et al., *Inhaled fluticasone reduces sputum inflammatory indices in severe bronchiectasis.* Am J Respir Crit Care Med, 1998. **158**(3): p. 723-7.
65. Tsang, K.W., et al., *Inhaled fluticasone in bronchiectasis: a 12 month study.* Thorax, 2005. **60**(3): p. 239-43.
66. Elborn, J.S., et al., *Inhaled steroids in patients with bronchiectasis.* Respir Med, 1992. **86**(2): p. 121-4.
67. Martinez-Garcia, M.A., et al., *Inhaled steroids improve quality of life in patients with steady-state bronchiectasis.* Respir Med, 2006. **100**(9): p. 1623-32.
68. Martinez-Garcia, M.A., et al., *Clinical efficacy and safety of budesonide-formoterol in non-cystic fibrosis bronchiectasis.* Chest, 2012. **141**(2): p. 461-8.
69. Llewellyn-Jones, C.G., et al., *In vivo study of indomethacin in bronchiectasis: effect on neutrophil function and lung secretion.* Eur Respir J, 1995. **8**(9): p. 1479-87.

70. Mandal, P., et al., *Atorvastatin as a stable treatment in bronchiectasis: a randomised controlled trial.* Lancet Respir Med, 2014. **2**(6): p. 455-63.
71. Zhang, P., et al., *Video-assisted thoracic surgery for bronchiectasis.* Ann Thorac Surg, 2011. **91**(1): p. 239-43.
72. Vallilo, C.C., et al., *Lung resection improves the quality of life of patients with symptomatic bronchiectasis.* Ann Thorac Surg, 2014. **98**(3): p. 1034-41.
73. McDonnell, M.J., et al., *Comorbidities and the risk of mortality in patients with bronchiectasis: an international multicentre cohort study.* Lancet Respir Med, 2016. **4**(12): p. 969-979.
74. Waters, E.M., et al., *Phage therapy is highly effective against chronic lung infections with Pseudomonas aeruginosa.* Thorax, 2017.
75. Milla, C.E., et al., *Anti-PcrV antibody in cystic fibrosis: a novel approach targeting Pseudomonas aeruginosa airway infection.* Pediatr Pulmonol, 2014. **49**(7): p. 650-8.
76. De Soyza, A., et al., *A randomised, placebo-controlled study of the CXCR2 antagonist AZD5069 in bronchiectasis.* Eur Respir J, 2015. **46**(4): p. 1021-32.
77. Stephenson, A.L., et al., *Survival Comparison of Patients With Cystic Fibrosis in Canada and the United States: A Population-Based Cohort Study.* Ann Intern Med, 2017.
78. Goss, C.H., et al., *Children and young adults with CF in the USA have better lung function compared with the UK.* Thorax, 2015. **70**(3): p. 229-36.
79. Mehta, G., M. Macek, Jr., and A. Mehta, *Cystic fibrosis across Europe: EuroCareCF analysis of demographic data from 35 countries.* J Cyst Fibros, 2010. **9 Suppl 2**: p. S5-s21.
80. Registry, U.C.F., *UK Cystic Fibrosis Registry 2014 Annual Report.* 2015.
81. Foundation, C.F., *Patient Registry Annual Data Report 2014.* 2015.
82. Registry, U.C.F., *UK Cystic Fibrosis Registry 2015 Annual Report.* 2016

83. Andersen, D.H., *Cystic fibrosis of the pancreas and its relation to celiac disease: A clinical and pathologic study.* Am J Dis Child, 1938. **56**(2): p. 344-399.
84. Andersen, D.H. and R.G. Hodges, *Celiac syndrome; genetics of cystic fibrosis of the pancreas, with a consideration of etiology.* Am J Dis Child, 1946. **72**: p. 62-80.
85. Rommens, J.M., et al., *Identification of the cystic fibrosis gene: chromosome walking and jumping.* Science, 1989. **245**(4922): p. 1059-65.
86. O'Sullivan, B.P. and S.D. Freedman, *Cystic fibrosis.* Lancet, 2009. **373**(9678): p. 1891-904.
87. Elborn, J.S., *Cystic fibrosis.* Lancet, 2016. **388**(10059): p. 2519-2531.
88. Sosnay, P.R., et al., *Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene.* Nat Genet, 2013. **45**(10): p. 1160-7.
89. Britto, M.T., et al., *Impact of recent pulmonary exacerbations on quality of life in patients with cystic fibrosis.* Chest, 2002. **121**(1): p. 64-72.
90. Sanders, D.B., et al., *Pulmonary exacerbations are associated with subsequent FEV1 decline in both adults and children with cystic fibrosis.* Pediatr Pulmonol, 2011. **46**(4): p. 393-400.
91. de Boer, K., et al., *Exacerbation frequency and clinical outcomes in adult patients with cystic fibrosis.* Thorax, 2011. **66**(8): p. 680-5.
92. Stephenson, A.L., et al., *A contemporary survival analysis of individuals with cystic fibrosis: a cohort study.* Eur Respir J, 2015. **45**(3): p. 670-9.
93. Burgel, P.R., et al., *Future trends in cystic fibrosis demography in 34 European countries.* Eur Respir J, 2015. **46**(1): p. 133-41.
94. Doring, G. and N. Hoiby, *Early intervention and prevention of lung disease in cystic fibrosis: a European consensus.* J Cyst Fibros, 2004. **3**(2): p. 67-91.
95. Smyth, A.R., et al., *European Cystic Fibrosis Society Standards of Care: Best Practice guidelines.* J Cyst Fibros, 2014. **13 Suppl 1**: p. S23-42.

96. Fuchs, H.J., et al., *Effect of aerosolized recombinant human DNase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis. The Pulmozyme Study Group.* N Engl J Med, 1994. **331**(10): p. 637-42.
97. Wark, P. and V.M. McDonald, *Nebulised hypertonic saline for cystic fibrosis.* Cochrane Database Syst Rev, 2009(2): p. Cd001506.
98. Bilton, D., et al., *Pooled analysis of two large randomised phase III inhaled mannitol studies in cystic fibrosis.* J Cyst Fibros, 2013. **12**(4): p. 367-76.
99. Alton, E.W., et al., *Repeated nebulisation of non-viral CFTR gene therapy in patients with cystic fibrosis: a randomised, double-blind, placebo-controlled, phase 2b trial.* Lancet Respir Med, 2015. **3**(9): p. 684-91.
100. Ramsey, B.W., et al., *A CFTR potentiator in patients with cystic fibrosis and the G551D mutation.* N Engl J Med, 2011. **365**(18): p. 1663-72.
101. Van Goor, F., et al., *Effect of ivacaftor on CFTR forms with missense mutations associated with defects in protein processing or function.* J Cyst Fibros, 2014. **13**(1): p. 29-36.
102. Davies, J., et al., *Assessment of clinical response to ivacaftor with lung clearance index in cystic fibrosis patients with a G551D-CFTR mutation and preserved spirometry: a randomised controlled trial.* Lancet Respir Med, 2013. **1**(8): p. 630-8.
103. Barry, P.J., et al., *Effects of ivacaftor in patients with cystic fibrosis who carry the G551D mutation and have severe lung disease.* Chest, 2014. **146**(1): p. 152-8.
104. Rowe, S.M., et al., *Clinical mechanism of the cystic fibrosis transmembrane conductance regulator potentiator ivacaftor in G551D-mediated cystic fibrosis.* Am J Respir Crit Care Med, 2014. **190**(2): p. 175-84.
105. Wainwright, C.E., et al., *Lumacaftor-Ivacaftor in Patients with Cystic Fibrosis Homozygous for Phe508del CFTR.* N Engl J Med, 2015. **373**(3): p. 220-31.
106. Boyle, M.P., et al., *A CFTR corrector (lumacaftor) and a CFTR potentiator (ivacaftor) for treatment of patients with cystic fibrosis who have a phe508del CFTR mutation: a phase 2 randomised controlled trial.* Lancet Respir Med, 2014. **2**(7): p. 527-38.
107. McDonnell, M.J., et al., *Non cystic fibrosis bronchiectasis: A longitudinal retrospective observational cohort study of Pseudomonas persistence and resistance.* Respir Med, 2014.
108. Nicotra, M.B., et al., *Clinical, pathophysiologic, and microbiologic characterization of bronchiectasis in an aging cohort.* Chest, 1995. **108**(4): p. 955-61.
109. Palwatwichai, A., et al., *Clinical, laboratory findings and microbiologic characterization of bronchiectasis in Thai patients.* Respirology, 2002. **7**(1): p. 63-6.
110. Finch, S., et al., *A Comprehensive Analysis of the Impact of Pseudomonas aeruginosa Colonization on Prognosis in Adult Bronchiectasis.* Ann Am Thorac Soc, 2015. **12**(11): p. 1602-11.
111. Evans, S., et al., *Lung function in bronchiectasis: the influence of Pseudomonas aeruginosa.* Eur Respir J, 1996. **9**(8): p. 1601-1604.
112. Davies, G., et al., *The effect of Pseudomonas aeruginosa on pulmonary function in patients with bronchiectasis.* Eur Respir J, 2006. **28**(5): p. 974-9.
113. Wilson, C.B., et al., *Effect of sputum bacteriology on the quality of life of patients with bronchiectasis.* Eur Respir J, 1997. **10**(8): p. 1754-60.
114. Miszkiel, K.A., et al., *Effects of airway infection by Pseudomonas aeruginosa: a computed tomographic study.* Thorax, 1997. **52**(3): p. 260-4.
115. Chalmers, J.D., et al., *Short- and long-term antibiotic treatment reduces airway and systemic inflammation in non-cystic fibrosis bronchiectasis.* Am J Respir Crit Care Med, 2012. **186**(7): p. 657-65.
116. Pedersen, B. and A. Kharazmi, *Inhibition of human natural killer cell activity by Pseudomonas aeruginosa alkaline protease and elastase.* Infection and immunity, 1987. **55**(4): p. 986-989.

117. Kharazmi, A., et al., *Pseudomonas aeruginosa exoproteases inhibit human neutrophil chemiluminescence.* Infection and immunity, 1984. **44**(3): p. 587-591.
118. Vasil, M.L., *Pseudomonas aeruginosa: biology, mechanisms of virulence, epidemiology.* J Pediatr, 1986. **108**(5 Pt 2): p. 800-5.
119. Wilson, R., et al., *Pyocyanin and 1-hydroxyphenazine produced by Pseudomonas aeruginosa inhibit the beating of human respiratory cilia in vitro.* J Clin Invest, 1987. **79**(1): p. 221-9.
120. Haworth, C.S., et al., *Inhaled colistin in patients with bronchiectasis and chronic Pseudomonas aeruginosa infection.* Am J Respir Crit Care Med, 2014. **189**(8): p. 975-82.
121. Henry, R.L., C.M. Mellis, and L. Petrovic, *Mucoid Pseudomonas aeruginosa is a marker of poor survival in cystic fibrosis.* Pediatr Pulmonol, 1992. **12**(3): p. 158-61.
122. Kerem, E., et al., *Pulmonary function and clinical course in patients with cystic fibrosis after pulmonary colonization with Pseudomonas aeruginosa.* J Pediatr, 1990. **116**(5): p. 714-9.
123. Kosorok, M.R., et al., *Acceleration of lung disease in children with cystic fibrosis after Pseudomonas aeruginosa acquisition.* Pediatr Pulmonol, 2001. **32**(4): p. 277-87.
124. Pamukcu, A., A. Bush, and R. Buchdahl, *Effects of pseudomonas aeruginosa colonization on lung function and anthropometric variables in children with cystic fibrosis.* Pediatr Pulmonol, 1995. **19**(1): p. 10-5.
125. Emerson, J., et al., *Pseudomonas aeruginosa and other predictors of mortality and morbidity in young children with cystic fibrosis.* Pediatr Pulmonol, 2002. **34**(2): p. 91-100.
126. Konstan, M.W., et al., *Risk factors for rate of decline in forced expiratory volume in one second in children and adolescents with cystic fibrosis.* J Pediatr, 2007. **151**(2): p. 134-9, 139.e1.
127. Courtney, J.M., et al., *Predictors of mortality in adults with cystic fibrosis.* Pediatr Pulmonol, 2007. **42**(6): p. 525-32.
128. Qvist, T., et al., *Comparing the harmful effects of nontuberculous mycobacteria and Gram negative bacteria on lung function in patients with cystic fibrosis.* J Cyst Fibros, 2016. **15**(3): p. 380-5.
129. Burns, J.L., et al., *Longitudinal assessment of Pseudomonas aeruginosa in young children with cystic fibrosis.* J Infect Dis, 2001. **183**(3): p. 444-52.
130. Gaspar, M.C., et al., *Pseudomonas aeruginosa infection in cystic fibrosis lung disease and new perspectives of treatment: a review.* Eur J Clin Microbiol Infect Dis, 2013. **32**(10): p. 1231-52.
131. De Soyza, A., et al., *Molecular epidemiological analysis suggests cross-infection with Pseudomonas aeruginosa is rare in non-cystic fibrosis bronchiectasis.* Eur Respir J, 2014. **43**(3): p. 900-3.
132. Parkins, M.D. and R.A. Floto, *Emerging bacterial pathogens and changing concepts of bacterial pathogenesis in cystic fibrosis.* J Cyst Fibros, 2015. **14**(3): p. 293-304.
133. Flight, W.G., et al., *Rapid detection of emerging pathogens and the loss of microbial diversity associated with severe lung disease in cystic fibrosis.* J Clin Microbiol, 2015.
134. Hendricks, M.R. and J.M. Bomberger, *Digging through the Obstruction: Insight into the Epithelial Cell Response to Respiratory Virus Infection in Patients with Cystic Fibrosis.* J Virol, 2016. **90**(9): p. 4258-61.
135. Chattoraj, S.S., et al., *Rhinovirus infection liberates planktonic bacteria from biofilm and increases chemokine responses in cystic fibrosis airway epithelial cells.* Thorax, 2011. **66**(4): p. 333-9.
136. Stevens, D.A., et al., *Allergic bronchopulmonary aspergillosis in cystic fibrosis--state of the art: Cystic Fibrosis Foundation Consensus Conference.* Clin Infect Dis, 2003. **37 Suppl 3**: p. S225-64.

137. Moss, R.B., *Fungi in cystic fibrosis and non-cystic fibrosis bronchiectasis.* Semin Respir Crit Care Med, 2015. **36**(2): p. 207-16.
138. Iversen, M., et al., *Aspergillus infection in lung transplant patients: incidence and prognosis.* Eur J Clin Microbiol Infect Dis, 2007. **26**(12): p. 879-86.
139. Kondori, N., et al., *Development of IgG antibodies to Exophiala dermatitidis is associated with inflammatory responses in patients with cystic fibrosis.* J Cyst Fibros, 2014. **13**(4): p. 391-9.
140. Williams, K.P., et al., *Phylogeny of gammaproteobacteria.* J Bacteriol, 2010. **192**(9): p. 2305-14.
141. Mathee, K., et al., *Dynamics of Pseudomonas aeruginosa genome evolution.* Proc Natl Acad Sci U S A, 2008. **105**(8): p. 3100-5.
142. Klockgether, J., et al., *Pseudomonas aeruginosa Genomic Structure and Diversity.* Front Microbiol, 2011. **2**: p. 150.
143. Tummler, B., et al., *Advances in understanding Pseudomonas.* F1000Prime Rep, 2014. **6**: p. 9.
144. Todar, K. *Todar's Online Textbook of Bacteriology*. Available from: http://www.textbookofbacteriology.net/index.html.
145. Romling, U., et al., *Worldwide distribution of Pseudomonas aeruginosa clone C strains in the aquatic environment and cystic fibrosis patients.* Environ Microbiol, 2005. **7**(7): p. 1029-38.
146. Eisele, N.A. and D.M. Anderson, *Host Defense and the Airway Epithelium: Frontline Responses That Protect against Bacterial Invasion and Pneumonia.* J Pathog, 2011. **2011**: p. 249802.
147. Locke, L.W., et al., *Pseudomonas infection and mucociliary and absorptive clearance in the cystic fibrosis lung.* Eur Respir J, 2016. **47**(5): p. 1392-401.
148. Tsang, K.W., et al., *Interaction of Pseudomonas aeruginosa with human respiratory mucosa in vitro.* Eur Respir J, 1994. **7**(10): p. 1746-53.
149. Plotkowski, M.C., et al., *Differential adhesion of Pseudomonas aeruginosa to human respiratory epithelial cells in primary culture.* J Clin Invest, 1991. **87**(6): p. 2018-28.
150. Plotkowski, M.C., et al., *Adherence of Pseudomonas aeruginosa to respiratory epithelium and the effect of leucocyte elastase.* J Med Microbiol, 1989. **30**(4): p. 285-93.
151. Kanthakumar, K., et al., *Mechanisms of action of Pseudomonas aeruginosa pyocyanin on human ciliary beat in vitro.* Infection and immunity, 1993. **61**(7): p. 2848-2853.
152. Deligianni, E., et al., *Pseudomonas aeruginosa cystic fibrosis isolates of similar RAPD genotype exhibit diversity in biofilm forming ability in vitro.* BMC Microbiol, 2010. **10**: p. 38.
153. Cullen, L., et al., *Phenotypic characterisation of an international Pseudomonas aeruginosa reference panel: Strains of cystic fibrosis origin show less in vivo virulence than non-CF strains.* Microbiology, 2015.
154. Smith, E.E., et al., *Genetic adaptation by Pseudomonas aeruginosa to the airways of cystic fibrosis patients.* Proc Natl Acad Sci U S A, 2006. **103**(22): p. 8487-92.
155. Mahenthiralingam, E., M.E. Campbell, and D.P. Speert, *Nonmotility and phagocytic resistance of Pseudomonas aeruginosa isolates from chronically colonized patients with cystic fibrosis.* Infect Immun, 1994. **62**(2): p. 596-605.
156. D'Argenio, D.A., et al., *Growth phenotypes of Pseudomonas aeruginosa lasR mutants adapted to the airways of cystic fibrosis patients.* Mol Microbiol, 2007. **64**(2): p. 512-33.
157. Oberhardt, M.A., et al., *Metabolic network analysis of Pseudomonas aeruginosa during chronic cystic fibrosis lung infection.* J Bacteriol, 2010. **192**(20): p. 5534-48.
158. Kenna, D.T., et al., *Hypermutability in environmental Pseudomonas aeruginosa and in populations causing pulmonary infection in individuals with cystic fibrosis.* Microbiology, 2007. **153**(Pt 6): p. 1852-9.

159. Oliver, A., et al., *High frequency of hypermutable Pseudomonas aeruginosa in cystic fibrosis lung infection.* Science, 2000. **288**(5469): p. 1251-4.
160. Gutierrez, O., et al., *Lack of association between hypermutation and antibiotic resistance development in Pseudomonas aeruginosa isolates from intensive care unit patients.* Antimicrob Agents Chemother, 2004. **48**(9): p. 3573-5.
161. Macia, M.D., et al., *Hypermutation is a key factor in development of multiple-antimicrobial resistance in Pseudomonas aeruginosa strains causing chronic lung infections.* Antimicrob Agents Chemother, 2005. **49**(8): p. 3382-6.
162. Matsui, H., et al., *A physical linkage between cystic fibrosis airway surface dehydration and Pseudomonas aeruginosa biofilms.* Proc Natl Acad Sci U S A, 2006. **103**(48): p. 18131-6.
163. Imundo, L., et al., *Cystic fibrosis epithelial cells have a receptor for pathogenic bacteria on their apical surface.* Proc Natl Acad Sci U S A, 1995. **92**(7): p. 3019-23.
164. Worlitzsch, D., et al., *Effects of reduced mucus oxygen concentration in airway Pseudomonas infections of cystic fibrosis patients.* J Clin Invest, 2002. **109**(3): p. 317-25.
165. White, L., et al., *Outcomes of Pseudomonas eradication therapy in patients with non-cystic fibrosis bronchiectasis.* Respir Med, 2012. **106**(3): p. 356-60.
166. Orriols, R., et al., *Eradication Therapy against Pseudomonas aeruginosa in Non-Cystic Fibrosis Bronchiectasis.* Respiration, 2015. **90**(4): p. 299-305.
167. Langton Hewer, S.C. and A.R. Smyth, *Antibiotic strategies for eradicating Pseudomonas aeruginosa in people with cystic fibrosis.* Cochrane Database Syst Rev, 2014. **11**: p. Cd004197.
168. Dhar, R., et al., *Efficacy of nebulised colomycin in patients with non-cystic fibrosis bronchiectasis colonised with Pseudomonas aeruginosa.* Thorax, 2010. **65**(6): p. 553.
169. Steinfort, D.P. and C. Steinfort, *Effect of long-term nebulized colistin on lung function and quality of life in patients with chronic bronchial sepsis.* Intern Med J, 2007. **37**(7): p. 495-8.
170. Barker, A.F., et al., *Tobramycin solution for inhalation reduces sputum Pseudomonas aeruginosa density in bronchiectasis.* Am J Respir Crit Care Med, 2000. **162**(2 Pt 1): p. 481-5.
171. Schelstraete, P., et al., *Pseudomonas aeruginosa in the home environment of newly infected cystic fibrosis patients.* Eur Respir J, 2008. **31**(4): p. 822-9.
172. Genevois, A., et al., *Bacterial colonization status of cystic fibrosis children's toothbrushes: A pilot study.* Arch Pediatr, 2015. **22**(12): p. 1240-6.
173. Ranganathan, S.C., et al., *Geographical differences in first acquisition of Pseudomonas aeruginosa in cystic fibrosis.* Ann Am Thorac Soc, 2013. **10**(2): p. 108-14.
174. Goeminne, P.C., et al., *Proximity to blue spaces and risk of infection with Pseudomonas aeruginosa in cystic fibrosis: A case–control analysis.* J Cyst Fibros, 2015. **14**(6): p. 741-747.
175. Pirnay, J.P., et al., *Global Pseudomonas aeruginosa biodiversity as reflected in a Belgian river.* Environ Microbiol, 2005. **7**(7): p. 969-80.
176. Collaco, J.M., et al., *Respiratory pathogens mediate the association between lung function and temperature in cystic fibrosis.* J Cyst Fibros, 2016. **15**(6): p. 794-801.
177. Psoter, K.J., et al., *Association of meteorological and geographical factors and risk of initial Pseudomonas aeruginosa acquisition in young children with cystic fibrosis.* Epidemiol Infect, 2016. **144**(5): p. 1075-83.
178. Knibbs, L.D., et al., *Viability of Pseudomonas aeruginosa in cough aerosols generated by persons with cystic fibrosis.* Thorax, 2014. **69**(8): p. 740-5.
179. Saiman, L., et al., *Infection prevention and control guideline for cystic fibrosis: 2013 update.* Infect Control Hosp Epidemiol, 2014. **35 Suppl 1**: p. S1-s67.

180. Zimakoff, J., et al., *Epidemiology of Pseudomonas aeruginosa infection and the role of contamination of the environment in a cystic fibrosis clinic.* J Hospital Infect, 1983. **4**(1): p. 31-40.
181. Mayr, A., et al., *Nosocomial outbreak of extensively drug-resistant Pseudomonas aeruginosa associated with aromatherapy.* Am J Infect Control, 2016.
182. Schaffer, K., *Epidemiology of infection and current guidelines for infection prevention in cystic fibrosis patients.* J Hospital Infect, 2015. **89**(4): p. 309-313.
183. Clifton, I.J., et al., *An aerobiological model of aerosol survival of different strains of Pseudomonas aeruginosa isolated from people with cystic fibrosis.* J Cyst Fibros, 2010. **9**(1): p. 64-8.
184. Panagea, S., et al., *Environmental contamination with an epidemic strain of Pseudomonas aeruginosa in a Liverpool cystic fibrosis centre, and study of its survival on dry surfaces.* J Hosp Infect, 2005. **59**(2): p. 102-7.
185. Thomassen, M.J., et al., *Pseudomonas aeruginosa isolates: comparisons of isolates from campers and from sibling pairs with cystic fibrosis.* Pediatric pulmonology, 1985. **1**(1): p. 40-45.
186. Tummler, B., et al., *Nosocomial acquisition of Pseudomonas aeruginosa by cystic fibrosis patients.* J Clin Microbiol, 1991. **29**(6): p. 1265-7.
187. Ojeniyi, B., B. Frederiksen, and N. Hoiby, *Pseudomonas aeruginosa cross-infection among patients with cystic fibrosis during a winter camp.* Pediatr Pulmonol, 2000. **29**(3): p. 177-81.
188. Hunfeld, K.P., et al., *Risk of Pseudomonas aeruginosa cross-colonisation in patients with cystic fibrosis within a holiday camp--a molecular-epidemiological study.* Wien Klin Wochenschr, 2000. **112**(7): p. 329-33.
189. Hoogkamp-Korstanje, J.A., et al., *Risk of cross-colonization and infection by Pseudomonas aeruginosa in a holiday camp for cystic fibrosis patients.* J Clin Microbiol, 1995. **33**(3): p. 572-5.
190. Grothues, D., et al., *Genome fingerprinting of Pseudomonas aeruginosa indicates colonization of cystic fibrosis siblings with closely related strains.* J Clin Microbiol, 1988. **26**(10): p. 1973-7.
191. Somayaji, R., et al., *Infection control knowledge, beliefs and behaviours amongst cystic fibrosis patients with epidemic Pseudomonas aeruginosa.* BMC pulmonary medicine, 2015. **15**(1): p. 138.
192. Cheng, K., et al., *Spread of beta-lactam-resistant Pseudomonas aeruginosa in a cystic fibrosis clinic.* Lancet, 1996. **348**(9028): p. 639-42.
193. Pedersen, S.S., et al., *An epidemic spread of multiresistant Pseudomonas aeruginosa in a cystic fibrosis centre.* J Antimicrob Chemother, 1986. **17**(4): p. 505-16.
194. Aaron, S.D., et al., *Infection with transmissible strains of Pseudomonas aeruginosa and clinical outcomes in adults with cystic fibrosis.* Jama, 2010. **304**(19): p. 2145-53.
195. Armstrong, D.S., et al., *Detection of a widespread clone of Pseudomonas aeruginosa in a pediatric cystic fibrosis clinic.* Am J Respir Crit Care Med, 2002. **166**(7): p. 983-7.
196. Al-Aloul, M., et al., *Increased morbidity associated with chronic infection by an epidemic Pseudomonas aeruginosa strain in CF patients.* Thorax, 2004. **59**(4): p. 334-6.
197. Jones, A.M., et al., *Increased treatment requirements of patients with cystic fibrosis who harbour a highly transmissible strain of Pseudomonas aeruginosa.* Thorax, 2002. **57**(11): p. 924-5.
198. Jones, A.M., et al., *Clinical outcome for cystic fibrosis patients infected with transmissible pseudomonas aeruginosa: an 8-year prospective study.* Chest, 2010. **137**(6): p. 1405-9.
199. Edenborough, F.P., et al., *Genotyping of Pseudomonas aeruginosa in cystic fibrosis suggests need for segregation.* J Cyst Fibros, 2004. **3**(1): p. 37-44.
200. van Mansfeld, R., et al., *The Effect of Strict Segregation on Pseudomonas aeruginosa in Cystic Fibrosis Patients.* PLoS One, 2016. **11**(6): p. e0157189.

201. Shepherd, S.L., et al., *Counterpoint: does the risk of cross infection warrant exclusion of adults with cystic fibrosis from cystic fibrosis foundation events? No.* Chest, 2014. **145**(4): p. 680-3.
202. Schmid, J., et al., *Pseudomonas aeruginosa transmission is infrequent in New Zealand cystic fibrosis clinics.* Eur Respir J, 2008. **32**(6): p. 1583-90.
203. Speert, D.P., et al., *Epidemiology of Pseudomonas aeruginosa in cystic fibrosis in British Columbia, Canada.* Am J Respir Crit Care Med, 2002. **166**(7): p. 988-93.
204. Frederiksen, B., C. Koch, and N. Hoiby, *Changing epidemiology of Pseudomonas aeruginosa infection in Danish cystic fibrosis patients (1974-1995).* Pediatr Pulmonol, 1999. **28**(3): p. 159-66.
205. Ashish, A., et al., *Halting the spread of epidemic pseudomonas aeruginosa in an adult cystic fibrosis centre: a prospective cohort study.* JRSM Short Rep, 2013. **4**(1): p. 1.
206. Griffiths, A.L., et al., *Australian epidemic strain pseudomonas (AES-1) declines further in a cohort segregated cystic fibrosis clinic.* J Cyst Fibros, 2012. **11**(1): p. 49-52.
207. Jones, A.M., et al., *Prospective surveillance for Pseudomonas aeruginosa cross-infection at a cystic fibrosis center.* Am J Respir Crit Care Med, 2005. **171**(3): p. 257-60.
208. Griffiths, A.L., et al., *Effects of segregation on an epidemic Pseudomonas aeruginosa strain in a cystic fibrosis clinic.* Am J Respir Crit Care Med, 2005. **171**(9): p. 1020-5.
209. Robinson, P., et al., *Pseudomonas cross-infection from cystic fibrosis patients to non-cystic fibrosis patients: implications for inpatient care of respiratory patients.* J Clin Microbiol, 2003. **41**(12): p. 5741.
210. Pujana, I., et al., *Epidemiological analysis of sequential Pseudomonas aeruginosa isolates from chronic bronchiectasis patients without cystic fibrosis.* J Clin Microbiol, 1999. **37**(6): p. 2071-3.
211. Hilliam, Y., et al., *Pseudomonas aeruginosa adaptation and diversification in the non-cystic fibrosis bronchiectasis lung.* Eur Respir J, 2017. **49**(4).
212. Pirnay, J.P., et al., *Pseudomonas aeruginosa population structure revisited.* PLoS One, 2009. **4**(11): p. e7740.
213. Curran, B., et al., *Development of a multilocus sequence typing scheme for the opportunistic pathogen Pseudomonas aeruginosa.* J Clin Microbiol, 2004. **42**(12): p. 5644-9.
214. Wiehlmann, L., et al., *Population structure of Pseudomonas aeruginosa.* Proc Natl Acad Sci U S A, 2007. **104**(19): p. 8101-6.
215. Maatallah, M., et al., *Population structure of Pseudomonas aeruginosa from five Mediterranean countries: evidence for frequent recombination and epidemic occurrence of CC235.* PLoS One, 2011. **6**(10): p. e25617.
216. Vernez, I., et al., *Population genetic analysis of Pseudomonas aeruginosa using multilocus sequence typing.* FEMS Immunol Med Microbiol, 2005. **43**(1): p. 29-35.
217. Wiehlmann, L., N. Cramer, and B. Tummler, *Habitat-associated skew of clone abundance in the Pseudomonas aeruginosa population.* Environ Microbiol Rep, 2015. **7**(6): p. 955-60.
218. Romling, U., et al., *A major Pseudomonas aeruginosa clone common to patients and aquatic habitats.* Appl Environ Microbiol, 1994. **60**(6): p. 1734-8.
219. Rahme, L.G., et al., *Common virulence factors for bacterial pathogenicity in plants and animals.* Science, 1995. **268**(5219): p. 1899-902.
220. Kidd, T.J., et al., *Pseudomonas aeruginosa exhibits frequent recombination, but only a limited association between genotype and ecological setting.* PLoS One, 2012. **7**(9): p. e44199.
221. Dinesh, S.D., et al., *European-wide distribution of Pseudomonas aeruginosa clone C.* Clin Microbiol Infect, 2003. **9**(12): p. 1228-33.
222. Kidd, T.J., et al., *Shared Pseudomonas aeruginosa genotypes are common in Australian cystic fibrosis centres.* Eur Respir J, 2013. **41**(5): p. 1091-100.

223. Martin, K., et al., *Clusters of genetically similar isolates of Pseudomonas aeruginosa from multiple hospitals in the UK.* J Med Microbiol, 2013. **62**(Pt 7): p. 988-1000.
224. Scott, F.W. and T.L. Pitt, *Identification and characterization of transmissible Pseudomonas aeruginosa strains in cystic fibrosis patients in England and Wales.* J Med Microbiol, 2004. **53**(Pt 7): p. 609-15.
225. Panagea, S., et al., *PCR-based detection of a cystic fibrosis epidemic strain of Pseudomonas Aeruginosa.* Mol Diagn, 2003. **7**(3-4): p. 195-200.
226. Al-Aloul, M., et al., *Acute renal failure in CF patients chronically infected by the Liverpool epidemic Pseudomonas aeruginosa strain (LES).* J Cyst Fibros, 2005. **4**(3): p. 197-201.
227. McCallum, S.J., et al., *Superinfection with a transmissible strain of Pseudomonas aeruginosa in adults with cystic fibrosis chronically colonised by P aeruginosa.* Lancet, 2001. **358**(9281): p. 558-60.
228. Mohan, K., et al., *Empyema due to a highly transmissible Pseudomonas aeruginosa strain in an adult cystic fibrosis patient.* J Med Microbiol, 2010. **59**(Pt 5): p. 614-6.
229. Jones, A.M., et al., *Spread of a multiresistant strain of Pseudomonas aeruginosa in an adult cystic fibrosis clinic.* Lancet, 2001. **358**(9281): p. 557-8.
230. Bradbury, R., A. Champion, and D.W. Reid, *Poor clinical outcomes associated with a multi-drug resistant clonal strain of Pseudomonas aeruginosa in the Tasmanian cystic fibrosis population.* Respirology, 2008. **13**(6): p. 886-92.
231. Smith, D.J., et al., *Pseudomonas aeruginosa antibiotic resistance in Australian cystic fibrosis centres.* Respirology, 2016. **21**(2): p. 329-37.
232. Fothergill, J.L., M.J. Walshaw, and C. Winstanley, *Transmissible strains of Pseudomonas aeruginosa in cystic fibrosis lung infections.* Eur Respir J, 2012. **40**(1): p. 227-38.
233. Parkins, M.D., et al., *Twenty-five-year outbreak of Pseudomonas aeruginosa infecting individuals with cystic fibrosis: identification of the prairie epidemic strain.* J Clin Microbiol, 2014. **52**(4): p. 1127-35.
234. Somayaji, R., et al., *Long-term clinical outcomes of 'Prairie Epidemic Strain' Pseudomonas aeruginosa infection in adults with cystic fibrosis.* Thorax, 2017. **72**(4): p. 333-339.
235. Luna, R.A., et al., *Molecular epidemiological surveillance of multidrug-resistant Pseudomonas aeruginosa isolates in a pediatric population of patients with cystic fibrosis and determination of risk factors for infection with the Houston-1 strain.* J Clin Microbiol, 2013. **51**(4): p. 1237-40.
236. Mulet, X., et al., *Biological markers of Pseudomonas aeruginosa epidemic high-risk clones.* Antimicrob Agents Chemother, 2013. **57**(11): p. 5527-35.
237. Woodford, N., J.F. Turton, and D.M. Livermore, *Multiresistant Gram-negative bacteria: the role of high-risk clones in the dissemination of antibiotic resistance.* FEMS microbiology reviews, 2011. **35**(5): p. 736-755.
238. Wright, L.L., et al., *Dominance of international 'high-risk clones' among metallo-beta-lactamase-producing Pseudomonas aeruginosa in the UK.* J Antimicrob Chemother, 2015. **70**(1): p. 103-10.
239. Cholley, P., et al., *Population structure of clinical Pseudomonas aeruginosa from West and Central African countries.* PLoS One, 2014. **9**(9): p. e107008.
240. Edelstein, M.V., et al., *Spread of extensively resistant VIM-2-positive ST235 Pseudomonas aeruginosa in Belarus, Kazakhstan, and Russia: a longitudinal epidemiological and clinical study.* Lancet Infect Dis, 2013. **13**(10): p. 867-76.
241. Silva, F.M., et al., *SPM-1-producing Pseudomonas aeruginosa: analysis of the ancestor relationship using multilocus sequence typing, pulsed-field gel electrophoresis, and automated ribotyping.* Microb Drug Resist, 2011. **17**(2): p. 215-20.
242. Wi, Y.M., et al., *Emergence of colistin resistance in Pseudomonas aeruginosa ST235 clone in South Korea.* Int J Antimicrob Agents, 2017. **49**(6): p. 767-769.

243. Tenover, F.C., et al., *Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing.* J Clin Microbiol, 1995. **33**(9): p. 2233-9.
244. Mahenthiralingam, E., et al., *Random amplified polymorphic DNA typing of Pseudomonas aeruginosa isolates recovered from patients with cystic fibrosis.* J Clin Microbiol, 1996. **34**(5): p. 1129-35.
245. Suzuki, M., et al., *Applying a PCR-based open-reading frame typing method for easy genotyping and molecular epidemiological analysis of Pseudomonas aeruginosa.* J Appl Microbiol, 2016. **120**(2): p. 487-97.
246. Bukanov, N., et al., *Pseudomonas aeruginosa corneal ulcer isolates distinguished using the arbitrarily primed PCR DNA fingerprinting method.* Curr Eye Res, 1994. **13**(11): p. 783-90.
247. Kersulyte, D., et al., *Comparison of arbitrarily primed PCR and macrorestriction (pulsed-field gel electrophoresis) typing of Pseudomonas aeruginosa strains from cystic fibrosis patients.* J Clin Microbiol, 1995. **33**(8): p. 2216-9.
248. Reiter, R.S., et al., *Global and local genome mapping in Arabidopsis thaliana by using recombinant inbred lines and random amplified polymorphic DNAs.* Proc Natl Acad Sci U S A, 1992. **89**(4): p. 1477-81.
249. De Vos, D., et al., *Analysis of epidemic Pseudomonas aeruginosa isolates by isoelectric focusing of pyoverdine and RAPD-PCR: modern tools for an integrated anti-nosocomial infection strategy in burn wound centres.* Burns, 1997. **23**(5): p. 379-86.
250. Trautmann, M., et al., *Common RAPD pattern of Pseudomonas aeruginosa from patients and tap water in a medical intensive care unit.* Int J Hyg Environ Health, 2006. **209**(4): p. 325-31.
251. Hernandez, J., et al., *Arbitrary primed PCR fingerprinting and serotyping of clinical Pseudomonas aeruginosa strains.* FEMS Immunol Med Microbiol, 1997. **17**(1): p. 37-47.
252. Waters, V., et al., *Comparison of three typing methods for Pseudomonas aeruginosa isolates from patients with cystic fibrosis.* Eur J Clin Microbiol Infect Dis, 2012. **31**(12): p. 3341-50.
253. Power, E.G., *RAPD typing in microbiology--a technical review.* J Hosp Infect, 1996. **34**(4): p. 247-65.
254. Maiden, M.C., et al., *Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms.* Proc Natl Acad Sci U S A, 1998. **95**(6): p. 3140-5.
255. Johnson, J.K., et al., *Multilocus sequence typing compared to pulsed-field gel electrophoresis for molecular typing of Pseudomonas aeruginosa.* J Clin Microbiol, 2007. **45**(11): p. 3707-12.
256. Kidd, T.J., et al., *Comparison of three molecular techniques for typing Pseudomonas aeruginosa isolates in sputum samples from patients with cystic fibrosis.* J Clin Microbiol, 2011. **49**(1): p. 263-8.
257. Fothergill, J.L., et al., *Impact of Pseudomonas aeruginosa genomic instability on the application of typing methods for chronic cystic fibrosis infections.* J Clin Microbiol, 2010. **48**(6): p. 2053-9.
258. Garcia-Castillo, M., et al., *Emergence of a mutL mutation causing multilocus sequence typing-pulsed-field gel electrophoresis discrepancy among Pseudomonas aeruginosa isolates from a cystic fibrosis patient.* J Clin Microbiol, 2012. **50**(5): p. 1777-8.
259. Lopez-Causape, C., et al., *Clonal dissemination, emergence of mutator lineages and antibiotic resistance evolution in Pseudomonas aeruginosa cystic fibrosis chronic lung infection.* PLoS One, 2013. **8**(8): p. e71001.
260. Schmidt, K.D., B. Tummler, and U. Romling, *Comparative genome mapping of Pseudomonas aeruginosa PAO with P. aeruginosa C, which belongs to a major clone in cystic fibrosis patients and aquatic habitats.* J Bacteriol, 1996. **178**(1): p. 85-93.

261. Waine, D.J., et al., *Cross-sectional and longitudinal multilocus sequence typing of pseudomonas aeruginosa in cystic fibrosis sputum samples.* J Clin Microbiol, 2009. **47**(11): p. 3444-8.
262. Eusebio, N., et al., *SNaPaer: a practical single nucleotide polymorphism multiplex assay for genotyping of Pseudomonas aeruginosa.* PLoS One, 2013. **8**(6): p. e66083.
263. Syrmis, M.W., et al., *High-throughput single-nucleotide polymorphism-based typing of shared Pseudomonas aeruginosa strains in cystic fibrosis patients using the Sequenom iPLEX platform.* J Med Microbiol, 2013. **62**(Pt 5): p. 734-40.
264. Woo, P.C., et al., *Analysis of multilocus sequence typing schemes for 35 different bacteria revealed that gene loci of 10 bacteria could be replaced to improve cost-effectiveness.* Diagn Microbiol Infect Dis, 2011. **70**(3): p. 316-23.
265. Davis, R.J., et al., *Whole Genome Sequencing in Real-Time Investigation and Management of a Pseudomonas aeruginosa Outbreak on a Neonatal Intensive Care Unit.* Infect Control Hosp Epidemiol, 2015: p. 1-7.
266. Quick, J., et al., *Seeking the source of Pseudomonas aeruginosa infections in a recently opened hospital: an observational study using whole-genome sequencing.* BMJ open, 2014. **4**(11): p. e006278.
267. Snyder, L.A., et al., *Epidemiological investigation of Pseudomonas aeruginosa isolates from a six-year-long hospital outbreak using high-throughput whole genome sequencing.* Euro Surveill, 2013. **18**(42).
268. Pankhurst, L.J., et al., *Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study.* Lancet Respir Med, 2016. **4**(1): p. 49-58.
269. Cramer, N., et al., *Microevolution of the major common Pseudomonas aeruginosa clones C and PA14 in cystic fibrosis lungs.* Environ Microbiol, 2011. **13**(7): p. 1690-704.
270. Pages-Monteiro, L., et al., *Strong incidence of Pseudomonas aeruginosa on bacterial rrs and ITS genetic structures of cystic fibrosis sputa.* PLoS One, 2017. **12**(3): p. e0173022.
271. Purcell, P., et al., *Polymicrobial airway bacterial communities in adult bronchiectasis patients.* BMC Microbiol, 2014. **14**: p. 130.
272. Sibley, C.D., et al., *A polymicrobial perspective of pulmonary infections exposes an enigmatic pathogen in cystic fibrosis patients.* Proc Natl Acad Sci U S A, 2008. **105**(39): p. 15070-5.
273. Pompilio, A., et al., *Cooperative pathogenicity in cystic fibrosis: Stenotrophomonas maltophilia modulates Pseudomonas aeruginosa virulence in mixed biofilm.* Front Microbiol, 2015. **6**: p. 951.
274. Einarsson, G.G., et al., *Community dynamics and the lower airway microbiota in stable chronic obstructive pulmonary disease, smokers and healthy non-smokers.* Thorax, 2016. **71**(9): p. 795-803.
275. Sibley, C.D., et al., *Culture enriched molecular profiling of the cystic fibrosis airway microbiome.* PLoS One, 2011. **6**(7): p. e22702.
276. Chmiel, J.F., et al., *Antibiotic management of lung infections in cystic fibrosis. I. The microbiome, methicillin-resistant Staphylococcus aureus, gram-negative bacteria, and multiple infections.* Ann Am Thorac Soc, 2014. **11**(7): p. 1120-9.
277. Rogers, G.B., et al., *Determining cystic fibrosis-affected lung microbiology: comparison of spontaneous and serially induced sputum samples by use of terminal restriction fragment length polymorphism profiling.* J Clin Microbiol, 2010. **48**(1): p. 78-86.
278. Erb-Downward, J.R., et al., *Analysis of the lung microbiome in the "healthy" smoker and in COPD.* PloS one, 2011. **6**(2): p. e16384.
279. Willner, D., et al., *Comparison of DNA extraction methods for microbial community profiling with an application to pediatric bronchoalveolar lavage samples.* PLoS One, 2012. **7**(4): p. e34605.
280. Zhao, J., et al., *Impact of enhanced Staphylococcus DNA extraction on microbial community measures in cystic fibrosis sputum.* PLoS One, 2012. **7**(3): p. e33127.

281. Rogers, G.B., et al., *Respiratory microbiota: addressing clinical questions, informing clinical practice.* Thorax, 2015. **70**(1): p. 74-81.
282. Salter, S.J., et al., *Reagent and laboratory contamination can critically impact sequence-based microbiome analyses.* BMC Biol, 2014. **12**: p. 87.
283. Cuthbertson, L., et al., *Respiratory microbiota resistance and resilience to pulmonary exacerbation and subsequent antimicrobial intervention.* Isme j, 2016. **10**(5): p. 1081-91.
284. Rogers, G.B., et al., *Bacterial diversity in cases of lung infection in cystic fibrosis patients: 16S ribosomal DNA (rDNA) length heterogeneity PCR and 16S rDNA terminal restriction fragment length polymorphism profiling.* J Clin Microbiol, 2003. **41**(8): p. 3548-58.
285. Muyzer, G., E.C. de Waal, and A.G. Uitterlinden, *Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA.* Appl Environ Microbiol, 1993. **59**(3): p. 695-700.
286. Cox, M.J., et al., *Airway microbiota and pathogen abundance in age-stratified cystic fibrosis patients.* PLoS One, 2010. **5**(6): p. e11044.
287. van Dorst, J., et al., *Community fingerprinting in a sequencing world.* FEMS Microbiol Ecol, 2014. **89**(2): p. 316-30.
288. Borneman, J. and E.W. Triplett, *Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation.* Appl Environ Microbiol, 1997. **63**(7): p. 2647-53.
289. Ranjard, L., F. Poly, and S. Nazaret, *Monitoring complex bacterial communities using culture-independent molecular techniques: application to soil environment.* Res Microbiol, 2000. **151**(3): p. 167-77.
290. Ranjard, L., et al., *Characterization of bacterial and fungal soil communities by automated ribosomal intergenic spacer analysis fingerprints: biological and methodological variability.* Appl Environ Microbiol, 2001. **67**(10): p. 4479-87.
291. Danovaro, R., et al., *Comparison of two fingerprinting techniques, terminal restriction fragment length polymorphism and automated ribosomal intergenic spacer analysis, for determination of bacterial diversity in aquatic environments.* Appl Environ Microbiol, 2006. **72**(9): p. 5982-9.
292. Jami, E., N. Shterzer, and I. Mizrahi, *Evaluation of automated ribosomal intergenic spacer analysis for bacterial fingerprinting of rumen microbiome compared to pyrosequencing technology.* Pathogens, 2014. **3**(1): p. 109-20.
293. Baxter, C.G., et al., *Intravenous antibiotics reduce the presence of Aspergillus in adult cystic fibrosis sputum.* Thorax, 2013. **68**(7): p. 652-7.
294. Nazaret, S., et al., *RISA-HPLC analysis of lung bacterial colonizers of cystic fibrosis children.* J Microbiol Methods, 2009. **76**(1): p. 58-69.
295. Fox, G.E., et al., *Classification of methanogenic bacteria by 16S ribosomal RNA characterization.* Proc Natl Acad Sci U S A, 1977. **74**(10): p. 4537-41.
296. Woese, C.R. and G.E. Fox, *Phylogenetic structure of the prokaryotic domain: the primary kingdoms.* Proc Natl Acad Sci U S A, 1977. **74**(11): p. 5088-90.
297. Pace, N.R., J. Sapp, and N. Goldenfeld, *Phylogeny and beyond: Scientific, historical, and conceptual significance of the first tree of life.* Proc Natl Acad Sci U S A, 2012. **109**(4): p. 1011-8.
298. Huse, S.M., et al., *A core human microbiome as viewed through 16S rRNA sequence clusters.* PLoS One, 2012. **7**(6): p. e34242.
299. Hahn, A., et al., *Different next generation sequencing platforms produce different microbial profiles and diversity in cystic fibrosis sputum.* J Microbiol Methods, 2016. **130**: p. 95-99.
300. Peterson, J., et al., *The NIH Human Microbiome Project.* Genome Res, 2009. **19**(12): p. 2317-23.

301. Hilty, M., et al., *Disordered microbial communities in asthmatic airways.* PLoS One, 2010. **5**(1): p. e8578.
302. Dickson, R.P., et al., *The Microbiome and the Respiratory Tract.* Annu Rev Physiol, 2016. **78**: p. 481-504.
303. Dickson, R.P., J.R. Erb-Downward, and G.B. Huffnagle, *Towards an ecology of the lung: new conceptual models of pulmonary microbiology and pneumonia pathogenesis.* Lancet Respir Med, 2014. **2**(3): p. 238-46.
304. Segal, L.N., et al., *Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation.* Microbiome, 2013. **1**(1): p. 19.
305. Moran Losada, P., et al., *The cystic fibrosis lower airways microbial metagenome.* ERJ Open Res, 2016. **2**(2).
306. Charlson, E.S., et al., *Topographical continuity of bacterial populations in the healthy human respiratory tract.* Am J Respir Crit Care Med, 2011. **184**(8): p. 957-63.
307. Zakharkina, T., et al., *Analysis of the airway microbiota of healthy individuals and patients with chronic obstructive pulmonary disease by T-RFLP and clone sequencing.* PLoS One, 2013. **8**(7): p. e68302.
308. Huxley, E.J., et al., *Pharyngeal aspiration in normal adults and patients with depressed consciousness.* Am J Med, 1978. **64**(4): p. 564-8.
309. Syed, S.A., et al., *Reemergence of Lower-Airway Microbiota in Lung Transplant Patients with Cystic Fibrosis.* Ann Am Thorac Soc, 2016. **13**(12): p. 2132-2142.
310. Bassis, C.M., et al., *Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals.* MBio, 2015. **6**(2): p. e00037.
311. Dickson, R.P., et al., *Spatial Variation in the Healthy Human Lung Microbiome and the Adapted Island Model of Lung Biogeography.* Ann Am Thorac Soc, 2015. **12**(6): p. 821-30.
312. Klepac-Ceraj, V., et al., *Relationship between cystic fibrosis respiratory tract bacterial communities and age, genotype, antibiotics and Pseudomonas aeruginosa.* Environ Microbiol, 2010. **12**(5): p. 1293-303.
313. van der Gast, C.J., et al., *Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities.* Isme j, 2011. **5**(5): p. 780-91.
314. Zhao, J., et al., *Decade-long bacterial community dynamics in cystic fibrosis airways.* Proc Natl Acad Sci U S A, 2012. **109**(15): p. 5809-14.
315. Zhao, J., S. Murray, and J.J. Lipuma, *Modeling the impact of antibiotic exposure on human microbiota.* Sci Rep, 2014. **4**: p. 4345.
316. Bacci, G., et al., *Pyrosequencing Unveils Cystic Fibrosis Lung Microbiome Differences Associated with a Severe Lung Function Decline.* PLoS ONE, 2016. **11**(6): p. e0156807.
317. Burke, D.G., et al., *The altered gut microbiota in adults with cystic fibrosis.* BMC Microbiol, 2017. **17**(1): p. 58.
318. Mahboubi, M.A., et al., *Culture-Based and Culture-Independent Bacteriologic Analysis of Cystic Fibrosis Respiratory Specimens.* J Clin Microbiol, 2016. **54**(3): p. 613-9.
319. Cox, M.J., et al., *Longitudinal assessment of sputum microbiome by sequencing of the 16S rRNA gene in non-cystic fibrosis bronchiectasis patients.* PLoS One, 2017. **12**(2): p. e0170622.
320. Yang, D.H., et al., *Rapid Identification of Bacterial Species Associated with Bronchiectasis via Metagenomic Approach.* Biomed Environ Sci, 2014. **27**(11): p. 898-901.
321. Rogers, G.B., et al., *Clinical measures of disease in adult non-CF bronchiectasis correlate with airway microbiota composition.* Thorax, 2013. **68**(8): p. 731-7.
322. Rogers, G.B., C.J. van der Gast, and D.J. Serisier, *Predominant pathogen competition and core microbiota divergence in chronic airway infection.* Isme j, 2014.
323. Rogers, G.B., et al., *A novel microbiota stratification system predicts future exacerbations in bronchiectasis.* Ann Am Thorac Soc, 2014. **11**(4): p. 496-503.

324. Tunney, M.M., et al., *Lung microbiota and bacterial abundance in patients with bronchiectasis when clinically stable and during exacerbation.* Am J Respir Crit Care Med, 2013. **187**(10): p. 1118-26.
325. Martinez-Garcia, M.A., et al., *Multidimensional approach to non-cystic fibrosis bronchiectasis: the FACED score.* Eur Respir J, 2014. **43**(5): p. 1357-67.
326. Fothergill, J.L., et al., *Comparison of real time diagnostic chemistries to detect Pseudomonas aeruginosa in respiratory samples from cystic fibrosis patients.* J Cyst Fibros, 2013. **12**(6): p. 675-81.
327. Rudkjobing, V.B., et al., *True microbiota involved in chronic lung infection of cystic fibrosis patients found by culturing and 16S rRNA gene analysis.* J Clin Microbiol, 2011. **49**(12): p. 4352-5.
328. Cameron, S.J., et al., *Metagenomic Sequencing of the Chronic Obstructive Pulmonary Disease Upper Bronchial Tract Microbiome Reveals Functional Changes Associated with Disease Severity.* PLoS One, 2016. **11**(2): p. e0149095.
329. Fodor, A.A., 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): p. e45001.
330. Molyneaux, P.L., et al., *The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis.* Am J Respir Crit Care Med, 2014. **190**(8): p. 906-13.
331. Hafiane, A. and M. Ravaoarinoro, *Characterization of Pseudomonas aeruginosa strains isolated from cystic fibrosis patients by different typing methods.* Pathol Biol (Paris), 2011. **59**(5): p. e109-14.
332. Jolley, K.A. and M.C. Maiden, *BIGSdb: Scalable analysis of bacterial genome variation at the population level.* BMC Bioinformatics, 2010. **11**: p. 595.
333. Van Mansfeld, R., et al., *Pseudomonas aeruginosa genotype prevalence in Dutch cystic fibrosis patients and age dependency of colonization by various P. aeruginosa sequence types.* J Clin Microbiol, 2009. **47**(12): p. 4096-4101.
334. Connor, T., et al., *CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community.* Microbial Genomics, 2016. **2**(9).
335. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-20.
336. Wood, D.E. and S.L. Salzberg, *Kraken: ultrafast metagenomic sequence classification using exact alignments.* Genome Biol, 2014. **15**(3): p. R46.
337. Bankevich, A., et al., *SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.* J Comput Biol, 2012. **19**(5): p. 455-77.
338. Seemann, T., *Prokka: rapid prokaryotic genome annotation.* Bioinformatics, 2014. **30**(14): p. 2068-9.
339. Hecht, M., Y. Bromberg, and B. Rost, *Better prediction of functional effects for sequence variants.* BMC Genomics, 2015. **16 Suppl 8**: p. S1.
340. Choi, Y. and A.P. Chan, *PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels.* Bioinformatics, 2015. **31**(16): p. 2745-7.
341. Callahan, B.J., et al., *DADA2: High-resolution sample inference from Illumina amplicon data.* Nat Methods, 2016. **13**(7): p. 581-3.
342. Millar, F.A., N.J. Simmonds, and M.E. Hodson, *Trends in pathogens colonising the respiratory tract of adult patients with cystic fibrosis, 1985-2005.* J Cyst Fibros, 2009. **8**(6): p. 386-91.
343. Barker, A.F. and E.J. Bardana, Jr., *Bronchiectasis: update of an orphan disease.* Am Rev Respir Dis, 1988. **137**(4): p. 969-78.
344. Winsor, G.L., et al., *Enhanced annotations and features for comparing thousands of Pseudomonas genomes in the Pseudomonas genome database.* Nucleic Acids Res, 2016. **44**(D1): p. D646-53.

345. Ciofu, O., et al., *Occurrence of hypermutable Pseudomonas aeruginosa in cystic fibrosis patients is associated with the oxidative stress caused by chronic lung inflammation.* Antimicrob Agents Chemother, 2005. **49**(6): p. 2276-82.
346. Silbert, S., A.L. Barth, and H.S. Sader, *Heterogeneity of Pseudomonas aeruginosa in Brazilian cystic fibrosis patients.* J Clin Microbiol, 2001. **39**(11): p. 3976-81.
347. Aliberti, S., et al., *Clinical phenotypes in adult patients with bronchiectasis.* Eur Respir J, 2016. **47**(4): p. 1113-22.
348. Marvig, R.L., et al., *Genome analysis of a transmissible lineage of pseudomonas aeruginosa reveals pathoadaptive mutations and distinct evolutionary paths of hypermutators.* PLoS Genet, 2013. **9**(9): p. e1003741.
349. Marvig, R.L., et al., *Within-host microevolution of Pseudomonas aeruginosa in Italian cystic fibrosis patients.* BMC Microbiol, 2015. **15**: p. 218.
350. Markussen, T., et al., *Environmental heterogeneity drives within-host diversification and evolution of Pseudomonas aeruginosa.* MBio, 2014. **5**(5): p. e01592-14.
351. Feliziani, S., et al., *Coexistence and within-host evolution of diversified lineages of hypermutable Pseudomonas aeruginosa in long-term cystic fibrosis infections.* PLoS Genet, 2014. **10**(10): p. e1004651.
352. Chung, J.C., et al., *Genomic variation among contemporary Pseudomonas aeruginosa isolates from chronically infected cystic fibrosis patients.* J Bacteriol, 2012. **194**(18): p. 4857-66.
353. Maatallah, M., et al., *Four genotyping schemes for phylogenetic analysis of Pseudomonas aeruginosa: comparison of their congruence with multi-locus sequence typing.* PLoS One, 2013. **8**(12): p. e82069.
354. Chalmers, J.D., et al., *Cross-infection risk in patients with bronchiectasis: a position statement from the European Bronchiectasis Network (EMBARC), EMBARC/ELF patient advisory group and European Reference Network (ERN-Lung) Bronchiectasis Network.* Eur Respir J, 2018, **51**(1).
355. Mitchelmore, P.J., et al., *Molecular epidemiology of Pseudomonas aeruginosa in an unsegregated bronchiectasis cohort sharing hospital facilities with a cystic fibrosis cohort.* Thorax, 2017.
356. Sherrard, L.J., et al., *Within-host whole genome analysis of an antibiotic resistant Pseudomonas aeruginosa strain sub-type in cystic fibrosis.* PLoS One, 2017. **12**(3): p. e0172179.
357. Oliver, A., F. Baquero, and J. Blazquez, *The mismatch repair system (mutS, mutL and uvrD genes) in Pseudomonas aeruginosa: molecular characterization of naturally occurring mutants.* Mol Microbiol, 2002. **43**(6): p. 1641-50.
358. Oliver, A. and A. Mena, *Bacterial hypermutation in cystic fibrosis, not only for antibiotic resistance.* Clin Microbiol Infect, 2010. **16**(7): p. 798-808.
359. Conibear, T.C., S.L. Collins, and J.S. Webb, *Role of mutation in Pseudomonas aeruginosa biofilm development.* PLoS One, 2009. **4**(7): p. e6289.
360. Macia, M.D., et al., *Efficacy and potential for resistance selection of antipseudomonal treatments in a mouse model of lung infection by hypermutable Pseudomonas aeruginosa.* Antimicrob Agents Chemother, 2006. **50**(3): p. 975-83.
361. Waine, D.J., et al., *Association between hypermutator phenotype, clinical variables, mucoid phenotype, and antimicrobial resistance in Pseudomonas aeruginosa.* J Clin Microbiol, 2008. **46**(10): p. 3491-3.
362. Plasencia, V., et al., *Influence of high mutation rates on the mechanisms and dynamics of in vitro and in vivo resistance development to single or combined antipseudomonal agents.* Antimicrob Agents Chemother, 2007. **51**(7): p. 2574-81.
363. Davies, E.V., et al., *Evolutionary diversification of Pseudomonas aeruginosa in an artificial sputum model.* BMC Microbiol, 2017. **17**(1): p. 3.
364. Ferroni, A., et al., *Effect of mutator P. aeruginosa on antibiotic resistance acquisition and respiratory function in cystic fibrosis.* Pediatr Pulmonol, 2009. **44**(8): p. 820-5.

365. Auerbach, A., et al., *Is infection with hypermutable Pseudomonas aeruginosa clinically significant?* J Cyst Fibros, 2015. **14**(3): p. 347-52.
366. Bar-Meir, M., *Clinical relevance of Pseudomonas aeruginosa hypermutation in cystic fibrosis chronic respiratory infection: Response to Dr. Oliver.* J Cyst Fibros, 2015. **14**(4): p. e3.
367. Bryant, J.M., et al., *Whole-genome sequencing to identify transmission of Mycobacterium abscessus between patients with cystic fibrosis: a retrospective cohort study.* Lancet, 2013. **381**(9877): p. 1551-60.
368. Lutz, L., et al., *Hypermutable Pseudomonas aeruginosa in Cystic fibrosis patients from two Brazilian cities.* J Clin Microbiol, 2013. **51**(3): p. 927-30.
369. Mandsberg, L.F., et al., *Development of antibiotic resistance and up-regulation of the antimutator gene pfpI in mutator Pseudomonas aeruginosa due to inactivation of two DNA oxidative repair genes (mutY, mutM).* FEMS Microbiol Lett, 2011. **324**(1): p. 28-37.
370. Drevinek, P., et al., *Direct culture-independent Strain typing of Burkholderia cepacia complex in sputum samples from patients with cystic fibrosis.* J Clin Microbiol, 2010. **48**(5): p. 1888-91.
371. Doughty, E.L., et al., *Culture-independent detection and characterisation of Mycobacterium tuberculosis and M. africanum in sputum samples using shotgun metagenomics on a benchtop sequencer.* PeerJ, 2014. **2**: p. e585.
372. Goddard, A.F., et al., *Direct sampling of cystic fibrosis lungs indicates that DNA-based analyses of upper-airway specimens can misrepresent lung microbiota.* Proc Natl Acad Sci U S A, 2012. **109**(34): p. 13769-74.
373. Ott, S.J. and S. Schreiber, *Reduced microbial diversity in inflammatory bowel diseases.* Gut, 2006. **55**(8): p. 1207.
374. Ranjard, L., et al., *Heterogeneous Cell Density and Genetic Structure of Bacterial Pools Associated with Various Soil Microenvironments as Determined by Enumeration and DNA Fingerprinting Approach (RISA).* Microb Ecol, 2000. **39**(4): p. 263-272.
375. Kovacs, A., K. Yacoby, and U. Gophna, *A systematic assessment of automated ribosomal intergenic spacer analysis (ARISA) as a tool for estimating bacterial richness.* Res Microbiol, 2010. **161**(3): p. 192-7.
376. Crosby, L.D. and C.S. Criddle, *Understanding bias in microbial community analysis techniques due to rrn operon copy number heterogeneity.* Biotechniques, 2003. **34**(4): p. 790-4, 796, 798 passim.
377. Acosta, N., et al., *The Evolving Cystic Fibrosis Microbiome: A Comparative Cohort Study Spanning Sixteen Years.* Ann Am Thorac Soc, 2017.
378. Wang, Z., et al., *Sputum microbiome temporal variability and dysbiosis in chronic obstructive pulmonary disease exacerbations: an analysis of the COPDMAP study.* Thorax, 2018. 73(4):331-338
379. Taylor, S.L., et al., *Matrix Metalloproteinases Vary with Airway Microbiota Composition and Lung Function in Non-Cystic Fibrosis Bronchiectasis.* Ann Am Thorac Soc, 2015.
380. Huang, Y.J., et al., *Airway microbiota and bronchial hyperresponsiveness in patients with suboptimally controlled asthma.* Journal of Allergy and Clinical Immunology, 2011. **127**(2): p. 372-381. e3.
381. Ege, M.J., et al., *Exposure to environmental microorganisms and childhood asthma.* N Engl J Med, 2011. **364**(8): p. 701-9.
382. Garcia-Nunez, M., et al., *Severity-related changes of bronchial microbiome in chronic obstructive pulmonary disease.* J Clin Microbiol, 2014. **52**(12): p. 4217-23.
383. Byun, M.K., et al., *Differences of lung microbiome in patients with clinically stable and exacerbated bronchiectasis.* PLoS One, 2017. **12**(8): p. e0183553.
384. Wang, Z., et al., *Lung microbiome dynamics in COPD exacerbations.* Eur Respir J, 2016. **47**(4): p. 1082-92.

385. Dickson, R.P., et al., *Changes in the lung microbiome following lung transplantation include the emergence of two distinct Pseudomonas species with distinct clinical associations.* PLoS One, 2014. **9**(5): p. e97214.
386. Coman, I., et al., *Ralstonia mannitolilytica in cystic fibrosis: A new predictor of worse outcomes.* Respir Med Case Rep, 2017. **20**: p. 48-50.