Airway Microbiome in Chronic Lung Disease

By

Sylvia Adel Daniel Rofael

A thesis submitted to University College London (UCL) in fulfilment of the requirement for the degree of Doctor of Philosophy

Centre for Clinical Microbiology
Division of Infection and Immunity
University College London
Royal Free Campus
Rowland Hill Street
London
NW3 2PF
Declaration

I, Sylvia Adel Daniel Rofael confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature
Abstract

Chronic lung disease is one of the main causes of morbidity and mortality worldwide. The recent discovery of the lung microbiome has transformed our understanding of the pathophysiology of respiratory infections and chronic lung disease. In the presented PhD thesis, the hypothesis that the composition of the whole microbial community rather than individual pathogens, is critical in the pathogenesis of chronic lung disease has been investigated. The airway microbiome was studied in a spectrum of chronic lung diseases: non-cystic fibrosis bronchiectasis, chronic obstructive pulmonary disease (COPD), bronchopulmonary dysplasia (BPD) in adult survivors of extremely preterm birth and early pulmonary changes in people living with HIV (PLW-HIV) using culture independent approaches: next generation sequencing and quantitative polymerase chain reactions (qPCR).

In all forms of the chronic lung diseases studied, a characteristic pattern of bacterial dysbiosis was identified. This was characterised by a significant decline in the bacterial community biodiversity and a shift in the bacterial community composition away from phylum Bacteroidetes; particularly genus Prevotella whose relative abundance was correlated with an important lung function parameter. In PLW-HIV, some potential respiratory pathogens and gut bacteria were enriched in the airway microbiome which may place this population at higher risk to respiratory morbidities and pneumonia.

Chronic lung disease is a sector employing extensive antibiotic prescription practices either to treat acute exacerbations, or as prophylaxis therapy. Substantial scientific evidence currently supports the clinical usefulness of macrolide prophylaxis therapy in managing chronic respiratory conditions. In this thesis, I investigated the effect of antibiotics on the homeostasis of the bacterial communities in the airways and how it contributed to the of antimicrobial resistance (AMR) among microbiota.
The airway was found to harbour a rich source of AMR determinants and resistant microbiota. The AMR determinants were more related to the antibiotics used as rescue packs for prompt initiation of self-treatment of exacerbations.

Antibiotic prophylaxis therapy was associated with lower total bacterial load and suppressed recognised pathogenic bacteria in the airways with minimal effect on the homeostasis of the respiratory microbiota. The airway bacterial community was resilient towards the disturbances caused by antibiotics use. No definite directional shift in the microbiome compositions associated with prophylactic antibiotics was identified at the group level.
Impact Statement

Since the start of the Human Microbiome Project; our knowledge about the microbes inhabiting the human body (microbiota) and their amazing functional capacity has been expanding. Accumulating scientific evidence has demonstrated that microbiota are key players in maintaining homeostasis within the human body and many of them contribute to physiological functions of the body. Disturbance in the microbial community structure (microbial dysbiosis) is widely considered a possible underling basis for disease susceptibility in many chronic disease conditions in various human body sites.

Chronic lung disease imposes a great health and economic burden on individuals and health-care systems, in addition it is one of the leading cause of death worldwide. Most of the microbiome research has concentrated on the gut microbiome. Recently, science has refuted the old belief that the lungs and lower respiratory tract are sterile. In the presented thesis, I have studied the airway microbiome in a wide spectrum of chronic lung disease conditions in comparison to healthy control individuals. This work adds to the existing knowledge about the airway microbiome composition and contributes to developing our understanding around the microbial dysbiosis associated with chronic lung disease development and prognosis.

For the first time, the airway microbiome was investigated in adult survivors of extreme pre-term birth; through the unique EPICure Cohort, who were prone to higher risk of developing chronic respiratory disease. In this study, microbial signatures related to the development of chronic reparatory disease has been identified which may have clinical implications in the future. The findings of this study were published in the European Respiratory Journal (DOI: 10.1183/13993003.01225-2018)

In PLW-HIV with well controlled disease, the respiratory microbiome was found to be enriched with potential respiratory pathogens and gut bacteria which may explain the higher incidence of pneumonia and respiratory morbidities reported in
this population despite the antiretroviral therapy. These results were published in The Lancet: E-Clinical Medicine (DOI: 10.1016/j.eclinm.2020.100427)

Patients with chronic lung disease receive frequent prolonged courses of antibiotics to treat chronic infections, exacerbations and manage the disease. A growing body of scientific evidence is currently supporting the clinical usefulness of antibiotic prophylaxis therapy in managing chronic respiratory disease. In this thesis, I was interested to look at the impact of the antibiotic use on the homeostasis of the microbiota and how it may contribute to the prevalence of AMR in the human body. The results presented here support the added benefit of the antibiotic prophylaxis therapy in patients with chronic lung disease who experience frequent exacerbations. Results demonstrated that airway microbiota may act as a reservoir for AMR determinants. Established microbiome based diagnostic tools would promote personalised medicine in the future. Tailoring antimicrobial treatment based on the individual microbiome could improve the treatment outcome, prescribing practices of chemotherapeutic antimicrobial agents and would slow down the development of AMR in this sector.
Acknowledgments

I am deeply grateful to many people who have either helped me personally to accomplish my PhD journey or have contributed to the work presented in this thesis.

I would like to start with acknowledging my supervisory committee for their guidance that set me on the right path throughout my PhD journey. I am indebted to my principal supervisor Prof. Tim McHugh for his continual support, advice, kindness, generosity, and for all the invaluable opportunities he has offered me to grow and develop as a scientist. I am deeply grateful to my secondary supervisor Prof. John Hurst for challenging and encouraging my research, for his continual encouragement that kept me going through the hardest times, for giving me his time to review my thesis and for always responding quickly despite all his clinical duties especially during the COVID-19 pandemic. Many Thanks go to Prof. David Spratt for his insights into my research; and for passing his experience and enthusiasm about microbial communities to me. A special thank you for Prof. Marc Lipman for his clinical insights into my work, his encouragement, efforts and help in the recruitment of patients into the clinical study.

I would like to deeply thank the clinical staff in the Royal Free Hospital; especially, Dr. James Brown, Dr. Shumonta Quaderi and Ms. Elisha Pickett form UCL Respiratory Centre for their advice on the ethics application process and for helping in the clinical specimens’ collection. A special thank you for Dr. David Lowe for letting me use the sputum samples collected from his CVID patients. I am also very appreciative of the patients with chronic lung disease who volunteered to participate in the presented study and dedicate their time and energy into helping others combat this debilitating disease.

I would like to thank Prof. David Spratt’s team in the Eastman Dental Institute: Dr. Ann Tymon, Dr. Catryn Williams and Dr. Mehmet Davrandi, for their technical advices and for providing training at the beginning of my PhD. In addition, I would like to recognise Mr. Tony Brooks and Dr. Paola Niola from UCL genomics for their technical advices and help in running Miseq sequencing.

My lab and office mates in UCL Centre for Clinical Microbiology; Julio, Alp, Dewi, Anna, Kerry, Leanne, Helen, Priya, Linzy, Isobella, Alexandra, Clara, Mercedes, Eloise and Hanouf, who provided a fun working environment with whom it was wonderful to share scientific ideas and experiences; I have learnt something from each one and they made the PhD an enjoyable journey. A particular mention must be reserved for Dr. Virve Enne for her valuable advice on the AMR work and for allowing me to use her MinION station in CCM. I am also grateful for Mr. Alp Aydin.
and Mr. Dewi Owen for their help in the lab work and sharing their invaluable experience on the MinION sequencing. It was a great privilege to work with Dr Lucy Van Drop and Prof. Francois Balloux on the metagenomics project and learn more from them about metagenomics bioinformatics.

My family has been an integral part of my journey and without them I would not be in this position. My Husband Michael who has always been behind me supporting my decisions, thanks for his patience, tolerance and understanding; doing our doctoral degrees together at the same time was not an easy task but I enjoyed the whole journey having him by my beside. My son, Daniel though he has joined us near the end of the journey, having him was such a blessing which brought joy into our life. My daughter, Marley, who has endured a lot during this journey; leaving our extended family and friends in Egypt was not easy for any of us but her smile has always comforted me and gave me courage to push myself forward to accomplish my goal. Thanks to her for being such an amazing loving and caring sister to her baby brother, she has helped me a lot looking after him during writing up my thesis; To them both I dedicate this work, hoping that they may find pride in their mother’s success one day. Thanks to my mum and dad, for having faith in me, for their unconditional love and care, for offering me all kind of support throughout my life that helped me to reach my full potential and to my brother Michael for always being there for me in good and bad times.

I was lucky to have such amazing friends and neighbours in the UK; Afifa, Dina, Mona, Nora, Eriny, Dalia and Sara who have surrounded me with their love and care and have always been there for me whenever I needed them, I am deeply grateful for them all, my life would have been difficult without their help and support.

Finally, I would like to acknowledge my sponsors; the Newton Mosharafa team in the British Council Egypt and the missions sector in the Egyptian Ministry of Higher Education for sponsoring me and funding my doctoral degree in UCL. Also, I would like to thank the UCL Grand Challenges programme for granting me the UCL doctoral students grant to fund the final metagenomics project.
### Table of Contents

Declaration .................................................................................................................. 3

Abstract ....................................................................................................................... 4

Impact Statement ......................................................................................................... 6

Acknowledgments ......................................................................................................... 8

Abbreviations .............................................................................................................. 23

Chapter 1: Introduction ............................................................................................... 26
  1.1. Classical Microbiology versus Microbiome Perspectives ........................................ 26
  1.2. 16S rRNA ............................................................................................................. 27
  1.3. The Lung Microbiome ......................................................................................... 31
  1.4. Chronic Lung Disease ......................................................................................... 35
    1.4.1. Chronic Obstructive Pulmonary Disease (COPD) ........................................ 36
    1.4.2. Bronchiectasis ............................................................................................... 38
    1.4.3. Chronic infections in chronic lung disease .................................................... 39
    1.4.4. Acute Exacerbations .................................................................................... 40
  1.5. Prophylactic antibiotic therapy ........................................................................... 41
  1.6. Classes of Antibiotics ........................................................................................ 42
    1.6.1. Macrolides ..................................................................................................... 42
    1.6.2. Tetracyclines ................................................................................................. 49
    1.6.3. Quinolones .................................................................................................... 50
    1.6.4. ß-Lactams ..................................................................................................... 52
    1.6.5. Sulfonamides Combinations ......................................................................... 53
  1.7. Antimicrobial Resistance (AMR) ....................................................................... 54
  1.8. Sequencing Technologies .................................................................................... 55
  1.9. Bioinformatic Analysis ....................................................................................... 58
  1.10. Biodiversity in Microbiome Data ....................................................................... 59
    1.10.1. Alpha- diversity, Richness and Evenness metrics ........................................ 59
      1.10.1.1. Chao ......................................................................................................... 61
      1.10.1.2. Shannon-Entropy .................................................................................. 61
      1.10.1.3. Fisher-alpha .......................................................................................... 62
      1.10.1.4. PD whole Tree ...................................................................................... 62
    1.10.2. Beta Diversity .............................................................................................. 62

Aim of Thesis: ............................................................................................................ 64

Chapter 2: Materials and Methods .............................................................................. 65
  2.1. Sputum Sample Collection .................................................................................. 65
  2.2. Sputum Homogenisation ..................................................................................... 65
  2.3. Bacteriological cultures ...................................................................................... 65
2.4. Bacterial Identification using MALDI-TOF .......................................................... 66
2.5. Antimicrobial Susceptibility Test (Kirby-Bauer method) .................................. 67
2.6. DNA extraction .................................................................................................... 68
  2.6.1. DNA extraction from sputum samples using the LIAISON® automated
         extraction platform ......................................................................................... 68
  2.6.2. DNA extraction from EPICure induced sputum samples using Qiagen
         spin column ..................................................................................................... 68
  2.6.3. Measuring the DNA concentration and purity of DNA extracts .................. 69
2.7. Mock Community Preparation ........................................................................... 69
2.8. 16S rRNA gene sequencing ................................................................................. 70
  2.8.1 16S rRNA gene amplification ........................................................................ 70
  2.8.2. Library preparation ...................................................................................... 76
  2.8.3. NGS Sequencing .......................................................................................... 77
2.9. Bioinformatic and Statistical Analyses ................................................................. 77
2.10. Multiplex quantitative polymerase chain reaction (qPCR) for
      Respiratory pathogens ....................................................................................... 78
  2.10.1. Typical respiratory pathogen qPCR .............................................................. 78
  2.10.2. Standard Curves of typical Respiratory pathogens ..................................... 80
  2.10.3. P. aeruginosa and total bacterial load Diplex qPCR .................................... 81
  2.10.4 Standard Curves and Standards preparation of P. aeruginosa and
       16S rRNA ........................................................................................................ 83
2.11. Statistics ............................................................................................................. 84

Chapter 3: Microbiome Analysis Method Optimisation and Evaluation

Results ....................................................................................................................... 86
3.1. Introduction ........................................................................................................... 86
3.2. Methods ................................................................................................................ 87
  3.2.1. Preparation of Mock Communities .............................................................. 87
  3.2.2. Paired Sputum and Saliva Samples .............................................................. 88
  3.2.3. Determination of Limit of Detection (LOD) and Limit of Quantification
         (LOQ) ............................................................................................................... 89
3.3. Results and Discussion ....................................................................................... 89
  3.3.1. Choice of the lysing matrix in sputum homogenisation ................................. 89
  3.3.2. Mock communities as positive controls for sequencing ....................... 91
  3.3.3. Negative controls in sequencing ................................................................. 96
  3.3.4. Paired sputum and saliva samples ............................................................... 98
  3.3.5. Performance based specifications of typical respiratory multiplex
         qPCR ............................................................................................................... 109
  3.3.6. Performance based specifications of P. aeruginosa and total bacterial
         load duplex qPCR ......................................................................................... 112
  3.3.7. Comparison between qPCR and 16S rRNA sequencing ......................... 114
3.4. General Discussion ............................................................................................. 115
Chapter 4: Dynamics of Airway Microbiome in Chronic Lung Disease ...

4.1. Introduction ........................................................................................................ 123
4.2. Methods ............................................................................................................... 123
  4.2.1. Ethics Approvals ...................................................................................... 123
  4.2.2. Study Conduct ......................................................................................... 124
4.3. Results ............................................................................................................... 125
  4.3.1. Recruitment ............................................................................................ 125
  4.3.2. Pooling samples in Microbiome Analysis ............................................. 126
  4.3.3. Cohort characteristics .......................................................................... 128
  4.3.4. Stability of the airway microbiome over time ...................................... 131
  4.3.5. Factors impacting the airway microbiome ........................................... 138
    4.3.5.1. Exacerbation .................................................................................... 138
    4.3.5.2. Antibiotic treatment ....................................................................... 140
    4.3.5.3. Antibiotic prophylaxis ..................................................................... 146
  4.3.6. Resilience of airway Microbiome .......................................................... 150
4.4. Discussion ......................................................................................................... 154

Chapter 5: Microbiome and Antimicrobial Treatment ........................................ 159

5.1. Introduction ....................................................................................................... 159
5.2. Methods ............................................................................................................ 159
5.3. Results .............................................................................................................. 160
  5.3.1. Cohort’s characteristics ......................................................................... 160
  5.3.2. Richness and diversity of bacterial communities .................................. 164
  5.3.3. β-diversity .............................................................................................. 166
  5.3.4. Bacterial community composition ....................................................... 169
  5.3.5. Respiratory Pathogens .......................................................................... 173
  5.3.6. Ciprofloxacin treatment course and P. aeruginosa eradication ........ 176
  5.3.7. Covariates in microbiome data .............................................................. 176
  5.3.8. Bacteriological Culture results ............................................................... 182
  5.3.9 Antimicrobial susceptibility of bacterial isolates .................................... 184
5.4. Discussion ......................................................................................................... 188

Chapter 6: Airway Microbiome in Adult Survivors of Extremely Pre-term Birth (EPICure Study) .......................................................... 199

6.1. Introduction ....................................................................................................... 199
6.2. Methods ............................................................................................................ 200
  6.2.1. Ethics ..................................................................................................... 200
  6.2.2. Induced sputum samples collection ...................................................... 201
  6.2.3. Sputum homogenisation and processing ............................................. 201
6.3. Results .............................................................................................................. 202
  6.3.1. General Characteristics of the EPICure Study groups ....................... 202
  6.3.2. Airway Microbiome Diversity indices ................................................. 205
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.3.3. Airway bacterial community composition</td>
<td>209</td>
</tr>
<tr>
<td>6.3.4. Asthma</td>
<td>209</td>
</tr>
<tr>
<td>6.3.5. Loads and Prevalence of Pathogenic Airway Bacteria</td>
<td>214</td>
</tr>
<tr>
<td>6.4. Discussion</td>
<td>216</td>
</tr>
<tr>
<td>Chapter 7: Airway Microbiome of People Living with HIV (PLW-HIV) in comparison to HIV-uninfected Individuals</td>
<td>222</td>
</tr>
<tr>
<td>7.1. Introduction</td>
<td>222</td>
</tr>
<tr>
<td>7.2. Methods</td>
<td>223</td>
</tr>
<tr>
<td>7.2.1. Study design and population</td>
<td>223</td>
</tr>
<tr>
<td>7.3. Results</td>
<td>224</td>
</tr>
<tr>
<td>7.3.1. Clinical and demographic cohort characteristics</td>
<td>224</td>
</tr>
<tr>
<td>7.3.2. Respiratory status of participants</td>
<td>226</td>
</tr>
<tr>
<td>7.3.3. Richness and alpha Diversity of bacterial communities</td>
<td>227</td>
</tr>
<tr>
<td>7.3.4. Beta Diversity of the bacterial community</td>
<td>229</td>
</tr>
<tr>
<td>7.3.5. Composition of Bacterial Communities</td>
<td>233</td>
</tr>
<tr>
<td>7.3.6. Relation of the microbiology findings and the clinical status</td>
<td>236</td>
</tr>
<tr>
<td>7.3.7. Covariates in the microbiome data</td>
<td>236</td>
</tr>
<tr>
<td>7.3.8. Load and Prevalence of Airway Bacteria using qPCR</td>
<td>240</td>
</tr>
<tr>
<td>7.4. Discussion</td>
<td>242</td>
</tr>
<tr>
<td>Chapter 8: The Respiratory Resistome in Chronic Lung Disease</td>
<td>247</td>
</tr>
<tr>
<td>8.1. Introduction</td>
<td>247</td>
</tr>
<tr>
<td>8.2. Methods</td>
<td>248</td>
</tr>
<tr>
<td>8.3. Results</td>
<td>250</td>
</tr>
<tr>
<td>8.3.1. Cohort Characteristics</td>
<td>250</td>
</tr>
<tr>
<td>8.3.2. Human DNA Depletion</td>
<td>252</td>
</tr>
<tr>
<td>8.3.3. Prevalence of AMR genes</td>
<td>253</td>
</tr>
<tr>
<td>8.4. Discussion</td>
<td>256</td>
</tr>
<tr>
<td>Chapter 9: General Discussion and Conclusions</td>
<td>263</td>
</tr>
<tr>
<td>10. References</td>
<td>279</td>
</tr>
<tr>
<td>Appendices</td>
<td>312</td>
</tr>
</tbody>
</table>
List of Tables

Table 1.1: Randomized double blind placebo controlled clinical trials studying the usefulness of macrolides long term prophylaxis in COPD and non-cystic fibrosis bronchiectasis ................................................................. 46

Table 1.2: Comparison between Illumina Miseq Sequencer and Oxford Nanopore MinION Sequencer .................................................................................................................. 57

Table 2.1: NGS Primers for V3-V4 16S rRNA Amplification ......................... 72

Table 2.2: NGS Primers for V5-V7 16S rRNA Amplification .......................... 74

Table 2.3: The primers and probes for the three bacterial targets and the internal amplification control in multiplex qPCR for typical respiratory pathogens .......... 79

Table 2.4: Master-mix preparation of the typical respiratory pathogens .......... 80

Table 2.5: Primers and probe sequences of P. aeruginosa and total bacterial load Diplex qPCR ......................................................................................................................... 82

Table 2.6: Master-mix preparation of the P. aeruginosa and total bacterial load Diplex qPCR ......................................................................................................................... 83

Table 3.1: The OD and the bacterial count (BC) by Miles and Misra of the constituting bacteria in mock communities: D and (E/F) ........................................................................... 94

Table 3.2: The limit of quantification and limit of detection of typical respiratory qPCR ............................................................................................................................ 109

Table 3.3: Sensitivity, Specificity and Accuracy of the 16S rRNA sequencing .. 114

Table 4.1: The demographics, medical and clinical characteristics of the patients’ cohort with non-cystic fibrosis bronchiectasis and COPD .................................... 129

Table 5.1: Cohort’s demographics, medical and clinical characteristics with respect to the use of antibiotic prophylaxis therapy ......................................................... 162

Table 5.2: Lung function parameters and the inhaled respiratory medication used by participants in the antibiotic prophylaxis group and the comparator group .. 163

Table 6.1: Demographic and clinical data of whole EPICure participants and those with sequenced samples ................................................................................................. 204

Table 7.1: Demographics and clinical characteristics of the HIV study cohort .. 225

Table 7.2: Lung Function parameters and Inhalers use .................................... 226

Table 7.3: Comparison between the α-diversity indices in HIV positive and negative groups .......................................................................................................................... 228

Table 8.1: The sequence of primers and probe of the Human RNA polymerase II method. .......................................................................................................................... 248

Table 8.2: Clinical and demographic characteristics of participants in resistome analysis ......................................................................................................................... 251
List of Figures

Figure 1.1. A schematic diagram showing (A) the rRNA genesis; first the rrn operon in the bacterial chromosome is transcribed into a polycistronic transcript, which is then cleaved into the precursor rRNA molecules, these are further processed into the final mature rRNA molecules (Lu et al., 2009) (B) The mature rRNA together with the constituting proteins are assembled into the two ribosomal subunits (50S and 30S) which together form the prokaryotic 70S ribosome. .................................. 28

Figure 1.2. The secondary structure of 16S rRNA molecule (of Escherichia coli) highlighting the nine hypervariable regions of the molecule................................................. 29

Figure 1.3. A schematic diagram for the workflow of 16S rRNA sequencing for microbiome analysis.................................................................................................................. 30

Figure 1.4. The anatomy of the respiratory system.............................................................. 31

Figure 1.5. Pathological changes related to COPD demonstrating emphysematous lungs and inflammatory changes in the airways; airflow constriction and mucus accumulations, in comparison with healthy lungs................................................................. 36

Figure 1.6. Pathological changes related to bronchiectasis demonstrating the characteristic irreversibly damaged and hyper-diluted airways with poor mucus clearance in comparison to the normal healthy airways......................................................... 38

Figure 1.7. A schematic diagram for the workflow of the bioinformatic pipeline using QIIME ................................................................................................................................. 59

Figure 2.1. TapeStation Electrophoresis result for ‘run 4’ sequence library (V3-V4 16S rRNA gene) showing a single peak at 580 bp.......................................................... 76

Figure 2.2. Amplification plot of the internal amplification control (IAC); SPUD A, in tested sputum samples, Ct values< 30 are acceptable, however for any sample a Ct value of the IAC >30 indicates PCR inhibition .................................................. 78

Figure 3.1. Comparison between the performance of FastPrep® lysis matrices A and B in sputum homogenisation with and without heat killing steps on the bacterial loads........................................................................................................ 90

Figure 3.2. Mock community B composition by 16S rRNA sequencing, qPCR and the DNA concentration; g_: genus. ................................................................. 91

Figure 3.3. Mock community C: Comparison between the performance of the amplification of V5-V7 versus V3-V4 variable regions of 16S rRNA using (785F-1175R) and (341F-805R) primer sets respectively in representing bacterial community composition at (A) the phylum level (B) at the genus level. The reproducibility of the microbiome profile of mock community C using V3-V4 variable regions amplification over three 16S rRNA sequencing runs is demonstrated. Number represent the proportions; ........................................ 93

Figure 3.4. Mock community D, E and F microbiome profiles used to assess the efficiency of the automated DNA extraction platform and the effect of the heat killing step on the bacterial community composition representation, both mock communities: D and E were extracted from living bacterial mixture whereas Mock
community F was heat killed before the DNA extraction step. g_ : genus, f_ : family.

Figure 3.5. The number of reads, their percentage, and the microbiome profiles of the negative controls of sequencing. NCpcr: No-template 16S rRNA PCR negative control, NCext: Negative controls during samples processing and DNA extractions.

Figure 3.6. Microbiome profiles of paired sputum and saliva samples from ten patients (A) V3-V4 representation of the microbiome profiles of the ten sputum samples (the bioinformatics done through the microbiome helper pipeline) (B) V5-V7 representation of the microbiome profiles of the ten sputum/saliva pairs through the microbiome helper bioinformatic pipeline (C) V5-V7 representation of the microbiome profiles of the ten sputum/saliva pairs through the alternative bioinformatic pipeline. Sa: Saliva, Sp: Sputum, g_: genus, f_: family.

Figure 3.7. Comparing the richness and α-diversity indices of the microbiome profiles in the ten sputum samples between the variable regions: V3-V4 (green) and V5-V7 (blue) 16S rRNA sequencing.

Figure 3.8. P. aeruginosa loads (log_{10} CFU/mL) in the ten sputum samples measured by qPCR, error bars: SEM; Sp: Sputum sample.

Figure 3.9. PCoA three dimensional (3D) Plot of Bray Curtis β-Diversity showing the clustering of the ten pairs of saliva and sputum samples based on the sequenced 16S rRNA variable regions and the used bioinformatic pipeline. Each dot represents a microbiome profile of a sample.

Figure 3.10. Comparison of the richness and diversity of the bacterial community in the ten sputum and saliva paired samples (A) Richness estimated by the number of observed OTUs (p=0.082, by paired t-test) (B) α-diversity measured by Shannon diversity index (p=0.003, by paired t-test) (C) PCoA plot of unweighted UniFrac β-diversity (PERMANOVA p=0.031, ANOSIM p=0.019).

Figure 3.11. PCoA and cladogram based on weighted UniFrac β-diversity (PERMANOVA p=0.002, ANOSIM p=0.001) showing separation between the sputum samples (Sp, Blue, n=10) and saliva samples (Sa, yellow, n=10); V5-V7 16SrRNA amplicon sequencing; bioinformatics through the alternative pipeline.

Figure 3.12. Comparison between the bacterial loads (log_{10} CFU/mL) of the three respiratory pathogens: H. influenzae (Hi: pink), M. catarrhalis (Mc: green) and S. pneumoniae (Spn: navy blue) in the ten sputum and saliva paired samples measured by qPCR, Sa: Saliva, Sp: Sputum, error bars: SEM.

Figure 3.13. The amplification plots (A,C and E) and the standard curves (B,D and F) of qPCRs for typical respiratory pathogens.

Figure 3.14. Robustness and range of typical respiratory pathogens qPCR using three independent sets of laboratory prepared ten-fold serially diluted standards of (A) S. pneumoniae (B) H. influenzae (C) M. catarrhalis; each set is represented by different line shape, colours represent different bacterial concentrations.
Figure 3.15. The amplification plots (A,C) and the standard curves (B,D) of qPCR for P. aeruginosa and total bacterial load through measuring V3-V4 of 16S qPCR. ................................................................. 113

Figure 3.16. Correlation between the Relative Abundance (RA) results of V3-V4 16S rRNA sequencing and the quantitative loads of qPCR results (CFU/mL) in all tested sputum samples (n=182) (A) Correlation between relative abundance of genus Haemophilus and load of H. influenzae (Hi) (B) Correlation between relative abundance of H. influenzae OTUs and load of H. influenzae (C) Correlation between relative abundance of genus Moraxella and load of M. catarrhalis (Mc) (D) Correlation between relative abundance of genus Pseudomonas and load of P. aeruginosa (Pa). ρ: Spearman’s rho Correlation. .................................................. 115

Figure 4.1. Recruitment curve into the “Changes in Microbiome with Treatment in Chronic Lung Diseases” study. ........................................................................................................ 126

Figure 4.2. PCoA plot of weighted UniFrac β-diversity involving all tested sputum samples (A) No separation based on the sequencing run; red: run 3 (n=84), navy blue: run 4 (n=93) (p=0.768 by PERMANOVA, p=0.600 by ANOSIM). The mock communities from three different sequencing runs coincide in the highlighted green circle. (B) comparison between total CVID samples (pink, n=93) and samples from immunocompetent patients (purple, n=70), No separation based on the immune status (p=0.104 by PERMANOVA, p= 0.083 by ANOSIM). .................................................. 127

Figure 4.3. The stability of airway microbiome in sputum over time in participants with multiple longitudinal samples at baseline respiratory status demonstrated through the UniFrac β-diversity indices and Shannon α-diversity A. PCoA plot of weighted Unifrac β-diversity (p=0.001by PERMANOVA), B. PCoA plot of unweighted UniFrac β-diversity (p=0.001by PERMANOVA), C. Shannon α-diversity index (p=0.314 by paired t-test). Each data point represents sputum sample, samples from the same participant are coloured similarly. Patients with COPD: 2, 93 and 95; while the rest were bronchiectasis patients. ....................... 133

Figure 4.4. The stability of airway bacterial community composition through frequent sputum sampling from three patients: 3 (red), 27 (green) and 33 (blue) demonstrated by α and β diversity indices and microbiome profiles......................... 137

Figure 4.5. Effect of respiratory exacerbation on the airway microbiome .......... 140

Figure 4.6. Effect of antibiotic treatment courses on the airway microbiome .... 142

Figure 4.7. Dynamic changes in the airway microbiome of patient 82 with COPD after a two-week course of high dose of ciprofloxacin.................................................. 145

Figure 4.8. Impact of prolonged prophylactic courses of azithromycin (AZM) on the airway microbiome ................................................................. 147

Figure 4.9. Impact of switching the antibiotic prophylaxis from azithromycin (points B, w4 and w6) to ciprofloxacin (point w12) on the airway microbiome of patient 30 with bronchiectasis ................................................................. 149

Figure 4.10. Resilience of the airway microbiome in patient 32 with bronchiectasis after the exposure to an exacerbation event and antimicrobial treatment; indicated by the navy blue vector and recovery indicated by the green vector in.......... 151
Figure 4.11. Resilience of the airway microbiome in patient 13 with COPD after an exacerbation event and antimicrobial treatment

Figure 5.1. Comparison between α-diversity and richness of airway bacterial community in patients with chronic lung disease who were receiving antibiotic prophylaxis therapy (n= 47) and those who were not (n= 37)

Figure 5.2. Comparison between the Shannon α-diversity index within the group of patients receiving different regimes of antibiotic prophylaxis therapy

Figure 5.3. PCoA plot of weighted UniFrac (A and C) and unweighted UniFrac (B and D), samples are coloured with respect to: (A and B) being on prophylactic antibiotic therapy (navy blue, n=47) or not (comparator group, yellow, n=37) (C and D) the prophylactic antibiotic regime: azithromycin (blue, n=29), clarithromycin (light blue, n=2), doxycycline (orange, n=4), ciprofloxacin (green, n=2), β-lactam (red, n=3) co-trimoxazole (purple, n=7)

Figure 5.4. PCoA plot of weighted UniFrac (A and C) and unweighted UniFrac (B and D), samples are coloured with respect to:  (A and B) chronic respiratory condition (blue: bronchiectasis, n=61), (red: COPD, n=23) (C and D) Immunodeficiency status pink: immunodeficient, n=38) (purple: immunocompetent, n= 46)

Figure 5.5. Comparison between the average airway microbiome profile of patients on prophylactic antibiotic therapy (A) and those not (comparator group) (B) at the phylum level (p>0.05); (C) the RA of phylum Synergistetes was significantly lower in patients receiving antibiotic prophylaxis therapy (p=0.00247)

Figure 5.6. Airway microbiome profiles of patients with chronic lung disease who were receiving antibiotic prophylaxis therapy (n=47) in comparison to patients in the comparator group; not receiving antibiotic prophylaxis (n=37); NC: negative control, g_: genus, f_: family.

Figure 5.7. Suppression of potential pathogenic taxa in the sputum of patients on prophylactic antibiotic therapy (navy blue, n=47) compared to others in the comparator group (yellow, n=37) A. genus Pseudomonas (p=0.027) B. family Enterobacteriaceae (p=0.021), C. genus Moraxella (p=0.087), D. genus Staphylococcus (p=0.159)

Figure 5.8. Prevalence and Co-existence of P. aeruginosa (Pa), H. influenzae (Hi), S. pneumoniae (Spn) and M. catarrhalis (Mc) in patients who were receiving antibiotic prophylaxis therapy (AB: navy blue, n= 46) and those not

Figure 5.9. Bacterial Load of pathogenic airway bacteria in patients who were receiving antibiotic prophylaxis therapy (AB: navy blue, n= 46) and those not (comparator group, yellow, n=36) as determined by the multiplex qPCR

Figure 5.10. The relative abundance of Pseudomonas in sputum of patients who had received high dose of ciprofloxacin treatment course for the eradication of P. aeruginosa colonization (green, n=18) compared to others in the cohort

Figure 5.11. Comparison between the airway microbiome profiles of COPD and bronchiectasis patients at phylum level
Figure 5.12. Differential abundance of the following taxa in the airways of COPD (red, n=23) and bronchiectasis patients (blue, n=61) (A) Streptococcus (p=0.04) more abundant in COPD; whereas, in bronchiectasis, (B) Family Gemellaceae (p=0.029), (C) Klebsiella (p=0.003) (D) Burkholderia (p=0.032) were more abundant and/or frequent. .......................... 178

Figure 5.13. Differentially abundant taxa associated with inhaled respiratory medication (A) SABA was associated with increased RA of Prevotella intermedia (p=0.009) (SABA users (group 1): green, n=32, and comparator (group 0): blue, n=52), (B) LABA was associated with reduced RA of Neisseria (p=0.044), (LABA users (group 1): green, n=35, and comparator (group 0): blue, n=49) .......... 179

Figure 5.14. Differentially abundant taxa associated with the use of mucolytic agent carbocisteine on microbiome......................................................... 180

Figure 5.15. Differentially abundant taxa in immunocompromised patients (A) Prevotella intermedia (p=0.022), (B) Bulleidia (p=0.035) (C) Sphaerochaeta (p=0.046), (D) Filifactor (p=0.020), (E) Finegoldia (p=0.003).............................. 181

Figure 5.16. The bacteria isolated from the sputum of patients with chronic lung disease (COPD and bronchiectasis) and their prevalence among the patients who were receiving antibiotic prophylaxis therapy (n=36) and those not (comparator group, n=32). ................................................................. 183

Figure 5.17. Comparison between the AMR prevalence in the Gram positive and P. aeruginosa sputum isolates from patients receiving antibiotic prophylaxis therapy and those not................................................................. 185

Figure 5.18. Comparison between the AMR prevalence in the Gram negative sputum isolates from patients receiving antimicrobial prophylaxis and those not. ................................................................. 186

Figure 5.19. (A) Association between the AMR in pathogenic sputum isolates and the prescribed rescue antibiotics for the self-treatment of acute exacerbations (p=0.024 by Fisher test) (B) Comparison between the AMR in the commensal streptococci viridans group isolated from the sputum of patients receiving antibiotic prophylaxis therapy (group 1, n=18) and those not (comparator group) (group 0, n=24) (p<0.0001 by Chi Square test), the prevalence: percentage of the resistant sputum bacterial isolates, S: Sensitive (green), R: resistant (red), DO: doxycycline, FQ: fluoroquinolone, AZM: Azithromycin, LEV: Levofloxacin, TE: Tetracycline, AMP: Ampicillin, CTX: Cefotaxime, VA: Vancomycin........................................ 187

Figure 6.1. The total bacterial density in sputum samples of EPICure cohort indicated by 16S rRNA gene copies measured by qPCR (p=0.726 by ANOVA). ................................................................. 205

Figure 6.2. Comparison of the richness and α-diversity indices of bacterial communities in sputum between the pre-term birth survivors (EP), with and without neonatal Bronchopulmonary Dysplasia (BPD), and controls ......................... 207

Figure 6.3. PCoA of weighted UniFrac β diversity index (A and B) and unweighted UniFrac β diversity index (C and D) (A and C) comparing the whole EP group
(orange, n=37) and controls (blue, n=33) (B and D) comparing the three groups: EP+BPD (red, n=28), EP no BPD (yellow, n=9), Controls (n=33).  

Figure 6.4. Comparison of the airway microbiome profiles of individual sputum samples of EPICure cohort at (A) phylum level and (B) genus level.  

Figure 6.5. Comparison of relative abundancies of (A) Phylum Bacteroidetes (p=0.01), (B) Class: Bacteroidia (p=0.00729), (C) Order: Bacteroidales (p=0.00729), (D) Family: Prevotellaceae (p=0.00662), (E) Genus: Prevotella (p=0.00662) across the three study groups.  

Figure 6.6. Comparison between the relative abundance (RA) of Prevotella across the three study groups (A) the RA of genus Prevotella was significantly lower in both extremely preterm (EP) groups regardless the neonatal broncho-pulmonary dysplasia (BPD) status compared to controls (p=0.007 by KW) (B) the RA of genus Prevotella correlated with the FEV1 z-score (Spearman rho = 0.272, p = 0.02) (C) OTU 4458304 contributed most to the observed difference in the RA of genus Prevotella across the study groups (p=0.005, by KW) (D) Phylogenetic Tree of OTU 4458304 was 100% identical to Prevotella melaninogenica strains as obtained by BLAST Analysis of the representative sequence against NCBI 16S ribosomal RNA database.  

Figure 6.7. Comparison of weighted Unifrac β diversity index, Chao_1 α diversity index, the RA of genus Prevotella and the RA of OTU 4458304 identified as Prevotella melaninogenica in sputum samples from participants who were labelled with asthma (pink) and those who were not (blue) within the whole cohort (A, B, C and D respectively) and within the comparator group only (E, F, G and H respectively). No significant differences were detected. 1: p-values by T test, 2: p-values by MW.  

Figure 6.8. Prevalence and loads of pathogenic airway bacteria; H. influenzae (Hi), M. catarrhalis (Mc), S. pneumoniae (Spn) (A) Prevalence of the three pathogenic airway bacteria within the three study groups (p=0.176, Chi squared test) (B) Co-existence of the three bacteria in the sputum samples in each group (p=0.88, Fisher Exact test) (C) Mean bacterial load of each of the three bacteria as determined by the multiplex qPCR (p=0.179, MANOVA). Bacterial load.  

Figure 7.1. Comparison between the rates of air flow obstruction (FEV1/FVC) in the HIV positive (Orange, n=63) and HIV negative comparator group (Blue, n=38).  

Figure 7.2. Total bacterial density was significantly lower in the HIV positive group (orange, n=55) compared to the HIV negative group (blue, n=38); p=0.026 by MW; Total bacterial load was quantified by measuring the total 16S rRNA (V3-V4) gene copies/µL of reaction mixture.  

Figure 7.3. Richness and alpha diversity of airway bacterial communities in PLW-HIV with respect to the blood CD4 count; < 500 cells/ µL burgundy (n=22), ≥500 cells/ µL, green, (n=42), p-values by t-test.  

Figure 7.4. Comparing the microbiome profiles through β-diversity with respect to the HIV status (A) PCoA weighted Unifrac β-diversity index showing no separation
with respect to the HIV status (p=0.283 by PERMANOVA, p=0.273 by ANOSIM) (B) PCoA plot of unweighted Unifrac β-diversity in which samples clustered based on the HIV status (p=0.002 by PERMANOVA, p=0.031 by ANOSIM).

Figure 7.5. PCoA plot of unweighted Unifrac β-diversity (p=0.002 by PERMANOVA, p=0.031 by ANOSIM) demonstrating that the presence of some minor taxa may explain the dispersion of PLW-HIV samples; Group size: 38 HIV-ve group and 64 HIV +ve group

Figure 7.6. Airway microbiome profiles of HIV negative and HIV positive groups (A) Mean RA of bacterial phyla (all p-values >0.05 by White non-parametric t-test) (B) Individual microbiome profiles at the genus level

Figure 7.7. Comparison between the relative abundance of (A) genus Lactobacillus (p=0.029) (B) S. aureus (p=0.0016) (C) genus Pseudomonas (p=0.0093) (D) genus Klebsiella (p=0.024) (E) Family Enterobacteriaceae (p=0.04) (F) genus Bilophila (p<0.0001) in the individual sputum samples

Figure 7.8. Differential abundance of genus Lactobacillus with respect to (A) HIV status (p=0.029 by White t-test), (B) recreational drug use (red: drug users, n=30 while blue: non drug users, n=71) (p=0.023 by White t-test) and (C) smoking status: Ex-smokers (lilac) n=37, Non-smokers (green) n=35, Smokers (navy blue) n=29, (p=0.000834, by KW).

Figure 7.9. Differences in the airway microbiota associated with recreational drug use in the PLW-HIV cohort and the HIV negative comparator group

Figure 7.10. Differences in the airway microbiota associated with the smoking status in the PLW-HIV cohort and the HIV negative comparator group

Figure 7.11. Prevalence and loads of pathogenic airway bacteria; H. influenzae (Hi), M. catarrhalis (Mc), S. pneumoniae (Spn) within the PLW-HIV compared to the comparator group

Figure 8.1. Reduction in the human DNA load in the treated sample manifested by a shift of a mean of +7 in the Ct values of real-time PCR targeting human RNA polymerase II gene

Figure 8.2. Comparison between the relative abundance (expressed in part per million reads) of the total AMR genes detected in sputum samples from bronchiectasis patients on antibiotic prophylaxis (blue, n=8), patients having frequent exacerbations treated with antibiotics but not receiving antibiotic prophylaxis therapy (yellow, n=9) and relatively healthy participants in the comparator group (green, n=8) (p-value=0.331 by KW)

Figure 8.3. The prevalence of AMR genes in bronchiectasis patients on antibiotic prophylaxis (group 1, blue, n=8), bronchiectasis patients having frequent exacerbations treated with antibiotic rescue packs but not receiving antibiotic prophylaxis therapy (group 2, blue, n=9) and healthy participants in the comparator group (group 3, green, n=8) (All p-values >0.05 by KW)

Figure 9.1. Decline of α-diversity in chronic lung diseases (A) Comparison of Chao 1 α-diversity metric in COPD and bronchiectasis patients compared to the HIV negative comparator group, p<0.001 by ANOVA, V3-V4 16S rRNA sequencing (B)
Comparison of Chao 1 α-diversity metric in the extremely preterm born (EP) groups with and without BPD compared to the EPICure controls, p=0.016 by ANOVA, V5-V7 16S rRNA sequencing. ................................................................. 268

Figure 9.2. Airway microbiome is distinguished by the respiratory condition: COPD (red, n=23), bronchiectasis (blue, n=61), healthy controls (green, n=32) in the PCoA of (A) Unweighted UniFrac β-diversity (p=0.001 by PERMANOVA, p=0.001 by ANOSIM) (B) Weighted UniFrac β-diversity (p=0.001 by PERMANOVA, p=0.002 by ANOSIM) ......................................................................................................... 269

Figure 9.3. Bacterial dysbiosis at the phylum level in the airways of patients with COPD, bronchiectasis and adult survivors of extreme preterm birth with history of developing neonatal BPD. ........................................................................................................ 269

Figure 9.4. The relative abundance of genus Prevotella significantly declined in the presence of chronic respiratory disease and correlated with FEV1% predicted 270

Figure 9.5. Comparison between the total bacterial densities in sputum in chronic lung diseases expressed in the total 16S rRNA gene (copies/μL), EP: Extreme Pre-term birth survivors ............................................................................................................................................. 272
**Abbreviations**

<table>
<thead>
<tr>
<th></th>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>α-diversity</td>
<td>Alpha- diversity</td>
</tr>
<tr>
<td></td>
<td>AB</td>
<td>Antibiotic</td>
</tr>
<tr>
<td></td>
<td>ABC</td>
<td>ATP-binding cassette</td>
</tr>
<tr>
<td></td>
<td>AK</td>
<td>Amikacin</td>
</tr>
<tr>
<td></td>
<td>AMC</td>
<td>Amoxicillin/clavulanic acid</td>
</tr>
<tr>
<td></td>
<td>AMP</td>
<td>Ampicillin</td>
</tr>
<tr>
<td></td>
<td>AMR</td>
<td>Antimicrobial Resistance</td>
</tr>
<tr>
<td></td>
<td>ART</td>
<td>Antiretroviral Therapy</td>
</tr>
<tr>
<td></td>
<td>AZM</td>
<td>Azithromycin</td>
</tr>
<tr>
<td>B</td>
<td>β-diversity</td>
<td>Beta diversity</td>
</tr>
<tr>
<td></td>
<td>BAL</td>
<td>Broncho-Alveolar Lavage</td>
</tr>
<tr>
<td></td>
<td>BC</td>
<td>Bacterial Count</td>
</tr>
<tr>
<td></td>
<td>bd</td>
<td>Bi (twice) Daily</td>
</tr>
<tr>
<td></td>
<td>BH-FDR</td>
<td>Benjamini-Hochberg False Discovery Rate</td>
</tr>
<tr>
<td></td>
<td>BMI</td>
<td>Body Mass Index</td>
</tr>
<tr>
<td></td>
<td>bp</td>
<td>Base Pair</td>
</tr>
<tr>
<td></td>
<td>BPD</td>
<td>Bronchopulmonary Dysplasia</td>
</tr>
<tr>
<td></td>
<td>BTX</td>
<td>Bronchiectasis</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>Chloramphenicol</td>
</tr>
<tr>
<td></td>
<td>CAZ</td>
<td>Ceftazidime</td>
</tr>
<tr>
<td></td>
<td>CCM</td>
<td>UCL Centre for Clinical Microbiology</td>
</tr>
<tr>
<td></td>
<td>CI</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td></td>
<td>CIP</td>
<td>Ciprofloxacin</td>
</tr>
<tr>
<td></td>
<td>CF</td>
<td>Cystic Fibrosis</td>
</tr>
<tr>
<td></td>
<td>CFU</td>
<td>Colony Forming Unit</td>
</tr>
<tr>
<td></td>
<td>Ch</td>
<td>Chapter</td>
</tr>
<tr>
<td></td>
<td>CHOC</td>
<td>Chocolate Horse blood agar</td>
</tr>
<tr>
<td></td>
<td>CL 3</td>
<td>Containment level 3 laboratory</td>
</tr>
<tr>
<td></td>
<td>CN</td>
<td>Gentamycin</td>
</tr>
<tr>
<td></td>
<td>Conc</td>
<td>Concentration</td>
</tr>
<tr>
<td></td>
<td>COPD</td>
<td>Chronic Obstructive Pulmonary Disease</td>
</tr>
<tr>
<td></td>
<td>CT</td>
<td>Computed Tomography</td>
</tr>
<tr>
<td></td>
<td>CTX</td>
<td>Cefotaxime</td>
</tr>
<tr>
<td></td>
<td>CV</td>
<td>Coefficient of Variation</td>
</tr>
<tr>
<td></td>
<td>CVID</td>
<td>Common Variable Immuno-Deficiency</td>
</tr>
<tr>
<td>D</td>
<td>3D</td>
<td>Three Dimensional</td>
</tr>
<tr>
<td></td>
<td>DA</td>
<td>Clindamycin</td>
</tr>
<tr>
<td></td>
<td>ddNTP</td>
<td>di-deoxy-Nucleoside-Tri-Phosphates</td>
</tr>
<tr>
<td></td>
<td>DHFR</td>
<td>Di-Hydro-Folate Reductase</td>
</tr>
<tr>
<td>E</td>
<td>E</td>
<td>Erythromycin</td>
</tr>
<tr>
<td>F</td>
<td>EP</td>
<td>Extremely Preterm born</td>
</tr>
<tr>
<td>F</td>
<td>FEB</td>
<td>Cefepime</td>
</tr>
<tr>
<td></td>
<td>FeNO</td>
<td>Fractional exhaled Nitric Oxide</td>
</tr>
<tr>
<td></td>
<td>FEV1</td>
<td>Forced Expiratory Volume in 1 second</td>
</tr>
<tr>
<td></td>
<td>FMT</td>
<td>Faecal Microbiota Transplantation</td>
</tr>
<tr>
<td></td>
<td>FQ</td>
<td>Fluoroquinolones</td>
</tr>
</tbody>
</table>
SD Standard Deviation
secs seconds
SEM Standard Error of the Mean
SOP Standard Operating Procedure
Spn Streptococcus pneumoniae
SR Standardised Residuals
SXT Sulfamethoxazole/ Trimethoprim
T TE Tetracycline
TE buffer Tris-EDTA Buffer
TNF-α Tumour Necrosis Factor alpha
TZP Piperacillin/Tazobactam
U UK United Kingdom
V VA Vancomycin
VL Viral Load
W WHO World Health Organisation
wk week
Y yr year
Chapter 1: Introduction

The human body was found to harbour an average 10-100 trillion microbial cells that live-in symbiosis in/on various organs; the gut is inhabited by the vast majority of these microbes. (Turnbaugh et al., 2007) Over the past decade, microbiome research has captured the attention of the scientific community as microbiota were shown to play an important and active role in maintaining homeostasis within the body and regulating many physiological functions in various organs. (HMP 2019) The "microbiome" is defined as the assembly of microorganisms and their genomes into a diverse microbial community within an ecosystem, whereas, the term “microbiota” refers to the collective sum of microorganisms that occupies one habitat. (Faintuch and Faintuch, 2019, Ursell et al., 2012, Marchesi and Ravel, 2015, Berg et al., 2020)

1.1. Classical Microbiology versus Microbiome Perspectives

Classical microbiology has always concentrated on studying pathogens as discrete units. From the microbiome perspective, pathogens are regarded as members of a wider microbial community living in an ecosystem in which they interact with each other and with the host. They contribute to and are affected by the microenvironment in which they live.

Culture dependant techniques have long been the gold standard for infection and colonization studies. They are ideal for studying the phenotypic characteristics and antimicrobial susceptibility testing of bacteria. Nevertheless, traditional culture protocols were primarily designed for the isolation and identification of pathogenic bacteria. The sensitivity of bacteriological cultures also varies for different bacteria; in addition, many organisms inhabiting the human body cannot be effectively cultured, highlighting the limitations of culture dependant methods in exploring the human microbiome. (Rogers et al., 2009) Recent advances in molecular techniques and sequencing technologies have overcome these limitations, removed the boundaries set by the classical microbiology approaches and provided efficient and sensitive tools for studying the micro-organisms
comprehensively in their native environments without the need for prior isolation and cultivation in the laboratory. (Aho et al., 2015)

The advent of the Next Generation Sequencing (NGS) technologies during the first decade of the 21st century has reduced the cost of sequencing considerably and has provided efficient platforms to sequence multiple specimens with sufficient discrimination (explained later under 1.8. Sequencing Technologies). This facilitated and promoted Human microbiome research. (Fuks et al., 2018)

The Human Microbiome Project (HMP) was a program launched in 2007 by the American National Institute of Health with a total budget of $215 million over ten years with the aim of exploring microbial diversity in the human body. The first phase of the project (2008-2012) was designed to explore and understand the core healthy human microbiome in five anatomical sites: gut, oral cavity, nasal passages, skin, uro-genitals in 300 healthy individuals. It has revealed extensive microbial diversity that was previously unappreciated with culture-dependent methods. The second phase (2013-2016) aimed to understand how microbiome contributes to the pathogenesis of a range of diseases. In the second phase, the relation between: the gut microbiome and inflammatory bowel disease, the gut/nasal microbiome and type 1 diabetes mellitus, the association between the maternal vaginal microbiome and preterm birth were investigated. (HMP 2012, 2019) Since then, the accumulating scientific evidence from the HMP and wider research community suggests that the whole microbial community composition rather than specific pathogenic organisms plays an important role in the pathogenesis of conditions with chronic nature.

1.2. 16S rRNA

Microbial census and diversity studies have relied on sequencing the phylogenetic marker gene 16S rRNA, either directly from metagenomic DNA extracted from specimens (metagenomic sequencing) or after amplification (amplicon sequencing). 16S rRNA is an integral molecule in the small ribosomal subunit of prokaryotes (bacteria and archaea). (Figures 1.1 and 1.2) The gene which
Chapter 1: Introduction

encodes for the 16S rRNA molecule is highly preserved within the bacterial species. The features of the 16S rRNA gene qualified it to be an efficient bacterial genetic marker. The size of 16S rRNA gene range between 1400-1900 base-pairs (bp); the exact size differs between different bacterial species. It is composed of nine hypervariable regions flanked by conserved regions whose sequences are similar among most prokaryotes. However, the genetic code of the hypervariable regions is unique for each bacterial species. This feature allowed the design of universal primers that can amplify a specific region of the gene from all the bacterial members within the microbial population. Knowing the sequences of the hypervariable regions through sequencing the pooled amplicons then aligning them against reference databases for known bacterial and archaeal 16S rRNA sequences, can provide a consensus of the bacterial community. This represents the principal of microbiome studies (Figure 1.3). (Mizrahi-Man et al., 2013, 2018, Fuks et al., 2018)

![Figure 1.1](image)

**Figure 1.1.** A schematic diagram showing (A) the rRNA genesis; first the rRNA operon in the bacterial chromosome is transcribed into a polycistronic transcript, which is then cleaved into the precursor rRNA molecules, these are further processed into the final mature rRNA molecules (Lu et al., 2009) (B) The mature rRNA together with the constituting proteins are assembled into the two ribosomal subunits (50S and 30S) which together form the prokaryotic 70S ribosome.
Figure 1.2. The secondary structure of 16S rRNA molecule (of *Escherichia coli*) highlighting the nine hypervariable regions of the molecule. (Yarza et al., 2014)
Figure 1.3. A schematic diagram for the workflow of 16S rRNA sequencing for microbiome analysis * demonstrating the amplification of V5-V7 variable regions of the 16S rRNA gene from the metagenomic DNA using a pair of Universal Primers (785 Forward primer and 1175 Reverse primers) tagged by specific barcodes (represented by red segment) and attached to adaptors (green) which allow binding to complementary adaptors on the sequencing flow cells later during the sequencing run. Different samples are assigned to unique dual barcodes (represented here by different colours); Therefore, this allows multiplexing amplicons from different sources (samples) into one pooled library which is then sequenced on Illumina® MiSeq platform, samples are demultiplexed based on the barcodes and the primary analyses of the raw sequencing reads are done on the machine then all the data is exported into Illumina® Basespace cloud-based storage from where data can be retrieved for secondary analyses using specialised bioinformatics pipelines.

*: Diagram was adapted from (Del Chierico et al., 2015)
1.3. The Lung Microbiome

Until recently the healthy lower respiratory tract and lungs were thought to be normally sterile and that microorganisms were detected in lower airway specimens only during respiratory infections. This concept was well established. This misconception may be attributed to the fact that the respiratory system possesses many defence lines and intensive innate protective mechanisms to clear out microbes and foreign bodies. The defence mechanisms are represented in sneezing, coughing, mucin secreting goblet cells, muco-ciliary clearance, secreting antimicrobial compounds such as lysozyme, surfactants, fatty acids and lactoferrin; in addition to, the innate and adaptive immunity represented in: the alveolar macrophages, polymorphonuclear leukocytes, Immunoglobulins (IgG and IgA), opsonins, defensins which are present in the airways and alveolar lining fluids. (Wilson, 2005)

Nevertheless, looking at the anatomy of the respiratory system, it is formed of continuous connecting tracts without physical barriers starting with the oral cavity and nasal passages which are heavily colonized with microorganisms and ending with the supposedly sterile respiratory tissue: bronchioles and alveoli (Figure 1.4). Consequently, it is not unreasonable to imagine that microorganisms are continuously introduced into the lower respiratory compartments through mucosal dispersion or micro aspiration or carried by the inhaled air. (Dickson et al., 2016a)

Figure 1.4. The anatomy of the respiratory system (Tu, 2013)
The misconception of lung sterility was first challenged by (Hilty et al., 2010) in a study that amplified bacterial 16S rRNA sequences in bronchoalveolar lavage (BAL) samples and protected specimen brushings from chronic obstructive pulmonary disease (COPD) and asthmatic patients, both adults and children, as well as eight healthy volunteers. Their results provided evidence that the bronchial tree was not sterile and it contained a mean of 2,000 bacterial genomes per square centimetre. (Hilty et al., 2010) Since then; characterising respiratory microbiota in health and disease conditions has become a hot topic. (Dickson et al., 2016a) In many studies involving healthy participants, bacterial 16S rRNA sequences were detected in lower respiratory specimens like BAL or protected specimen brushing. (Morris et al., 2013, Bassis et al., 2015, Charlson et al., 2011, Dickson et al., 2017) Furthermore, bacterial sequences have been detected in human lung tissues not only in chronic lung disease (Pragman et al., 2018, Sze et al., 2012, Erb-Downward et al., 2011) but also in excised healthy lung tissues such as in case of malignancies (Yu et al., 2016).

Initially, scepticism has been expressed regarding the lung microbiome, some proposed that the bacterial sequences detected may belong to dead bacteria or it could be just contamination from the upper respiratory tract. The lung microbiome was not studied in the context of the Human Microbiome Project. This may be attributed to the view of lung sterility at that time or due to the challenges of sampling the lower respiratory system. The most favourable lower respiratory specimens would be those which by-pass the upper respiratory tract to avoid contamination from upper compartments. The best would be tissues biopsies, followed by bronchoscopic samples such as BAL and bronchial specimen brushing; but these specimens would involve invasive techniques. Spontaneously expectorated or induced sputum is an easy accessible non-invasive specimen that has been long used in respiratory microbiology to sample the lower respiratory tract. Nevertheless, concerns over microbial contamination via carryover of pharyngeal microbiota has compromised the respiratory microbiome research for sometime; even though multiple lines of evidence suggest that this effect is minimal. (Dickson et al., 2016a, Man et al., 2017)
In a study addressing the issue of contamination, rigorous procedures were applied during bronchoscopy to minimize sources of contamination during the passage of bronchoscope through the upper respiratory tract, revealed that the bacterial density and richness were significantly higher in paired BAL and protected specimen brushing samples from seven participants compared to all examined negative controls taken at every step to control for contamination. (Dickson et al., 2017) Using a special two-bronchoscope sampling technique, Charleston et al. excluded the possibility of upper airway contamination and demonstrated that even though the bacterial communities detected in BAL were indistinguishable from the upper airway, it was two to four logs lower in biomass. (Charlson et al., 2011) In another study although the oral and nasal microbiome profiles were distinct, the composition of bacterial communities in BAL samples was similar despite the route of bronchoscopy whether it was introduced through the mouth or nose. (Dickson et al., 2014a) Therefore the detected bacterial sequences cannot be just contamination from the upper respiratory tract.

To assess the viability of the detected bacteria, a study which applied extended bacterial cultures using a wide range of nutrient media under a variety of incubation conditions on 62 BAL samples collected from healthy volunteers, found that 61% of the detected 16S rRNA sequences by molecular techniques belong to viable and readily culturable bacteria. (Venkataraman et al., 2015) In another study, sequencing the reverse transcribed rRNA revealed that the lungs of mice harbour metabolically active microbiota that was similar in composition at the phylum and family levels in different murine models apart from the germ-free mice which were completely sterile. (Yun et al., 2014)

The human lungs were found to harbour an unexpected diversity of microorganisms. (Aho et al., 2015) The core respiratory microbiome was mainly constituted of the following bacterial phyla: Firmicutes, Proteobacteria, Bacteroidetes, Actinobacteria and Cyanobacteria. The most encountered genera in most studies characterizing the lung microbiome were Streptococcus, Haemophilus, Prevotella, Veillonella, Actinomyces, Fusobacterium, Porphyromonas, Pseudomonas, Enterobacteriaceae, Rothia, Neisseria and
Chapter 1: Introduction

*Gemella* which was commonly detected in bronchial specimens from healthy subjects as well as non-diseased lung tissue. (Erb-Downward et al., 2011, Morris et al., 2013, Charlson et al., 2011, Bassis et al., 2015, Pragman et al., 2018, Yu et al., 2016)

Bacterial sequences were even detected in the airways of new-born children soon after birth. On the first day of life, the microbiome was dominated by *Firmicutes* and *Proteobacteria*, other bacterial phyla like *Actinobacteria*, *Bacteroidetes*, *Fusobacterium*, and *Cyanobacteria* were also detected but contributed less to the structure of microbiome at birth. (Lal et al., 2016) The airway microbiome has been shown to be affected by mode of delivery and it develops during the first year of life influenced by number of factors that shape the microbiome such as living with pets, attending day-care, having siblings and acute respiratory illnesses and treatment with antibiotics. (Teo et al., 2015, Man et al., 2017, Bosch et al., 2016, Unger and Bogaert, 2017, Dominguez-Bello et al., 2010)

Studies have demonstrated that the source community of the lung microbiota is most probably the oral cavity. (Charlson et al., 2011, Venkataraman et al., 2015) Within 28 healthy participants, the microbiome profiles of BAL samples were found to be closer in composition to that of the corresponding oral washes from the same participants rather than the nasal microbiome. (Bassis et al., 2015) Similarly, another study reported a great similarity in the microbiome composition between the mouth and lung bacterial communities in paired BAL and protected specimen brushing samples. (Dickson et al., 2017) In a study involving lung tissues from patients with varying COPD severities found more oral bacteria and less nasal bacteria amongst the lung microbiota. This study highlighted that intra-subject similarity was greater than inter-subject similarity at the level of lung tissue. (Pragman et al., 2018)

The healthy airway microbiome was regarded as a dynamic assemblage of microbes that fits a neutral model which describes the dispersal assemblage of communities. In this context, microbes move in bidirectional ways, microbes continuously disperse and drift from the source communities which, in this case,
would be the oral cavity; while simultaneously microbes are constantly eliminated by muco-ciliary clearance, innate and adaptive immune response. An equilibrium between all processes maintains the microbial homeostasis in healthy airway. In contrast, in disease, the microbiome is distinguished by being under active selection that agrees more with niche community assembly model. In this case, alterations in the pathophysiology and physiochemical properties of the microenvironent are associated with a disturbance in the homeostasis of microbiota within the microbial population which defines the term ‘dysbiosis’. (Venkataraman et al., 2015, Dickson et al., 2017, Dickson et al., 2016a) In advanced COPD, huge significant differences in the bacterial community composition were found between the various micro-anatomic sites within the same lung. (Erb-Downward et al., 2011)

1.4. Chronic Lung Disease

The global burden of disease is shifting from communicable to non-communicable diseases with chronic conditions. According to WHO, these are now the chief causes of morbidity and mortality worldwide. (WHO, 2017, Alwan, 2010) The impact of obstructive lung diseases on health and quality of life is considerable, particularly in advanced cases. It has been suggested that symptom burden in this patient cohort is comparable to that of patients with cancer, although patients with obstructive lung diseases have a greater life expectancy. (Joshi et al., 2012) Obstructive lung diseases interfere with the patients’ productivity, some patients experience difficulties in normal physical exertion, social and family activities especially those with severe disease. (Uzaslan et al., 2012)

The chronic lung conditions considered in this study were COPD, and non-cystic fibrosis bronchiectasis. In addition, two cohorts of adult survivors of preterm birth with and without history of developing neonatal bronchopulmonary dysplasia (BPD) (Chapter 6) and another cohort of healthy People Living with HIV (PLW-HIV) (Chapter 7) who are more susceptible to chronic respiratory morbidities compared to the general population were studied in this context to understand the early onset
of bacterial dysbiosis that may be associated with development of chronic lung disease.

1.4.1. Chronic Obstructive Pulmonary Disease (COPD)

COPD disease is characterised by progressive airflow limitation and chronic inflammation of the airways and lungs in response to noxious particles and gases, causing respiratory impairment and distress manifested as dyspnoea, chronic cough and/or excessive mucus production, airway ciliary malfunction and repeated lower respiratory infections. The inflammatory process in COPD leads to irreversible abnormalities in the airways and/or alveoli structure (Figure 1.5). (GOLD, 2020)

Figure 1.5. Pathological changes related to COPD demonstrating emphysematous lungs and inflammatory changes in the airways; airflow constriction and mucus accumulations, in comparison with healthy lungs.

COPD was put on top of the respiratory diseases list which cause significant morbidity and mortality by the forum of International respiratory societies. It is estimated that 65 million people have moderate to severe COPD world-wide and around 3 million people die each year. This makes COPD currently the third leading
cause of death worldwide. More than 90% of COPD related deaths occurs in low and middle income countries. (WHO, 2018a, WHO, 2017) Since the numbers are increasing; therefore, it is projected to be the leading cause of death within 15 years. (Quaderi and Hurst, 2018) In the UK, COPD is considered the second most prevalent diagnosed lung disease, the second cause of death from lung disease after lung cancer and the fifth biggest killer. (Snell et al., 2016)

The major risk factors of COPD include: active and passive tobacco smoking and other forms of smoking like cigar, shisha, pipe and marijuana, exposure to indoor air pollution resulting from burning biomass fuel used in traditional stoves for cooking or for heating purposes in developing countries, and occupational exposure to organic and inorganic dusts, fumes and chemical agents. Severe hereditary deficiency of alpha-1-anti-trypsin is a genetic cause for COPD. (Hunt and Tuder, 2012) Other risk factors include old age, female sex, low socioeconomic standards, chronic bronchitis, abnormal lung development, asthma and severe respiratory infections during childhood. (GOLD, 2020)

Clinically, COPD diagnosis is made based on both the patient’s history of a potentially causative exposure and spirometry results showing persistent airflow limitation defined by post-bronchodilator ratio between forced expiratory volume in one second and forced vital capacity (FEV1/FVC ratio) less than 0.7. Guidelines recommend that COPD is managed through inhaled medication involving long-acting β2-agonists (LABA), inhaled corticosteroids (ICS), combined LABA and ICS or long-acting anti-muscarinic antagonists (LAMAs) which are effective in reducing the rate of decline of the lung function and may improve the quality of life. (Celli et al., 2015)
1.4.2. Bronchiectasis

Bronchiectasis is another chronic lung disease characterised by irreversibly damaged and dilated airways with poor mucus clearance and persistent bacterial colonization (Figure 1.6). (Pasteur et al., 2010) Bronchiectasis is diagnosed using computed tomography (CT) of the chest and it can range between focal disease or diffuse disease involving multiple lobes in the lung.

![Figure 1.6](image_url)  
Figure 1.6. Pathological changes related to bronchiectasis demonstrating the characteristic irreversibly damaged and hyper-diluted airways with poor mucus clearance in comparison to the normal healthy airways.

Bronchiectasis has various aetiologies; most commonly due to the exposure of the developing lung to serious and repeated insults such as severe infections with tuberculosis, pertussis and severe pneumonia. Non-tuberculous mycobacterial infections can cause nodular bronchiectasis. Immunological disorders such as primary immunodeficiency, inflammatory bowel syndrome, Crohn’s disease, and rheumatoid arthritis have been associated with bronchiectasis. Cystic fibrosis (CF) and primary ciliary dyskinesia are the underlying causes of special phenotypes of bronchiectasis. Gastric reflux and foreign body aspiration that damage airway can
also cause bronchiectasis. COPD and asthma may eventually lead to bronchiectasis. Nevertheless, a substantial proportion of bronchiectasis cases are idiopathic. (Flume et al., 2018, Kelly et al., 2018)

Bronchiectasis has not received sufficient attention in the past; therefore, the prevalence of the disease most probably has been underestimated. However, with the increasing awareness of the disease and improved access to radiological facilities, bronchiectasis diagnosis has been increasing world-wide. According to the British Lung Foundation, it is estimated that 212,000 people were living with non-CF bronchiectasis in the United Kingdom (UK) in 2012 this figure suggests that the prevalence of the disease increased by 20% in the period between 2008 and 2012 in the UK. Nevertheless, these figures may be attributed to improved diagnosis rather than a true rise in the incidence of the disease. Females and the elderly above 70 years old are at greater risk for developing bronchiectasis. (Snell et al., 2019) The mortality rate of bronchiectasis is markedly high, around 2.2 times higher than the general population in the UK. (Quint et al., 2016) In a Belgian prospective cohort study, the mortality rate was 20% over 5 years among non-CF bronchiectasis. (Goeminne et al., 2014)

1.4.3. Chronic infections in chronic lung disease

An overlap between COPD and bronchiectasis sometimes exists. Bronchiectasis can be seen in advanced COPD. On the other hand, as bronchiectasis progresses, patients are presented with increasing degree of chronic airflow obstruction and airway inflammation which meet diagnostic criteria for COPD. (Flume et al., 2018) Airway colonization with the following pathogenic respiratory bacteria: Streptococcus pneumoniae, Haemophilus influenzae, Moraxella catarrhalis and Pseudomonas aeruginosa is common at the advanced stage in both conditions. In many cases, these pathogens establish chronic respiratory infections which are difficult to eradicate, this leads to more damage, chronic cough, excessive sputum production often associated with breathlessness. (GOLD, 2020, Hill et al., 2018) Chronic infection with P. aeruginosa is the most concerning since it has been related to poor prognosis, rapid decline in lung function, more frequent
Chapter 1: Introduction

exacerbations and impaired quality of life. (Wilson et al., 2016, Yum et al., 2014, Rogers et al., 2014a)

1.4.4. Acute Exacerbations

Patients with both chronic lung conditions, periodically suffer periods of acute deterioration of their respiratory condition, away from their stable state and beyond normal day-to-day variations, usually extending over several days which necessitate a change in regular medication. These events are known as “Acute Exacerbation”. (Pasteur et al., 2010, Burge and Wedzicha, 2003) Exacerbation events vary in their severity. In addition to causing significant morbidity, exacerbations may accelerate disease progression as a result of the vicious cycle of; airway obstruction, bacterial colonization, more inflammation, and progressive tissue destruction. (Cole, 1986)

Frequent exacerbations have been associated with progressive lung damage, faster decline in the lung function and poorer quality of life. (Donaldson et al., 2002, Martinez-García et al., 2007) In addition, exacerbations account for a substantial share in the economic burden on the health care system; for exacerbations are a major cause for primary care visits and severe events may require hospitalization. (Andersson et al., 2002, Brill et al., 2015b, Polatli et al., 2012)

In COPD, 50–70% of exacerbations are due to respiratory infections with bacteria, atypical organisms and respiratory viruses, 10 % are due to environmental pollution, and up to 30 % are of unknown aetiology. (Sapey and Stockley, 2006, Segreti et al., 2014) In bronchiectasis, it is mostly infective exacerbations due to chronic airway bacterial colonization; nevertheless, in one cohort study respiratory viruses were implicated in 49% of exacerbations. (Gao et al., 2015)

Currently in clinical practice, moderate to severe acute exacerbation events are empirically managed with antibiotics and/or oral corticosteroids. The frequently prescribed first and second lines antibiotics in the UK include β-lactams (amoxicillin, co-amoxiclav, ceftriaxone), tetracyclines (doxycycline), macrolides (clarithromycin) and fluoroquinolones (ciprofloxacin). (Hill et al., 2018, Hodges,
In bronchiectasis exacerbations; a 14-day antibiotic course is recommended for the treatment of acute exacerbations. For COPD patients, a five to seven day course of antibiotics is recommended as it may shorten the recovery period, hospitalisation duration and reduce the risk of early relapse and treatment failure. (GOLD, 2020, Burns et al., 2011) It is recommended by the British Thoracic Society that individual patients have a self-management plan for acute exacerbation in place in which antibiotics are prescribed and kept with patients as rescue pack at home to allow prompt start of antibiotic treatment to ameliorate the complications of the exacerbation events. The decision on antibiotic choice should be based on the individual’s clinical condition, previous sputum bacteriology culture results and the antimicrobial susceptibility profile (antibiogram) of sputum pathogenic isolates. (Hill et al., 2018)

Nevertheless, the benefit of antibiotic in the treatment of acute exacerbation has remained a matter of controversy for many years and the evidence supporting this practice is limited. (Puhan et al., 2007, GOLD, 2020) This may be attributed to the difficulty in identifying the infectious aetiology underlying the acute exacerbation. (Fantin, 2010) Theoretically, treatment with antibiotics would be beneficial only in the management of bacterial infective exacerbations. Concerns over the excessive use of antibiotics have been raised as it will not be without adverse effects in terms of direct side effects for patients, adds to the healthcare cost and contributes to antimicrobial resistance. (Wenzel et al., 2012) The British Thoracic Society has suggested a guide to identify exacerbation of bacterial origin and the patient who would benefit from antibiotic prescription, when three of the following symptoms co-exist: first, increased sputum volume or viscosity; second, increased sputum purulence; third, worsening local symptoms (increased cough, wheeze, breathlessness) or systemic upset. (Pasteur et al., 2010)

1.5. Prophylactic antibiotic therapy

Due to the dangerous implications of acute exacerbations, reducing the rate of exacerbations has been an ultimate aim in order to slow the progression of chronic lung diseases. (Pasteur et al., 2010) Prolonged prophylactic antibiotic therapy has
been a successful strategy in reducing exacerbation rate and managing chronic lung disease. Numerous clinical trials were conducted testing various antibiotics and different regimes: continuous, intermittent and pulsed. Currently, scientific evidence based on systematic reviews and meta-analyses of randomised placebo controlled clinical trials revealed the clinical usefulness of the prophylactic use of antibiotics in reducing the frequency of exacerbations and improving the quality of life in COPD (Herath et al., 2018, Wentao et al., 2015, Donath et al., 2013) and bronchiectasis (Kelly et al., 2018, Hnin et al., 2015, Fan et al., 2015). Most supporting evidence is derived mainly from studies involving macrolides particularly azithromycin. Nevertheless, concerns around the potential adverse effects due to the prolonged antibiotic treatment and the emergence of antimicrobial resistance have been expressed. Consequently, the option of prophylactic antibiotic therapy is usually reserved for patients with advanced chronic lung disease who experience frequent and/or severe exacerbations. In addition, the balance between the benefits and the risks should be considered on individual basis. (Hill et al., 2018, Wenzel et al., 2012, Miravitlles and Anzueto, 2014)

1.6. Classes of Antibiotics

In the following section, the properties, mechanism of action and resistance of the various classes of antibiotics which are prescribed for the treatment of acute exacerbations in chronic lung disease are discussed. Furthermore, the potential antibiotic candidates for prophylactic antibiotic therapy which showed promising results in randomised clinical trials are highlighted.

1.6.1. Macrolides

Macrolides are a class of antibiotics characterised by having a macrocyclic lactone ring to which one or more deoxy sugar moieties are attached. Macrolides are classified according to the size of lactone ring in their structure into 14, 15 and 16 membered rings. (Zuckerman et al., 2011) Erythromycin which is a 14-membered lactone ring was the prototype first discovered in 1952 and it was derived from
Chapter 1: Introduction

Saccharopolyspora erythraea which belongs to actinobacteria (Lohsen and Stephens, 2019) Second generation macrolides such as Clarithromycin (14-membered lactone) and azithromycin (15-membered azalide; i.e. with an extra nitrogen atom in the lactone ring) are semi-synthesized compounds which have more favourable pharmacokinetics and pharmacodynamics properties compared to erythromycin. More advanced macrolides are generally more tolerable and have broader spectrum of activity. Both azithromycin and clarithromycin have exceptionally high penetration power into the tissues of respiratory system. (Tamaoki, 2004) In addition, macrolides concentrate intracellularly at significantly higher concentration inside macrophages and neutrophils compared to in the tissue fluid and serum levels. (Hodges, 2011) Azithromycin has an exceptionally long serum half-life (up to 60 hr) which together with the characteristic post-antibiotic effect of macrolides, helps the antibiotic to persist in the target tissue at a concentration above the therapeutic level (minimum inhibitory concentration for most bacteria) for weeks post-antibiotic treatment. Consequently, these pharmacodynamic and pharmacokinetic features allow a short treatment course with reduced dosing frequency; which promotes patient compliance especially in children. (Girard et al., 2005) Therefore, macrolides have been favoured by patients and have been extensively used in the treatment of respiratory infections world-wide. (Zuckerman et al., 2011, Wilson et al., 2016)

Macrolide antibacterial activity is by the inhibition of protein biosynthesis by reversibly binding to 23S rRNA in the prokaryotic large (50S) ribosomal sub-unit; therefore, interfering with the translocation step in the protein biosynthesis process. (Lambert, 2011) Macrolides are mostly bacteriostatic agents at therapeutic doses with broad spectrum antibacterial activity; earlier macrolides are mainly active against Gram positive bacteria; whereas; advanced macrolides exhibit extended antibacterial activity against Gram negative bacteria such as H. influenzae, M. catarrhalis and Neisseria. Enterobacteriaceae are susceptible to azithromycin only. Furthermore, macrolides exhibit antibacterial activity against atypical bacteria such as Chlamydia, Legionella, Mycoplasma and non-tuberculous mycobacteria (NTM). (Hodges, 2011) Although macrolides have poor antibacterial activity against
Chapter 1: Introduction

*P. aeruginosa*, azithromycin was found to inhibit its biofilm formation through interfering with essential quorum sensing pathways. (Köhler et al., 2010, Burr et al., 2016)

Macrolide resistance determinants are highly transmissible. (Serisier, 2013) The mechanisms of resistance include active efflux which is mediated by macrolide specific efflux pumps (Mef) and other multidrug resistance (MDR) pumps. However, this mechanism confers low level resistance. Modification of the drug binding site in the bacterial ribosome is another mechanism which confers high level resistance. This mechanism includes either mutation or expression of ribosomal protection proteins or the methyl-transferase enzymes (Erm: Erythromycin ribosome methylation) which methylate 23S rRNA at the drug binding site; therefore, hindering the binding of macrolides to their target sites. The Erm enzymes are sometimes constitutively expressed or rapidly induced by the presence of macrolides and they cause cross resistance to all other macrolides in addition to lincosamide and Streptogramin B leading to Macrolide-Lincosamide-Streptogramin B (MLS\textsubscript{B}) resistant phenotype. (Zuckerman et al., 2011) The third mechanism is antibiotic inactivation mediated by various classes of enzymes such as macrolide phosphotransferases (Mph) which phosphorylate the sugar moiety in the compound, and esterases (Ere) that hydrolyse the lactone ring. (Alcock et al., 2020, Roberts)

In addition to antibacterial activity, macrolides exhibit other favourable pharmacological activities such as anti-inflammatory, immunomodulatory and pro-kinetic effects. Macrolides were found to suppress the secretion of pro-inflammatory chemokines and cytokines from bronchial epithelial cells, inhibit neutrophil activation and leucocyte chemotaxis, reduce oxidative stress and maintain the integrity of the airway lining epithelial cells. (Tan et al., 2016, Tamaoki, 2004, Halldorsson et al., 2010, Yamaya et al., 2012) Furthermore, macrolides were found to inhibit rhinovirus infections which are implicated in COPD exacerbations and reduce cytokines secretions from airway epithelial cells in response to the viral infection. (Suzuki et al., 2002, Molyneaux et al., 2013) In addition, azithromycin was found to reduce the mucus secretion. (Lohsen and Stephens, 2019)
Therefore, theoretically speaking macrolides would be effective as a prophylactic therapy in chronic lung disease as it interferes with multiple stages in the chronic lung disease vicious cycle. (Yamaya et al., 2012) Clinically, several randomised clinical trials have proven the efficacy of macrolides prolonged prophylactic treatment in managing COPD (Wentao et al., 2015, Donath et al., 2013) and bronchiectasis (Kelly et al., 2018). Table 1.1 summarises the main findings of the large randomised placebo control clinical trials involving erythromycin or azithromycin prophylaxis therapy in adult patients with COPD and non-cystic fibrosis bronchiectasis.

Generally, macrolides are well tolerated the most commonly reported side effect was gastrointestinal intolerance including diarrhoea and abdominal pain especially with erythromycin. (Zuckerman et al., 2011, Kelly et al., 2018) Nevertheless, clinical trials of prolonged macrolide prophylactic therapy have reported some rare severe adverse effects such as reversible hearing impairment, vertigo, tinnitus, cardiac arrhythmias, prolonged QTc intervals and cardiac arrest in some patients. (Wenzel et al., 2012) Furthermore, reports from some clinical trials suggest increased prevalence of macrolide resistance has been associated with prolonged prophylactic therapy.
Table 1.1: Randomized double blind placebo controlled clinical trials studying the usefulness of macrolides long term prophylaxis in COPD and non-cystic fibrosis bronchiectasis

<table>
<thead>
<tr>
<th>Study</th>
<th>Place</th>
<th>N (Controls/Treated)</th>
<th>Condition</th>
<th>Antibiotic</th>
<th>Dose</th>
<th>Duration</th>
<th>Main Findings</th>
<th>Main Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Seemungal et al., 2008)</td>
<td>UK</td>
<td>109(56/53)</td>
<td>COPD</td>
<td>Erythromycin</td>
<td>250 mg twice a day</td>
<td>12 mo.</td>
<td>f(AE)</td>
<td>125 81 0.003 • AE Rate ratio for treated compared to controls was 0.648</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>f(AE)/ patient-yr</td>
<td>1.83 1.48 0.01 • Shorter duration exacerbations</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Patients% with sig. quality of life improvement</td>
<td>36% 43% 0.03 • No effect on airway or systemic inflammatory markers</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>f (Macrolides R)</td>
<td>41% 81% &lt;0.01 • No association between response and each of; sex, chronic bronchitis, oxygen use, or concomitant therapy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Adverse effects: increase in hearing decrements by 5% in treated group</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Frequency of macrolides resistance was doubled</td>
</tr>
<tr>
<td>(Albert et al., 2011)</td>
<td>multicentre (12) in USA</td>
<td>1117 (559/558)</td>
<td>COPD</td>
<td>Azithromycin</td>
<td>250 mg once daily</td>
<td>12 mo.</td>
<td>median time to 1st AE</td>
<td>174 days 266 days &lt;0.001 • AZM decreased the frequency of exacerbations</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>f(AE)/ patient-yr</td>
<td>1.83 1.48 0.01 • Improved quality of life</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Patients% with sig. quality of life improvement</td>
<td>36% 43% 0.03 • AZM was most effective in AE requiring both antibiotic and steroid, older age, milder stages</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>f (Macrolides R)</td>
<td>41% 81% &lt;0.01 • No association between response and each of; sex, chronic bronchitis, oxygen use, or concomitant therapy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Adverse effects: increase in hearing decrements by 5% in treated group</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Frequency of macrolides resistance was doubled</td>
</tr>
</tbody>
</table>
### COLUMBUS Study (Uzun et al., 2014)

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Size (COPD/total)</th>
<th>Disease</th>
<th>Treatment</th>
<th>Duration</th>
<th>Outcome</th>
<th>Adjusted Rate (AE)</th>
<th>Rate Ratio</th>
<th>CI</th>
<th><strong>P</strong></th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Netherlands</td>
<td>92 (36/41) COPD</td>
<td>Azithromycin 500 mg thrice/wk</td>
<td>12 mo</td>
<td>Δ Adjusted Rate (AE)</td>
<td>129</td>
<td>84</td>
<td>-0.58</td>
<td>0.001</td>
<td></td>
<td>Lower AE Rate in the treatment arm&lt;br&gt;• Sig higher adverse event (diarrhoea) in treatment arm</td>
</tr>
</tbody>
</table>

### EMBRACE Study (Wong et al., 2012)

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Size (COPD/total)</th>
<th>Disease</th>
<th>Treatment</th>
<th>Duration</th>
<th>Outcome</th>
<th>Rate (AE)/ patient</th>
<th>Prebronchodilator FEV1 (L)</th>
<th>SGRQ total score</th>
<th><strong>P</strong></th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Multicentre (3) New Zealand</td>
<td>141 (70/71) Bronchiectasis</td>
<td>Azithromycin 500 mg thrice/wk</td>
<td>6 mo</td>
<td>Rate Ratio</td>
<td>1.57</td>
<td>0.59</td>
<td>&lt;10^-4</td>
<td>0.25</td>
<td>0.108</td>
<td>Lower AE Rate&lt;br&gt;• Increased time to the first AE event&lt;br&gt;• Benefits persisted for 6 months after completion of treatment.&lt;br&gt;• No Sig. change in the health-related quality of life&lt;br&gt;• No Sig. change in the pre-bronchodilator FEV1</td>
</tr>
</tbody>
</table>

### (Altenburg et al., 2013)

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Size (COPD/total)</th>
<th>Disease</th>
<th>Treatment</th>
<th>Duration</th>
<th>Outcome</th>
<th>Hazard Ratio</th>
<th>No of individual having at least 1 exc/yr</th>
<th>Δ Predicted FEV1over 3 mo</th>
<th>Gastrointestinal adverse effects</th>
<th>Macrolides Resistance Rate in participants</th>
<th><strong>P</strong></th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Netherlands</td>
<td>83 (40/43) Bronchiectasis</td>
<td>Azithromycin 250 mg daily</td>
<td>12 mo</td>
<td>Median number of AE</td>
<td>2</td>
<td>0</td>
<td>0.29</td>
<td>(95% CI: 0.16 - 0.51)</td>
<td>80%</td>
<td>46%</td>
<td>&lt;0.001</td>
<td>Lower AE rate&lt;br&gt;• Improved quality of life&lt;br&gt;• No change in the microbiological profile&lt;br&gt;• Higher resistance levels&lt;br&gt;• Higher gastrointestinal adverse effects</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Hazard Ratio</th>
<th>No of individual having at least 1 exc/yr</th>
<th>Δ Predicted FEV1over 3 mo</th>
<th>Gastrointestinal adverse effects</th>
<th>Macrolides Resistance Rate in participants</th>
<th><strong>P</strong></th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Chapter 1: Introduction**

<table>
<thead>
<tr>
<th>Study</th>
<th>Australia</th>
<th>Bronchiectasis</th>
<th>Erythromycin</th>
<th>12 mo.</th>
<th>Mean f(AE)/ patient-yr AE Rate Ratio</th>
<th>Median Δ in 24 hr sputum production</th>
<th>Mean Δ FEV1 post-bronchodilator</th>
<th>Difference in median Δ in Macrolides R</th>
<th>AE Rate Ratio</th>
<th>Median Δ in 24 hr sputum production</th>
<th>Mean Δ FEV1 post-bronchodilator</th>
<th>Difference in median Δ in Macrolides R</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLESS Study (Serisier et al., 2013)</td>
<td>117 (58/59)</td>
<td></td>
<td></td>
<td>400 mg twice daily</td>
<td>1.97</td>
<td>1.29</td>
<td>0.57</td>
<td>-4.3g</td>
<td>+2.2 % Predicted</td>
<td>+25.5%</td>
<td>&lt;0.05</td>
<td>• Reduced AE Frequency</td>
<td>• Reduced 24 hr sputum production</td>
</tr>
</tbody>
</table>

N: total the number of participants  
wk: week  
R: resistance  
Δ: change  
SGRQ: St Georges Respiratory Questionnaire  
FEV1: Forced Expiratory Volume in the first second  
AE: Acute Exacerbation  
mot.: months  
AZM: azithromycin  
yr: year  
sig.: significant
1.6.2. Tetracyclines

Tetracyclines are another class of antibiotics that inhibit protein synthesis by binding to the small (30S) ribosomal subunit, sterically interfering with the docking of aminoacyl-tRNA into the ribosome-mRNA complex; therefore, halting the elongation process of protein synthesis. (Lambert, 2011)

Tetracyclines are characterised by having four cyclic rings in their structure. The clinically useful tetracyclines such as doxycycline are semisynthetic derivatives of natural antimicrobial products made by *Streptomyces* species. (Hodges, 2011, Grossman, 2016) Generally, tetracyclines are the second most commonly prescribed class of antibiotics after β-lactams for ambulatory outpatients in the UK. (Hodges, 2011, WHO, 2018b) Doxycycline is usually a second-line antibiotic used as an alternative option for β-lactam in the treatment of exacerbations of chronic lung disease and an alternative for macrolides as a prophylactic therapy when the first-line option is intolerable or ineffective. (Hill et al., 2018)

Tetracyclines are broad spectrum bacteriostatic antibiotics that are active against both Gram positive and Gram negative bacteria; however, *P. aeruginosa* are intrinsically less sensitive to this class of antibiotics. (Hodges, 2011)

In addition to the antibacterial activity, doxycycline has been reported to exhibit antioxidant and anti-inflammatory activity through the inhibition of matrix metalloproteinases which are a group of host proteolytic enzymes secreted by a various host cells such as macrophages and neutrophils, and contributes to the inflammation process. These enzymes are activated under oxidative stress leading to uncontrolled extracellular matrix degradation and extensive tissue damage. (Griffin et al., 2010) Therefore, this favourable pharmacological activity may be beneficial in the management of chronic lung disease in which inflammation is a cornerstone in pathogenesis. A recent randomised clinical trial involving 60 COPD patients randomised between a group receiving the standard inhaled therapy of COPD and another with an additional daily doxycycline therapy for three months,
Chapter 1: Introduction

has showed that doxycycline treatment was associated with reductions in the plasma levels of inflammatory cytokines, oxidative stress biomarkers and metalloproteinases. The observed effects were associated with better lung function and quality of life in the doxycycline treatment group. (Singh et al., 2019)

The decline in the utility of tetracyclines in clinical practice in the past may be attributed to the global dissemination of tetracycline resistance which is highly transmissible between bacteria. Resistance mechanisms to tetracyclines include ribosomal protection and modification of the antibiotic target site through the expression of ribosomal protection proteins, reducing the cellular antibiotic uptake through altering the cell wall permeability, enzymatic inactivation of tetracyclines molecules which is mediated by *tet (X)*, and active efflux through expression of tetracycline specific efflux pumps (Tet efflux pumps) and/or MDR efflux pumps extruding the antibiotic molecules outside the bacterial cells and reducing its intracellular concentrations. (Smith, 2011, Grossman, 2016)

The major adverse effects of tetracyclines include gastrointestinal intolerance manifested as nausea, vomiting and diarrhoea. Since tetracyclines chelate divalent cations such as calcium and magnesium; these compounds deposit in the bones and teeth; therefore, this class is not recommended for paediatric use or during pregnancy. (Hodges, 2011)

1.6.3. Quinolones

Quinolones are entirely synthetic compounds. The second-generation quinolones which include ciprofloxacin are also known as fluoroquinolones. These are bactericidal agents which have exceptional antibacterial activity against Gram negative bacteria including the *Enterobacteriaceae* and *P. aeruginosa*. The third generation (e.g. moxifloxacin) has in addition extended antibacterial activity against Gram positive bacteria and anaerobes. (Hodges, 2011)

Quinolones act by disrupting DNA replication by selectively inhibiting the bacterial topoisomerases II (DNA gyrase) and IV which eventually lead to DNA damage and cell death. Both enzymes are composed of two subunits. The function of the former
(gyrase) is to unwind the supercoils in the bacterial chromosome and it is the primary target for fluoroquinolones in the Gram negative bacteria; while the latter (Topoisomerase IV) is essential for DNA strand separations during replication. Both enzymes are targeted in Gram positive bacteria. (Lambert, 2011)

Bacteria have developed various strategies to resist quinolones. The first mechanism is through protecting or modifying the quinolone targets in the bacterial cell either by mutations of which the most clinically significant would be in the subunit A of gyrase and ParC in topoisomerase IV resulting in less affinity to antibiotic binding, Gram positive tend to mutate to resistance at higher rate compared to Gram negatives. Enterobacteriaceae on the other hand, express penta-peptide repeat proteins known as Qnr that protect the DNA-topoisomerases complex from the inhibitory effects of quinolones. This mechanism is a plasmid mediated. (Strahilevitz et al., 2009) The second mechanism is active efflux usually mediated through MDR efflux pumps. The third mechanism is altering the outer membrane permeability of the Gram negative porins to reduce the antibiotic uptake into bacterial cells. (Smith, 2011) Recently a new mechanism has been identified in P. aeruginosa which confers resistance to ciprofloxacin through the expression of inactivating phosphotransferase enzyme encoded by crpP. (Chávez-Jacobo et al., 2018, Alcock et al., 2020)

The adverse effects related to fluoroquinolones are gastro-intestinal disturbances (nausea, vomiting, diarrhoea and dyspepsia), CNS side-effects (headache, dizziness, insomnia, blurred vision, tremors and seizures), tendinopathy (tendonitis and sometimes tendon rapture), arthropathy and phototoxicity. Like macrolides, QTc prolongation which can lead to ventricular arrhythmias has been associated with fluoroquinolones treatment. Fluoroquinolones are also not recommended for paediatric use or during pregnancies. (Roberts, 2008)

Ciprofloxacin is widely used for the treatment of acute exacerbations in patients with chronic lung disease colonised by P. aeruginosa. (Hill et al., 2018) Prolonged course of ciprofloxacin are usually prescribed after first isolation of P. aeruginosa.
Chapter 1: Introduction

attempting to eradicate colonization before chronic infection is established. (Wilson et al., 2016)

A randomised placebo-controlled trial that involved 738 participants assessed intermitted pulsed moxifloxacin regime; 400 mg once daily for five days repeated every eight weeks with a total of six courses over 48 weeks, as antimicrobial prophylaxis therapy in COPD. The study found that the odds of exacerbation dropped by 25% in the treated participants compared to the placebo control arm. The benefit of the antibiotic prophylaxis was particularly obvious in cases with purulent sputum production at baseline. However, the moxifloxacin treatment group reported a significantly higher incidence of adverse events which were mainly gastrointestinal intolerance. (Sethi et al., 2010)

1.6.4. ß-Lactams

ß-Lactams are a class of antibiotics that encompass penicillins and cephalosporins. This class is the most widely prescribed antibiotics in the clinical practice in the UK for ambulatory outpatients and the first line antibiotic for the treatment of respiratory infections and acute exacerbations. ß-Lactams are generally safe and well tolerated and can be used for paediatrics and during pregnancy. Nevertheless, hypersensitivity skin allergies sometimes happen in up to 10% of the patients, more severe fatal anaphylactic shock happens in 0.05% of cases. (Hodges, 2011)

ß-Lactams are bactericidal agents that bind to penicillin biding proteins (PBP) (carboxy-transpeptidase) in the bacterial cell wall structure; therefore, inhibiting a late stage (transpeptidation) in the peptidoglycan biosynthesis which leads to impaired cell wall structure. (Lambert, 2011)

ß-Lactam resistance is widespread. The mechanisms of resistance involve PBP alteration through mutations which result in reduced affinity to ß-lactams binding, active efflux and altering the membrane permeability therefore reducing the cellular drug up-take. Nevertheless, the most clinically significant mechanism involves the expression of the inactivating enzymes ß-lactamases which hydrolyse the ß-lactam
ring in the compounds. The expression of β-lactamases can be either constitutive or inducible in bacteria. A strategy to overcome β-lactamase resistance is through using combinations of penicillin and β-lactamase inhibitor; for example, amoxicillin/clavulanic acid (co-amoxiclav); in this example clavulanic acid acts as a suicide β-lactam that has a high affinity for β-lactamase; therefore, it binds to the inactivating enzyme sparing the amoxicillin (the antibacterial agent) to carry out its function. (Smith, 2011)

1.6.5. Sulfonamides Combinations

Sulfonamides are synthetic drugs that interfere with the folic acid biosynthesis in bacteria by competing with para-amino-benzoic acid for dihydropteroate synthetase enzyme; since folic acid is a necessary vitamin for the manufacture of essential nucleic acids; antifolates eventually lead to the deprivation of the bacterial cell of folate; an essential vitamin for bacterial growth. (Lambert, 2011)

Sulfonamides are frequently used in synergistic combination with diamino-pyrimidines e.g. co-trimoxazole which is a combination of trimethoprim and sulfamethoxazole. Trimethoprim is a diamino-pyrimidine that binds to and selectively inhibit the bacterial dihydrofolate reductase (DHFR) enzyme. Both agents exhibit bacteriostatic broad spectrum antibacterial activity against Gram positive and negative bacteria; in addition to antiprotozoal activity against malaria and toxoplasma. The combination of the two antibiotics results in a bactericidal synergy as it targets successive steps in the folate biosynthesis achieving double interference within the same metabolic pathway. This combination is commonly used for the treatment of urinary tract infections and respiratory infections to a lesser extent. (Lambert, 2011, Hodges, 2011) Prolonged co-trimoxazole therapy is prescribed for patients with impaired immune system such as in cases of primary immunodeficiency, Human Immunodeficiency Virus (HIV) infection, and patients receiving immunosuppressant medications as prophylaxis for Pneumocystis jiroveci (previously known as Pneumocystis carnii) which is an opportunistic fungus causing a special kind of pneumonia; Pneumocystis pneumonia, mainly among immunocompromised patients. (Stern et al., 2014)
Chapter 1: Introduction

The most common mechanism for resistance for both classes of antibiotics involves: overproduction of the target host enzyme, or target modification through mutations in the host enzyme encoding genes, or target replacement by acquiring sul genes which encodes for resistant form of dihydropteroate synthetase the target enzyme for sulphonamides and dfr gene which encodes a resistant form of the DHFR enzyme the target enzyme for trimethoprim. (Sköld, 2000, Smith, 2011) The most commonly reported adverse events are gastrointestinal disturbances; nausea, vomiting and anorexia, and skin hypersensitivity reactions such as rashes, itching and hives. (Epps et al., 2019)

1.7. Antimicrobial Resistance (AMR)

The prevalence of AMR has been increasing all over the globe. (WHO, 2020) WHO has recognised the problem as one of the greatest threats to public health worldwide and called for a global action. (WHO, 2014) It is generally accepted within the scientific community that antibiotic misuse and/or overuse is a major contributing factor to the emergence and dissemination of AMR in the population. (Goossens et al., 2005) For example, research revealed a relation between the increasing prevalence of macrolide resistance in USA in consistency with the increasing global annual sales of azithromycin. (Serisier, 2013)

Despite the growing scientific evidence of the favourable clinical outcomes of the antibiotic prophylaxis therapy in chronic lung disease, concerns have been expressed over the development of AMR in this sector and its spread in the general population. (Hnin et al., 2015, Leung et al., 2011) Furthermore, fears over the development of cross-resistance between various classes of antibiotics that may result from sharing multidrug resistance determinants between bacteria in response to the selective pressure exerted by the extensive antibiotic use. (Fan et al., 2015) For example, a previous study has shown that penicillin resistance was more strongly associated with macrolide consumption rather than the β-lactam’s consumption in S. pneumoniae isolates. (Dias and Caniça, 2008, García-Rey et al., 2002)
Most clinical trials did not report on AMR induction in response to prolonged antibiotic therapy. In the studies that have addressed the issue the data on the risk of AMR induction in response to the prolonged prophylactic antibiotic therapy has been controversial (discussed in Chapter 5). Furthermore, clinical trials have concentrated on reporting the resistance among respiratory bacterial pathogens as per the standard conventional microbiology protocols. Nevertheless, it is easy to imagine that the indigenous uncultivable bacteria could play a role in the acquisition and spread of antibiotic resistance to pathogenic bacteria within the microbial community. Therefore, the microbiota may act a reservoir for antibiotic resistance within the body. (Sommer et al., 2009, Penders et al., 2013) In many studies, the human gut microbiota has been shown to act as a reservoir of antibiotic resistance genes. The resistant bacteria can persist for years, after the administration of short-term antibiotic courses and can transfer resistance genes to other bacteria, even if those bacteria are just passing through the intestine. (Hu et al., 2013, Jernberg et al., 2010, Salyers et al., 2004, Zaura et al., 2015) While the gut microbiota has been extensively researched in this regard, the respiratory resistome has not been widely explored yet.

The ‘resistome’ is the collective sum of all the antimicrobial resistance determinants in a microbial ecosystem. (Wright, 2007) Studying the resistome can reveal the diversity of genetic AMR determinants present within the bacterial community with the possibility of identifying the source organisms, the origin genetic material which can predict the transmissibility between bacteria and a potential for predicting the antimicrobial susceptibility to antibiotic treatment.

1.8. Sequencing Technologies

Sanger Sequencer was the first sequencer invented in 1977, its principal was a based-on chain termination sequencing by synthesis of single molecules (up to 1000 bp). Next Generation Sequencing (NGS) refers to the second-generation sequencers which caused a revolution in the sequencing technologies and transformed the biological sciences. The first NGS platform was Roche 454 sequencer and it emerged in 2005; after that many more NGS platforms like
Chapter 1: Introduction

Illumina, Ion Torrent, Pacific Biosciences (PacBio) followed. All these technologies facilitated and promoted microbiome research in various environmental and biological fields. Currently, Illumina is the most popular sequencing platform used world-wide and the one which dominates the NGS market especially for amplicon sequencing due to its accuracy, efficiency, simplicity, and flexibility in library preparation. The principal of Illumina’s technology is based on sequencing by synthesis using fluorescently labelled reversible terminators to detect bases as they are incorporated into the growing strand. (Schirmer et al., 2016) Nevertheless, all NGS technologies are limited by the short read lengths (< 500 bp) that can be generated by most platforms; therefore, imposing a challenge on the decision of which hyper-variable regions in the 16S rRNA gene to be targeted than can achieve the required universality and at the same time maximises resolution power; i.e. the ability to distinguish between bacteria for taxonomic classification and phylogenetic analyses. Third generation sequencing technologies represented by the MinION presented by Oxford Nanopore Technologies Ltd (ONP) first emerged in 2014. MinION is a small portable device, in the size of a hand palm, that is connected to any computer or laptop through USB 3.0 connection. Its principal is based on protein nanopores which are immersed in conductive electrolyte solution and a voltage is applied across the pores. As a single stranded nucleic acid molecule (DNA or RNA) transverses the nanopore, one nucleotide at a time, it modulates the current in a characteristic way related to the conformation of each nucleotide. Consequently, the modulation of the ionic current across each nanopore can then be translated into the genetic code sequence. This technology provides a promising sequencing tool as it allows real time sequencing and generates long sequence reads (typically around 3000 bp but it was reported that it can reach up to 200,000 bp). However, these technologies are currently prone to the highest sequencing error rates (10-20%), compared to NGS platforms. Nevertheless, the technology has been rapidly and continuously evolving. (Malla et al., 2018, Almeida, 2017, Lavezzo et al., 2016, Fuks et al., 2018) A comparison between the two sequencing platform: Illumina Miseq and ONT MinION which have been used in the presented thesis is summarised in Table 1.2.
### Table 1.2: Comparison between Illumina Miseq Sequencer and Oxford Nanopore MinION Sequencer

<table>
<thead>
<tr>
<th>Points of Comparison</th>
<th>Illumina Miseq</th>
<th>ONT MinION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequencing Generation Year</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
</tr>
<tr>
<td>Principal</td>
<td>paired end sequencing by synthesis using Fluorescent labelled terminators (ddNTPs)</td>
<td>Single molecule Real-Time Sequencing based on modulations of the ionic current across nanopores as template nucleic acid strand transverses</td>
</tr>
<tr>
<td>Sequencing Error Rate</td>
<td>&lt;0.1%</td>
<td>10-20%</td>
</tr>
<tr>
<td>Number of samples to be</td>
<td>Up to 96</td>
<td>Up to 12</td>
</tr>
<tr>
<td>multiplexed in one run</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kits</td>
<td>Standard V.2 Kit</td>
<td>Rapid Barcoding Kit</td>
</tr>
<tr>
<td>Read length</td>
<td>2x 250 bp</td>
<td>Up to 200 kbp</td>
</tr>
<tr>
<td>Throughput</td>
<td>12.5 GB</td>
<td>Up to 10 GB</td>
</tr>
<tr>
<td>Approximate total number of</td>
<td>25 Million</td>
<td>6 Million</td>
</tr>
<tr>
<td>reads</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Run-Time</td>
<td>48 hrs</td>
<td>Real-time Up to 48 hrs depending on the required depth of sequencing</td>
</tr>
</tbody>
</table>
1.9. Bioinformatic Analysis

Bioinformatics provides tools to analyse the big data sets produced by the high-throughput sequencing and mine for useful and valid information about the microbial communities, enabling the interpretation of microbiome data. New computational tools are being continually developing to provide more efficient ways of analysing the data. QIIME “quantitative insights into microbial ecology”, the programme used in analysing data in the presented thesis, is an open source pipeline developed primarily for processing 16S rRNA sequencing data. (Navas-Molina et al., 2013, Kuczynski et al., 2011) The workflow of the applied bioinformatic pipeline is explained in (Figure 1.7). (Langille, 2017) After the initial quality control steps to exclude the low-quality reads, the sequence reads were sorted and clustered based on 97% sequence similarity into discrete clusters known as operational taxonomic units (OTUs). This threshold is the default setting in the OTU picking step in the QIIME pipeline (Kuczynski et al., 2011) and may be derived from a previous empirical study that showed most species had 97% 16S rRNA sequence similarity. (Konstantinidis and Tiedje, 2005) Despite the current inherent limitations of the OTU picking step and the controversy regarding the definition of OTU, currently it is generally accepted that each OTU can be regarded as an equivalent to species. A single sequence is then randomly picked to represent the whole OTU. This representative sequence is used to assign taxonomy to each OTU by aligning against the reference database which is the Greengenes in the default settings of QIIME. Then, an OTU table is constructed and a phylogenetic tree is created to infer the phylogenetic relationships relating the sequences. QIIME offers three methods for OTU picking: first, closed reference OTU picking, in this method, the sequences that align with the reference database are only retained and those which did not are discarded. The second is de-novo OTU picking in which the sequences are clustered against one another without using an external reference database, this method is useful in discovering novel sequences; however, it can be computationally exhaustive and needs high computational power. Finally, Open reference OTU picking which is the used method in the current work, is a combination of the previous two methods it starts
as closed reference and the sequences that failed to hit the reference database carry on with the de-novo clustering. (Kuczynski et al., 2011)

Figure 1.7. A schematic diagram for the workflow of the bioinformatic pipeline using QIIME (Langille, 2017). Metadata in the mapping file include all the clinical and demographical variables that may categorize the participants within the study. Downstream analyses include: estimating and comparing $\alpha$ and $\beta$ diversity, and differential abundances of OTUs/taxa across the samples.

1.10. Biodiversity in Microbiome Data

As microbiome data is large and multidimensional in nature showing the consensus of hundreds of taxa, therefore, efficient summarising measures would be useful in this context for the sake of comparisons. (Cheng et al., 2019) Microbiologists adopted traditional biodiversity measures used in the field of environmental ecology to describe and explore microbial ecology. In the following section, the biodiversity terms and measures used in describing and comparing microbiome in the results section are discussed.

1.10.1. Alpha- diversity, Richness and Evenness metrics

Alpha- diversity ($\alpha$-diversity) is used to describe within sample diversity. Richness is the simplest measure of diversity as it simply counts the numbers of entities (species, taxa or OTUs) within an assemblage representing the microbiome of a
sample. Richness measures assume all entities are equivalent discrete units disregarding the relative abundance information. This means that richness metrics regard all OTUs within the OTU table are equally related and are given the same weight regardless of the number of sequences per OTU. Evenness, on the other hand, measures how homogenously the individuals (or sequences in case of microbiome data) are distributed among the taxa. The more the evenness values, the more uniform the distribution among the taxa, in contrast, lower values indicate the predominance of the community by a few taxa. Evenness is inherently incorporated in compound $\alpha$-diversity indices rather than calculated independently. (Hughes et al., 2001)

Composite $\alpha$-diversity metrics incorporate both richness i.e. the number of taxa in an assemblage; in addition to evenness information. Many $\alpha$-diversity indices exist; each metric has its own strengths and weaknesses. Some are parametric measures like Fisher alpha index which assume a certain species abundance distribution, others are non-parametric like Chao and Shannon indices. Some $\alpha$-diversity like PD whole tree incorporate extra information about the phylogenetic relatedness between taxa within an assemblage set into consideration when estimating $\alpha$-diversity. This sometimes provides a deeper insight and illuminate a new aspect in the biodiversity of the studied communities. An assemblage of closely related species is less phylogenetically diverse compared to an assemblage of distantly related species, provided that all other parameters affecting $\alpha$-diversity are equivalent in the two communities. (Gotelli and Chao, 2013)

A debate around the most useful and informative $\alpha$-diversity index that could be applied in microbial ecology has been going on since the emergence of microbiome research, but has not been concluded yet as there is no perfect index that adequately summarises all aspects of the concept. Therefore, it was advised to compare biodiversity using various indices; at least two, in order to be able to draw robust conclusions on biodiversity of the studied communities. (Morris et al., 2014) In the following section, $\alpha$-diversity indices that were used to describe biodiversity of the studied airway bacterial communities in the results section are reviewed.
1.10.1.1. Chao

Chao is a richness measure calculated by adding a correction factor to the total number of observed species. This index is particularly useful for data sets skewed toward the low-abundance classes, which is likely to be the case in microbial communities in chronic lung disease. (Hughes et al., 2001)

\[ S_{\text{Chao1}} = S_{\text{obs}} + \frac{n_1^2}{2n_2} \]

where \( S_{\text{obs}} \): is the number of observed species, \( n_1 \): is the number of singletons (species captured once), and \( n_2 \): is the number of doubletons (species captured twice).

1.10.1.2. Shannon-Entropy

Shannon entropy is a non-parametric index which does not assume any particular species' abundance distribution pattern in the population. It quantifies the uncertainty in the species identity of a randomly chosen individual in the assemblage. In a highly diverse (and evenly distributed) system, an unknown individual could belong to any species, leading to a high uncertainty in predictions of its identity. In a less diverse system dominated by one or a few species, it is easier to predict the identity of unknown individuals and there is less uncertainty in the prediction. Shannon is equally sensitive to rare and abundant taxa. Despite its abstract concept, it is very popular metric in ecological literature. (Hughes et al., 2001)

\[ H(A) = - \sum_{i=1}^{n} p_i \log_2 p_i \]

Where \( H \): is Shannon entropy, \( p_i \): is the proportional abundance of species

Shannon index has been recommended by some researchers as it is less sensitive to sample size variations and it performs well with small incomplete sample sizes. It also shows superior discrimination ability between sites and a good indicator for

1.10.1.3. Fisher-alpha

Fisher-alpha assume that most communities are composed of a few species that tend to be abundant while most are rare species represented by only a few individuals (sequences). It is a parametric index that assume that the abundance of species follows log-series distribution

\[ \alpha x, \alpha x^2/2, \alpha x^3/3,..., \alpha x^n/n \]

where each term gives the number of species predicted to have 1, 2, 3, ..., n individuals in the sample. The index is the alpha parameter which is calculated by iteration

1.10.1.4. PD whole Tree

Taxa are placed on a tree (a cladogram) that describes their evolutionary relationships. The base of the tree represents ancestral taxon, the branching forks (nodes) represent speciation or divergence events, the branch tips represent the species (OTUs). An assemblage in which all the species are closely related i.e. concentrated in one region of the tree is less diverse than an assemblage in which the same number of species is widely distributed among distant branch tips of the tree. PD-whole tree is calculated by adding the branch lengths of a phylogeny connecting all species in the target assemblage. Nevertheless, species abundances are not considered in estimating this \( \alpha \)-diversity metric. Therefore, phylogenetic \( \alpha \)-diversity measures capture unique information and provide deeper insight into biodiversity within communities when considered along-side other diversity indices. (Gotelli and Chao, 2013)

1.10.2. Beta Diversity

\( \beta \)-diversity compares the microbiome profiles between samples by partitioning the biological diversity among environments or along a gradient so that it distinguishes the unique entities (species or OTUs) and defines the number of common ones
shared between environments. β-diversity measures can be either qualitative or quantitative. Qualitative measures account on the presence and absence of constituting taxa that are all considered equivalent. These measures are useful in revealing the unique taxa that may distinguish some microbial communities; especially when the difference or change involves minor taxa. On the other hand, the quantitative measures incorporate the information of relative abundance of constituting taxa in its calculations and they are more informative in distinguishing communities based on the relative abundances of taxa for example enrichment or decline in the abundance of specific taxa in response to a certain variable. Nevertheless, since taxa are weighted with respect to their relative abundances hence the quantitative β-diversity measures are driven mainly by variations in the highly abundant taxa while significant differences and/or variations within the minor ones may be obscured. (Lozupone et al., 2007, Anderson et al., 2011, Parks and Beiko, 2012)

Many β-diversity measures that are used by ecologists on macro-organisms such as Jaccard, Bray-Curtis, and Euclidean can be applied on microbes as well. Most of the adopted β-diversity measures are multivariate measures based on pairwise comparisons between sets of assemblages (representing the samples' microbiome) to determine resemblance between them in the form of either: similarities, dissimilarities or distance scores. Nevertheless, the previously mentioned measures treat entities (species or OTUs) as equally related. (Anderson et al., 2011) UniFrac is a recently developed tool to be specifically used in microbial ecology. It is a distance matrix based tool that measures the difference between two sets of taxa as the amount of evolutionary history that is unique to either of the two samples. It has been proven to be efficient in testing whether the phylogenetic lineages between microbial communities are significantly different, and in clustering samples through multivariate statistical techniques or principal coordinate analysis (PCoA). It has two versions; unweighted UniFrac (qualitative) and weighted UniFrac (quantitative) that complement each other in exploring biodiversity variations across samples. (Lozupone and Knight, 2005, Lozupone et al., 2010)
**Chapter 1: Introduction**

**Aim of Thesis:**

The hypotheses on which the presented work based were:

- The whole microbial community composition contributes to the pathogenesis of chronic respiratory conditions rather than individual pathogens.
- Prolonged antibiotic prophylaxis therapy may reduce the total bacterial load and negatively impact the diversity of the bacterial population in the airways.
- The extensive use of antibiotics may contribute to the enrichment of AMR determinants among microbiota which may act as a reservoir for AMR genes in the body.

The aims of this thesis were to:

- Characterize the airway microbial dysbiosis associated with different forms of chronic lung conditions in comparison to matched healthy comparator groups.
- Identify bacterial signatures in sputum that are associated with chronic lung disease.
- Understand the dynamics of airway microbiome in sputum over time (Chapter 4).
- To study the impact of therapeutic interventions on the individual microbiome and resistome in patients (Chapter 5).

Various cohorts were included in the presented thesis that represent a spectrum of chronic lung diseases: non-cystic fibrosis bronchiectasis and COPD, bronchopulmonary dysplasia (BPD) in adult survivors of extremely preterm birth (Chapter 6) and early pulmonary changes in people living with HIV (PLW-HIV) (Chapter 7). The hypotheses were addressed using both standard bacteriological culture techniques and culture independent molecular approaches. Specific targeted qPCRs for respiratory pathogenic bacteria and 16S rRNA sequencing using Illumina MiSeq platform were performed on sputum samples collected from the study cohorts. The AMR was studied using the standard antimicrobial susceptibility testing on sputum isolates and metagenomics sequencing using Oxford Nanopore MinION platform on a subset of sputum samples to compare resistome between the different cohorts with varying exposure extents to chemotherapeutic antimicrobial treatment.
Chapter 2: Materials and Methods

2.1. Sputum Sample Collection

All sputum samples, except the EPICure samples, were spontaneously expectorated and collected by patients (Ch.4.2.2, Appendix 3.1). For the EPICure study (Chapter 6), sputum was induced in participants using nebulised hypertonic saline by the clinical team of the EPICure study ([Rofael et al., 2019]. Sputum samples were stored in the -70°C freezer in Centre for Clinical Microbiology (CCM), Royal Free Campus within 3 hrs (±1 hr) of receipt. The samples collected in the context of “Changes in Microbiome with Treatment in Chronic Lung Diseases” study (Appendix 1, Ch.4, 4.2, Ch.5, 5.3.8.) were processed immediately upon receipt; then bacteriology cultures were performed on them before being stored at 70°C.

2.2. Sputum Homogenisation

All sputum samples were processed in the containment level 3 (CL 3) laboratory. The samples that were stored at -70°C, were allowed first to thaw at room temperature, then they were treated with an equal volume of freshly diluted Sputasol (Oxoid Ltd, UK) by diluting 750 µL in 10 mL DNA free molecular grade water. The samples were thoroughly mixed on vortex (Clifton™ Cyclone vortex mixer, Nickel-Electro Ltd, UK) for 30 secs and incubated at room temperature (24-25°C) for 15 min with occasional mixing (two to three times) on vortex. The sputum was then aliquoted into 1 mL aliquots and stored at -70°C.

2.3. Bacteriological cultures

Following the standard bacteriological culture protocol for respiratory samples of the Clinical Microbiology department, Royal Free NHS Foundation trust (Lyons, 2014), 10 µL loopful of digested sputum were streaked over 2/3 of the surface of the following agar plates (Oxoid Ltd, UK): Columbia agar with chocolate horse blood agar (CHOC) (an enrichment culture medium), Columbia CNA agar
(selective culture medium to Gram positive cocci), and CLED agar (selective and differential culture medium for Gram negative bacteria based on lactose fermentation) in a three-phase streak pattern using sterile plastic inoculation loops (Nunc™ Stervino Life Science, Denemark). Optochin (5 µg) and bacitracin (10U) discs (Oxoid Ltd, UK) were placed at the junction of the pool and the first zone streaks of CNA and CHOC agar plates respectively to test for the presence of *S. pneumoniae* which is sensitive to optochin (inhibition zone diameter should be ≥14 mm) and *H. influenzae* which is resistant to bacitracin (Tille, 2017). The sputum was then diluted to 10⁻⁵ by inoculating 10 µL of digested sputum into 9 mL Ringer solution ¼ strength (Oxoid Ltd, UK); then 10 µL of the diluted sample was spread over the 3rd quadrant of each plate as a semi-quantitative method. The plates were incubated overnight at 37°C; the CLED at ambient atmosphere; while the CHOC and CNA in the presence of 5% carbon dioxide using anaerobic jars and CO₂ Gen™ sachets (Thermo Scientific, UK). If a bacterial pathogen was present at ≥20 colonies in the dilution quadrant, this was considered active infection. All morphologically distinct isolated colonies on the three plates were picked up and identified using Matrix Assisted Laser Desorption/Ionization – Time Of Flight (MALDI-TOF) mass spectrometry.

### 2.4. Bacterial Identification using MALDI-TOF

A single isolated colony was picked using a wooden pick stick and was spread over two spots on the MALDI-TOF plate, proteins were extracted using 1 µL of 100% formic acid for each spot on the plate, then after air drying the spots were overlaid with 1 µL of daily freshly reconstituted ionization matrix and allowed to air dry. The ionization matrix is composed of crystallized molecules of 2.5 mg α-cyano-4-hydroxycinnamic acid (alpha-matrix) (Bruker Daltonik, Germany) reconstituted in a 250 µL solution made of highly purified water, 50% v/v acetonitrile and 2.5% v/v trifluoroacetic acid (TFA) (all previous solvents purchased from Sigma-Aldrich, UK). Then the plate is loaded into the MALDI-TOF platform (Bruker MALDI Biotyper; Bruker Daltonik, Germany). A laser beam strikes the targeted spots with the analyte/matrix hundreds of times. A mass spectrum is generated for each
analyte and is compared with spectra in the reference database; therefore, an identity of the bacterial isolate (analyte) is given with a confidence score based on the unique proteomic fingerprint of each organism (a sample of MALDI-TOF report is included in Appendix 3.5). If the confidence score is $\geq 2$ then the species identification is accepted; however, if the score is $\geq 1.7$ and $<2$ then the identification to genus level is accepted, $<1.7$ the identification is considered not reliable. (Gorton, 2014)

2.5. Antimicrobial Susceptibility Test (Kirby-Bauer method)

Antimicrobial susceptibility of bacterial isolates isolated from the sputum samples was tested following the standardised disc diffusion technique by European Committee on Antimicrobial Susceptibility Testing (EUCAST, 2017). First the bacterial isolates were grown on the appropriate culture medium (CHOC or Columbia horse blood agar) (Oxoid, UK) in pure culture, then a few colonies were suspended in 3 mL sterile normal saline (0.9%) (Oxoid, UK) to obtain a suspension visually equivalent to 0.5 McFarland Opacity Standard (Oxoid, UK), the optical density (OD) was also adjusted to absorbance reading 0.1 at 600 nm using cell density meter (WPA Biochrom Ltd., UK). A sterile cotton swab was dipped into the adjusted bacterial suspension and the excess suspension was removed by pressing the swab gently against the walls of the test tube. The swab was streaked all over the surface of the recommended Müller-Hinton agar plates, depending on the tested organism, in three directions at an angle of 60º to obtain an even distribution of the inoculum. The plates were then left to dry at room temperature for a few minutes with the covers closed. Selected antibiotic discs were then placed on the surface of the inoculated agar plate using an Oxoid™ disk dispenser (Thermo Scientific, UK). Antibiotics tested were selected to represent the major classes of widely used antibiotics in clinical practice in the light of the recommendations of EUCAST and the ‘Clinical and Laboratory Standards Institute’ (CLSI) for each group of organisms. The plates were then incubated at 37ºC for 18 hr. After incubation, the diameters of the inhibition zones developed around the discs were measured manually in millimetres in three different directions and the
Chapter 2: Materials and Methods

average of three readings was compared to the breakpoints in susceptibility tables published in (EUCAST, 2017) and/or (CLSI, 2017).

2.6. DNA extraction

2.6.1. DNA extraction from sputum samples using the LIAISON® automated extraction platform

A one millilitre aliquot of each sputum sample was taken out of -70°C and allowed to thaw at room temperature in the CL 3 laboratory. All samples were centrifuged at 20,000 xg then the supernatant was discarded and the pellets were heated at 95°C for 30 min. The pellets were then disrupted with approximately equal volume of silica beads (FastPrep® lysis matrix B) (MP Biomedicals™, Fisher Scientific, UK) on Fast-Prep®-24 Instrument (MP Biomedicals™, Fisher Scientific, UK) for 45 secs followed by treatment with 10 µL proteinase K and 240 µL of DNA Pre-treatment buffer 2; both were provided with Arrow™ DNA extraction kit (DiaSorin Ltd, Ireland). Samples were incubated at 56°C for at least 10 min. Metagenomic DNA was then extracted on the automated LIAISON® Ixt extraction platform (DiaSorin Ltd, Ireland) using Arrow™ DNA extraction kit (DiaSorin Ltd, Ireland) and DNA extraction programme V2 as per manufacturer instructions. Extracted DNA was eluted into 100 µL. Extraction negative controls of diluted Sputasol were extracted along-side each batch to test for potential contamination.

2.6.2. DNA extraction from EPICure induced sputum samples using Qiagen spin column

Metagenomic DNA was extracted from 500 µL of Sputasol® treated sputum samples after being heated at 95°C for 30 min using Qiagen DNeasy® Blood and Tissue kit (Qiagen, UK) following the user group developed protocol for purification of total DNA from viscous samples using the DNeasy® Blood & Tissue Kit ; however, the N-acetyl L-cysteine (NALC) pre-treatment step was omitted since Sputasol served to liquefy the samples. Extraction negative controls of sterile water and the sputum induction matrix were also performed alongside the processed samples. Extracted DNA was eluted twice in 100 µL AE buffer each; to maximize the amount of DNA eluted.
2.6.3. Measuring the DNA concentration and purity of DNA extracts

The DNA concentration of all DNA extracts was measured in 1 µL using a NanoDrop™ 2000 spectrophotometer (Thermo Scientific, UK). The purity of the DNA extracts was checked by the $A_{260}/A_{280}$ ratio; a reading given by NanoDrop™ 2000 which should be between 1.8 and 2. If the ratio <1.8 this indicates contamination with proteins and if >2 this indicates RNA contamination (Manchester, 1995). The integrity of the extracted DNA was checked through gel electrophoresis in which 1 µL aliquots were run in 0.5 g% w/v agarose gel at 100 V electric field for 45 min. The gel should reveal a single high molecular weight DNA band without smearing which may indicate fragmentation. The DNA extracts were then stored at -20°C.

2.7. Mock Community Preparation

The first set of mock communities (A, D, E and F) were to test the efficacy of the DNA extraction method and were prepared by mixing equal proportions of approximately equally dense bacterial suspensions in sterile normal saline.

Mock community A was prepared by mixing 500 µL aliquots of *S. pneumoniae* ATCC 49619 (OD 0.3 at 600 nm) *H. influenzae* NCTC 8468 (OD 0.27) and *Moraxella catarrhalis* ATCC 25240 (OD 0.31) in 3500 µL artificial sputum base (Appendix 3.2). The bacterial mixture was then aliquoted into 500 µL aliquots and DNA extracted LIAISON® Ixt extraction platform with and without the heat killing step.

Mock communities: D and (E/F) were prepared by mixing 1 mL of each of the following ten bacteria: *Streptococcus pneumoniae* ATCC 49619, *Streptococcus parasanguinis* clinical isolate, *Staphylococcus aureus* ATCC 29213, *Haemophilus influenzae* NCTC 8468, *Moraxella catarrhalis* ATCC 25240, *Neisseria subflava* clinical isolate, *Pseudomonas aeruginosa* clinical isolate, *Escherichia coli* ATCC 25922, *Klebsiella pneumoniae* NCTC 13438, and *Acinetobacter baumannii* clinical isolate. The OD of each bacterial suspension was adjusted at around 0.1 for mock community D and at 0.3 for mock communities: E and F. The bacterial viable count
Chapter 2: Materials and Methods

was determined for each using Miles and Misra technique. (Miles et al., 1938) Both mock communities: E and F were prepared from the same bacterial suspension mixture and were treated similarly and simultaneously, the only difference was that Mock F was heat killed before DNA extraction; while Mock E was not. Mock community D was prepared independently. Then, the whole mixture in 1mL aliquots of each mock community was extracted using the same protocol as samples (section 2.6.1); mock community F with the heat killing step; while mock D and mock E without.

The second set of mock communities (mock B and C) was prepared from the genomic DNA extracts of the individual bacterial components. Mock community B was prepared by diluting 2 µL of DNA extracts of each of Streptococcus pneumoniae ATCC 49619 (0.79 ng/mL), Haemophilus influenzae NCTC 8468 (17 ng/mL) and Moraxella catarrhalis ATCC 25240 (2.67 ng/mL) pure cultures, in 196 µL Tris-EDTA (TE) Buffer. This mock community was used as positive control in the typical respiratory pathogens multiplex qPCR (2.10.1).

Mock C was included in every 16S rRNA sequencing to control the inter-run variations between the sequencing runs. It was prepared by pooling equal DNA concentrations (122 ng/µL) of DNA extracts of pure bacterial suspensions of each of the following nine bacteria: S. pneumoniae ATCC 49619, H. influenzae NCTC 8468, Streptococcus pyogenes ATCC 19615, E. coli ATCC 27853, K. pneumoniae NCTC 13438, S. aureus ATCC 29213, A. baumannii clinical isolate, P. aeruginosa clinical isolate, M. catarrhalis clinical isolate. Such that the final concentration of each is 13.5 ng/µL in the final mixture. All DNA concentrations were measured using Qubit™ dsDNA HS kit and Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, UK). All mock communities were aliquoted in 5 µL aliquots and stored at -20 °C.

2.8. 16S rRNA gene sequencing

2.8.1 16S rRNA gene amplification

Following the UCL Genomics 16S Metagenomic Profiling Assay Protocol (Appendix 3.3, Personal Communication), a sequence library was created by
amplification of either V3-V4 or V5-V7 regions of the bacterial 16S \textit{rrna} gene through conventional PCR using 341 forward primer and 805 reverse primer (Klindworth et al., 2013, Hugerth et al., 2014) or 785 forward primer and 1175 reverse primer (as per protocol) respectively. The Miseq primers were designed with standard Illumina® dual indices and adaptors (P5 and P7 in the forward and reverse primers respectively) attached. The NGS primers were purchased from Sigma-Aldrich, UK, HPLC purified. The details of NGS primers are given in Tables 2.1 and 2.2 (personal communication with Dr. Mehmet Davrandi). Each sample was assigned a unique pair combination of standard Illumina® dual indexed primers. The extraction negative control and a no-template PCR control (water) were also run throughout the amplification and library preparation steps as negative controls to allow for the evaluation of potential contamination. Mock community C was also run in every sequencing run as a positive control and to evaluate the inter-run variation.

The PCR master-mix per reaction was composed of; 0.2 µM forward and reverse primers each, 10 µL 2.5x basic Master-mix and 0.8 µL MolTaq 16S DNA polymerase (Mastermix 16S basic, cat no. S-040-0250, Molzym, VH Bio Limited, UK). The amount of DNA template added was adjusted such that the final DNA input per reaction was 300 ng. The final volume of PCR reaction was adjusted with DNA-free PCR-grade water (Molzym, VH Bio Limited, UK) to 25 µL. The thermocycling conditions were initial denaturing step: at 95°C for 5 min, followed by 30 cycles of denaturing at 95°C for 30 secs, annealing at 55°C for 40 sec and extension at 72°C for 1 min, in addition to a final extension phase at 72°C for 10 min.

The PCR products were checked by gel electrophoresis in which 2 µL aliquots from each reaction post PCR were run in 1 g% w/v agarose gel in 120 V electric field for 50 mins. Single DNA band should appear per reaction at the expected amplicon size; 580 bp in case of V3-V4 or 504 bp in case of V5-V7.
### Table 2.1: NGS Primers for V3-V4 16S rRNA Amplification

<table>
<thead>
<tr>
<th>Primer ID</th>
<th>P5 (29bp)</th>
<th>Index (8bp)</th>
<th>Tm Padding (11bp)</th>
<th>Forward Pr 341F (17bp)</th>
<th>Complete Sequence</th>
<th>bp</th>
</tr>
</thead>
<tbody>
<tr>
<td>P5_V3_FW 01</td>
<td>AGATACGGCACCCCTCTACAGTAA</td>
<td>GTAG</td>
<td>AGTACGTA</td>
<td>CCTACGGGNGG</td>
<td>AATGATACGGCGACCACCCCTCTACAGTAA</td>
<td>65</td>
</tr>
<tr>
<td>P5_V3_FW 02</td>
<td>CCGAGATCTACAC</td>
<td>TCGC</td>
<td>CGT</td>
<td>CCWGCAG</td>
<td>GTACGTACGTCTACGGGNGGCGCCWGCAG</td>
<td></td>
</tr>
<tr>
<td>P5_V3_FW 03</td>
<td>CCGAGATCTACAC</td>
<td>CTCT</td>
<td>AGTACGTA</td>
<td>CCTACGGGNGG</td>
<td>AATGATACGCGACCACCCCTCTACATCCTCTCTATATCGCAG</td>
<td>65</td>
</tr>
<tr>
<td>P5_V3_FW 04</td>
<td>CCGAGATCTACAC</td>
<td>TTCT</td>
<td>CGT</td>
<td>CCWGCAG</td>
<td>GTACGTACGTCTACGGGNGGCGCCWGCAG</td>
<td></td>
</tr>
<tr>
<td>P5_V3_FW 05</td>
<td>CCGAGATCTACAC</td>
<td>TAGA</td>
<td>AGTACGTA</td>
<td>CCTACGGGNGG</td>
<td>AATGATACGCGACCACCCCTCAGTTAG</td>
<td>65</td>
</tr>
<tr>
<td>P5_V3_FW 06</td>
<td>CCGAGATCTACAC</td>
<td>GTTA</td>
<td>AGTACGTA</td>
<td>CCTACGGGNGG</td>
<td>AATGATACGCGACCACCCCTCAGTTAG</td>
<td>65</td>
</tr>
<tr>
<td>P5_V3_FW 07</td>
<td>CCGAGATCTACAC</td>
<td>ACTG</td>
<td>AGTACGTA</td>
<td>CCTACGGGNGG</td>
<td>AATGATACGCGACCACCCCTCAGTTAG</td>
<td>65</td>
</tr>
<tr>
<td>P5_V3_FW 08</td>
<td>CCGAGATCTACAC</td>
<td>AAGG</td>
<td>AGTACGTA</td>
<td>CCTACGGGNGG</td>
<td>AATGATACGCGACCACCCCTCAGTTAG</td>
<td>65</td>
</tr>
</tbody>
</table>
### Table 2.1 Continue: NGS Primers for V3-V4 16S rRNA Amplification

#### Reverse 16S rRNA amplification Primers

<table>
<thead>
<tr>
<th>Primer ID</th>
<th>P7 (24bp)</th>
<th>Index (8bp)</th>
<th>Tm Padding (15bp)</th>
<th>Reverse Pr 805R (21bp)</th>
<th>Complete Sequence b p</th>
</tr>
</thead>
<tbody>
<tr>
<td>P7_V4_Rv</td>
<td>CAAGCAGAAGACGGCA</td>
<td>TCGC</td>
<td>GCGAGTCA</td>
<td>GACTACHVGGG</td>
<td>CAAGCAGAAGACGGCATACGAGATTCGCTTACGCGAG 68</td>
</tr>
<tr>
<td>01</td>
<td>TACGAGAT</td>
<td>CTGA</td>
<td>GTCAGCG</td>
<td>TATCTAATCC</td>
<td>TCAAGTCAGCGGACTACHVGGGTATCTAATCC</td>
</tr>
<tr>
<td>P7_V4_Rv</td>
<td>CAAGCAGAAGACGGCA</td>
<td>CTAG</td>
<td>GCGAGTCA</td>
<td>GACTACHVGGG</td>
<td>CAAGCAGAAGACGGCATACGAGATCTAGTACGCGAG 68</td>
</tr>
<tr>
<td>02</td>
<td>TACGAGAT</td>
<td>TACG</td>
<td>GTCAGCG</td>
<td>TATCTAATCC</td>
<td>TCAAGTCAGCGGACTACHVGGGTATCTAATCC</td>
</tr>
<tr>
<td>P7_V4_Rv</td>
<td>CAAGCAGAAGACGGCA</td>
<td>TTCT</td>
<td>GCGAGTCA</td>
<td>GACTACHVGGG</td>
<td>CAAGCAGAAGACGGCATACGAGATTTCGCTTACGCGAG 68</td>
</tr>
<tr>
<td>03</td>
<td>TACGAGAT</td>
<td>GCTG</td>
<td>GTCAGCG</td>
<td>TATCTAATCC</td>
<td>TCAAGTCAGCGGACTACHVGGGTATCTAATCC</td>
</tr>
<tr>
<td>P7_V4_Rv</td>
<td>CAAGCAGAAGACGGCA</td>
<td>GCTG</td>
<td>GCGAGTCA</td>
<td>GACTACHVGGG</td>
<td>CAAGCAGAAGACGGCATACGAGATGCTCAGGAGCGAG 68</td>
</tr>
<tr>
<td>04</td>
<td>TACGAGAT</td>
<td>AGGA</td>
<td>GTCAGCG</td>
<td>TATCTAATCC</td>
<td>TCAAGTCAGCGGACTACHVGGGTATCTAATCC</td>
</tr>
<tr>
<td>P7_V4_Rv</td>
<td>CAAGCAGAAGACGGCA</td>
<td>GTAG</td>
<td>GCGAGTCA</td>
<td>GACTACHVGGG</td>
<td>CAAGCAGAAGACGGCATACGAGATGTAGAGAGGCGAG 68</td>
</tr>
<tr>
<td>05</td>
<td>TACGAGAT</td>
<td>AGGA</td>
<td>GTCAGCG</td>
<td>TATCTAATCC</td>
<td>TCAAGTCAGCGGACTACHVGGGTATCTAATCC</td>
</tr>
<tr>
<td>P7_V4_Rv</td>
<td>CAAGCAGAAGACGGCA</td>
<td>CAGC</td>
<td>GCGAGTCA</td>
<td>GACTACHVGGG</td>
<td>CAAGCAGAAGACGGCATACGAGATCAGCCTCGGCGAG 68</td>
</tr>
<tr>
<td>06</td>
<td>TACGAGAT</td>
<td>CTGC</td>
<td>GTCAGCG</td>
<td>TATCTAATCC</td>
<td>TCAAGTCAGCGGACTACHVGGGTATCTAATCC</td>
</tr>
<tr>
<td>P7_V4_Rv</td>
<td>CAAGCAGAAGACGGCA</td>
<td>GGCG</td>
<td>GCGAGTCA</td>
<td>GACTACHVGGG</td>
<td>CAAGCAGAAGACGGCATACGAGATAGCGTAGCGCGAG 68</td>
</tr>
<tr>
<td>07</td>
<td>TACGAGAT</td>
<td>AGGA</td>
<td>GTCAGCG</td>
<td>TATCTAATCC</td>
<td>TCAAGTCAGCGGACTACHVGGGTATCTAATCC</td>
</tr>
<tr>
<td>P7_V4_Rv</td>
<td>CAAGCAGAAGACGGCA</td>
<td>CAGT</td>
<td>GCGAGTCA</td>
<td>GACTACHVGGG</td>
<td>CAAGCAGAAGACGGCATACGAGATCAGCCTCGGCGAG 68</td>
</tr>
<tr>
<td>08</td>
<td>TACGAGAT</td>
<td>AGGA</td>
<td>GTCAGCG</td>
<td>TATCTAATCC</td>
<td>TCAAGTCAGCGGACTACHVGGGTATCTAATCC</td>
</tr>
<tr>
<td>P7_V4_Rv</td>
<td>CAAGCAGAAGACGGCA</td>
<td>CAGT</td>
<td>GCGAGTCA</td>
<td>GACTACHVGGG</td>
<td>CAAGCAGAAGACGGCATACGAGATCAGCCTCGGCGAG 68</td>
</tr>
<tr>
<td>09</td>
<td>TACGAGAT</td>
<td>CTGC</td>
<td>GTCAGCG</td>
<td>TATCTAATCC</td>
<td>TCAAGTCAGCGGACTACHVGGGTATCTAATCC</td>
</tr>
<tr>
<td>P7_V4_Rv</td>
<td>CAAGCAGAAGACGGCA</td>
<td>CAGC</td>
<td>GCGAGTCA</td>
<td>GACTACHVGGG</td>
<td>CAAGCAGAAGACGGCATACGAGATCAGCCTCGGCGAG 68</td>
</tr>
<tr>
<td>10</td>
<td>TACGAGAT</td>
<td>CTGC</td>
<td>GTCAGCG</td>
<td>TATCTAATCC</td>
<td>TCAAGTCAGCGGACTACHVGGGTATCTAATCC</td>
</tr>
<tr>
<td>P7_V4_Rv</td>
<td>CAAGCAGAAGACGGCA</td>
<td>TGCC</td>
<td>GCGAGTCA</td>
<td>GACTACHVGGG</td>
<td>CAAGCAGAAGACGGCATACGAGATTTCGCTTACGCGAG 68</td>
</tr>
<tr>
<td>11</td>
<td>TACGAGAT</td>
<td>TCTT</td>
<td>GTCAGCG</td>
<td>TATCTAATCC</td>
<td>TCAAGTCAGCGGACTACHVGGGTATCTAATCC</td>
</tr>
<tr>
<td>P7_V4_Rv</td>
<td>CAAGCAGAAGACGGCA</td>
<td>TCGC</td>
<td>GCGAGTCA</td>
<td>GACTACHVGGG</td>
<td>CAAGCAGAAGACGGCATACGAGATTTCGCTTACGCGAG 68</td>
</tr>
<tr>
<td>12</td>
<td>TACGAGAT</td>
<td>CTAC</td>
<td>GTCAGCG</td>
<td>TATCTAATCC</td>
<td>TCAAGTCAGCGGACTACHVGGGTATCTAATCC</td>
</tr>
</tbody>
</table>

#### Sequencing Primers

<table>
<thead>
<tr>
<th>Primer ID</th>
<th>Complete Sequence b p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read 1 Primer_V3</td>
<td>AGTACGTAGCTCTACGGGNGGCWGGCAG 28</td>
</tr>
<tr>
<td>Read 2 Primer_V4</td>
<td>GCGAGTCAGTGCAAGGACTACHVGGGTATCTAATCC 36</td>
</tr>
<tr>
<td>Index i7 read Primer_V4 (Reverse complement of Read 2 Primer_V4)</td>
<td>GAGATTAGATACCBDGTAGTCGCCTGACTGACTCGC 36</td>
</tr>
</tbody>
</table>
## Chapter 2: Materials and Methods

### Table 2.2: NGS Primers for V5-V7 16S rRNA Amplification

<table>
<thead>
<tr>
<th>Primer ID</th>
<th>P5 (29 bp)</th>
<th>Index i5 (8 bp)</th>
<th>Tm Padding (12 bp)</th>
<th>Forward Pr 785F</th>
<th>Complete Sequence</th>
<th>bp</th>
</tr>
</thead>
<tbody>
<tr>
<td>P5_FW 01</td>
<td>AATGATACGGCGACCAACC</td>
<td>TAGATC GCCTCTCT AT</td>
<td>ACGTACGTA CGT</td>
<td>GAATTAGATACCBR GGTAGTC</td>
<td>AATGATACGGCGACCAACCTGCTGATTAGATACCCBR GGTAGTC</td>
<td>70</td>
</tr>
<tr>
<td>P5_FW 02</td>
<td>AATGATACGGCGACCAACC</td>
<td>TAGATC GCCTCTCT AT</td>
<td>ACGTACGTA CGT</td>
<td>GAATTAGATACCBR GGTAGTC</td>
<td>AATGATACGGCGACCAACCTGCTGATTAGATACCCBR GGTAGTC</td>
<td>70</td>
</tr>
<tr>
<td>P5_FW 03</td>
<td>AATGATACGGCGACCAACC</td>
<td>TAGATC GCCTCTCT AT</td>
<td>ACGTACGTA CGT</td>
<td>GAATTAGATACCBR GGTAGTC</td>
<td>AATGATACGGCGACCAACCTGCTGATTAGATACCCBR GGTAGTC</td>
<td>70</td>
</tr>
<tr>
<td>P5_FW 04</td>
<td>AATGATACGGCGACCAACC</td>
<td>TAGATC GCCTCTCT AT</td>
<td>ACGTACGTA CGT</td>
<td>GAATTAGATACCBR GGTAGTC</td>
<td>AATGATACGGCGACCAACCTGCTGATTAGATACCCBR GGTAGTC</td>
<td>70</td>
</tr>
<tr>
<td>P5_FW 05</td>
<td>AATGATACGGCGACCAACC</td>
<td>TAGATC GCCTCTCT AT</td>
<td>ACGTACGTA CGT</td>
<td>GAATTAGATACCBR GGTAGTC</td>
<td>AATGATACGGCGACCAACCTGCTGATTAGATACCCBR GGTAGTC</td>
<td>70</td>
</tr>
<tr>
<td>P5_FW 06</td>
<td>AATGATACGGCGACCAACC</td>
<td>TAGATC GCCTCTCT AT</td>
<td>ACGTACGTA CGT</td>
<td>GAATTAGATACCBR GGTAGTC</td>
<td>AATGATACGGCGACCAACCTGCTGATTAGATACCCBR GGTAGTC</td>
<td>70</td>
</tr>
<tr>
<td>P5_FW 07</td>
<td>AATGATACGGCGACCAACC</td>
<td>TAGATC GCCTCTCT AT</td>
<td>ACGTACGTA CGT</td>
<td>GAATTAGATACCBR GGTAGTC</td>
<td>AATGATACGGCGACCAACCTGCTGATTAGATACCCBR GGTAGTC</td>
<td>70</td>
</tr>
<tr>
<td>P5_FW 08</td>
<td>AATGATACGGCGACCAACC</td>
<td>TAGATC GCCTCTCT AT</td>
<td>ACGTACGTA CGT</td>
<td>GAATTAGATACCBR GGTAGTC</td>
<td>AATGATACGGCGACCAACCTGCTGATTAGATACCCBR GGTAGTC</td>
<td>70</td>
</tr>
</tbody>
</table>

**Notes:**
- The indices are added to the 5' end of the primer.
- The Tm Padding is added to the 3' end of the primer.
- The Complete Sequence includes the index, Tm Padding, and forward Pr 785F sequence.
#### Table 2.2 Continue: NGS Primers for V5-V7 16S rRNA Amplification

<table>
<thead>
<tr>
<th>Primer ID</th>
<th>Primer Name</th>
<th>Seq (24 bp)</th>
<th>Index i7 (8 bp)</th>
<th>Tm Padding (12 bp)</th>
<th>Reverse Pr 1175 R (20 bp)</th>
<th>Complete Sequence</th>
<th>bp</th>
</tr>
</thead>
<tbody>
<tr>
<td>P7_Rv 01</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>TCGCCT</td>
<td>AGTCAGTCA</td>
<td>ACGTCRTCCCDCTCCT</td>
<td>CAAGCAGAAGACGGCATAACGTCAGTCAAGTCRTCCCDCTCCT</td>
<td>64</td>
</tr>
<tr>
<td>P7_Rv 02</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>TAGCTTA</td>
<td>AGTCAGTCA</td>
<td>ACGTCRTCCCDCTCCT</td>
<td>CAAGCAGAAGACGGCATAACGTCAGTCAAGTCRTCCCDCTCCT</td>
<td>64</td>
</tr>
<tr>
<td>P7_Rv 03</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>TTCTGC</td>
<td>AGTCAGTCA</td>
<td>ACGTCRTCCCDCTCCT</td>
<td>CAAGCAGAAGACGGCATAACGTCAGTCAAGTCRTCCCDCTCCT</td>
<td>64</td>
</tr>
<tr>
<td>P7_Rv 04</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>GTAAGGA</td>
<td>AGTCAGTCA</td>
<td>ACGTCRTCCCDCTCCT</td>
<td>CAAGCAGAAGACGGCATAACGTCAGTCAAGTCRTCCCDCTCCT</td>
<td>64</td>
</tr>
<tr>
<td>P7_Rv 05</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>CATGCTCC</td>
<td>AGTCAGTCA</td>
<td>ACGTCRTCCCDCTCCT</td>
<td>CAAGCAGAAGACGGCATAACGTCAGTCAAGTCRTCCCDCTCCT</td>
<td>64</td>
</tr>
<tr>
<td>P7_Rv 06</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>GTAGAGA</td>
<td>AGTCAGTCA</td>
<td>ACGTCRTCCCDCTCCT</td>
<td>CAAGCAGAAGACGGCATAACGTCAGTCAAGTCRTCCCDCTCCT</td>
<td>64</td>
</tr>
<tr>
<td>P7_Rv 07</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>CCGCTG</td>
<td>AGTCAGTCA</td>
<td>ACGTCRTCCCDCTCCT</td>
<td>CAAGCAGAAGACGGCATAACGTCAGTCAAGTCRTCCCDCTCCT</td>
<td>64</td>
</tr>
<tr>
<td>P7_Rv 08</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>AGCTG</td>
<td>AGTCAGTCA</td>
<td>ACGTCRTCCCDCTCCT</td>
<td>CAAGCAGAAGACGGCATAACGTCAGTCAAGTCRTCCCDCTCCT</td>
<td>64</td>
</tr>
<tr>
<td>P7_Rv 09</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>CGATG</td>
<td>AGTCAGTCA</td>
<td>ACGTCRTCCCDCTCCT</td>
<td>CAAGCAGAAGACGGCATAACGTCAGTCAAGTCRTCCCDCTCCT</td>
<td>64</td>
</tr>
<tr>
<td>P7_Rv 10</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>CGAGCG</td>
<td>AGTCAGTCA</td>
<td>ACGTCRTCCCDCTCCT</td>
<td>CAAGCAGAAGACGGCATAACGTCAGTCAAGTCRTCCCDCTCCT</td>
<td>64</td>
</tr>
<tr>
<td>P7_Rv 11</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>TCGCTTG</td>
<td>AGTCAGTCA</td>
<td>ACGTCRTCCCDCTCCT</td>
<td>CAAGCAGAAGACGGCATAACGTCAGTCAAGTCRTCCCDCTCCT</td>
<td>64</td>
</tr>
<tr>
<td>P7_Rv 12</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>CAAGCAGAAGACGGCATA</td>
<td>ATCCCTG</td>
<td>AGTCAGTCA</td>
<td>ACGTCRTCCCDCTCCT</td>
<td>CAAGCAGAAGACGGCATAACGTCAGTCAAGTCRTCCCDCTCCT</td>
<td>64</td>
</tr>
</tbody>
</table>

**Sequencing Primers**

- **Read 1 Primer_v5**
  - Read 2 Primer_v7
  - Index i7 read Primer_v7 (Reverse complement of Read 2 Primer_v7)
2.8.2. Library preparation

The PCR products were cleaned up using Agencourt AMPure XP beads (Beckman Coulter, UK) with a binding buffer of 2.5 M sodium chloride and 20 g% w/v PEG-8000, 80% freshly prepared ethanol and EB Buffer® (Qiagen, UK) to remove nonspecific amplicons <200 bp and primer dimers. DNA in the cleaned products was then quantified using Qubit™ dsDNA HS kit and Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, UK). The samples were pooled in an equimolar ratio between 2 and 10 nM into one library. The pooled library concentration was checked again by Qubit™ dsDNA HS kit and the size of DNA fragments on TapeStation Automated Electrophoresis (Agilent, UK) (Figure 2.1).

![TapeStation Electrophoresis result for ‘run 4’ sequence library (V3-V4 16S rRNA gene) showing a single peak at 580 bp.](image)

Figure 2.1. TapeStation Electrophoresis result for ‘run 4’ sequence library (V3-V4 16S rRNA gene) showing a single peak at 580 bp.

Sequencing of the library was performed by UCL Genomics unit. In brief; the library was diluted to 4 nM using EB buffer (Qiagen, UK) based on the actual reading of Qubit®. Following Illumina protocol A in the Miseq system denature and dilute libraries guide (Illumina, 2018), 5 µL of library in denatured using 5 µL of 0.2 N sodium hydroxide at room temperature for 5 min then the reaction was promptly stopped with 990 µL pre-chilled HT1 (Illumina Cambridge, Ltd, UK). After that 405 µL of denatured library (20 pM) was diluted with 360 µL pre-chilled HT1 down to 8 pM (1000 µL). Similarly, PhiX control V3 KIT (cat no. FC-110-3001) which is an internal sequencing control for the sequencing run (Illumina Cambridge, Ltd, UK) is denatured and diluted to 12.5 pM then it was spiked into the sequence library with final concentration 10% v/v.
2.8.3. NGS Sequencing

Sequencing was performed on Illumina MiSeq Platform using MiSeq® Reagent Kit v2 (500 cycles) (cat no. MS-102-2003, Illumina Cambridge, Ltd, UK) following Miseq System Custom Primers Guide (Illumina, 2016). The cassette of the kit was removed from -20°C and was left to thaw in water bath at room temperature. Then the 1 ml 8 pM sequence library was transferred into Well 17. The contents of wells 12, 13 and 14 were taken out and were mixed with the costume sequencing primers: read 1, read 2 and index i7 (Tables 2.1 and 2.2) respectively in microcentrifuge tubes then were transferred back to their corresponding wells.

Next the flow cell was removed from the preservative solution and rinsed with distilled water and wiped well with 100% Ethanol. Finally, the flow cell, cassette and sample sheet of the library were loaded into the Miseq machine and sequencing run is started. The run took up to 48 hrs.

2.9. Bioinformatic and Statistical Analyses

The sequencing reads were primary analysed and demultiplexed on the Miseq machine and were exported to Illumina cloud-based BaseSpace®. For bioinformatic analysis, the workflow established by the Microbiome Helper was adopted. (Comeau et al., 2017) The paired end reads were stitched together using PEAR. (Zhang et al., 2014) The low-quality reads with quality score <30 over 10% of its bases and length less than 400 bp in case of V3-V4 runs or 350 in V5-V7 run were filtered out using FASTX-toolkit (v.0.0.14). (Gordon, 2009) The reads were then screened for possible chimeras that may have resulted from PCR using VSEARCH (v1.11.1). (Rognes et al., 2016) The subsequent steps were through QIIME pipeline v1.9.1 (Caporaso et al., 2010) where the sequences were clustered based on 97% similarity into Operational Taxonomic Units (OTU) and taxonomic classification was assigned to OTUs using open reference OTU picking against Greengenes database version 13_8. All samples within an OTU table were rarefied to the same depth of reads depending on each run. Alpha and beta diversity indices were calculated on the rarefied OTU tables using QIIME. The appropriate statistical significance tests were calculated using SPSS V. 25 (IBM, 2017) or QIIME wrapper
scripts. STAMP (v2.1.3) (Parks et al., 2014) was used to visualize the results and explore the OTUs showing significant differences across the study groups. The significance of the differential abundances of the constituting taxa/OTUs in the BIOM table was tested using either White’s non-parametric test (for comparing two groups) (White et al., 2009) or Kruskal-Wallis test (for comparing multiple groups >2) which are implemented in the STAMP bioinformatics programme, all p-values were corrected for multiple comparisons using Benjamini-Hochberg False Discovery Rate (BH-FDR).

2.10. Multiplex quantitative polymerase chain reaction (qPCR) for Respiratory pathogens

2.10.1. Typical respiratory pathogen qPCR

A multiplex qPCR was used for identification and quantification of three respiratory bacterial pathogens; *Haemophilus influenzae* (Hi), *Moraxella catarrhalis* (Mc), *Streptococcus pneumoniae* (Spn) targeting P4 lipoprotein gene, CopB outer membrane-protein gene and *spn9802* gene fragments respectively using TaqMan® dual probes. (Garcha et al., 2012)

An Internal amplification control (IAC), Spud A, that targets the *phyB* gene from the potato species *Solanum tuberosum* was used to test for PCR inhibition. (Nolan et al., 2006) Inhibition is indicated by a delayed Ct value of the IAC amplification as shown in Figure 2.2.

![Figure 2.2](image-url)

**Figure 2.2.** Amplification plot of the internal amplification control (IAC); SPUD A, in tested sputum samples, Ct values < 30 are acceptable, however for any sample a Ct value of the IAC > 30 indicates PCR inhibition.
SPUD A is a 101 base amplicon (TIB® MOLBIOL, Berlin, Germany) which is added in the master-mix at a fixed final concentration of 0.04 pM along with the SPUD primers and probe. The sequences of Spud A, primers and probes of the three targets and the internal control are described in Table 2.3. The specificity of the used primers and probes to their corresponding bacterial targets were checked in-silico through running Primer BLAST against NCBI RefSeq Representative Genome Database. (Ye et al., 2012) In addition, the specificity were previously checked against number of isolates during the in-house validation of the method in UCL Centre for Clinical Microbiology by Dr Clare Ling (Garcha, 2014). Primers and probes were ordered from Eurofins genomics, Germany.

Table 2.3: The primers and probes for the three bacterial targets and the internal amplification control in multiplex qPCR for typical respiratory pathogens

<table>
<thead>
<tr>
<th>Targets</th>
<th>5′–3′ Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>H. influenzae</strong></td>
<td></td>
</tr>
<tr>
<td>Hi Forward primer</td>
<td>CCGGGTGC GGTTAGAAT TTAATA</td>
</tr>
<tr>
<td>Hi Reverse primer</td>
<td>CTGATTTTTCA GTGCCTTTTG</td>
</tr>
<tr>
<td>Hi Probe</td>
<td>6FAM-ACAGCCACAACGGTA AGTGTTTCTACGT-DB</td>
</tr>
<tr>
<td><strong>M. catarrhalis</strong></td>
<td></td>
</tr>
<tr>
<td>Mc Forward primer</td>
<td>GTGAGTGCCGCTTTT ACAACC</td>
</tr>
<tr>
<td>Mc Reverse primer</td>
<td>TGTATCGCCTGC CAAAGC AA</td>
</tr>
<tr>
<td>Mc Probe</td>
<td>6JOE-TGCTTTT GCAGCTTTAGCCAGCCTAA-BHQ1</td>
</tr>
<tr>
<td><strong>S. pneumoniae</strong></td>
<td></td>
</tr>
<tr>
<td>Spn Forward primer</td>
<td>AGTCTTCCAAGGTAACAAGTCT</td>
</tr>
<tr>
<td>Spn Reverse primer</td>
<td>ACCAACTCGACCACCTT TT</td>
</tr>
<tr>
<td>Spn Probe</td>
<td>ROX-TACATGTAGGA AACTATTTTCCTCACA A-BHQ2</td>
</tr>
<tr>
<td><strong>IAC</strong></td>
<td></td>
</tr>
<tr>
<td>Spud Forward primer</td>
<td>AACTTGCGTTT AATGGACCTCAA</td>
</tr>
<tr>
<td>Spud Reverse primer</td>
<td>ACATTCACTTTACATGGACA</td>
</tr>
<tr>
<td>Spud Probe</td>
<td>Cy5-TGCACAAGCTATGGGAACACCACGT-BBQ</td>
</tr>
<tr>
<td>SpudA</td>
<td>AACTTGCGTTT AATGGACCTCAAATTTTGAGTGTCGACCAAGCTATGGGAACACCACGT-BBQ</td>
</tr>
</tbody>
</table>
The master-mix was prepared using Platinum® quantitative PCR Supermix-UDG (Thermo-Fisher Scientific, UK) and additional magnesium chloride at final concentration of 3 mM as shown in Table 2.4. Each sample was run in duplicate in addition to 1-in-10 dilution to check for inhibition. Each run also included one standard for each organism in duplicate, a positive control containing a mixture of the three organisms at the limit of detection (Mock B) and a negative control. The thermo-cycling conditions were 95°C for 3 min followed by 40 cycles of 95°C for 10 secs and 60°C for 45 sec as carried out on Qiagen Rotor– gene® 6000 real-time PCR machine (Corbett Research UK, Cambridgeshire, UK). The signal of *H. influenzae* was captured on the green channel, *M. catarrhalis* on the yellow channel, *S. pneumoniae* on the orange channel and IAC on the red channel. Load of each tested organism was calculated in colony forming unit (CFU) per mL of sputum and average of the duplicates was taken.

### Table 2.4: Master-mix preparation of the typical respiratory pathogens

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Stock Conc</th>
<th>Volume per reaction (µL)</th>
<th>Final Conc per reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Platinum qPCR Supermix</td>
<td>X1</td>
<td>12.5</td>
<td></td>
</tr>
<tr>
<td>MgCl₂</td>
<td>50 mM</td>
<td>1.5</td>
<td>3 mM</td>
</tr>
<tr>
<td>Bacterial Forward Primers (x3)</td>
<td>50 µM</td>
<td>0.075 each (0.225)</td>
<td>150 nM each</td>
</tr>
<tr>
<td>Bacterial Reverse Primers (x3)</td>
<td>50 µM</td>
<td>0.125 each (0.375)</td>
<td>250 nM each</td>
</tr>
<tr>
<td>Bacterial Probes (x3)</td>
<td>50 µM</td>
<td>0.125 each (0.375)</td>
<td>250 nM each</td>
</tr>
<tr>
<td>IAC Forward Primer</td>
<td>50 µM</td>
<td>0.125</td>
<td>250 nM</td>
</tr>
<tr>
<td>IAC Reverse Primer</td>
<td>50 µM</td>
<td>0.125</td>
<td>250 nM</td>
</tr>
<tr>
<td>IAC Probe</td>
<td>50 µM</td>
<td>0.1</td>
<td>200 nM</td>
</tr>
<tr>
<td>IAC Amplicon</td>
<td>100 µM (DF: 10⁻⁸)</td>
<td>1</td>
<td>0.04 pM</td>
</tr>
<tr>
<td>PCR Grade Water</td>
<td>N/A</td>
<td>3.675</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>N/A</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Templet DNA extract</td>
<td>N/A</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

### 2.10.2. Standard Curves of typical Respiratory pathogens

For quantification of the bacterial loads of the three typical respiratory pathogens, standard curves were created for each organism. *S. pneumoniae* ATCC 49619,
H. influenzae NCTC 8468, and M. catarrhalis ATCC 25240 were sub-cultured on appropriate culture media; on Columbia horse blood agar (Spn) and CHOC (Hi and Mc). For each tested organism, 5-7 loopfuls of bacterial colonies were suspended in 1 mL sterile normal saline, to produce a bacterial suspension equivalent to the 4 McFarland standard. Each bacterial suspension was diluted 1:1 in sterile artificial sputum base (Appendix 3.2) then it was split into two 1mL aliquots. The first 1 mL aliquot was used for DNA extraction as per samples (as discussed under section 2.6.1.),

The other 1 mL aliquot was used to determine the bacterial count using the Miles and Misra method (Miles et al., 1938) in which the bacterial suspension is 10-fold serially diluted down to $10^{-7}$ then three 20 µL aliquots from each dilution were plated on appropriate air dried (inside class II laminar flow cabinet) culture medium according to the organism. After the drops were all absorbed into the agar, the plates were incubated overnight at 37°C in the presence of 5% CO₂. Next day the dilution between 10 and 50 colonies, were counted and an average number of colonies per drop is calculated. The exact initial bacterial viable count (CFU/mL) is calculated according to the following equation:

Viable count (CFU/mL) = average number of colonies/drop $\times 50 \times$ dilution factor

Then the DNA extract was serially diluted by 10-fold down to $10^{-6}$. These dilutions were used as standards in the multiplex qPCR in 5 replica to generate a standard curve for each organism (Ch. 3, Figure 3.13). The standards’ dilutions were then aliquoted and stored at -20°C. One standard for each organism was added in all subsequent samples runs; this allowed importation of the standard curve prepared earlier and adjustment of the new run with the standard curve run to overcome inter-run variations. The results of the qPCR were reported as CFU/mL of original sputum sample.

2.10.3. P. aeruginosa and total bacterial load Diplex qPCR

The total bacterial load (copies/µL) and load of P. aeruginosa (CFU/mL) were determined in sputum samples using another Taq-Man® assay diplex qPCR
targeting a 466 bp fragment of 16S rRNA gene (V3-V4 hypervariable regions) using the universal bacterial primers Bact340F and Bact806R and a 65 bp fragment of the \textit{regA} gene in \textit{P. aeruginosa}. The protocol employed has been adopted from two previously published qPCR methods for 16S rRNA (Erb-Downward et al., 2011) and for \textit{P. aeruginosa} (Stressmann et al., 2011) modified and optimized into one diplex qPCR. The IAC was also employed in this method to test for PCR inhibition as have discussed earlier (under section 2.10.1). The specificity of \textit{P. aeruginosa} primers and probe was checked in-silico through running Primer BLAST against NCBI RefSeq Representative Genome Database. (Ye et al., 2012) The specificity, coverage, matches and mismatches of 16S rRNA primers and probes were checked in-silico by running TestPrime and TestProbe on SILVA against ssu-132 reference database. (Klindworth et al., 2013) The primers and probes were ordered from Eurofins Genomics (Germany) the sequences of primers and probes of both targets are given in Table 2.5.

| Table 2.5: Primers and probe sequences of \textit{P. aeruginosa} and total bacterial load Diplex qPCR |
|---------------------------------|---------------------------------|
| **5’-3’ Sequence**              | **Sequence**                   |
| Ps Forward Primer               | TGCTGTTGGCAGCAGGACAT            |
| Ps Reverse Primer               | TTGGTTGGTCAGGCCTCTGTTG          |
| Ps probe                        | [FAM] CAGATGCTTTGCTCAAA [TAM]   |
| q16S rRNA Forward Primer (Bact340F) | TCCTACGGGAAGGCAGGCT            |
| q16S rRNA Reverse Primer (Bact806R) | GGACTACCAGGGTATCTAATCTT       |
| q16S rRNA probe                 | [ROX] CGTATTACCGCGCTGCTGGCAC [BHQ2] |

The master-mix was prepared using Platinum® quantitative PCR Supermix-UDG (Thermo-Fisher Scientific, UK) as shown in Table 2.6. Each sample was diluted 1-in-10 and run in duplicate in addition to 1-in-100 dilution to check for inhibition. Each run also included a negative control and three standards for each target, each was run in triplicate to generate an internal standard curve for each target within each run (Appendix 3.4). The thermo-cycling conditions were 95°C for 10 min...
followed by 40 cycles of 95°C for 15 sec and 60°C for 60 sec as carried out on Qiagen Rotor– gene® 6000 real-time PCR machine (Corbett Research UK, Cambridgeshire, UK). The signal of \textit{P. aeruginosa} was captured on the green channel, 16S rRNA on the orange channel and IAC on the red channel. Load of \textit{P. aeruginosa} was calculated in colony forming unit (CFU) per mL of sputum while the load of 16S rRNA was calculated as copies/ µL.

\textbf{Table 2.6: Master-mix preparation of the \textit{P. aeruginosa} and total bacterial load Diplex qPCR}

<table>
<thead>
<tr>
<th>Reagents</th>
<th>Volume (µL) per reaction</th>
<th>Final Concentration per reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Platinum qPCR supermix</td>
<td>12.5</td>
<td>1x</td>
</tr>
<tr>
<td>Ps forward primer (50 µM)</td>
<td>0.5</td>
<td>1000 nM</td>
</tr>
<tr>
<td>Ps reverse primer (50 µM)</td>
<td>0.5</td>
<td>1000 nM</td>
</tr>
<tr>
<td>Ps probe (10 µM)</td>
<td>0.75</td>
<td>300 nM</td>
</tr>
<tr>
<td>16S forward primer (10 µM)</td>
<td>0.25</td>
<td>100 nM</td>
</tr>
<tr>
<td>16S reverse primer (10 µM)</td>
<td>0.25</td>
<td>100 nM</td>
</tr>
<tr>
<td>16S probe (10 µM)</td>
<td>0.25</td>
<td>100 nM</td>
</tr>
<tr>
<td>IAC forward primer (50 µM)</td>
<td>0.125</td>
<td>250 nM</td>
</tr>
<tr>
<td>IAC reverse primer (50 µM)</td>
<td>0.125</td>
<td>250 nM</td>
</tr>
<tr>
<td>IAC Probe (50 µM )</td>
<td>0.1</td>
<td>200 nM</td>
</tr>
<tr>
<td>SPUD 10^{-8}</td>
<td>1</td>
<td>0.04 pM</td>
</tr>
<tr>
<td>H2O</td>
<td>5.65</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Template DNA</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

\textbf{2.10.4 Standard Curves and Standards preparation of \textit{P. aeruginosa} and 16S rRNA}

\textit{P. aeruginosa} standards were prepared using \textit{P. aeruginosa} NCTC 13437 sub-cultured on Columbia horse blood agar incubated overnight at 37°C in ambient atmosphere as discussed earlier (under section 2.10.2). The load of \textit{P. aeruginosa} was calculated in CFU/mL of sputum.

The 16S rRNA standards were prepared by amplifying the 16s \textit{rrna} gene from DNA extracts obtained from pure cultures of \textit{P. aeruginosa} NCTC 13437, \textit{S. aureus} ATCC 29213, \textit{S. pneumoniae} ATCC 49619, \textit{M. catarrhalis} ATCC 25240,
Chapter 2: Materials and Methods

*H. influenzae* NCTC 8468 and *A. baumannii* clinical isolate using the universal bacterial primers 27F (AGAGTTTGATCMTGGCTCAG) and 1492R (TACGGYTACCTTGTTACGACTT). The PCR products were purified using QIAquick PCR purification kit (Qiagen, UK) as per manufacturer protocol. The cleaned PCR products were then pooled, and the DNA concentration was measured using Qubit™ dsDNA HS kit and Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, UK). The number of gene copies/ µL was then calculated using the following equation:

\[
\text{Copies/ } \mu\text{L} = \frac{\text{DNA Concentration (ng/ } \mu\text{L)} \times 6.022 \times 10^{23}}{660 \times \text{amplicon size} \times 10^9}
\]

\[
6.022 \times 10^{23} \text{ is Avogadro’s number}
\]

\[
660 \text{ (g/mol) is the average molecular weight per base pair and } \frac{1}{10^9} \text{ is g/ng}
\]

\[
\text{amplicon size is } 1466 \text{ bp}
\]

Since the Qubit reading was 7.32 ng/ µL, therefore the number of 16S rRNA gene copies:

\[
\text{Copies/ } \mu\text{L} = \frac{7.32 \times 6.022 \times 10^{23}}{660 \times 1466 \times 10^9} = 4.556 \times 10^9
\]

The 16S rRNA standard was then serially diluted 10-fold to $10^{-7}$. The standards' dilutions of both targets were run in 5 replicates each to produce a standard curve (Ch.3, Figure 3.15).

2.11. Statistics

All studies described here are observational study, the primary outcome was defining the associations between the microbiome parameters (α and β diversity indices, and differential abundances of OTUs/taxa) and: disease status and
treatment regimes. Since there was no prior work published on which formal power calculations could be based on; therefore, the sample size was established on a pragmatic recruitment rate from clinics in each study within the time frame of the doctoral degree.

For continuous data, the significance of the observed differences in the results was tested using the parametric tests: t-test (when comparing two groups) or one way analysis of variance (ANOVA), followed by Tukey post hoc test (comparisons involving more than two groups) for normally distributed variables. The alternative non-parametric tests: Mann-Whitney (when comparing two groups) or Kruskal-Wallis test (comparisons involving more than two groups) were used when the underlying assumptions of the parametric tests were not fulfilled. Paired t-test (parametric) and Wilcoxon Rank test (non-parametric) were used to test the significance in paired observations results.

The distribution of each continuous variable had been checked prior to the selection of the most appropriate significance test. Normal distribution was confirmed when Shapiro-Wilk test was insignificant (p>0.05), and skewness and kurtosis z-scores were within ±1.96 while equal variance assumption was confirmed when the Levene’s test was insignificant (p>0.05).

Fisher exact or Chi squared tests were used to test significance in comparisons involving categorical data as appropriate. The statistical analysis was performed using IBM SPSS Statistics, Version 25.0. (IBM, 2017). Significance was set at p<0.05. (Petrie, 2019)
Chapter 3: Microbiome Analysis Method Optimisation and Evaluation Results

3.1. Introduction

Recent advances in molecular techniques have provided efficient tools for both the detection of individual microorganisms and the characterization of the whole microbial community composition. The NGS high throughput sequencing technologies have offered rapid and cost-effective tools that allow simultaneous sequencing and analysis of many samples with sufficient depth within the same run; this has promoted large microbiome studies. (Aho et al., 2015, Malla et al., 2018) Real-time PCR provides robust, specific and sensitive tools for the detection and quantification of specific marker genes within mixed microbial communities and has been widely used in microbial ecology studies to quantify specific functional or taxonomic genes. The specificity of the qPCR method is determined by the primer design and it is greatly augmented when combined with target specific fluorescently labelled Taqman™ probes. (Smith and Osborn, 2009)

Nevertheless, modern molecular assays have some limitations that need to be considered. The quality and efficiency of nucleic acid extraction varies considerably across the different extraction methods. Within the same method, the efficiency of DNA extraction from different classes of bacteria may also differ depending on the structural characteristics of their cell walls. The performance of the extraction method also depends on the sample type; sputum is a challenging sample type given the consistency and complexity of its matrix, nuclease activity and the heterogeneity of microbial community’s composition. Therefore, the selection of the nucleic acid extraction method is crucial in microbial ecology studies as it can introduce bias in the final outcomes. (Marotz et al., 2017, Terranova et al., 2018, Hang et al., 2014)

Bioinformatics can be another source of bias. The Microbiome Quality Control project showed that differences in the computational pipeline, even on the same data, could lead to large differences in the inferred outcomes at levels from the
species to the phylum. (Allaband et al., 2019) Some elements in the bioinformatic pipeline such as the reference database and taxonomic classifier may introduce bias into the final representation of taxa in the microbiome profiles. (Comeau et al., 2017)

As second-generation sequencing technologies are limited by the length of sequence reads which can be generated; in the case of the Illumina Miseq it is 2x250 bp with the version 2 kits used. (Barb et al., 2016, Illumina, 2020) The entire sequence of 16S rRNA gene cannot be practically sequenced with NGS; therefore, the resolution of microbiome profiling is inherently limited and hardly goes beyond the genus level for most bacterial identification. A big challenge would lie in choosing universal PCR primers which can amplify variable regions of the gene from most of the bacterial 16S rRNA sequences under the constraint of the read length. In addition, the amplified variable regions of 16S rRNA should supply sufficient sequence diversity to maximize the phylogenetic resolution and accurately identify most bacteria throughout the various taxonomic classes providing a reliable representation of the studied microbial community. (Orna et al., 2013, Fuks et al., 2018)

In this chapter, the technical challenges, the specifications of the methods used and their optimisation are presented.

3.2. Methods

3.2.1. Preparation of Mock Communities

Details on the preparation of mock communities are described in Chapter 2, 2.7.

Briefly, mock community A was prepared by mixing equal volumes of approximately equally dense bacterial suspension of three bacteria: *S. pneumoniae* ATCC 49619 (OD 0.3), *H. influenzae* NCTC 8468 (OD 0.27) and *Moraxella catarrhalis* ATCC 25240 (OD 0.31) in 3.5 mL artificial sputum base. DNA was extracted with and without the heat killing step.
Mock community B was a 50-folds diluted mixture of genomic DNA of three bacteria: *S. pneumoniae* ATCC 49619, *H. influenzae* NCTC 8468 and *M. catarrhalis* ATCC 25240 in TE buffer.

Mock community C was a mixture of equal concentrations of genomic DNA of nine bacteria; *S. pneumoniae* ATCC 49619, *H. influenzae* NCTC 8468, *Streptococcus pyogenes* ATCC 19615, *E. coli* ATCC 27853, *K. pneumoniae* NCTC 13438, *S. aureus* ATCC 29213, *A. baumannii* clinical isolate, *P. aeruginosa* clinical isolate and *M. catarrhalis* clinical isolate. All were adjusted to the same final concentration of 13.5 ng/µL in the community mixture.

Mock Communities E and F were prepared by combining equal proportions of approximately equally dense bacterial suspensions of the constituting bacteria in sterile normal saline. The mixture was then aliquoted into 1 mL portions. One portion (Mock F) was heated at 95°C for 30 min prior to the DNA extraction step while mock E was not heat killed. Mock community D was an independent replicate of mock E; it was prepared similarly without the heat killing step. Then the 3 mock communities were simultaneously extracted by LIAISON® Ixt extraction platform and processed similarly through amplifying and sequencing the V3-V4 variable regions of 16S rRNA.

### 3.2.2. Paired Sputum and Saliva Samples

Paired saliva and sputum, spontaneously expectorated, samples were collected from ten patients with confirmed diagnosis of bronchiectasis or COPD in the respiratory clinic in a pilot study. The samples collected were frozen at -70°C within two hours; mean while they were kept at 8°C. Sputum samples were homogenized with Sputasol aliquoted and heated at 95°C for 30min then DNA was extracted from 1 mL aliquots on the automated LIAISON® Ixt extraction platform (Ch2, 2.6.1.).

The saliva samples were first treated with 500 µL of phosphate buffer saline, mixed thoroughly by vortexing and heated at 95°C for 30min. Aliquots of 500 µL of each sample were digested with 10 µL proteinase K at 56°C for 10 minutes then they
Chapter 3: Microbiome Analysis Method Optimisation and Evaluation

were extracted on the automated LIAISON® lxt extraction platform as the sputum samples but they were eluted into 50 µL.

The variable regions V5-V7 of the 16S rRNA were amplified using 785F and 1175R primer set in the ten pairs of sputum and saliva samples. In addition, the variable regions V3-V4 of the 16S rRNA were amplified using 341F and 805R primer set in the ten sputum samples only. Amplicons were sequenced in two Miseq sequencing runs. Bioinformatic analysis was done according to the published microbiome helper workflow through QIIME (Comeau et al., 2017). All samples in the biome table were rarefied at 3000 reads to compare α and β diversity. The impact of some initial quality control steps on the raw sequencing reads such as quality and length filtering and chimera filtering steps were assessed. Details on sequencing, bioinformatics and statistical analyses are described in Ch.2, 2.8, 2.9.

3.2.3. Determination of Limit of Detection (LOD) and Limit of Quantification (LOQ)

Details on the typical respiratory pathogens qPCR, the P. aeruginosa and total bacterial load qPCR and the preparation of qPCR standard are described in Ch.2, 2.10. For the determination of LOD and LOQ, the standards down the dynamic range for each target were 2-fold serially diluted; each dilution was run in ten replicates and a standard curve was then constructed. LOQ was determined at the lowest standard concentration showing a coefficient of variation < 20% between replica; while the LOD was determined at the lowest standard concentration at which all replicates produced positive signals that can be distinguished from the background noise.

3.3. Results and Discussion

3.3.1. Choice of the lysing matrix in sputum homogenisation

The performance of two FastPrep® lysis matrices (A and B) (MP Biomedicals™, Fisher Scientific, UK) in sputum homogenisation during the bead beating step employed in the DNA extraction of mock community A were compared. The first
Chapter 3: Microbiome Analysis Method Optimisation and Evaluation

was FastPrep® lysing matrix A which is composed of garnet matrix and a 1/4 inch zirconium banded sphere (referred to as 'satellite') the second was lysing matrix B which is composed of 0.1 mm silica spheres. In addition, the impact of the heat killing step before the DNA extraction on the quality of extracted DNA and bacterial loads in mock community A was assessed.

Upon measuring the DNA yield in the extracts with NanoDrop™, with lysing matrix A DNA yield was much higher 1358 ng/µL versus 8.8 ng/µL in case of lysing matrix B. However, when the bacterial load was quantified using the typical respiratory pathogens qPCR, in the presence of the zirconium satellite, the DNA was completely sheared and no signal for any of the three bacteria was detected. By removing the zirconium satellite and doing the bead beating with the garnet matrix only, signals for the three bacteria were detected; nevertheless, the signals were still lower than that detected with using the lysing matrix B (silica) (p=0.028 by Wilcoxon Signed Ranks test) (Figure 3.1).

Generally, the heat killing step slightly increased the DNA yield measured by NanoDrop™ in each pair, however, in the qPCR the signals detected were slightly lower (p=0.028 by Wilcoxon Signed Ranks test).

Therefore, lysing matrix B was chosen to be implemented in the DNA extraction protocol for processing the sputum samples in the following studies.

Figure 3.1. Comparison between the performance of FastPrep® lysis matrices A and B in sputum homogenisation with and without heat killing steps on the bacterial loads of the constituting bacteria in mock community A measured by qPCR. The results are the mean of triplicates; the error bars are the standard deviations. L: Living (i.e. extracted without the heat killing step), HK: Heat Killed, Hi: H. influenzae, Mc: M. catarrhalis, Spn: S. pneumoniae.
3.3.2. Mock communities as positive controls for sequencing

Mock B is composed of the genomic DNA extracts of the three typical respiratory pathogens: *S. pneumoniae*, *H. influenzae* and *M. catarrhalis*. Figure 3.2. shows the relative abundance (RA) of each constituent measured by V5-V7 16S rRNA sequencing and qPCR methods in comparison to the concentrations of each bacterium’s DNA concentration in the mixture. There was a good concordance in the RA of *Moraxella* in the 3 profiles; However, *Haemophilus* was slightly overestimated at the expense of the RA of *Streptococcus* in the genus level representation of the sequencing results.

Nevertheless, it is difficult to know exactly which of the three profiles is the true representation due to the variability in the number of copies of 16S rRNA and the genome size. *H. influenzae* for example has more copies of 16S rRNA gene per genome which is six in average in contrast to four copies in average in both *S. pneumoniae* and *M. catarrhalis* strains. (Vetrovsky and Baldrian, 2013) Concerning the DNA concentration, the average genome size of *H. influenzae* is smaller than that of *S. pneumoniae*; the former is 1.8906 Mb while the latter is 2.096 Mb (ENA accession number ERS450003 and GenBank accession number AP018938.1 respectively). Therefore, it is expected that *S. pneumoniae* would yield more DNA and this explains the higher proportion of *Streptococcus* in the DNA concentration representation in Figure 3.2.

![Figure 3.2. Mock community B composition by 16S rRNA sequencing, qPCR and the DNA concentration; g_: genus.](image-url)
Mock C was prepared to contain equal DNA concentrations of nine bacteria. This mock community was used as positive control for sequencing and was included in every sequencing run. Mock C community was used to compare the V3-V4 versus V5-V7 16S rRNA sequencing on the resulting microbiome profile and evaluate the performance of two bioinformatic SOPs through QIIME.

Following the microbiome helper work-flow (Comeau et al., 2017, Douglas, 2017), *Proteobacteria* was under-represented at the phylum level in the V5-V7 representation. Specifically, the RA of *Haemophilus* was greatly diminished: 2% vs the expected 11%; in favour of *Streptococcus* whose RA was almost doubled (41% vs the expected 22%), *Pseudomonas* was completely missed out and *Acinetobacter baumannii* OTU was over-represented (20% vs the expected 11%) whose identity was resolved down to the family level only. This may be attributed to some of the initial quality control steps such as quality filtering and chimera filtering in which 42% of the initial reads had been filtered out. Omitting the chimera filtering step and filtering out the paired reads with quality scores <20 instead of 30 gave a more accurate representation for V5-V7 as shown in Figure 3.3.

On the other hand, V3-V4 16S rRNA presented a more precise and robust representation of the mock community composition both at the phylum and genus levels following the microbiome helper work-flow (Douglas, 2017) (Figure 3.3). A highly consistent representation of the mock community composition was obtained for the independent triplicates over three sequencing runs where the coefficient of variation in the RA of the individual taxa ranged between 11% and 1%. The RA of genus *Streptococcus* was more accurately represented in the V3-V4 representation compared to the former primer set. However, both genus *Haemophilus* in *Proteobacteria* and *Staphylococcus* in *Firmicutes* were overestimated, where their RA was almost twice the expected values. Also, V3-V4 lacked the resolution to differentiate between *Escherichia* and *Klebsiella* in the *Enterobacteriaceae* family unlike in V5-V7 representation in which *Klebsiella* was slightly represented (Figure 3.3). Both variable regions were not able to differentiate between the different *Streptococcus* species: *S. pneumoniae* ATCC 49619 and *S. pyogenes* ATCC 19615.
Chapter 3: Microbiome Analysis Method Optimisation and Evaluation

A. Figure 3.3. Mock community C: Comparison between the performance of the amplification of V5-V7 versus V3-V4 variable regions of 16S rRNA using (785F-1175R) and (341F-805R) primer sets respectively in representing bacterial community composition at (A) the phylum level (B) at the genus level. The reproducibility of the microbiome profile of mock community C using V3-V4 variable regions amplification over three 16S rRNA sequencing runs is demonstrated. Number represent the proportions; g_: genus, f_: family.
Mock communities: D, E and F were prepared to assess the efficiency of the automated DNA extraction platform and the effect of the heat killing step on the microbial community composition representation. Both mock communities: E and F were prepared from the same bacterial suspension mixture and were treated similarly and simultaneously, the only difference was that Mock F was heat killed before DNA extraction. The optical density (OD) of bacterial suspensions that are used in the constitution of the bacterial mixtures for the preparation of mock communities: D and (E/F), and the bacterial viable count are presented in Table 3.1.

Table 3.1: The OD and the bacterial count (BC) by Miles and Misra of the constituting bacteria in mock communities: D and (E/F)

<table>
<thead>
<tr>
<th></th>
<th>Mock D</th>
<th>Mock E &amp; F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OD&lt;sub&gt;600&lt;/sub&gt;</td>
<td>BC (CFU/mL)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>S. pneumoniae</em> ATCC 49619</td>
<td>0.09</td>
<td>3.17E+05</td>
</tr>
<tr>
<td><em>S. parasanguinis</em> clinical isolate</td>
<td>0.09</td>
<td>5.50E+07</td>
</tr>
<tr>
<td><em>S. aureus</em> ATCC 29213</td>
<td>0.12</td>
<td>7.30E+06</td>
</tr>
<tr>
<td><em>H. influenzae</em> NCTC 8468</td>
<td>0.12</td>
<td>1.50E+08</td>
</tr>
<tr>
<td><em>M. catarrhalis</em> ATCC 25240</td>
<td>0.11</td>
<td>7.49E+05</td>
</tr>
<tr>
<td><em>N. subflava</em> clinical isolate</td>
<td>0.1</td>
<td>9.60E+07</td>
</tr>
<tr>
<td><em>P. aeruginosa</em> clinical isolate</td>
<td>0.11</td>
<td>1.85E+07</td>
</tr>
<tr>
<td><em>E. coli</em> ATCC 25922</td>
<td>0.11</td>
<td>7.70E+07</td>
</tr>
<tr>
<td><em>K. pneumoniae</em> NCTC 13438</td>
<td>0.1</td>
<td>1.10E+08</td>
</tr>
<tr>
<td><em>A. baumannii</em> clinical isolate</td>
<td>0.11</td>
<td>1.20E+08</td>
</tr>
</tbody>
</table>

* Optical density measured at wave length 600 nm

As shown in Figure 3.4, DNA extraction protocol using the automated LIAISON® Ixt extraction (Ch.2, 2.6.1.) was an efficient method to extract DNA from both Gram positive and Gram negative bacteria in the mock community mixture without biases. The heat killing step did not cause a significant impact on the resulting microbiome profile it produced a more accurate representation for most of the taxa apart from *Haemophilus* which was overestimated; nevertheless, the primer set used to amplify the 16S rRNA V3-V4 variable regions could be contributing in this bias as it was demonstrated in Figure 3.3.
Figure 3.4. Mock community D, E and F microbiome profiles used to assess the efficiency of the automated DNA extraction platform and the effect of the heat killing step on the bacterial community composition representation, both mock communities: D and E were extracted from living bacterial mixture whereas Mock community F was heat killed before the DNA extraction step. g_: genus, f_: family.
3.3.3. Negative controls in sequencing

To be able to evaluate the potential contamination, a negative extraction control was run alongside each batch of sputum samples processed and was treated the same throughout the whole process from the Sputasol treatment to sequencing. Another no-template PCR control (water) was also run alongside each batch of samples amplified and it was treated the same throughout the whole process from PCR amplification step to sequencing. However, during library preparation, all the negative controls of the same kind which were amplified using the same barcoded primers set, were combined; to save slots for samples in the sequencing run.

Most negative controls, except the extraction negative control in run 2, showed very few sequencing reads which ranged between 2 to 436 reads, these represented <0.75% of the mean number of reads of the clinical samples in their corresponding sequencing runs. Most of the taxa recovered in these negative controls were main constituents of the microbiome profiles of sputum samples which were processed along-side and pooled together in the same library for 16S rRNA sequencing (Figure 3.5). In run 2 which included the samples from the HIV cohort (Chapter 7), there was significant contamination probably introduced during sputum processing with a bacterium belonging to the family Burkholderiaceae; however, this contamination did not impact the key results in this study; as shown later (Ch.7, Figure 7.6.).

In the mock communities, < 0.5% of the reads (range 261 – 2 reads/sample) were false positives. These were classified as Neisseria, Prevotella, Veillonella, Rothia, Actinomyces, and Granulicatella. The source of these taxa can be either contamination during library preparation steps or cross contamination between the multiplexed samples within the same library due to trivial bleed through across the barcoded primers especially that these taxa are frequently observed with high RA in the sputum samples which were multiplexed with the controls in the sequencing libraries.
Figure 3.5. The number of reads, their percentage, and the microbiome profiles of the negative controls of sequencing. NCpcr: No-template 16S rRNA PCR negative control, NCext: Negative controls during samples processing and DNA extractions, R: run. g_: genus, f_: family.
3.3.4. Paired sputum and saliva samples

Figure 3.6 compares the microbiome profiles of the individual sputum samples, sequencing the 16S rRNA variable regions V5-V7 presented richer bacterial composition; this was confirmed by the number of observed OTUs which was significantly higher in V5-V7 representation compared to that of V3-V4 presentation (p=0.007) for the same samples (Figure 3.7.A). Also, the α-diversity indices Shannon and PD-whole tree were slightly but not significantly higher in the V5-V7 representation (Figure 3.7). Nevertheless, V5-V7 was more vulnerable to the stringent quality control filtering steps implemented in the microbiome helper workflow. As was noted with Mock C community, some taxa such as genus *Pseudomonas* which was dominant in the V3-V4 presentation of microbiome profiles of sputum samples Sp1, Sp4, Sp6 and Sp7 were greatly diminished in the V5-V7 presentation. The presence of *P. aeruginosa* was confirmed in these sputum samples in addition to Sp2 and Sp10 by specific qPCR which showed that the load of *P. aeruginosa* ranged between 8.83 and 6.96 log_{10} CFU/mL (Figure 3.8).

Repeating the bioinformatic analysis but with a reduced threshold of quality filtering of the raw sequencing reads to Q20 rather than Q30 and omitting the length filtering, and the chimera filtering steps (referred to here as the alternative bioinformatic pipeline) gave a better representation of the relative abundances of the taxa that agrees with the V3-V4; however, this increased the gap in the richness and alpha diversity between both representations such that all richness and α-diversity indices were significantly higher (p<0.01 by paired-test).

In sputum sample Sp2, although the microbiome profile in both V3-V4 and V5-V7 representations was dominated mainly by the genus *Haemophilus* (RA around 90%) and *Pseudomonas* which was detected with RA 2.5%. Specific qPCR methods found the loads of these two bacteria to be 7.65 and 8.37 log_{10} CFU/mL respectively (Figures 3.8 and 3.12.) which indicates that 16S rRNA sequencing method may have underestimated the relative abundance of genus *Pseudomonas* in samples.
In Figure 3.9, the principal coordinate (PCoA) analysis of Bray Curtis \(\beta\)-diversity index shows that the samples significantly clustered with respect to the amplified and sequenced 16S rRNA variable regions whether being the V3-V4 or V5-V7 (\(p=0.001\) by both PERMANOVA and ANOSIM). More importantly, the bioinformatic pipeline was an important covariate in samples clustering (\(p=0.001\) by ADONIS). Therefore, to be able to compare microbiome data from different cohorts, samples should not only be processed similarly, but also it is important to use one primer set for the amplification of variable regions of the 16S \textit{rrna} gene and to analyse all samples within the same bioinformatic pipeline.
Figure 3.6. Microbiome profiles of paired sputum and saliva samples from ten patients (A) V3-V4 representation of the microbiome profiles of the ten sputum samples (the bioinformatics done through the microbiome helper pipeline) (B) V5-V7 representation of the microbiome profiles of the ten sputum/saliva pairs through the microbiome helper bioinformatic pipeline (C) V5-V7 representation of the microbiome profiles of the ten sputum/saliva pairs through the alternative bioinformatic pipeline. Sa: Saliva, Sp: Sputum, g_: genus, f_: family.
Figure 3.7. Comparing the richness and $\alpha$-diversity indices of the microbiome profiles in the ten sputum samples between the variable regions: V3-V4 (green) and V5-V7 (blue) 16S rRNA sequencing.

(A) Number of observed OTUs was significantly higher in the V5-V7 representation ($p=0.007$ by paired t-test)
(B) Shannon $\alpha$-diversity index ($p=0.057$ by paired t-test)
(C) PD-whole tree ($p=0.66$ by paired t-test)
Figure 3.8. *P. aeruginosa* loads (log$_{10}$ CFU/mL) in the ten sputum samples measured by qPCR, error bars: SEM; Sp: Sputum sample.
Figure 3.9. PCoA three-dimensional (3D) Plot of Bray Curtis β-Diversity showing the clustering of the ten pairs of saliva and sputum samples based on the sequenced 16S rRNA variable regions and the used bioinformatic pipeline. Each dot represents a microbiome profile of a sample.

**red:** Sputum samples (Sp), V3-V4 representation, microbiome helper bioinformatic pipeline  
**Blue:** Sputum samples (Sp), V5-V7 representation, microbiome helper bioinformatic pipeline  
**Green:** Sputum samples (Sp), V5-V7 representation, the alternative bioinformatic pipeline  
**Orange:** Saliva samples (Sa), V5-V7 representation, microbiome helper bioinformatic pipeline  
**Yellow:** Saliva samples (Sa), V5-V7 representation, the alternative bioinformatic pipeline.
Comparing each saliva and sputum pair in V5-V7 presentation processed through the alternative bioinformatic pipeline, the richness and diversity of microbial communities was generally higher in the saliva compared to corresponding sputum samples. The mean (SD) of Shannon α-diversity index of saliva samples was 4.19 (0.48) versus to 2.76 (1.19) in sputum samples (mean difference in each saliva and sputum pair was 1.43, \( p=0.003 \) by paired t-test) (Figure 3.10). Other richness and alpha diversity indices; e.g. number of observed OTUs, chao1, PD whole tree, and Fisher alpha showed similar trend but were not statistically significant \( (p>0.05 \) by paired t-test).

In the PCoA of the unweighted UniFrac which is more sensitive to the presence and absence of taxa rather than the weight of their relative abundances, the sputum samples were separated from the saliva samples (PERMANOVA \( p\text{-value}=0.031 \), ANOSIM \( p\text{-value}=0.019 \)); however, the saliva and sputum pairs of the 5\(^{th}\), 9\(^{th}\), 8\(^{th}\) and 10\(^{th}\) patients were closer to each other rather than to the sample’s type cluster i.e. whether being saliva or sputum. This shows that many taxa were shared between the saliva and sputum samples in these pairs. This was also evident in the microbiome profiles of the paired sputum and saliva samples in Figure 3.6.

Similarly, in the former PCoA analysis of Bray Curtis β-diversity index (Figure 3.9) sputum samples were separated from the corresponding saliva samples (PERMANOVA \( p=0.003 \), ANOSIM \( p=0.001 \)). The sputum samples with distinctive microbiome profiles in which specific pathogenic taxa dominated such as the case in Sp1, Sp2, and Sp3 clustered despite the sequenced variable regions and the applied bioinformatic pipeline.

In PCoA and cladogram of weighted UniFrac (Figure 3.11) all the saliva samples from ten individuals were clustered together and they were distinct from their corresponding sputum pairs in most of the cases (60%). The four sputum samples: Sp5, Sp8, Sp9, Sp10 which were clustered with the saliva samples are those which were dominated with streptococci (Figure 3.6.C). From the qPCR results, two of these sputum samples Sp5 and Sp9 contained a mean of 5.6 and 3.5 log10
CFU/mL of *S. pneumoniae* respectively (Figure 3.12). But the 16S rRNA sequencing could not distinguish between the pathogenic *S. pneumoniae* and the commensal streptococci in samples. Nevertheless, looking more closely at this cluster in the cladogram (Figure 3.11), apart from the 8th patient, the sputum and saliva pairs of 5th, 9th and 10th patients were not adjacent i.e. not closely related (Figure 3.11).

Although the microbiome profiles of the paired saliva and sputum samples of the 5th, 8th, 9th and 10th patients shared the top major taxa, especially in the 5th and 8th patients, there were considerable differences in the RA of the common taxa. Furthermore, some minor taxa were detected in the saliva were completely absent in the corresponding sputum and vice-versa (Figure 3.6.C). For example; the 5th sputum sample showed 20% higher RA of *Streptococcus*; but 19% less RA of *Rothia* and almost similar *Veillonella* RA (around 9%) compared to the corresponding saliva sample. On the other hand, Sp10 had almost similar Streptococcus RA (around 50%) but 22% higher RA of *Rothia* and 12% less *Veillonella* to the corresponding saliva sample Sa10. In contrast, the Sp7 showed 15% lower RA of *Streptococcus, Rothia* and *Veillonella* compared to the corresponding saliva sample Sa7. The qPCR results confirmed that Sp5 contained 2.5 log higher load of *S. pneumoniae* compared to the corresponding saliva sample and the saliva Sa7 contained 2.3 log_{10} CFU/mL *S. pneumoniae* which was completely absent in the corresponding sputum sample (Figure 3.12.). *Pseudomonas* was detected in the following sputum samples; Sp1 (34%), Sp2 (2.5%), Sp4 (89%), Sp6 (86%) and Sp7 (61%), their corresponding saliva samples revealed very low relative abundance (<2%) of *Pseudomonas* in the oral microbiome. *Pseudomonas*, was detected in the Sp5 and Sp10 with RA <2% but was completely absent in the corresponding saliva.
Chapter 3: Microbiome Analysis Method Optimisation and Evaluation

Figure 3.10. Comparison of the richness and diversity of the bacterial community in the ten sputum and saliva paired samples (A) Richness estimated by the number of observed OTUs (p=0.082, by paired t-test) (B) α-diversity measured by Shannon diversity index (p=0.003, by paired t-test) (C) PCoA plot of unweighted UniFrac β-diversity (PERMANOVA p=0.031, ANOSIM p=0.019). Blue: sputum samples (Sp); n=10, yellow: saliva samples (Sa); n=10. V5-V7 16SrRNA amplicon sequencing; bioinformatic through the alternative pipeline.
Figure 3.11. PCoA and cladogram based on weighted UniFrac β-diversity (PERMANOVA p=0.002, ANOSIM p=0.001) showing separation between the sputum samples (Sp, Blue, n=10) and saliva samples (Sa, yellow, n=10); V5-V7 16SrRNA amplicon sequencing; bioinformatics through the alternative pipeline.
In qPCR results whenever a typical respiratory pathogen was detected in both sputum and saliva pair a trend was observed in which the bacterial loads were higher in the sputum samples compared to the corresponding saliva pairs (Figure 3.12). *H. influenzae* was detected in the sputum of the 8th patient but not in the corresponding saliva. Also, both *H. influenzae* and *S. pneumoniae* were detected in the sputum of the 9th patient but not in the corresponding saliva from the same patient. On the other hand, *S. pneumoniae* and *H. influenzae* were detected in the saliva of the 3rd and 7th patients respectively but not the corresponding sputum pairs.

**Figure 3.12.** Comparison between the bacterial loads (log_{10} CFU/mL) of the three respiratory pathogens: *H. influenzae* (Hi: pink), *M. catarrhalis* (Mc: green) and *S. pneumoniae* (Spn: navy blue) in the ten sputum and saliva paired samples measured by qPCR, Sa: Saliva, Sp: Sputum, error bars: SEM.
3.3.5. Performance based specifications of typical respiratory multiplex qPCR

The standard curves and amplification plots of *S. pneumoniae*, *H. influenzae* and *M. catarrhalis* are presented in Figure 3.13. The efficiencies were 96%, 98% and 97% respectively and the correlation coefficients (R) were >0.99 and the residual sum of squares (R²) were >99.5% for the three bacteria. Figure 3.14 shows the robustness, repeatability and range of the qPCR method using three independent sets of laboratory prepared standards (Chapter 2.10.2). The sensitivity of the method expressed by the LOQ which is defined as the lowest amount of bacteria in a sample which can be quantitatively determined with a suitable accuracy and precision and the LOD which is defined as the lowest amount of bacteria in a sample which can be detected are presented in Table 3.2. (ICH, 1995, Oshiro, 2013)

The specificity of the primers and probes used were checked in-silico through running Primer BLAST against NCBI RefSeq Representative Genome Database (Ye et al., 2012) and every pair of primers and probes picked only their intended corresponding bacterial targets (reports of in-silico analysis are included in the Appendix 3.7).

Table 3.2: The limit of quantification and limit of detection of typical respiratory qPCR

<table>
<thead>
<tr>
<th></th>
<th><em>S. pneumoniae</em></th>
<th><em>H. influenzae</em></th>
<th><em>M. Catarrhalis</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LOQ</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25,000 CFU/mL</td>
<td>4.4 log₁₀ CFU/mL</td>
<td>20,000 CFU/mL</td>
<td>2500 CFU/mL</td>
</tr>
<tr>
<td>4.3 log₁₀ CFU/mL</td>
<td>250 CFU/reaction</td>
<td>4.3 log₁₀ CFU/mL</td>
<td>3 log₁₀ CFU/mL</td>
</tr>
<tr>
<td>≡ 0.5 pg/µl</td>
<td>≡ Ct 31</td>
<td>≡ 0.13 pg/µl</td>
<td>≡ Ct 34</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>≡ 0.02 pg/µl</td>
</tr>
<tr>
<td><strong>LOD</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3700 CFU/mL</td>
<td>3.57 log₁₀ CFU/mL</td>
<td>1000 CFU/mL</td>
<td>500 CFU/mL</td>
</tr>
<tr>
<td>3 log₁₀ CFU/mL</td>
<td>13 CFU/reaction</td>
<td>2.7 log₁₀ CFU/mL</td>
<td>7 CFU/reaction</td>
</tr>
<tr>
<td>≡ 0.07 pg/µl</td>
<td>≡ 0.006 pg/µl</td>
<td>≡ 0.003 pg/µl</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 3: Microbiome Analysis Method Optimisation and Evaluation

*S. pneumoniae*
A. Amplification plot

*B. Standard Curve*

*H. influenzae*
C. Amplification plot

*D. Standard Curve*

*M. catarrhalis*
E. Amplification plot

*F. Standard Curve*

Figure 3.13. The amplification plots (A,C and E) and the standard curves (B,D and F) of qPCRs for typical respiratory pathogens.
Chapter 3: Microbiome Analysis Method Optimisation and Evaluation

A. *S. pneumoniae*

![Graph showing robustness and range of *S. pneumoniae*](image)

B. *H. influenzae*

![Graph showing robustness and range of *H. influenzae*](image)

C. *M. catarrhalis*

![Graph showing robustness and range of *M. catarrhalis*](image)

Figure 3.14. Robustness and range of typical respiratory pathogens qPCR using three independent sets of laboratory prepared ten-fold serially diluted standards of (A) *S. pneumoniae* (B) *H. influenzae* (C) *M. catarrhalis*; each set is represented by different line shape, colours represent different bacterial concentrations.
3.3.6. Performance based specifications of *P. aeruginosa* and total bacterial load duplex qPCR

The standard curves and amplification plots of *P. aeruginosa* and V3-V4 of 16S rRNA are presented in Figure 3.15. The efficiencies were 94% and 101% respectively and the correlation coefficients (R) were >0.99 and the residual sum of squares (R²) were >99.5% in both. The 16S rRNA qPCR was tested against *S. pneumoniae, H. influenzae, M. catarrhalis, P. aeruginosa, S. aureus, K. pneumoniae, Ac. baumannii* and Bacillus Calmette Guerin (BCG) and it amplified 16S rRNA gene from all. Both the LOD and LOQ of *P. aeruginosa* qPCR was found to be 700 CFU/mL which is equivalent to 2.85 log10 CFU/mL, 10 CFU/reaction and Ct-value of 31.

The specificity of *P. aeruginosa* primers and probe was checked in-silico through running Primer BLAST against NCBI RefSeq Representative Genome Database (Appendix 3.7 G-H). (Ye et al., 2012) The coverage of 16S rRNA primers and probes were checked in-silico by running TestPrime and TestProbe on SILVA against ssu-132 reference database (Klindworth et al., 2013) (reports are included in the Appendix 3.6).
Chapter 3: Microbiome Analysis Method Optimisation and Evaluation

113

**P. aeruginosa**
A. Amplification plot

![Amplification plot for P. aeruginosa](image1)

B. Standard curve

![Standard curve for P. aeruginosa](image2)

16S rRNA
C. Amplification plot

![Amplification plot for 16S rRNA](image3)

D. Standard curve

![Standard curve for 16S rRNA](image4)

![Log values](image5)

Figure 3.15. The amplification plots (A,C) and the standard curves (B,D) of qPCR for *P. aeruginosa* and total bacterial load through measuring V3-V4 of 16S qPCR.
3.3.7. Comparison between qPCR and 16S rRNA sequencing

With short read sequencing technologies, 16S rRNA sequencing can reliably identify OTUs to the genus level; however, in a few cases some sequences could be resolved to the species level. For example, within genus *Haemophilus* some sequences were classified as *H. influenzae*. The specificity of multiplex qPCR was determined by in-silico analysis of the primers specificity and validated by screening against a range of airway bacteria while the sensitivity was expressed by the LOD determined through the calibration curves using prepared standards. Here the sensitivity, specificity and accuracy of 16S rRNA sequencing were determined by comparing its results versus qPCR results. Sensitivity in this context is defined as true positive rate and specificity as the true negative rate. (Lalkhen and McCluskey, 2008) While accuracy was defined as true positive and true negative rate. (Baratloo et al., 2015) The sensitivity, specificity and accuracy of the 16S rRNA sequencing in detecting *P. aeruginosa*, *H. influenzae* and *M. catarrhalis* are presented in Table 3.3. Such comparisons would not be valid in the case of *S. pneumoniae* because all sputum samples contain significant quantities of commensal streptococci which cannot be distinguished from *S. pneumoniae* using 16S rRNA.

<table>
<thead>
<tr>
<th>taxa</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Pseudomonas</em></td>
<td>90.5%</td>
<td>77.2%</td>
<td>80.7</td>
</tr>
<tr>
<td><em>Moraxella</em></td>
<td>82.5%</td>
<td>84.1%</td>
<td>83.9</td>
</tr>
<tr>
<td><em>Haemophilus</em></td>
<td>100%</td>
<td>0%</td>
<td>41.2%</td>
</tr>
<tr>
<td><em>H. influenzae</em></td>
<td>92.7%</td>
<td>81.1%</td>
<td>85.8%</td>
</tr>
</tbody>
</table>

While qPCR methods were optimized to give absolute quantitative data, 16S rRNA sequencing produces semi-quantitative data represented by the relative abundance (RA) which is the proportion of sequences that were classified within one OTU relative to the total number of sequence reads representing the whole sample. The correlation between the sequencing results of V3-V4 variable regions of 16S rRNA gene against qPCR results for *P. aeruginosa*, *H. influenzae*, *M. catarrhalis* and *S. pneumoniae* are presented in Figures 3.16.
Chapter 3: Microbiome Analysis Method Optimisation and Evaluation

Figure 3.16. Correlation between the Relative Abundance (RA) results of V3-V4 16S rRNA sequencing and the quantitative loads of qPCR results (CFU/mL) in all tested sputum samples (n=182) (A) Correlation between relative abundance of genus Haemophilus and load of H. influenzae (Hi) (B) Correlation between relative abundance of H. influenzae OTUs and load of H. influenzae (C) Correlation between relative abundance of genus Moraxella and load of M. catarrhalis (Mc) (D) Correlation between relative abundance of genus Pseudomonas and load of P. aeruginosa (Pa). ρ: Spearman’s rho Correlation.

3.4. General Discussion

In this chapter, the technical challenges and the method specifications of microbiome analysis were investigated to understand the method performance and limitations. In the pilot study, ten pairs of sputum and saliva from ten patients with COPD and bronchiectasis were collected with the aim of optimising the protocol of sputum sample processing, sequencing, the bioinformatic pipeline and assessment of the impact of microbial contamination in sputum from the upper respiratory tract.
The 16S rRNA gene is around 1600 bp long and it is composed of nine variable regions flanked by highly conserved regions. (Chakravorty et al., 2007) Due to NGS read length limitations only a subset of the gene can be sequenced imposing a challenge on the choice of the primer set that would provide the desired universality in amplifying a specific region from most of the 16S rRNA sequences and at the same time the targeted hypervariable regions should offer a sufficient sequence diversity that enables the differentiation of the various bacteria present. (Fuks et al., 2018)

However, no single hypervariable region can distinguish all bacteria. (Chakravorty et al., 2007) Most recent studies have focused on universal 16S rRNA primer sets for the amplification of variable regions including V4 region either alone, or in combination of V3 or V5 for Illumina paired-end sequencing platforms. (Caporaso et al., 2011, Mizrahi-Man et al., 2013, Sinclair et al., 2015, Walters et al., 2016) This is because the V4 region has been shown to be the best to capture most of the information in the full 16S rRNA gene and to have the best sensitivity for bacterial and phylogenetic analyses. (Yang et al., 2016, Pettengill and Rand, 2017)

Nevertheless, V4 may not provide good resolution at the lower taxonomic levels (species). (Pettengill and Rand, 2017) On the other hand, V3 were shown to be the most suitable for distinguishing all bacterial species at least to the genus level except for closely related Enterobacteriaceae; whereas, V6 could distinguish among most bacterial species except for Enterobacteriaceae. (Chakravorty et al., 2007)

Upon comparing the two primer sets: 785F and 1175R (recommended in the protocol of UCL Genomics institute) which amplifies the V5-V7 hypervariable regions of the 16S rRNA gene and 341F and 805R (suggested by Prof. Spratt’s research group at UCL Eastman Dental institute) which amplifies the V3-V4 variable regions in representing the microbiome profiles of mock community “C” (Figure 3.3) and the ten sputum samples in the pilot study (Figure 3.6), different results were obtained from the same starting material depending on the targeted variable regions of 16S rRNA. Although V5-V7 representation showed more diversity, it was less sensitive to some respiratory pathogens such as P.
aeruginosa. On the other hand, V3-V4 provided a better representation of the taxa that agrees with what expected in mock community “C”; in addition, it was robust and reproducible over three independent sequencing runs. Similarly, Mizrahi-Man et al. suggested that the most effective study design utilized paired end sequencing of V3 or V4 amplicons. (Mizrahi-Man et al., 2013) The HMP consortium found that the data produced by the V3–V5 primers was more accurate compared to that generated using two other primer pairs V1–V3 and V6–V9. (Jumpstart Consortium Human Microbiome Project Data Generation Working Group, 2012)

Including positive and negative controls is invaluable in all scientific methods. Negative controls were included in every sequencing run to reveal and evaluate potential contamination from the laboratory environment, reagents which potentially can contain microbial DNA (de Goffau et al., 2018), possible cross-contamination between samples and the bleed through during sequencing from the indexed primers. (Hornung et al., 2019) As shown in Figure 3.5, the contamination was trivial and negligible in most runs as the negative controls showed <0.75% of the mean number of reads representing the sputum samples of interest.

Mock communities have been used as positive controls for sequencing in the present work. (Hornung et al., 2019) Mock communities A, D, E and F were mixtures of bacterial suspensions and were used to assess the initial steps of sputum homogenisation and DNA extraction. As shown in Figure 3.1 lysing matrix B in the bead beating step performed better than the rest. The risk of destroying the integrity of the nucleic acids of some microorganisms in the microbial community by the heat killing step that is integrated in the workflow of sputum processing protocols implemented in the Centre for Clinical Microbiology (CCM) has been evaluated. Nevertheless, in both Figure 3.1 and 3.4 the heat killing step did not cause significant differences. The heat killing step is the standard of practice in processing sputum in diagnostic settings due to the risk of the presence of containment level three organisms, such as Mycobacterium tuberculosis in sputum. Therefore, it was decided to implement the heat killing step in the applied workflow by heating the pellet after the centrifugation step to reduce the DNA loss
in the discarded supernatant before the sputum samples can be transferred out to containment level two laboratory for DNA extraction. The DNA extraction protocol applied used the automated LIAISON® lxt extraction platform which is based on magnetic beads extraction technology was shown to be an efficient method for metagenomic DNA extraction from mixtures of Gram positive and negative bacteria without significant biases and yielded sufficient high quality DNA from sputum samples (Figures 3.4 and 3.6)

Mock communities: B and C are mixtures of genomic DNA and were used to control the sequencing and bioinformatic processes. This type of mock community helps to distinguish the bias introduced by PCR amplification. Mock B was a simpler combination of genomic DNA of three bacterial species (three genera) while mock C was a combination of nine bacterial species (representing eight genera) (Figures 3.2 and 3.3 respectively). As demonstrated there were some biases in the representation of the relative abundances of some taxa compared to what was expected. In the literature, three prime reasons have been highlighted through in silico and in-vitro experimentation that may explain such biases. First, the varying number of copies of 16S rRNA gene per genome in various bacterial species/strains that ranges from one up to 15. (Vetrovsky and Baldrian, 2013) Second, some sequences may amplify more easily or poorly than average this is due to various primer binding efficiencies at the annealing sites or various amplification efficiencies due to possible downstream secondary structure formation hindering polymerase extension, and leading to misrepresentations of different sequences in the final pool. Third, not all taxonomic groups have the same degree of resolution in each hypervariable region; therefore, even if amplification is successful, downstream clustering into operational taxonomic units (OTUs) (at a set identity level of 97%) may be limited in some groups. (Comeau et al., 2017) Nevertheless, mock community C which was included in every sequence to control for inter-run variations revealed that V3-V4 16S rRNA sequencing was highly reproducible over three independent runs (Figure 3.3).

In Bioinformatics, sequencing artefacts such as chimeric sequences and sequence errors can inflate estimates. The current standard practice is to filter sequencing
reads prior to clustering. (Mizrahi-Man et al., 2013) Nevertheless, rigorous initial quality control steps could waste meaningful data as demonstrated in (Figures 3.3, 3.6 and 3.9) in which *Pseudomonas* sequences were missed out in run 1 V5-V7 representation. This highlights the importance of running the sequencing data of the mock community along-side the samples in the bioinformatic pipeline as positive controls in bioinformatics.

Principally alpha diversity depends on sampling effort; the number of types of organisms observed increases as the sample size increases until all types has been sampled. (Hughes et al., 2001) Therefore, it is important to normalise the OTU tables to the same depth to allow biodiversity comparisons across samples. Rarefaction is a statistical interpolation method of normalising the sequence reads of all samples to the same depth then randomly discarding reads from larger samples until the number of remaining samples is equal to the set threshold. (Gotelli and Chao, 2013) Despite the recent debate around the validity of rarefaction because of wasting a considerable amount of the sequence data, it is still the standard approach to eliminate biases that may be introduced due to variable library sizes. (McMurdie and Holmes, 2014, Willis, 2019) In the presented work, all the samples in the OTU tables were rarefied to be equivalent to the smallest sample in the corresponding sequenced library, the range was of 4000 to 20,000 reads per sample. It was suggested that 2,000 Illumina single-end reads are sufficient to recapture the same relationships among samples that we observe with the full dataset. (Lozupone et al., 2010, Kuczynski et al., 2011, Caporaso et al., 2011)

There is always concern over the risk of sputum contamination with the oropharynx microbiota. Nevertheless, sputum is the simplest and the least invasive way to sample the respiratory system. In the pilot study estimating clustering through the phylogenetic based ß-diversity weighted UniFrac, the samples were grouped in two main clusters (Figure 3.11); one included all saliva samples in addition to four sputum samples and the second cluster contained six sputum samples which were dominated by one or two pathogenic taxa such *Pseudomonas*, *Enterobacteriaceae*, *Haemophilus* or *Moraxella* (Figure 3.6). Similar results were
obtained in a previous study comparing paired sputum microbiome to oral wash. (Pragman et al., 2019) The four sputum samples which are clustered with the saliva samples shared the top taxa as the corresponding saliva; they were mainly dominated with *Streptococcus*. However, there were remarkable differences in the relative abundance of the shared taxa between the paired saliva and sputum. No consistent trend (whether being increase or decrease) in the relative abundance of the common taxa in the sputum compared to the corresponding saliva was observed to define the direction of contamination. Also, three of these sputum samples were not closely related to their corresponding saliva pairs in the cladogram (Figure 3.11). In addition, some bacteria which were detected in the saliva were completely absent in the corresponding sputum and vice versa.

The possibility that the sputum samples are contaminated with the saliva cannot be ruled out; especially in the samples clustered with the saliva. However, it should be taken into account that similarity between lung and oral microbiota is evident from the previous lung microbiome studies; especially in healthy status, and that the presence of the observed shared bacterial taxa in sputum is not peculiar and were reported in other lower respiratory tract samples such as bronchoalveolar lavage in other studies as discussed earlier in Chapter 1.3 (Faner et al., 2017, Bassis et al., 2015, Charlson et al., 2011, Erb-Downward et al., 2011, Morris et al., 2013)

In a previous study, the respiratory microbiome has been compared in four respiratory samples: sputum, bronchial aspirates, BAL and bronchial biopsies from the six patients with moderate COPD, found that sputum showed significantly lower diversity compared to the other samples, while BAL and bronchial biopsies were very close in the microbiome; nevertheless, the core microbiome taxa were shared between the samples (reference Figure 4 and Suppl table 1 in (Cabrera-Rubio et al., 2012)) In literature, sputum has been used in many studies to investigate the lower respiratory microbiome in cystic fibrosis, bronchiectasis, COPD and asthma. These studies revealed significant and relevant associations with biologically and clinically meaningful indices such as: severity of illness, airway inflammation, and antibiotic exposure. (Tunney et al., 2013, Dickson et al., 2014b, Huang et al., 2014,
Rogers et al., 2014b, Zhang et al., 2016) This means that even if oropharyngeal microbial contamination is present in the sputum, it still does not obscure the meaningful microbial signal in sputum from lower respiratory tract.

*S. pneumoniae, H. influenzae,* and *M. catarrhalis* conventionally have been classified as typical respiratory pathogens as they are the most common causes of community acquired pneumonia. These organisms normally reside harmlessly within the human nasopharynx. (Sulikowska et al., 2004) These organisms have been frequently isolated and cultured from lower respiratory samples in chronic lung conditions at stable state as well as exacerbations. (Bisgaard et al., 2007, De Schutter et al., 2012, Wu et al., 2014) Numerous studies have demonstrated the implication of these organisms in COPD (Patel, 2002, Garcha et al., 2012) and asthma (Hilty et al., 2010, Hans Bisgaard et al., 2011) exacerbations. Another problematic bacterium is *P. aeruginosa* which has been associated with progressive chronic lung disease conditions; COPD, bronchiectasis and cystic fibrosis, and it has been implicated in chronic recurrent respiratory infections. (Yum et al., 2014) Therefore these organisms are regarded as the common airway bacterial colonisers in chronic lung disease.

Since, 16S rRNA sequencing is a semi-quantitative method in which the abundance of the taxa is expressed in terms of relative abundance. Therefore, an increase or decrease in relative sequence abundances does not necessarily mean that the corresponding bacterial group occurs at higher or lower cell densities. Furthermore, in most cases, OTUs can be reliably classified to the genus level only; therefore, 16S rRNA sequencing could not differentiate between various species belonging to the same genus. Nevertheless, the airways are inhabited by various commensal species which belongs to genus *Streptococcus* and *Haemophilus*. Therefore, qPCR provide unparalleled specificity and sensitivity to target sequences present within a mixed community background; it may be a more appropriate method to detect specific pathogenic bacteria. (Smith and Osborn, 2009) The typical respiratory bacteria multiplex qPCR and *P. aeruginosa* qPCR provide an easy, rapid, specific, and sensitive methods for estimating the loads and prevalence of these pathogenic respiratory bacteria in our cohorts. Garcha et
al. have previously showed that typical respiratory bacteria multiplex qPCR is more discriminatory than culture based techniques which has diagnostic cut-off $10^5$ CFU/mL at detecting these three airway bacteria during acute infections. (Garcha et al., 2012) The 16S rRNA qPCR was used in the present work to measure the total bacterial density and compare it between samples. It amplifies V3-V4 variable regions of the 16S rRNA gene using the universal primers 340F and 806R. However, one should be cautious in interpreting the results as the gene’s copy number is not equivalent to the bacterial cell count as the number of copies of 16S rRNA gene varies among the various bacterial species. (Smith and Osborn, 2009, Vetrovsky and Baldrian, 2013)

To sum up, 16S rRNA sequencing is a very useful semi-quantitative method to study microbial ecology in the human body and reveal poly-microbial signatures that are associated with health and disease condition. Thanks to the great advent in sequencing technology which made microbiome research feasible. However, it is important to be aware of the technology’s limitations and the sources of biases. Standardisation of the whole process from DNA extraction to bioinformatic analysis and incorporation of appropriate positive and negative controls throughout the process are indispensable to allow meta-analysis of microbiome data from various sources. Quantitative PCR provides specific and sensitive complementary tool to accurately quantify specific targets of interest within mixed microbial community in the background. Sputum is a non-invasive respiratory sample that reflects significant microbial signatures in the lower respiratory tract and can be used to study the changes in the airway microbiome associated with chronic lung disease.
Chapter 4: Dynamics of Airway Microbiome in Chronic Lung disease

4.1. Introduction

The airway microbiome may be considered a dynamic assemblage of microorganisms in which microorganisms are continuously introduced into the respiratory system though inhaled air, micro-aspiration and direct mucosal dispersion. On the other hand, microorganisms are continuously eliminated through muco-ciliary clearance, cough, innate and adaptive immunity. (Dickson et al., 2016a) The stability and dynamics of the airway microbiome has not been thoroughly investigated. To date, most lung microbiome studies have been cross-sectional in nature; analysing samples collected at single time points which can be confounded by various variables at the time of collection and does not provide insights into the stability of the microbial community composition over an extended period of time.

Therefore, collecting multiple samples from a single patient cohort over a year in the context of a longitudinal study could be invaluable to allow better assessment and understanding of the dynamics in the airway bacterial communities. In this Chapter, the stability and resilience of airway microbiome and the factors that impact the airway microbiome are discussed.

4.2. Methods

4.2.1. Ethics Approvals

The study protocol: “Changes in Microbiome with Treatment in Chronic Lung Diseases”, version 1.5.3 dated 7/7/16 (IRAS no: 202883, sponsor Ref. no: 16/0178, REC Ref no.: 16/LO/1490, Royal Free Ref no.: 9927) received UCL sponsorship on 26/7/2016, Harrow Research Ethics Committee (REC) approval on 25/10/16, Health Research Authority (HRA) approval on 14/11/2016, Royal Free NHS Trust confirmed capacity and capability on 22/2/2017. The letter of HRA
approval, a copy of the approved: study protocol, participant information sheet, participant consent, the approved IRAS application and a sample of the daily respiratory symptoms detailing diary are included in Appendix 1.

4.2.2. Study Conduct

Participants with a confirmed diagnosis of bronchiectasis (CT) or COPD (spirometry) were invited to join the study from respiratory clinics at the Royal Free hospital. A medical history of participants was taken by interviewing participants during the first visit after consent and reviewing the participants’ health records documented on Electronic Document Retrieval Methods (EDRM) and Cerner Electronic Medical Records bioinformatic systems which are adopted by the Royal Free NHS Trust. Sputum samples were collected and a symptom diary booklet was handed out during the first visit. The participants were followed up for one year from consent date. Morning sputum samples were requested from participants on the day of their normal follow up appointment in the respiratory clinic which was every three to six months. Specimen collection kits and guidance for correct specimen collection (Appendix 3.1) were sent to participants by post. Samples were collected by patients at home and brought to the clinic in the afternoon. Samples were promptly collected and homogenized with Sputasol (Ch.2, 2.2.). Bacteriological cultures were carried out on the sputum samples (Ch.2, 2.3.) then the processed sputum samples were aliquoted into 1 mL portions and were stored in -70°C freezers for further analysis. Antimicrobial susceptibility testing was carried out on the pathogenic bacteria identified by MALDI-TOF from the bacteriological cultures after incubation (Ch.2, 2.4., 2.5.) and results were reported to the treating clinicians. At the end of the study, the participants were asked to return their diary booklets and the medical records of each participant was rechecked at the end of the study period for exacerbation events, hospitalization events, prescribed antimicrobial treatment during the period of participation and the most recent lung function (spirometry) test.

In addition, repeated sputum samples which were collected in the context of another prospective longitudinal observational cohort study led by Miss Alexandra
Kondratiuk (UCL, CCM) and Dr. David Lowe (UCL, I&I) to investigate the respiratory microbiome in common variable immunodeficiency (CVID) patients with chronic lung disease and on prolonged prophylactic antimicrobial therapy recruited from UCL Institute of Immunity and Transplantation were included in the presented microbiome analysis (The study proposal included in the Appendix 2).

The inclusion criteria include: confirmed primary diagnosis of COPD or non-cystic fibrosis bronchiectasis confirmed by respiratory clinicians and documented by lung function results or high resolution computed tomographic scan respectively, history of frequent exacerbations ≥2/year, ability to spontaneously expectorate sputum and consent to participate. The exclusion criteria include: inability to produce sputum, being on immunosuppressant medication not including steroids, confirmed diagnosis of cystic fibrosis or lung cancer or history of lung transplantation, and known tuberculosis or HIV infections at the time of recruitment.

The Methods are explained in detail in Chapter 2. Briefly, metagenomic DNA was extracted from sputum samples using the automated LIAISON® lxt extraction platform (Ch.2, 2.6.1.). The hypervariable regions V3-V4 of the 16S rRNA gene were amplified using 341 forward primer and 805 reverse primer and amplicons were sequenced on Illumina® MiSeq platform over two runs (Ch.2, 2.8.). In bioinformatic analysis, the biom table was rarefied at the depth of 5000 reads per sample. All p-values of the differential RA were corrected using Benjamini-Hochberg False Discovery Rate (BH-FDR) method for multiple comparisons. Total number of copies of 16S rRNA gene, and the bacterial loads of typical respiratory pathogens and P. aeruginosa were detected in sputum using qPCR (Ch.2, 2.10.).

4.3. Results

4.3.1. Recruitment

The respiratory clinic lists were screened for potential participants, in the period between February 2017 and February 2018, in total 103 patients were approached; and forty patients were recruited in the first described study “Changes in
Microbiome with Treatment in Chronic Lung Diseases” study from two outpatient respiratory clinics led by Prof. Hurst and Prof. Lipman (Figure 4.1). In the period between February 2017 and July 2018, a total of 64 sputum samples were collected in the context of the longitudinal study from participants with a median of two samples per patient.

From the CVID cohort, 99 sputum samples were collected from 39 patients in the period between June 2017 and May 2018. Overall, a range of 1 to 8 samples per patient were collected; with a median of two sputum samples per patient.

Figure 4.1. Recruitment curve into the “Changes in Microbiome with Treatment in Chronic Lung Diseases” study.

4.3.2. Pooling samples in Microbiome Analysis

In 16S rRNA sequencing, 163 sputum samples were analysed over two sequencing runs; all sequencing reads were analysed simultaneously through the bioinformatic pipeline. Run was not a significant covariate in weighted UniFrac (p=0.600 by ANOSIM, p=0.768 by PERMANOVA) and the mock communities included in both runs clustered in the PCoA plot of the β-diversity index (Figure 4.2A); therefore, the inter-run variation was minimal and the data from both runs could be combined for comparative analysis.

Upon comparing the microbiome profiles in all sputum samples of the CVID patients (immunocompromised) and that of the other immunocompetent participants, immunodeficiency was not found to be a significant covariate in
weighted UniFrac $\beta$-diversity index ($p=0.083$ by ANOSIM, $p=0.104$ by PERMANOVA) (Figure 4.2B). Therefore, it was decided that the samples could be pooled together to power the comparisons in the subsequent analyses.

A.

![PCoA plot of weighted UniFrac $\beta$-diversity involving all tested sputum samples](image)

B.

![Comparison between total CVID samples and samples from immunocompetent patients](image)

**Figure 4.2.** PCoA plot of weighted UniFrac $\beta$-diversity involving all tested sputum samples (A) No separation based on the sequencing run; red: run 3 ($n=84$), navy blue: run 4 ($n=93$) ($p=0.768$ by PERMANOVA, $p=0.600$ by ANOSIM). The mock communities from three different sequencing runs coincide in the highlighted green circle. (B) comparison between total CVID samples (pink, $n=93$) and samples from immunocompetent patients (purple, $n=70$), No separation based on the immune status ($p=0.104$ by PERMANOVA, $p=0.083$ by ANOSIM).
### 4.3.3. Cohort characteristics

The clinical and demographic characteristics of the chronic lung disease patients from whom sputum samples were collected and included in the current study are summarised in Table 4.1. Sixty-one patients had bronchiectasis and twenty-three had COPD. Most patients in both chronic lung disease groups reported daily cough and sputum production on a daily basis. No significant differences were noted between the two groups regarding age, sex, body mass index (BMI), ethnicity or number of exacerbations per year. Nevertheless, a significantly higher proportion of CVID patients \((n=33, 54\%)\) were among the bronchiectasis patients, in contrast to 22\% \((5 \text{ participants})\) among the COPD patients \((p=0.008)\). Also, a higher proportion of smokers and ex-smokers were found within the COPD group.

In total, 79\% of the participants were prescribed a rescue pack of antibiotics either with corticosteroids \((17\%)\) or without to keep at home to promptly self-administer in case of exacerbation events. Furthermore, 47 participants \((56\%)\) were on antimicrobial prophylaxis regimes; most commonly azithromycin, co-trimoxazole \((\text{Septrin®})\), doxycycline, ciprofloxacin (discussed in detail in Chapter 5).

Concerning the other medication, no significant differences in the proportions of patients who are prescribed the mucolytic agent carbocisteine or oral corticosteroids. However, a significantly higher proportion of COPD patients were prescribed short acting \(\beta2\) agonists \((\text{SABA})\), long acting \(\beta2\) Agonists \((\text{LABA})\), long acting muscarinic receptor antagonists \((\text{LAMA})\) and inhaled corticosteroids \((\text{ICS})\). Also, significantly more COPD patients were receiving oxygen therapy compared to the bronchiectasis patients.

The results of the most recent lung function tests done for the participants during their routine health care in the respiratory clinics were collected from their medical records and summarised in Table 4.1. Around 32\% of the bronchiectasis patients had some degree of air flow obstruction which was defined as \(\text{FEV1/FVC} <0.7\) and 10\% had restrictive airflow which was defined as \(\text{FEV1/FVC} >0.7\) and \(\text{FVC} <80\%\) predicted.
Table 4.1: The demographics, medical and clinical characteristics of the patients' cohort with non-cystic fibrosis bronchiectasis and COPD

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Bronchiectasis (n=61)</th>
<th>COPD (n=23)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>66.5 (56-74)</td>
<td>66 (54-74)</td>
<td>0.819¹</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>26 (43%)</td>
<td>14 (61%)</td>
<td>0.135²</td>
</tr>
<tr>
<td>Females</td>
<td>35 (57%)</td>
<td>9 (39%)</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>18 (49%)</td>
<td>9 (64%)</td>
<td>0.530³</td>
</tr>
<tr>
<td>Black</td>
<td>3 (8%)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>16 (43%)</td>
<td>5 (36%)</td>
<td></td>
</tr>
<tr>
<td>BMI (Kg/m²)</td>
<td>27.2 (23.9-34.8)</td>
<td>24.2 (22-29)</td>
<td>0.189¹</td>
</tr>
<tr>
<td>Daily Cough</td>
<td>41 (79%)</td>
<td>14 (78%)</td>
<td>1.000³</td>
</tr>
<tr>
<td>Daily Sputum</td>
<td>47 (83%)</td>
<td>17 (94%)</td>
<td>0.279³</td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smokers</td>
<td>1 (2%)</td>
<td>5 (28%)</td>
<td>&lt;0.001³</td>
</tr>
<tr>
<td>Ex-smokers</td>
<td>11 (18%)</td>
<td>9 (50%)</td>
<td></td>
</tr>
<tr>
<td>Non-smokers</td>
<td>49 (80%)</td>
<td>9 (50%)</td>
<td></td>
</tr>
<tr>
<td>Passive smoking</td>
<td>11 (38%)</td>
<td>11 (73%)</td>
<td>0.204²</td>
</tr>
<tr>
<td>Immunocompromised</td>
<td>33 (54%)</td>
<td>5 (22%)</td>
<td>0.008²</td>
</tr>
<tr>
<td>No of exacerbations per year</td>
<td>3 (2-5)</td>
<td>3 (2-6)</td>
<td>0.539¹</td>
</tr>
<tr>
<td>Prophylactic Antibiotics</td>
<td>37 (61%)</td>
<td>10 (44%)</td>
<td>0.157²</td>
</tr>
<tr>
<td>Macrolides</td>
<td>26 (43%)</td>
<td>7 (33%)</td>
<td></td>
</tr>
<tr>
<td>Azithromycin</td>
<td>24 (39%)</td>
<td>7 (33%)</td>
<td></td>
</tr>
<tr>
<td>Clarithromycin</td>
<td>2 (3%)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>FQ</td>
<td>2 (3%)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Tetracyclines</td>
<td>3 (5%)</td>
<td>1 (4%)</td>
<td></td>
</tr>
<tr>
<td>β-lactams</td>
<td>3 (5%)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Amoxicillin, Penicillin V cephalexin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sulfonamides</td>
<td>5 (8%)</td>
<td>2 (10%)</td>
<td>0.242³</td>
</tr>
<tr>
<td>Rescue Pack</td>
<td>50 (82%)</td>
<td>16 (70%)</td>
<td></td>
</tr>
<tr>
<td>Rescue Pack Antibiotics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β-lactams</td>
<td>26 (43%)</td>
<td>5 (22%)</td>
<td></td>
</tr>
<tr>
<td>Amoxicillin/clavulanic</td>
<td>26 (43%)</td>
<td>5 (22%)</td>
<td></td>
</tr>
<tr>
<td>Amoxicillin</td>
<td>7 (11%)</td>
<td>3 (13%)</td>
<td></td>
</tr>
<tr>
<td>FQ</td>
<td>26 (43%)</td>
<td>5 (22%)</td>
<td></td>
</tr>
<tr>
<td>Ciprofloxacin, Levofloxacin, Moxifloxacin</td>
<td>8 (13%)</td>
<td>2 (10%)</td>
<td></td>
</tr>
<tr>
<td>Tetracyclines</td>
<td>10 (16%)</td>
<td>5 (24%)</td>
<td></td>
</tr>
<tr>
<td>Macrolides</td>
<td>4 (7%)</td>
<td>1 (5%)</td>
<td></td>
</tr>
<tr>
<td>Azithromycin</td>
<td>4 (7%)</td>
<td>1 (5%)</td>
<td></td>
</tr>
<tr>
<td>Clarithromycin</td>
<td>3 (5%)</td>
<td>1 (5%)</td>
<td></td>
</tr>
<tr>
<td>Rescue Pack Corticosteroids</td>
<td>8 (13%)</td>
<td>6 (26%)</td>
<td>0.085³</td>
</tr>
</tbody>
</table>
Continue Table 4.1: The demographics, medical and clinical characteristics of the patients' cohort with non-cystic fibrosis bronchiectasis and COPD

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Bronchiectasis (n=61)</th>
<th>COPD (n=23)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inhaled medication†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SABA</td>
<td>19 (31%)</td>
<td>13 (57%)</td>
<td>0.033²</td>
</tr>
<tr>
<td>LABA</td>
<td>20 (33%)</td>
<td>15 (65%)</td>
<td>0.007²</td>
</tr>
<tr>
<td>LAMA</td>
<td>8 (13%)</td>
<td>14 (61%)</td>
<td>&lt;0.001²</td>
</tr>
<tr>
<td>ICS</td>
<td>23 (38%)</td>
<td>15 (65%)</td>
<td>0.024²</td>
</tr>
<tr>
<td>Oral Corticosteroids†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 (16%)</td>
<td>5 (22%)</td>
<td>0.542³</td>
<td></td>
</tr>
<tr>
<td>Carbocisteine‡</td>
<td>14 (23%)</td>
<td>3 (13%)</td>
<td>0.378³</td>
</tr>
<tr>
<td>Oxygen therapy‡</td>
<td>2 (3%)</td>
<td>5 (24%)</td>
<td>0.015³</td>
</tr>
<tr>
<td><strong>Lung Function</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEV1 (L)§</td>
<td>1.68</td>
<td>1.13</td>
<td>0.141¹</td>
</tr>
<tr>
<td>(1.20 – 2.94)</td>
<td>(0.78 – 2.06)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEV1 % predicted§</td>
<td>87%</td>
<td>50%</td>
<td>0.014¹</td>
</tr>
<tr>
<td>(65 – 106%)</td>
<td>(31 – 68%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEV1 z-score§</td>
<td>-0.81</td>
<td>-3.31</td>
<td>0.003¹</td>
</tr>
<tr>
<td>(-1.94 – 0.28)</td>
<td>(-3.60 – -2.18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC (L)§</td>
<td>2.81</td>
<td>2.93</td>
<td>0.922¹</td>
</tr>
<tr>
<td>(2.03 – 3.59)</td>
<td>(1.79 – 3.56)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC % predicted§</td>
<td>106</td>
<td>84%</td>
<td>0.159¹</td>
</tr>
<tr>
<td>(83 – 122%)</td>
<td>(54 – 111%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC z-score §</td>
<td>0.24</td>
<td>-1.21</td>
<td>0.179¹</td>
</tr>
<tr>
<td>(-1.33 – -1.02)</td>
<td>(-3.14 – -0.57)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEV1/FVC §</td>
<td>0.73</td>
<td>0.50</td>
<td>0.005¹</td>
</tr>
<tr>
<td>(0.62 – 0.78)</td>
<td>(0.31 – 0.64)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(FEV1/FVC) % predicted §</td>
<td>93%</td>
<td>66%</td>
<td>0.005¹</td>
</tr>
<tr>
<td>(77 – 99%)</td>
<td>(39 – 87%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(FEV1/FVC) z-score §</td>
<td>-0.63</td>
<td>-2.83</td>
<td>0.087¹</td>
</tr>
<tr>
<td>(-2.37 – -0.05)</td>
<td>(-5.15 – -0.64)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Airflow Obstruction*‡</td>
<td>10 (32%)</td>
<td>12 (100%)</td>
<td>&lt;0.001²</td>
</tr>
<tr>
<td>Restrictive Airflow**‡</td>
<td>3 (10%)</td>
<td>0</td>
<td>0.540³</td>
</tr>
</tbody>
</table>

† Mean (SD) § Median (IQR) ‡ n (%) ¹ p-value by Mann-Whitney Test ² p-value by Chi square ³ p-value by Fisher Exact Test

FQ: Fluoroquinolones
SABA: Short Acting β2 Agonists  LABA: Long Acting β2 Agonists
LAMA: Long Acting Muscarinic Receptor Antagonists
ICS: Inhaled Corticosteroids  SR: Standardised Residuals
FEV1: Forced Expiratory volume in 1 sec  FVC: Forced Vital Capacity
* FEV1/FVC <0.7  **FEV1/FVC ≥0.7 and FVC< 80% predicted
4.3.4. Stability of the airway microbiome over time

The longitudinal samples collected from participants over the one year follow up period demonstrated that airway microbiome at baseline was stable in most subjects. The samples from each participant clustered together (p-value <0.001 by PERMANOVA in both weighted and unweighted UniFrac). Figure 4.3 shows the clustering of baseline sputum samples per participant in the PCoA plot of both weighted and unweighted UniFrac β-diversity. The time interval between the collected sputum samples ranged between one month and up to one year. The Shannon α-diversity was very similar across the repeated samples from the same subjects with the mean difference (SD) of 0.131 (0.503) between the paired samples (p=0.314 by paired t-test). Similarly, other α-diversity indices; Chao1, PD whole tree and Fisher alpha showed less intra-individual variations (all p-values >0.05 by paired t-test) i.e. between the repeat samples from the same subject at baseline, compared to the inter-individual variations.
Figure 4.3. The stability of airway microbiome in sputum over time in participants with multiple longitudinal samples at baseline respiratory status demonstrated through the UniFrac $\beta$-diversity indices and Shannon $\alpha$-diversity

A. PCoA plot of weighted Unifrac $\beta$-diversity ($p=0.001$ by PERMANOVA),
B. PCoA plot of unweighted Unifrac $\beta$-diversity ($p=0.001$ by PERMANOVA),
C. Shannon $\alpha$-diversity index ($p=0.314$ by paired t-test).

Each data point represents sputum sample, samples from the same participant are coloured similarly. Patients with COPD: 2, 93 and 95; while the rest were bronchiectasis patients.
In Figure 4.4, the stability of airway microbiome is demonstrated through multiple more frequent sputum samples collected from three patients with bronchiectasis: 3, 27 and 33. The time intervals between the subsequent sputum collection ranged between two weeks up to six months covering a total period of 9, 4 and 3 months respectively, the exact dates are given in Figure 4.4 legend. The samples from the same patient clustered together in the PCoA of weighted and unweighted UniFrac β-diversity. Apart from sample C of patient 27 in which both the RA of genus *Haemophilus* (in 16S rRNA sequencing, Figure 4.4 G) and the load of *H. influenzae* by qPCR surged. In qPCR, *H. influenzae* was detected only in sample C with load 4.68E+04 CFU/mL while it was absent in all other samples from this patient. Also, sample G of patient 3, in which some upper respiratory infection symptoms were reported by the patient. The richness and alpha-diversity indices: Chao1, Shannon, PD whole tree and Fisher alpha across the repeated samples demonstrated stability over time. This was also evident in the microbiome profile of the repeated sputum samples, despite some fluctuations in the relative abundances of some taxa, the overall microbiome profile was highly reproducible within the same patient. These three microbiome profiles demonstrate three different stages of bronchiectasis severity that ranges from a relatively healthier more diverse and richer respiratory microbiome to a poorer less diverse microbiome which was predominated by a few taxa (Figure 4.4G).
Chapter 4: Dynamics of Airway Microbiome

C. Observed OTU

D. Shannon

E. PD-whole tree

F. Fisher alpha

G. 136
Figure 4.4. The stability of airway bacterial community composition through frequent sputum sampling from three patients: 3 (red), 27 (green) and 33 (blue) demonstrated by α and β diversity indices and microbiome profiles (A) the PCoA plot of weighted Unifrac β-diversity (B) the PCoA plot of unweighted Unifrac β-diversity (C) Richness expressed by the total number of observed OTUs (D) Shannon α-diversity index (E) PD-whole tree α-diversity index (F) Fisher-alpha diversity index (G) Relative abundances of the bacterial taxa (H) Chronic colonization of patient 33 with H. influenzae (Hi); the loads in the subsequent samples were measured by qPCR. Samples dates: Patient 3: B: 25/8/17, C: 6/9/17, D (w1): 22/3/18, E (w2): 31/3/18, F (w6)18/4/18, G (w8) 4/5/18, H (w10):15/5/18, I (w12): 29/5/18; Patient 33: B: 1/3/18, C (w3): 20/3/18, D (w6): 11/4/18, E (w8): 24/4/18, F (w10): 11/5/18; Patient 27 A: 25/7/17, B: 16/8/17, C: 22/8/17, D: 6/9/17, E: 27/9/17, F:19/10/17.
4.3.5. Factors impacting the airway microbiome

The airway microbiome could be impacted by various factors that resulted in alterations in the individual’s airway microbiome from the stable (baseline) status. The following factors were observed to influence the airway microbiome.

4.3.5.1. Exacerbation

Exacerbations negatively impacted the diversity of the airway microbiome. Figure 4.5 demonstrates how the airway microbiome changed during exacerbations in paired samples collected from three patients with bronchiectasis: 24, 76 and 94, and one patient with COPD: 9, during acute exacerbation events without administration of antibiotics. In the PCoA plot of weighted UniFrac β-diversity, there is a significant shift in the patient’s position in the spatial distribution in one direction towards less diversity (Figure 4.5A). The Shannon α-diversity was significantly lowered during exacerbations (p= 0.032, by paired t-test) (Figure 4.5B), similarly a decline was observed in the other α-diversity indices with exacerbation. The richness in microbiome was also considerably reduced though it has not reached statistical significance (p= 0.07, by paired t-test) (Figure 4.5C). The microbiome profiles, revealed that the bacterial phylum Proteobacteria (p= 0.022 by White’s t-test) particularly class Gamma-Proteobacteria (p= 0.023 by White’s t-test) were enriched in airways during exacerbations at the expense of Actinobacteria and Firmicutes whose relative abundances significantly dropped (p=0.027 and 0.03 respectively, by White’s t-test) (Figure 4.5D). This involved reductions in the relative abundances of multiple bacterial genera that normally make up the airway microbial community such as Prevotella, Veillonella, Actinomyces. (Figure 4.5E).
Chapter 4: Dynamics of Airway Microbiome

A.

B.

C.

D.

95% confidence intervals

<table>
<thead>
<tr>
<th>p</th>
<th>p-Proteobacteria</th>
<th>p-Actinobacteria</th>
<th>p-Firmicutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean proportion (%)</td>
<td>0.022</td>
<td>0.027</td>
<td>0.030</td>
</tr>
<tr>
<td>95% confidence intervals</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 4.5. Effect of respiratory exacerbation on the airway microbiome (A) Shifts in the spatial distribution in the PCoA plot of weighted Unifrac β-diversity; orange vector represents exacerbation shift while the green vector represents recovery (B) Decline in Shannon α-diversity index (p=0.032) (C) Reduction in the total number of observed OTUs which is a richness measure (p=0.07). Differential abundance of bacterial phyla (D) and genera (E) between base-line samples (green) and samples collected during exacerbation (orange).

4.3.5.2. Antibiotic treatment

The administration of short courses of antimicrobial chemotherapeutic agents prescribed either for the treatment of respiratory infections or other infections elsewhere in the body impacts the airway microbiome. Figure 4.6. compares the airway microbiome in six patients with bronchiectasis after the administration of prescribed courses of antibiotics with their corresponding baseline samples. Patient 71 was taking amoxicillin/clavulanic acid for seven days, patient 8 had taken amoxicillin/clavulanic acid for 14 days followed by clarithromycin for 5 days and doxycycline for 2 days which was stopped due to side effects, patient 74 was prescribed flucloxacillin for 9 days, patients 79 and 23 were on amoxicillin for 14 days and patient 81 had just finished a course of cephalexin in addition to prednisolone for seven days. Notable shifts in different directions were observed in the PCoA of both weighted and unweighted Unifrac β-diversity after the antibiotic courses (Figure 4.6A). There was not a consistent directional change in the α-diversity of the microbial communities in response to the antibiotic treatment courses. In case of patient 23, the isolated H. influenzae from sputum was found
resistant to multiple β-lactams: ampicillin and cefotaxime; this may explain the observed increase in the RA of *Haemophilus* after the administration of amoxicillin antibiotics.

A.

B.
Figure 4.6. Effect of antibiotic treatment courses on the airway microbiome showing shifts in the spatial position, indicated by vectors, in the PCoA plot of weighted Unifrac (A) and unweighted UniFrac (B) β-diversity indices, changes in Shannon α-diversity index (C) (p=0.59 by paired t-test) and (D) shifts in the RA of some genera in the microbiome profiles (A: before antibiotic administration, B: after, for Pt23; B: Before, C: after) Antibiotics involved; 71: amoxi/clav., 8: amoxi/clav., clarithromycin and doxycycline, 74: flucloxacillin, 79 and 23: amoxicillin, 81: cephalexin.
Chapter 4: Dynamics of Airway Microbiome

Focusing on patient 82 who had advanced COPD, the bacteriological culture and the microbiome results of the first sputum sample collected (A) revealed a developing respiratory infection with *P. aeruginosa* along with *H. influenzae* which was not evident in the patient’s previous medical records. Consequently, the patient was prescribed a two-weeks course of ciprofloxacin 750mg twice daily to eradicate the *P. aeruginosa* infection (sample B). Afterwards, the patient reported a general improvement in the respiratory status and decline in the frequency of the respiratory exacerbation symptoms. The third sample (C) collected seven months later, the microbiome profile of which revealed that *Pseudomonas* had not been eradicated and its relative abundance was almost the same as in the first sample; however, the relative abundance of the *Haemophilus* dropped significantly (p<0.0001). This observation was also confirmed by the respiratory qPCR (Figure 4.7E). Despite this, the microbiome was generally shifted towards a more diverse healthier microbiome after five months free of exacerbations or any further antibiotic treatments, there was a decline in the total bacterial density compared to the previous two samples; in addition, a significant rise in all α-diversity indices was noted with the enrichment of 23 bacterial genera (p<0.016) (Figure 4.7).
Chapter 4: Dynamics of Airway Microbiome

A.

![Principal Component Analysis (PCA) plot showing the distribution of samples along PC1 and PC2 axes.]

B.

- **Observed number of OTUs**
  - Samples A, B, C

- **Shannon Index**
  - Samples A, B, C

C.

- **RA (% of total reads)**
  - Samples 82A, 82B, 82C

Legend:
- g__Actinomyces
- g__Corynebacterium
- g__Porphyromonas
- g__Prevotella
- f__Gemellaceae; g__Gemella
- g__Lactobacillus
- g__Streptococcus
- g__Haemophilus
- g__Neisseria
- f__Enterobacteriaceae; g__Acinetobacter
- g__Rothia
- g__Staphylococcus
- g__Granulicatella
- g__Veillonella
- g__Fusobacterium
- g__Neisseria
- g__Pseudomonas

144
Figure 4.7. Dynamic changes in the airway microbiome of patient 82 with COPD after a two-week course of high dose of ciprofloxacin (sample B) (A) The vector indicates the shift in the patient’s microbiome position in the PCoA plot of weighted Unifrac β-diversity (B) Richness and α-diversity of the bacterial communities measured by observed number of OTUs and Shannon α-diversity index (respectively) (C) Microbiome profile at the genus level (D) Total bacterial density by measuring the total 16S rRNA gene copies/µL qPCR (E) P. aeruginosa and H. influenzae loads by qPCR. Sample Dates: A: 24/7/17, B: 11/11/17, C: 30/04/18.
4.3.5.3. Antibiotic prophylaxis

Antibiotic prophylaxis therapy impact the individual’s airway microbiome in different ways. In our cohort, two patients with bronchiectasis started azithromycin prophylaxis therapy. Patient 18 was prescribed azithromycin prophylaxis therapy 500 mg thrice weekly; the first sample (A) was collected just prior to starting the prophylaxis therapy whereas the second one (B) was collected three months later. Patient 75 was prescribed 500 mg azithromycin thrice weekly; the first sputum sample was collected just before starting the regime and the second one was collected one year after. Patients: 5 and 96 had been on azithromycin prophylaxis therapy; 250 mg thrice weekly, for one and three years respectively; the regime included taking the antibiotic over the winter months from October to May and breaks over the summer months. The first sputum sample of patient 96 was at the end of the break period whereas the second sputum sample was collected eight months later at the end of the treatment period. For patient 5, one sample was collected during the break time one month before restarting the prophylaxis therapy and the other was collected two month after, while being on the prophylaxis therapy. In the PCoA plot of weighted UniFrac (Figure 4.8A) notable shifts were observed towards a relatively less diverse microbiome after the prolonged course of the prophylactic antibiotic. This was evident in the significant decline in the PD-whole tree (p=0.03 by paired t-test) (Figure 4.8C). However, such decline in the alpha diversity was not captured by Shannon index (Figure 4.5D). The richness of the bacterial community expressed by the total number of observed OTU in the 16S rRNA sequencing data was also reduced however it was not statistically significant (p=0.107 by paired t-test) (Figure 4.8B). At the genus level of the microbiome profile, *Pseudomonas* was significantly enriched in the paired samples while being on or just after prolonged azithromycin prophylaxis therapy (p=0.039) (Figure 4.8E).
Figure 4.8. Impact of prolonged prophylactic courses of azithromycin (AZM) on the airway microbiome (A) Shifts in the spatial distribution in the PCoA plot of weighted Unifrac $\beta$-diversity (indicated by vectors) (B) Richness expressed by the total number of observed OTUs ($p=0.107$ by paired t-test) (C) $\alpha$-diversity expressed by PD-whole tree ($p=0.03$ by paired t-test) (D) Shannon $\alpha$-diversity index ($p=0.59$ by paired t-test) (E) change in the differential abundance of Pseudomonas between samples collected while being on azithromycin (blue) and off the prophylactic treatment (green). Before: Before the antibiotic course, After: after the antibiotic course.
The bronchiectasis patient; 30, had been taking azithromycin for prophylaxis when the first three sputum samples were collected, then the patient was shifted to ciprofloxacin when the last sputum sample (w12) was collected (Figure 4.9). Therefore, a significant shift in the position of w12 sample from the patient’s cluster was observed in the PCoA plot of weighted Unifrac β-diversity (Figure 4.9A) which can be explained by the significant reduction in the bacterial richness expressed by the total number of observed OTUs Figure 4.9B. The α-diversity did not significantly change (Figure 4.9C). The microbiome profile revealed a significant enrichment in some taxa like *Neisseria*, *Capnocytophaga* and *Porphyromonas* and decline in others such as *Streptococcus* and *Corynebacterium* (Figure 4.9E).
Figure 4.9. Impact of switching the antibiotic prophylaxis from azithromycin (points B, w4 and w6) to ciprofloxacin (point w12) on the airway microbiome of patient 30 with bronchiectasis showing (A) The vector indicates the shift in the patient’s microbiome position in the PCoA plot of weighted Unifrac β-diversity at week12 (B) Decline in the richness measured by observed number of OTUs (C) No significant change in the α-diversity of the bacterial community measured by Shannon and PD-whole tree diversity indices (D) Microbiome profiles at the genus level (E) Differential abundances of the genera that were affected by the shift to ciprofloxacin prophylaxis (light blue).
4.3.6. Resilience of airway Microbiome

The airway microbiome was highly resilient since it quickly returned to the original base-line position after the influence of the driving factor diminished. Some individual examples of patients are discussed in the following sections to demonstrate the resilience of airway microbiome.

Patient 32 had bronchiectasis and was on ciprofloxacin prophylaxis; 250 mg twice daily, six sputum samples were collected over three months from February to May 2018. Most of the samples collected from this patient clustered together in the PCoA plot of weighted UniFrac (Figure 4.10A) and the \( \alpha \)-diversity of the airway microbial communities was relatively stable over this time (Figure 4.10B). However, the sample collected at week 5 (w5) was when the patient had just finished a course of Doxycycline 100mg for 10 days for the treatment of an exacerbation; even though, from the microbiome profile there was no evidence of bacterial infection. The antimicrobial treatment impacted the airway microbiome, as shown in Figure 4.10, this sample (w5) appeared as a satellite point in the PCoA plot of weighted UniFrac. Furthermore, a remarkable decline in the richness and \( \alpha \)-diversity happened at this point. Nevertheless, the microbiome recovered quickly from the effect of the antimicrobial treatment and returned to normal patient's normal cluster within three weeks.
Figure 4.10. Resilience of the airway microbiome in patient 32 with bronchiectasis after the exposure to an exacerbation event and antimicrobial treatment; indicated by the navy blue vector and recovery indicated by the green vector in (A) the PCoA plot of weighted UniFrac $\beta$-diversity, (B) the timeline for samples collection and the associated conditions, (C) the richness and $\alpha$-diversity changes of the bacterial community measured by observed number of OTUs and Shannon $\alpha$-diversity index (respectively) (D) the microbiome profile of the individual samples at the genus level. Sample Dates: B: 28/2/18, w2: 14/3/18, w5: 6/4/18, w8: 25/4/18, w10: 10/5/18, w12: 24/5/18.
Chapter 4: Dynamics of Airway Microbiome

Patient 13 was a CVID patient who had severe COPD as well as an element of bronchiectasis. The patient had been on azithromycin prophylaxis therapy 250 mg thrice weekly. Samples collected at time points B and C were at respiratory base-line status, while the satellite point A was during an exacerbation event; this is evident from the microbiome profile Figure 4.11D where an active infection with *Acinetobacter* and *Pseudomonas* was obvious. Then the patient recovered within a month to return to the base-line status B and C which were one month apart. The sample collected later at w1 was four months from point C and by that time the patient had completed an eight-days course of piperacillin/tazobactam for treating a non-respiratory infection; nevertheless, this course had an impact on the respiratory microbiota and resulted in a shift in the spatial position of this sample in the PCoA of weighted UniFrac Figure 4.11A. The further samples which were collected within the following month revealed that the airway microbiome recovered to a slightly shifted new position (light blue cluster) in which the patient was shifted to Amoxicillin/clavulanic acid prophylaxis instead. Looking at the richness and alpha diversity indices (Figure 4.11B and C respectively), there were two levels across the two phases of prophylaxis therapy, in which a lower diversity was attained with Amoxicillin/clavulanic compared to the azithromycin. Furthermore, the richness as well as the alpha diversity showed 2 troughs where they dropped considerably during exacerbation (point A) and after the broad-spectrum piperacillin/tazobactam course (point w1).
Figure 4.11. Resilience of the airway microbiome in patient 13 with COPD after an exacerbation event and antimicrobial treatment, A) showing the PCoA plot of weighted Unifrac β-diversity, B) the timeline for samples collection and the associated conditions, (orange vector indicates directional shift due to exacerbation and the green ones indicate recovery) B) the richness and α-diversity changes of the bacterial community measured by observed number of OTUs and Shannon α-diversity index (respectively) C) the microbiome profiles of the individual samples at the genus level. Sample Dates: A: 7/9/17, B: 3/10/17, C: 5/11/17, w1: 27/3/18, w6: 12/4/18, w8: 1/5/18, w12: 22/5/18.
4.4. Discussion

In this chapter, the dynamics of the airway microbiome were studied in multiple sputum samples collected from participants in the context of a longitudinal study. The results revealed that, overall, the airway bacterial community was stable over the one year follow up period; even though there might be some temporal changes, it reverted to its norm as soon as the influencer ceased. The Inter-individual variability was greater than the intra-individual variability. This was evident by the clustering of samples with respect to the patient ID in the PCoA plot of β-diversity. Similarly, Carmody et al. found that airway microbial communities were relatively stable during periods of clinical stability over a period of 22 to 27 days despite some day to day fluctuations in cystic fibrosis patients. (Carmody et al., 2015) Moreover, Woo et al. reported that the airway microbiome remained stable for 4-16 years in non-cystic fibrosis bronchiectasis cohort. (Woo et al., 2019)

Some factors like exacerbations, antibiotic treatment and antibiotic prophylaxis had an impact on the diversity of the respiratory microbial communities and resulted in temporal shifts in the patient’s position in the PCoA plot of UniFrac β-diversity which reflects a change in the phylogenetic make-up of the bacterial communities in the airways.

Changes during exacerbation involved a significant rise in the relative abundance of Proteobacteria particularly gamma-Proteobacteria which is a bacterial class that encompasses most of the Gram negative respiratory pathogens, at the expense of the relative abundances of Actinobacteria and Firmicutes. Actinobacteria comprise a large group of metabolically diverse organisms that are prolific producers of secondary metabolites, including those with antimicrobial activity and may play a role in maintaining the homeostasis in the microbial ecosystem. (Ventura et al., 2007, Huang et al., 2014) While Firmicutes includes Gram positive bacteria and represents around 45% of the sequence reads in the results of relatively healthy subjects (as shown later in the data presented in Chapter 7, Figure 7.6). Similar compositional shifts in the microbial community structure upon exacerbation events in cystic fibrosis and bronchiectasis were reported in other studies.
Chapter 4: Dynamics of Airway Microbiome

(Carmody et al., 2015, Tunney et al., 2013) Carmody et al. study which looked at daily fluctuations in the airway microbiome in sputum specimens of four cystic fibrosis patients revealed that the onset of exacerbation may be heralded by marked shifts in airway bacterial communities which highlight the potential of the airway microbiome signatures as biomarkers for exacerbation prediction. (Carmody et al., 2015)

A study of a COPD cohort (Huang et al., 2014) found that despite inter-subject variability in the magnitude of microbial community changes at exacerbation, it was generally characterised by an increase in the relative abundance of Proteobacteria and reductions in relative abundances of Actinobacteria, Clostridia, and Bacteroidia. Furthermore, the predicted metagenomes suggested that the members involved in such compositional shifts might cause an increase in pathogen-elicited inflammation and loss of potentially key protective functions in the microbiome such as the synthesis of antimicrobial and anti-inflammatory products; these effects were reversed upon clinical recovery. Nevertheless, it is not clear whether the observed changes in the microbial community associated with an acute exacerbation event in the data presented and other studies were the response to the influence of other drivers in the microbial ecosystem or Proteobacteria are directly involved in the causal pathway of the exacerbation pathogenesis.

Even when the cause of exacerbation is viral infection, shifts in the bacterial community composition would be expected. Molyneaux et al. demonstrated in their study that fifteen days after the inoculation with rhinovirus intra-nasally, the total bacterial load was raised significantly in induced sputum samples from participants with COPD. The relative abundance of Proteobacteria; particularly H. influenzae, surged and was maintained high for up to 42 days. On the contrary, the airway microbiome of the healthy controls who were also inoculated with the virus, was relatively stable. (Molyneaux et al., 2013)

The changes in the airway bacterial community composition associated with short antibiotic treatment courses were temporal and differed among the various
chemotherapeutic antimicrobial courses. This may be attributed to the differences in the spectrum of activity of the various classes of antibiotics, duration of therapy, co-administration of corticosteroids and the sensitivity of the microbiota to the administered antibiotics. Haung et al. also found significant variations in the response to treatment for exacerbation depending on the therapy regimens. (Huang et al., 2014) At the individual level, the prolonged prophylactic antibiotic courses resulted in bigger and more persistent shifts. Jiangchao et al. who studied the airway microbiome in a cystic fibrosis cohort over a decade suggested that antibiotic use, rather than patient age or lung function, was the primary driver of decreasing diversity. (Jiangchao et al., 2012) It was noticed that in two patients with bronchiectasis who were on azithromycin prophylaxis therapy and were colonized with P. aeruginosa, the relative abundance of Pseudomonas increased (mean difference was 41%, p=0.039) while being on the prophylactic therapy compared to samples collected during the break period from the antimicrobial therapy. This agrees with the findings of the BLESS trial in some bronchiectasis patients where erythromycin 400 mg twice daily for 12 months promoted the enrichment of Pseudomonas at the expense of Haemophilus and other more susceptible bacteria in the airway bacterial communities, where the median difference in Pseudomonas relative abundance was 6.6%, 95% (CI 0.1-37). (Rogers et al., 2014b) Nevertheless, in terms of quantitative data, the absolute load of P. aeruginosa did not show a similar trend. It is noteworthy that the microbiome results are semi-quantitative in which results are expressed as the proportion of reads classified into one taxa relative to the total number of sequence reads that represent the whole microbiome of a sample. In the work presented, reduced total bacterial density in the airways were associated with prolonged antibiotic prophylaxis therapy as demonstrated in the next chapter (Ch.5, Figure 5.9C). This may suggest that the antimicrobial treatment had rather suppressed the susceptible bacteria in the microbial community; while, the more antimicrobial tolerant bacteria like Pseudomonas persisted, therefore, it relatively predominated the bacterial community.
Chapter 4: Dynamics of Airway Microbiome

It is a common clinical practice to treat patients with chronic lung disease with an intensive prolonged ciprofloxacin course after the first isolation of *P. aeruginosa* from sputum in an attempt to eradicate the bacteria and halt colonization. (Hurley and Smyth, 2012) Despite the favourable outcome in the patient with COPD (PT 82, Figure 4.7) who received the ciprofloxacin treatment course; where the patient reported a decline in the frequency of exacerbations that required further antibiotic treatment in the following six months; in addition, a more diverse microbiome profile, reduced total bacterial load and a decline in the relative abundance of *H. influenzae* in sputum were noted after the treatment, the respiratory colonization with *P. aeruginosa* was not eradicated. Several published reports have highlighted the difficulty of *P. aeruginosa* eradication, in a previous study, only one in fifteen bronchiectasis patients with chronic *P. aeruginosa* colonization had the organism eradicated; despite the aggressive antibiotic treatment prescribed. (Martínez-García et al., 2007)

The airway microbial communities were resilient; although the diversity of airway bacterial communities was influenced by some factors, they soon reverted to their normal base-line composition after the influence of the driver had been halted. Similar findings were found in cystic fibrosis patients where airways communities showed both short- and long-term resilience after antibiotic perturbation. (Jiangchao et al., 2012, Leah et al., 2015) Moreover, Bacci *et al.* found that not only the taxa but also the functional metagenome was highly resilient over time and after antibiotic treatment in the sputum of cystic fibrosis patients. (Bacci et al., 2020)

A limitation related to this work was the collection of sputum samples in the context of the longitudinal study which was not systematic; but rather opportunistic. This was attributed to some logistic limitations of the study; the samples were collected during the routine clinical follow up appointments of the patients in the respiratory clinics; nevertheless, many patients in the cohort attended the clinics every 6 months; in contrast the CVID patients attended hospital every two to three weeks for immunoglobulin replacement therapy which allowed more frequent samples collection. Further systematic studies that collect a larger set of samples at regular
intervals and more frequently around the time of exacerbations are needed to assess completely the potential of airway microbiome signatures to predict exacerbation which may inform the prescription of the most appropriate treatment in a timely manner to halt the exacerbations; therefore, slow down the progression of the disease and the development of antimicrobial resistance.

To sum up, the airway microbiome is highly individualized and resilient. It responded to the influence of several factors like exacerbations, antibiotic treatment or antimicrobial prophylaxis. This is usually associated with shifts in the bacterial community composition towards less diversity. However, it eventually returned to its norm in most of the cases after the influence of the driving factor ceased.
Chapter 5: Microbiome and Antimicrobial Treatment

5.1. Introduction

Scientific evidence has demonstrated the clinical benefits of long-term prophylactic use of antibiotics; especially macrolides, in reducing the frequency of exacerbations, and improving quality of life in both COPD (Herath et al., 2018, Donath et al., 2013, Wentao et al., 2015) and bronchiectasis (Hnin et al., 2015, Kelly et al., 2018, Fan et al., 2015) (Ch1: Table 1.1). Recently, the clinical utility of macrolides; especially azithromycin in reducing exacerbation over one year in COPD patients has been recognised as evidence A in the Global Initiative for Chronic Obstructive Lung disease guidance. (GOLD, 2020, Burns et al., 2011) Nevertheless, the adverse effects of the long-term use of antimicrobials in such chronic cases has not yet been fully examined. There is a concern around the selection pressure that a prolonged antibiotic regime may exert on the airway (and other) microbiota that may alter the microbial homeostasis in the body. In addition, there is a risk in the emergence and spread of antimicrobial resistance in the wider population. (Leung et al., 2011, Wenzel et al., 2012)

In this chapter, the impact of prolonged antibiotic prophylaxis therapy on the airway microbiome in patients with chronic lung disease and how it may influence AMR levels in bacterial isolates have been studied. In addition, the differences in the airway microbiome between COPD and bronchiectasis patients; as well as, how the airway microbiome in chronic lung disease patients with an underlying immunodeficiency may differ from other immunocompetent patients have been investigated.

5.2. Methods

Bacteriological cultures incubated at aerobic and high carbon-dioxide (5%) conditions on enriched and selective culture media were carried out on 116 spontaneous sputum samples collected from 68 patients with chronic lung disease.
Chapter 5: Microbiome and Antimicrobial Treatment

All morphologically different colonies were isolated, purified, identified using MALDI-TOF and stored at -70°C (Ch.2, 2.4.). Antimicrobial susceptibility testing was carried out on isolates using the disc diffusion method (Ch.2.5.).

A single sputum sample per patient that fulfilled the following criteria was selected to best represent the individual’s airway microbiome at stable baseline respiratory status in the cross-sectional comparisons. No significant change in the following respiratory symptoms: change in the colour, consistency and/or the amount of sputum, fever, worse cough, and wheeze; nor antibiotic treatment course was permitted in the previous fortnight and the week after sample’s collection. Whenever possible for patients who were on an antibiotic prophylaxis regime, the sample that fulfilled the above criteria and was collected while the patient was on or had been taking the prophylactic antibiotics for some time was selected.

Bioinformatic analysis was completed through the QIIME pipeline (Ch.2, 2.9.). Alpha and ß-diversity analyses were carried out on the OTU-table in which all samples were rarefied at the depth of 5000 reads. The significance of the differential abundances in the constituting taxa was tested using either White’s non-parametric test or Krüskal-Wallis tests which are implemented in STAMP bioinformatics programme as appropriate, p-values were corrected for multiple comparisons using Benjamini-Hochberg correction factor, otherwise is specified.

5.3. Results

5.3.1. Cohort’s characteristics

Forty-seven participants from the previously described cohort in Chapter 4 were prescribed antibiotic prophylaxis therapy in the context of their routine medical care; 66% were prescribed macrolides; specifically, 29 patients were on azithromycin 250 mg thrice weekly and eight were on 500 mg thrice weekly; while two were on clarithromycin (250 mg twice daily). Other prescribed antibiotics were co-trimoxazole (Septrin®; 960 mg once daily or 1920 mg, divided into two doses, thrice weekly), ciprofloxacin (250 mg twice daily), doxycycline (100 mg once daily), amoxicillin (250 mg twice daily), cephalexin (500 mg twice daily) and phenoxy-
methyl penicillin (250 mg twice daily). The clinical and demographic characteristics of the participants who were on prescribed antibiotic prophylaxis therapy in comparison with other participants in the cohort are presented in Table 5.1. There were no significant differences in the proportions of COPD and bronchiectasis patients between the two groups (p=0.157 by Chi Square test); in the antibiotic prophylaxis group, 79% had bronchiectasis and 21% had COPD; while in the comparator group, 65% had bronchiectasis and 35% had COPD. No significant differences were noted between the two groups regarding age, sex, BMI, ethnicity or smoking status. All participants were regularly receiving the influenza vaccine and had received streptococcal vaccines. Nevertheless, a significantly higher proportion of CVID immunodeficient patients was in the antibiotic prophylaxis group (66%) while only seven were in the comparator group (19%) (p<0.0001 by Chi Square test). In the cohort, 90% of the CVID patients were on immunoglobulin replacement therapy.

Seventy nine percent of the participants were prescribed a rescue pack of antibiotics to keep at home for self-administration in case of exacerbation events (Ch.1, 1.4.4); 47% of the rescue packs contained amoxicillin/clavulanic acid, 20% had doxycycline, 14% had quinolones either ciprofloxacin, levofloxacin or moxifloxacin, 15% had amoxicillin and 7.5% had macrolides either azithromycin of clarithromycin. There were no significant differences in the proportions of patients who have rescue packs among bronchiectasis and COPD patients. Nevertheless, the number of participants who had antibiotic rescue packs at home was higher in the antibiotic prophylaxis group 89% versus 65% in the comparator group (p=0.007 by Chi Square test). Also, a significantly higher proportion of patients in the antibiotic prophylaxis group had oral corticosteroids prescribed with the antibiotic rescue packs (20%) in comparison to 11% of comparator group (p=0.025 by Chi Square test).
### Table 5.1: Cohort’s demographics, medical and clinical characteristics with respect to the use of antibiotic prophylaxis therapy

<table>
<thead>
<tr>
<th></th>
<th>On Prophylactic Antibiotic treatment (n=47)</th>
<th>Comparator group (n=37)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prophylactic Antibiotics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Macrolides</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Azithromycin</td>
<td>29 (62%)</td>
<td>2 (4%)</td>
<td></td>
</tr>
<tr>
<td>Clarithromycin</td>
<td>2 (4%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quinolones</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ciprofloxacin</td>
<td>2 (4%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tetracyclines</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doxycycline</td>
<td>4 (9%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ß-lactams</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amoxicillin, Penicillin</td>
<td>3 (6%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenoxymethyl cephalaxin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sulfonamides</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Co-trimoxazole</td>
<td>7 (15%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age</strong> (years)</td>
<td>67 (12)</td>
<td>62 (19)</td>
<td>0.388¹</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>22 (47%)</td>
<td>18 (49%)</td>
<td>0.867²</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>13 (50%)</td>
<td>14 (56%)</td>
<td>0.669³</td>
</tr>
<tr>
<td>Black</td>
<td>1 (4%)</td>
<td>2 (8%)</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>12 (46%)</td>
<td>9 (36%)</td>
<td></td>
</tr>
<tr>
<td><strong>BMI</strong> (Kg/m²)</td>
<td>27.79 (4.40)</td>
<td>27.48 (9.14)</td>
<td>0.298¹</td>
</tr>
<tr>
<td><strong>Chronic lung disease</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bronchiectasis</td>
<td>37 (79%)</td>
<td>24 (65%)</td>
<td>0.157²</td>
</tr>
<tr>
<td>COPD</td>
<td>10 (21%)</td>
<td>13 (35%)</td>
<td></td>
</tr>
<tr>
<td><strong>Immunocompromised</strong></td>
<td></td>
<td></td>
<td>&lt;0.0001²</td>
</tr>
<tr>
<td>31 (66%)</td>
<td>7 (19%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Smoking</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smokers</td>
<td>1 (2%)</td>
<td>5 (14%)</td>
<td>0.140³</td>
</tr>
<tr>
<td>Ex-smokers</td>
<td>11 (23%)</td>
<td>9 (24%)</td>
<td></td>
</tr>
<tr>
<td>Never-smokers</td>
<td>35 (74%)</td>
<td>23 (62%)</td>
<td></td>
</tr>
<tr>
<td>Passive smoking</td>
<td>10 (63%)</td>
<td>12 (60%)</td>
<td>0.878²</td>
</tr>
<tr>
<td>Rescue Pack</td>
<td>42 (89%)</td>
<td>24 (65%)</td>
<td>0.007²</td>
</tr>
<tr>
<td><strong>Rescue Pack Antibiotics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ß-lactams</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Co-amoxiclav</td>
<td>23 (49%)</td>
<td>8 (22%)</td>
<td></td>
</tr>
<tr>
<td>Amoxicillin</td>
<td>3 (6%)</td>
<td>7 (19%)</td>
<td></td>
</tr>
<tr>
<td>FQ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ciprofloxacin, Levofloxacin</td>
<td>7 (15%)</td>
<td>2 (5%)</td>
<td></td>
</tr>
<tr>
<td>Moxifloxacin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tetracyclines</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doxycycline</td>
<td>6 (13%)</td>
<td>7 (19%)</td>
<td></td>
</tr>
<tr>
<td>Macrolides</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Azithromycin</td>
<td>3 (6%)</td>
<td>2 (5%)</td>
<td></td>
</tr>
<tr>
<td>Clarithromycin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rescue Pack Corticosteroids</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 (19%)</td>
<td>5 (14%)</td>
<td></td>
<td>0.025²</td>
</tr>
<tr>
<td>Prescribed CIP course for <em>P. aeruginosa</em> infection</td>
<td>11 (23%)</td>
<td>7 (19%)</td>
<td>0.619²</td>
</tr>
<tr>
<td>Oral Corticosteroids</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 (21%)</td>
<td>5 (14%)</td>
<td></td>
<td>0.356²</td>
</tr>
<tr>
<td>Carbocisteine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 (21%)</td>
<td>7 (19%)</td>
<td></td>
<td>0.789²</td>
</tr>
<tr>
<td>Oxygen therapy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 (9%)</td>
<td>3 (8%)</td>
<td></td>
<td>1.00¹</td>
</tr>
</tbody>
</table>

¹ Mean (SD) ² n (%) ³ FQ: Fluoroquinolones

¹ p-value by Mann-Whitney ² p-value by Chi square ³ p-value by Fisher Exact Test
Concerning the other medication, there were no significant differences in the proportions of patients who were prescribed the mucolytic agent carbocisteine, SABA, LABA, LAMA, inhaled corticosteroids, oral corticosteroids and oxygen therapy between the two studied groups (Table 5.1 and 5.2).

Lung function was comparable between the two study groups. Spirometry parameters were not significantly different between participants who were on antibiotic prophylaxis therapy and the other participants (Table 5.2). Airflow obstruction, defined as FEV1/FVC <0.7, was identified in 39% of the antibiotic prophylaxis groups and 59% of the comparator group (p=0.181 by Chi Square test). Restrictive airflow, defined as FEV1/FVC >0.7 and FVC <80% predicted, was found in 9% of the antibiotic prophylaxis groups and 5% of the comparator group (p=1.000 by Fisher Exact test).

Table 5.2: Lung function parameters and the inhaled respiratory medication used by participants in the antibiotic prophylaxis group and the comparator group

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>On Prophylactic Antimicrobial Treatment (n=47)</th>
<th>No Prophylactic Antimicrobial Treatment (n=37)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEV1 (L)§</td>
<td>1.96 (1.06–2.91)</td>
<td>1.52 (0.94–2.72)</td>
<td>0.371</td>
</tr>
<tr>
<td>FEV1 % predicted§</td>
<td>80.5% (43.8–109%)</td>
<td>75% (46–92%)</td>
<td>0.558</td>
</tr>
<tr>
<td>FEV1 z-score§</td>
<td>-1.00 (-3.21–0.44)</td>
<td>-1.57 (-2.98–0.34)</td>
<td>0.584</td>
</tr>
<tr>
<td>FVC (L)§</td>
<td>2.99 (2.31–3.59)</td>
<td>2.29 (1.72–3.72)</td>
<td>0.224</td>
</tr>
<tr>
<td>FVC % predicted§</td>
<td>105.9% (70.8–122%)</td>
<td>91% (73–122%)</td>
<td>0.454</td>
</tr>
<tr>
<td>FVC z-score§</td>
<td>0.26 (-1.99–1.26)</td>
<td>-0.63 (-1.54–0.97)</td>
<td>0.555</td>
</tr>
<tr>
<td>FEV1/FVC§</td>
<td>0.73 (0.50–0.80)</td>
<td>0.67 (0.51–0.74)</td>
<td>0.578</td>
</tr>
<tr>
<td>(FEV1/FVC)% predicted§</td>
<td>90.3% (68-98)</td>
<td>88% (68-98%)</td>
<td>0.633</td>
</tr>
<tr>
<td>FEV1/FVC z-score§</td>
<td>-0.63 (-2.30-0.19)</td>
<td>-1.6 (-3.02–0.10)</td>
<td>0.271</td>
</tr>
<tr>
<td>Airflow Obstruction*‡</td>
<td>9 (39%)</td>
<td>13 (59%)</td>
<td>0.181</td>
</tr>
<tr>
<td>Restrictive Airflow***‡</td>
<td>2 (9%)</td>
<td>1 (5%)</td>
<td>1.000</td>
</tr>
<tr>
<td>Inhaled respiratory medication‡</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SABA‡</td>
<td>16 (34%)</td>
<td>16 (43%)</td>
<td>0.389</td>
</tr>
<tr>
<td>LABA‡</td>
<td>19 (40%)</td>
<td>16 (43%)</td>
<td>0.795</td>
</tr>
<tr>
<td>LAMA‡</td>
<td>9 (19%)</td>
<td>13 (35%)</td>
<td>0.098</td>
</tr>
<tr>
<td>ICS‡</td>
<td>21 (45%)</td>
<td>17 (46%)</td>
<td>0.908</td>
</tr>
</tbody>
</table>

§ Median (IQR) □ n (%) □ p-value by Mann-Whitney ² p-value by Chi square ³ p-value by Fisher Exact
FEV1: Forced Expiratory volume in 1 sec  FVC: Forced Vital Capacity
* FEV1/FVC <0.7 **FEV1/FVC ≥0.7 and FVC < 80% predicted
5.3.2. Richness and diversity of bacterial communities

The PD-whole tree value was significantly lower in the group of patients who were on prophylactic antibiotic treatment (difference of means (SD)= 1.4 (0.56), p=0.014 by t-test) (Figure 5.1). Similarly, other richness and alpha diversity indices such as the observed number of OTUs, Shannon, Chao-1 and Fisher-alpha were numerically lower in the antibiotic prophylaxis group; however, these results did not achieve statistical significance (p>0.05 by t-test).

**Figure 5.1.** Comparison between α-diversity and richness of airway bacterial community in patients with chronic lung disease who were receiving antibiotic prophylaxis therapy (n= 47) and those who were not (n= 37) (A) α-diversity expressed by PD-whole tree (p=0.014 by t-test) and (B) richness expressed by total number of observed OTUs (p=0.078 by t-test).
Within the group of patients who were on prophylactic antibiotic therapy, those prescribed doxycycline, co-trimoxazole and cephalexin had the lowest Shannon α-diversity index (p=0.014 by KW) (Figure 5.2).

No significant differences were found in any of the α-diversity indices with respect to the chronic lung condition being bronchiectasis or COPD, immunodeficiency status, use of inhaled medication: SABA, LABA, LAMA or ICS, or administration of oral steroids. None of the spirometry parameters correlated with any α-diversity index.

**Figure 5.2.** Comparison between the Shannon α-diversity index within the group of patients receiving different regimes of antibiotic prophylaxis therapy (p=0.007 by KW), No AB: no prophylactic Antibiotic (n=37), AZM: Azithromycin (n=29), CLR: Clarithromycin (n=2), DO: Doxycycline (n=4), CIP: Ciprofloxacin(n=2), SXT: Co-trimoxazole(n=7), AMX: Amoxicillin, PV: Phenoxyethyl Penicillin, CEX: Cephalexin.
5.3.3. β-diversity

In the weighted UniFrac β-diversity index, antibiotic prophylaxis therapy was the only significant covariate (p=0.038 by PERMANOVA) (Figure 5.3). The chronic lung disease condition, immunodeficiency status, use of antibiotic rescue packs, sequencing runs, smoking, BMI, airflow obstruction and airflow restriction, severity of airflow obstruction based on both FEV1 z-score and FEV1/FVC ratio, use of inhaled respiratory medication: SABA, LABA, LAMA and ICS, oral corticosteroids, and carbocisteine were not significant covariates (all p-values >0.05 by PERMANOVA). Nevertheless, the samples did not cluster based on any of studied covariates in the clinical metadata: (all p-values resulted from ANOSIM test were >0.05).

In the unweighted UniFrac β-diversity index, antibiotic prophylaxis therapy (p-value= 0.012 by PERMANOVA and 0.02 by ANOSIM) (Figure 5.3) and chronic lung disease condition (diagnosis) (p-value =0.01 by PERMANOVA and 0.01 by ANOSIM), were significant covariates but not immunodeficiency status (p-value= 0.06 by PERMANOVA and 0.118 by ANOSIM). (Figure 5.4) SABA (p=0.01 by PERMANOVA), LABA (p=0.017 by PERMANOVA), carbocisteine (p=0.042 by PERMANOVA) were also significant covariates in the unweighted UniFrac β-diversity distance matrix.
Figure 5.3. PCoA plot of weighted UniFrac (A and C) and unweighted UniFrac (B and D), samples are coloured with respect to: (A and B) being on prophylactic antibiotic therapy (navy blue, n=47) or not (comparator group, yellow, n=37) (C and D) the prophylactic antibiotic regime: azithromycin (blue, n=29), clarithromycin (light blue, n=2), doxycycline (orange, n=4), ciprofloxacin (green, n=2), β-lactam (red, n=3) co-trimoxazole (purple, n=7).
Chapter 5: Microbiome and Antimicrobial Treatment

Figure 5.4. PCoA plot of weighted UniFrac (A and C) and unweighted UniFrac (B and D), samples are coloured with respect to: (A and B) chronic respiratory condition (blue: bronchiectasis, n=61), (red: COPD, n=23) (C and D) Immunodeficiency status pink: immunodeficient, n=38) (purple: immunocompetent, n= 46).
5.3.4. Bacterial community composition

At the phylum level, the relative abundance of the major phyla: *Firmicutes*, *Proteobacteria*, *Bacteroidetes*, *Actinobacteria*, and *Fusobacteria* were similar in both groups. However, the phylum *Synergistetes* which is a minor taxon present at RA <0.5% in all samples was significantly less abundant in the group of patients on prophylactic antibiotic therapy (p=0.00247) (Figure 5.5.C).

![Comparison between the average airway microbiome profile of patients on prophylactic antibiotic therapy (A) and those not (comparator group) (B) at the phylum level (p>0.05); (C) the RA of phylum Synergistetes was significantly lower in patients receiving antibiotic prophylaxis therapy (p=0.00247); yellow: comparator group (n=37), navy blue: antibiotic prophylactic group (n=47).](image-url)
The structure of bacterial communities in sputum, at the genus level, for patients who were receiving antibiotic prophylaxis therapy in comparison to the other participants, is presented in Figure 5.7. The following potentially pathogenic taxa were significantly less abundant in patients on prophylactic antibiotic therapy: *Pseudomonas* (p=0.027), *Enterobacteriaceae* (p=0.021), *Klebsiella* (p=0.046), *Pasteurella* (p=0.012), and *Morganella* (p <0.0001). The genera *Moraxella* and *Staphylococcus* appeared less abundant in the patients on prophylactic antibiotic therapy; however, the observed difference did not reach statistical significance (p=0.087 and 0.159 respectively by White t-test) (Figure 5.7). Some minor taxa with RA <1% such as *Bacteroides* (p=0.02) *Schwartzia* (p=0.001) *Sphaerochaeta* (p=0.01) were also significantly lower in the antibiotic prophylaxis group.
Figure 5.6. Airway microbiome profiles of patients with chronic lung disease who were receiving antibiotic prophylaxis therapy (n=47) in comparison to patients in the comparator group; not receiving antibiotic prophylaxis (n=37); NC: negative control, g_.: genus, f_.: family.
Figure 5.7. Suppression of potential pathogenic taxa in the sputum of patients on prophylactic antibiotic therapy (navy blue, n=47) compared to others in the comparator group (yellow, n=37) A. genus *Pseudomonas* (p=0.027) B. family *Enterobacteriaceae* (p=0.021), C. genus *Moraxella* (p=0.087), D. genus *Staphylococcus* (p=0.159).
5.3.5. Respiratory Pathogens

qPCR was used to compare the prevalence and loads of *P. aeruginosa*, *H. influenzae*, *S. pneumoniae* and *M. catarrhalis* between patients receiving antibiotic prophylaxis therapy and those not. *M. catarrhalis* was least detected among the four surveyed pathogens, it was significantly less frequent in the antibiotic prophylaxis group (4%) compared to 20% in the comparator group (p=0.039 by Fisher Exact test). The prevalence of *H. influenzae*, *S. pneumoniae* and *P. aeruginosa* was similar in both groups with no significant differences in the frequency of detection of any of these pathogens noticed between groups (all p>0.05 by Chi square test) (Figure 5.8A).

Co-existence of two respiratory pathogens was detected in 39% of the examined sputum samples (Figure 5.8B). The most common combinations of co-existence were *P. aeruginosa* and *H. influenzae* in 12% of samples (n=10) and *H. influenzae* and *S. pneumoniae* in 9% (n=8). The co-existence of the three pathogens *P. aeruginosa*, *H. influenzae* and *S. pneumoniae* were detected in 7% (n=7) of the samples. Four pathogens were detected in only one bronchiectasis patient who was not receiving prophylactic antibiotic treatment. However, there were no significant differences in the coexistence of these respiratory pathogens in the sputum of patients who were on antibiotic prophylaxis therapy and those not (p=0.203 by Fisher Exact test).

![Graph A](image1.png)  ![Graph B](image2.png)

**Figure 5.8.** Prevalence and Co-existence of *P. aeruginosa* (Pa), *H. influenzae* (Hi), *S. pneumoniae* (Spn) and *M. catarrhalis* (Mc) in patients who were receiving antibiotic prophylaxis therapy (AB: navy blue, n=46) and those not (comparator group, yellow, n=36) (A) Percentage of samples tested positive for the tested pathogens *p<0.05* (B) Number of organisms tested positive within the same sputum samples (p=0.203 by Fisher Exact).
The loads of *H. influenzae*, *S. pneumoniae* and *M. catarrhalis* were not significantly different between study groups (p= 0.189 by MANOVA) (Figure 5.9A) but the load *P. aeruginosa* was significantly reduced in the antibiotic prophylaxis group; the median load (IQR) was 2.89 (2.62-3.64) log$_{10}$ CFU/mL versus 7.23 (3.44-7.95) log$_{10}$ CFU/mL in the comparator group (p=0.001 by MW) (Figure 5.9B).

The bacterial density indicated by total 16S *rrna* gene load measured by qPCR was significantly lower in the patients who were receiving antibiotic prophylaxis (median (IQR)= 6.20 (5.54-6.73)), compared to the rest of the cohort (median (IQR)= 6.60 (6.17-6.92)), (p=0.029 by MW) (Figure 5.9C); even though bronchiectasis patients represented 80% of the total number of patients who were on prophylactic antibiotic treatment (Table 5.1) and the bacterial density was found to be significantly higher in bronchiectasis patients compared to COPD patients; the median (IQR) were 6.63 (0.80) versus 6.18 (1.47) log$_{10}$ copies/µL respectively (p=0.042 by MW) (Figure 5.9D). The loads and prevalence of the four respiratory pathogens were similar across COPD and bronchiectasis patients (all p>0.05 by Chi square test).

Comparing the immunodeficient patients with the others, no significant differences were noticed either in the prevalence or the loads of *H. influenzae*, *S. pneumoniae* and *M. catarrhalis*. The total bacterial density was also similar (p=0.387 by t-test). Nevertheless, *P. aeruginosa* was more frequently detected by qPCR in 65% of immunodeficient patients compared to 41% of the immunocompetent patients (p=0.033 by Chi square test). Nevertheless, the detected loads in the immunodeficient patients were significantly lower (p<0.0001 by MW).
Chapter 5: Microbiome and Antimicrobial Treatment

Figure 5.9. Bacterial Load of pathogenic airway bacteria in patients who were receiving antibiotic prophylaxis therapy (AB: navy blue, n= 46) and those not (comparator group, yellow, n=36) as determined by the multiplex qPCR; (A) Mean bacterial loads of *H. influenzae* (Hi), *M. catarrhalis* (Mc), *S. pneumoniae* (Spn) were similar in both groups (p=0.189, MANOVA). (B) *P. aeruginosa* load was significantly lower in the prophylactic antibiotic group (p=0.001 by MW) (C) total 16S rRNA load which is a proxy to bacterial density in the airways was significantly lower in the prophylactic antibiotic group (p=0.029 by MW) (D) total 16S rRNA load in the airways was significantly higher in the bronchiectasis patients (blue, n=50) compared to COPD patients (red, n=23) (p=0.042 by MW). Bacterial load was measured in CFU/mL of original sputum sample for each bacterium, bacterial density was measured in number of 16S rRNA gene copies per µL of reaction mixture. Samples which gave negative results for a given bacteria were excluded from the analysis. Error bars show ±1 SD.*p<0.05, **p<0.01.
5.3.6. Ciprofloxacin treatment course and *P. aeruginosa* eradication

In the cohort, 18 participants (38%) had previously received an intensive course of ciprofloxacin (750 mg twice daily for 2 weeks) to eradicate *P. aeruginosa* airway colonization. In all except one patient, *P. aeruginosa* was still detected by qPCR and/or 16S rRNA sequencing. Nevertheless, in 56% (10 participants) the *Pseudomonas* relative abundance was greatly diminished (RA<1%) and the qPCR detected loads were < 2.5 logs; it is noteworthy that eight participants out of the ten were on prophylactic antibiotic treatment.

![Figure 5.10.](image)

**Figure 5.10.** The relative abundance of *Pseudomonas* in sputum of patients who had received high dose of ciprofloxacin treatment course for the eradication of *P. aeruginosa* colonization (green, n=18) compared to others in the cohort (blue, n= 66) (p=0.368).

5.3.7. Covariates in microbiome data

The chronic lung disease condition was a significant covariate in unweighted UniFrac β-diversity index (p-value =0.01 by PERMANOVA and 0.01 by ANOSIM) (Figure 5.4). At phylum level, there were slight differences between COPD and bronchiectasis patients. A trend was observed in which the bacterial community composition shifted towards more *Proteobacteria* (p=0.041) in bronchiectasis patients while in COPD patients it shifted towards more *Firmicutes* (p=0.048). *Bacteroidetes* was slightly lower in the COPD patients (p=0.049) (Figure 5.11A). Similarly, airflow obstruction indicated by spirometry parameters was associated with reduced RA of phylum *Bacteroidetes* (p=0.046) in the current data (Figure 5.11B). At the genus level, *Streptococcus* (p=0.04) was more abundant in COPD; while *Klebsiella* (p=0.003), *Burkholderia* (p=0.027) and family *Gemellaceae*...
(p=0.024) were significantly more abundant and/or frequent in bronchiectasis (Figure 5.12).

Figure 5.11. Comparison between the airway microbiome profiles of COPD and bronchiectasis patients at phylum level (A) In COPD, *Firmicutes* (p=0.048) was significantly higher while *Bacteroidetes* was lower (p=0.049). On the other, hand *Proteobacteria* (p=0.041) was significantly higher in bronchiectasis (B) *Bacteroidetes* was significantly lower in patients with airflow obstruction (p=0.046) (purple (group 1): airflow obstruction, n= 22), (green (group 0): no airflow obstruction, n= 23).
Figure 5.12. Differential abundance of the following taxa in the airways of COPD (red, n=23) and bronchiectasis patients (blue, n=61) (A) *Streptococcus* (p=0.04) more abundant in COPD; whereas, in bronchiectasis, (B) Family *Gemellaceae* (p=0.029), (C) *Klebsiella* (p=0.003) (D) *Burkholderia* (p=0.032) were more abundant and/or frequent.
SABA (p=0.01 by PERMANOVA), LABA (p=0.017 by PERMANOVA), carbocistene (p=0.042 by PERMANOVA) were significant covariates in unweighted UniFrac β-diversity index. SABA was associated with increased RA of *Prevotella intermedia* (p=0.009), while, LABA was associated with reduced RA of *Neisseria* (p=0.044) (Figure 5.13). Carbocistene was associated with significant reductions in the RA of the following taxa: phylum: *Spirochaetes* (p=0.026); particularly genus *Treponema* (p=0.029), *Porphyromonas* (p=0.013), *Aggregatibacter* (p=0.024), *Butyrivibrio* (p=0.016) and *Pseudoramibacter_Eubacterium* (p=0.012) (Figure 5.14).

Though Immunodeficiency was not statistically significance in multivariate analysis of the unweighted UniFrac β-diversity distance matrix (p=0.06 by PERMANOVA, 0.118 by ANOSIM), upon comparing the microbiome profiles of immunodeficient patients with that of the immunocompetent patients in the cohort, the following minor taxa were found to be differentially abundant. *Finegoldia* (p=0.003) was more abundant and frequent in the immunocompromised patients. While, *Prevotella intermedia* (p=0.022), *Bulleidia* (RA<0.5%, p=0.035), *Sphaerochaeta* (RA<1%, p=0.046) and *Filifactor* (RA<0.5%, p=0.020) were less abundant in the immunocompromised group (Figure 5.15)

**Figure 5.13.** Differentially abundant taxa associated with inhaled respiratory medication

(A) SABA was associated with increased RA of *Prevotella intermedia* (p=0.009) (SABA users (group 1): green, n=32, and comparator (group 0): blue, n=52), (B) LABA was associated with reduced RA of *Neisseria* (p=0.044), (LABA users (group 1): green, n=35, and comparator (group 0): blue, n=49).
Figure 5.14. Differentially abundant taxa associated with the use of mucolytic agent carbocisteine on microbiome.

(A) *Treponema* (p=0.029),
(B) *Porphyromonas* (p=0.013),
(C) *Aggregatibacter* (p=0.024),
(D) *Butyribrio* (p=0.016)
(E) *Pseudoraminibacter_Eubacterium* (p=0.012)

Carbocisteine users (green, n= 17)
No Carbocisteine (light blue, n= 67)
Figure 5.15. Differentially abundant taxa in immunocompromised patients (A) *Prevotella intermedia* (p=0.022), (B) *Bulleidia* (p=0.035) (C) *Sphaerochaeta* (p=0.046), (D) *Filifactor* (p=0.020), (E) *Finegoldia* (p=0.003). pink: immunocompromised (n = 38), purple: immunocompetent (n = 46).
5.3.8. Bacteriological Culture results

In total, 187 distinct bacterial isolates were isolated from sputum cultures and identified by MALDI-TOF (an example of MALDI-TOF result report in Appendix 3.5). Several species were repeatedly isolated from the serial sputum samples collected from the same patient; therefore, for the current analysis repeated isolates was just included once.

The most frequently isolated species in > 90% of samples which are conventionally regarded as commensal bacteria, were viridans streptococci such as: *S. mitis*, *S. parasanguinis*, *S. sanguinis*, *S. gordonii*, *S. oralis*, *S. anginosus*, *S. constellatus*, *S. salivarius*, *S. vestibularis*, and *S. cristatus*; in addition to *Rothia* species most commonly: *R. mucilaginosa* and *R. dentocariosa*, and *Neisseria* species such as: *N. subflava*, *N. flavescens* and *N. perflava*. Less frequently, commensal *Haemophilus* species (30% of samples) such as *H. parainfluenzae*, *H. haemolyticus* and *H. parahaemolyticus* were isolated from sputum. The most frequently isolated species that are regarded as respiratory pathogens were *P. aeruginosa* (22% of samples), *H. influenzae* (17%), *S. pneumoniae* (13%), *M. catarrhalis* (9%), *Klebsiella* species (7%) particularly: *K. pneumoniae*, *K. oxytoca*, *K. variicola*, and *K. aerogenes*, *E. coli* (7%), *S. pseudopneumoniae* (3%), *Proteus mirabilis* (3%) and *S. aureus* (3%). Other less commonly isolated bacteria were: *Neisseria meningitidis*, *Neisseria macacae*, *Neisseria elongata*, *Moraxella bovis*, *Morganella morganii*, *Achromobacter xylosoxidans*, *Corynebacterium striatum*, *Pasteurella canis*, *Aggregatibacter segnis*, *Citrobacter koseri*, *Staphylococcus epidermidis*, *Staphylococcus hominis*, *Enterococcus faecalis*, *Candida albicans* and *Candida tropicalis*.

The prevalence of the bacterial species that were isolated from the sputum of 68 patients with chronic lung disease, 36 of them were receiving antibiotic prophylaxis therapy and 32 not (comparator group) are presented in Figure 5.16. Although there were no significant differences in the prevalence of *P. aeruginosa* and *H. influenzae* between both study groups, the variety and frequency of Gram
negative bacteria isolated from the sputum of patients on-prophylactic antibiotic therapy were significantly less compared to those not receiving prophylactic antibiotics in the comparator group; 26% versus 48% of the Gram negative bacterial isolates respectively (p=0.011 by Chi square).

**Figure 5.16.** The bacteria isolated from the sputum of patients with chronic lung disease (COPD and bronchiectasis) and their prevalence among the patients who were receiving antibiotic prophylaxis therapy (n=36) and those not (comparator group, n=32).
5.3.9 Antimicrobial susceptibility of bacterial isolates

Generally, resistance to azithromycin, tetracycline and ampicillin was high among most tested isolates: the pathogenic bacteria and the commensal viridans streptococci.

Among pathogenic bacteria, azithromycin resistance was detected in 64% of Gram positive isolates and 29% of the Gram-negative isolates. Tetracycline resistance was detected in 45% of Gram positive isolates and 34% of the Gram-negative isolates. Ampicillin resistance was detected in 10% of Gram positive isolates and 67% of the Gram-negative isolates. Ciprofloxacin and cefotaxime (third generation cephalosporin) resistance was detected in 21% and 17% respectively of the Gram-negative isolates only.

The prevalence of AMR was significantly higher among *P. aeruginosa* (p=0.03 by Fisher exact test) and Gram positive bacteria (p=0.014 by Fisher exact test) isolated from the group of patients on-prophylactic antibiotics compared to the others in the comparator group. Specifically, *P. aeruginosa* showed significantly higher AMR to fluoroquinolones; ciprofloxacin and levofloxacin, and 3rd generation cephalosporins (cefotaxime). Whereas, the Gram-positive isolates showed significantly higher AMR to macrolides (azithromycin) and tetracyclines (tetracycline) (Figure 5.17). There were no significant differences in the prevalence of AMR among Gram-negative isolates within the two groups (p=0.390 by Chi square test) (Figure 5.18). Nevertheless, the observed AMR in the pathogenic isolates was associated with the corresponding prescribed antibiotics in the rescue packs for self-treatment of acute exacerbations (p=0.024 by Fisher test) (Figure 5.19 A).

In the commensal viridans streptococci isolates, resistance to multiple antibiotics such as: azithromycin, levofloxacin, tetracycline, cefotaxime and vancomycin was common; however, it was significantly more frequently detected in the antibiotic prophylaxis group compared to the comparator group (p<0.0001 by Chi Square test) (Figure 5.19B).
A.

<table>
<thead>
<tr>
<th>AB</th>
<th>S. pneumoniae</th>
<th>S. aureus</th>
<th>E. faecalis</th>
</tr>
</thead>
<tbody>
<tr>
<td>AZM</td>
<td>R</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>E</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>CIP</td>
<td>S</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>TE</td>
<td>R</td>
<td>S</td>
<td>R</td>
</tr>
<tr>
<td>AMP</td>
<td>S</td>
<td>S</td>
<td>R</td>
</tr>
<tr>
<td>CTX</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>C</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>RIF</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>VA</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>OX</td>
<td>R</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>DA</td>
<td>R</td>
<td>R</td>
<td>R</td>
</tr>
</tbody>
</table>

Figure 5.17. Comparison between the AMR prevalence in the Gram positive and P. aeruginosa sputum isolates from patients receiving antibiotic prophylaxis therapy and those not (A) Antibiogram of Gram positive bacteria: S. pneumoniae (n=8), S. aureus (n=2) and E. faecalis (n=1) (p=0.014 by Fisher exact test) and (B) Antibiogram of P. aeruginosa isolates (n=14) (p=0.03 by Fisher exact test); antibiotic prophylaxis group (Blue) and the comparator group (orange). R: resistant (red), I: intermediate sensitivity (peach), S: sensitive (green), AB: antibiotic, AZM: azithromycin, E: erythromycin, CIP: ciprofloxacin, LEV: levofloxacin, TE: tetracycline, AMP: ampicillin, TZP: piperacillin/tazobactam, CTX: cefotaxime, CAZ: ceftazidime, MEM: meropenem, C: chloramphenicol, RIF: rifampicin, VA: vancomycin, OX: oxacillin, DA: clindamycin, AK: amikacin, grey shade: data not available.
**Chapter 5: Microbiome and Antimicrobial Treatment**

**Figure 5.18.** Comparison between the AMR prevalence in the Gram negative sputum isolates from patients receiving antimicrobial prophylaxis and those not. Each column represents antibiogram of a Gram negative isolate: H. influenzae (Hi, n=12), H. parainfluenzae (H. para, n=3), Pasteurella canis (Pc., n=1), M. catarrhalis (Mc, n=5) Proteus mirabilis (Pr, n=2), Citrobacter koseri (Ck, n=1), K. pneumoniae (K.pn, n=2), K. variicola (Kv, n=1), K. oxytoca (Ko, n=1), K. aerogenes (K.aer, n=1), M. morganii (M. mor, n=1): isolated from the sputum of chronic lung disease patients who were receiving antimicrobial prophylaxis (Blue) and those not (orange) (p=0.390 by Chi square test)

<table>
<thead>
<tr>
<th>Antibiotics</th>
<th>H. influenzae</th>
<th>H. para</th>
<th>Pc</th>
<th>Mc</th>
<th>M. catarrhalis</th>
<th>E. coli</th>
<th>E. coli</th>
<th>Pr</th>
<th>Pr</th>
<th>Ck</th>
<th>K. pn</th>
<th>Kv</th>
<th>Ko</th>
<th>K. aer</th>
<th>Mn</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIP</td>
<td>R S S S S S S S S S R S S S S R R R I S S S R S S S S S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TE</td>
<td>S R S S S S S S I S S S S R R R R R R R R R R R R R R R R R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AZM</td>
<td>R S S S S S S R S S S S R R R R R R R R R R R R R R R R R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMP</td>
<td>S R R R R R R R R S S S S R R R R R R R R R R R R R R R R R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMC</td>
<td>S S S S S S S S S S S S S S S S S S S S S S S S S S S S S S S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAZ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CXT</td>
<td>S S S R S S S S S S S S S S S S S S S S S S S S S S S S S S S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEM</td>
<td>S S S S S S S S S S S S S S S S S S S S S S S S S S S S S S S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R: resistant (red), I: intermediate sensitivity (peach), S: sensitive (green), grey shade: data not available.

Chapter 5: Microbiome and Antimicrobial Treatment

Figure 5.19. (A) Association between the AMR in pathogenic sputum isolates and the prescribed rescue antibiotics for the self-treatment of acute exacerbations (p=0.024 by Fisher test) (B) Comparison between the AMR in the commensal streptococci viridans group isolated from the sputum of patients receiving antibiotic prophylaxis therapy (group 1, n=18) and those not (comparator group) (group 0, n=24) (p<0.0001 by Chi Square test), the prevalence: percentage of the resistant sputum bacterial isolates, S: Sensitive (green), R: resistant (red), DO: doxycycline, FQ: fluoroquinolone, AZM: Azithromycin, LEV: Levofloxacin, TE: Tetracycline, AMP: Ampicillin, CTX: Cefotaxime, VA: Vancomycin.
5.4. Discussion

In this chapter, the impact of the prolonged antibiotic prophylaxis therapy on the airway microbiome and AMR in sputum bacterial isolates has been evaluated. Most participants (62%) in the antibiotic prophylaxis group were on an intermittent azithromycin regime (thrice weekly). However, other prophylactic antibiotic regimes like continuous and intermittent co-trimoxazole (15%) and continuous: doxycycline (9%), β-lactam (7%), clarithromycin (4%) and ciprofloxacin (4%) were also included. Most participants (90%) in the antibiotic prophylaxis group were on the prophylactic antibiotics prescribed for over a year prior to joining the study.

The antibiotic prophylaxis therapy was associated with lower α-diversity of the airway bacterial community and decline in the total bacterial density in sputum. The reduced α-diversity was significantly reflected in the PD-whole tree index which is a quantitative measure of the phylogenetic diversity within an ecosystem. (Faith and Baker, 2007) Doxycycline, co-trimoxazole and cephalexin showed the lowest Shannon α-diversity indices (although sample numbers here were low). This agrees with the findings of two small clinical trials; the first (Segal et al., 2017) was involving 20 COPD patients with emphysema randomised to receive either 250 mg of azithromycin daily or placebo for 8 weeks, in which azithromycin reduced α-diversity in BAL samples significantly, decreasing 11 low-abundant taxa compared to the placebo control group; however, no change in the total bacterial burden was found in this study. The second (Slater et al., 2014) reported that the richness and Shannon alpha diversity significantly declined in BAL samples post-six-weeks of 250 mg daily azithromycin in five patients with moderate/severe asthma.

In weighted UniFrac β-diversity, the samples did not cluster based on any of the studied covariates (all ANOSIM p-values >0.05), nevertheless, antibiotic prophylaxis therapy remained the only significant covariate (PERMANOVA p-value =0.038). PERMANOVA is a non-parametric multivariate analysis test that partitions a distance matrix among sources of variation in order to describe the strength and significance that a categorical covariate has in determining variation of distances; while, ANOSIM is a non-parametric hypothesis test that tests whether two or more
groups of samples are significantly different based on a categorical covariate in the metadata. (Jari Oksanen et al., 2018)

The microbiome profiles at the phylum level were similar within the two study groups. However, Synergistetes which is minor phylum whose RA was <0.5% was significantly less abundant in the antibiotic prophylaxis group. Synergistetes is a recently recognised bacterial phylum which has not been adequately characterised due to culture limitations, and most of our knowledge about its members originates from culture independent studies. However, currently this phylum encompasses 40 Gram negative anaerobic rod/vibrioid shaped bacteria which can ferment amino-acids and whose cell wall lacks lipopolysaccharides. It has been detected in various body site microbiomes such as the oral cavity, gut, umbilicus and vagina. Since its presence has been associated with disease in sites such as periodontitis, abscesses and cysts; Synergistetes are considered opportunistic pathogens. (Bhandari and Gupta, 2012)

Moving down the hierarchy of taxonomy towards higher resolution, there were no significant differences in the relative abundances of any of the genera that would normally defined as microbiota, between the patients on antibiotic prophylaxis therapy and those not. Similarly, Rogers et al. reported that even though significant within-patient change in microbiota composition was found in the treatment arm (erythromycin group) over the trial period, when change in microbiota composition was compared between groups rather than within individuals, there were no significant differences between the treatment and placebo arms. (Rogers et al., 2014b)

A trend was observed in which some potentially pathogenic taxa belonging to the class Gamma-Proteobacteria such as Pseudomonas, Enterobacteriaceae, Klebsiella, Pasteurella, and Morganella were significantly less abundant in the patients receiving antibiotic prophylaxis therapy compared to those not. Moraxella and Staphylococcus were also less abundant in the antibiotic prophylaxis group; however statistical significance was not achieved with these two taxa. Nevertheless, specific qPCR confirmed that M. catarrhalis was significantly less
prevalent in the antibiotic prophylaxis group. The load of *P. aeruginosa* measured by qPCR was also significantly lower in the patients on prophylactic antibiotic therapy. However, the prevalence and load of both *H. influenzae* and *S. pneumoniae* were similar in both groups. In addition, in bacterial cultures, pathogenic bacteria; especially the Gram-negative species, were significantly less frequently isolated from the sputum of patients on prophylactic antibiotic therapy compared the other participants.

The current findings are in line with the microbiological outcomes reported in clinical trials that evaluated the efficacy of antibiotic prophylaxis therapy in various chronic lung disease where the respiratory colonization with some typical respiratory bacterial pathogens was eliminated or inhibited post the prophylactic antibiotic course. It is noteworthy that these trials were applying different antibiotic regimes and varying durations; in most studies, microbiology was a secondary outcome rather than being a primary outcome and many of them their results were based on culture-dependant techniques which are less comprehensive compared to the microbiome approach and can be insensitive to detect bacteria present at low loads.

In a large American study involving 1142 participants, the acquisition of new respiratory pathogens such as *S. aureus*, *S. pneumoniae*, and *Moraxella* species in the nasopharynx was significantly less among the COPD patients in the treatment arm who had received 250 mg daily azithromycin for one year compared to those in the placebo control group. (Albert et al., 2011) Similarly, in the Dutch COLOMBUS study, the COPD patients in the treatment group receiving 500 mg azithromycin thrice weekly for 12 months were less likely to become colonised with the respiratory pathogens: *H. influenzae*, *S. pneumoniae* and *P. aeruginosa* compared to those in the placebo group. (Uzun et al., 2014) Also Berkhof et al. reported a reduction of the carriage of respiratory pathogens among the COPD patients who had received 250 mg azithromycin thrice weekly for 12 weeks. (Berkhof et al., 2013)

In antimicrobial susceptibility testing, *P. aeruginosa* and Gram-positive pathogens isolated from the sputum of patients in the antibiotic prophylaxis group showed
significantly higher AMR to (fluoroquinolones and 3rd generation cephalosporins) and (macrolides and tetracyclines) respectively compared to the corresponding isolates from the comparator group. In the Gram-negative isolates, the AMR frequencies were comparable between both groups; however, the prevalence of penicillin (54%), tetracycline (34%), and macrolide (29%) resistance was high among the Gram negative isolates. The observed AMR was linked to antibiotic-rescue packs prescribed for self-treatment of exacerbations in which β-lactams and fluoroquinolones were prescribed in 62% and 14% of cases.

In literature, there were contradictory findings regarding the development of AMR in repose to the antibiotic prophylaxis therapy. In Albert et al., the incidence of resistance to macrolides in newly acquired pathogens during the trial was significantly higher in the azithromycin arm; almost the double (81% versus 41%); nevertheless, it was reported in the same study that the acquisition of new respiratory pathogen was significantly less in the treatment arm. (Albert et al., 2011) Similarly, the Dutch BAT clinical trial involving 250 mg daily azithromycin for one year in bronchiectasis patients reported an increase in macrolide resistance expressed by respiratory pathogens in the treatment arm compared to placebo (88% versus 26%). (Altenburg et al., 2013) The Brill et al. study, included 99 COPD patients randomised between three treatment arms moxifloxacin 400 mg daily for 5 days every 4 weeks, doxycycline 100 mg daily, and azithromycin 250 mg thrice weekly or one placebo tablet daily for 3 months. The mean inhibitory concentrations of cultured isolates including commensal microbiota increased by at least three times over placebo in all treatment arms. (Brill et al., 2015a)

On the other hand, the COLOMBUS study found that the acquisition of new macrolide resistant bacteria at the end of the trial was significantly less in the treatment group compared to the placebo control group. (Uzun et al., 2014) Seemungal et al. found that there were no significant differences in the detection rates of the antibiotic resistant H. influenzae, S. pneumoniae and M. catarrhalis in COPD patients across the treatment arm (250 mg Erythromycin twice daily for 12 months) and the placebo control arm at any time during the follow up time points. (Seemungal et al., 2008)
Some studies have suggested that the resistance acquired during the antibiotic prophylaxis therapy might be temporary. Sethi et al. applied intermittent pulsed therapy with moxifloxacin (400 mg once daily for 5 days which was repeated every 8 weeks for a total of six courses) in COPD patients over a 48-week treatment period and further 24 weeks’ follow-up, a reduction in the isolation of pathogenic bacteria was noted in the treatment arm compared to the placebo arm. Although, this regime was associated with increased MICs in some S. pneumoniae, S. aureus and P. aeruginosa isolates; these resistant strains did not persist during the subsequent follow up period. Moreover, the median MIC of moxifloxacin for screened bacteria isolated from rectal swabs collected from participants in both arms did not reveal major consistent change. (Sethi et al., 2010) In a study assessing macrolide resistance in children with bronchiectasis who had received azithromycin 30 mg/Kg weekly for 24 weeks in the context of a previous multicentres randomized controlled trial in Australia and New Zealand, the carriage of H. influenzae and M. catarrhalis but not S. pneumoniae in the nasopharynx was reduced in the treatment arm. They reported that a major determinant in acquiring macrolide resistance was nonadherence to treatment. However, the resistance acquired during the course of treatment did not persist. (Hare et al., 2015)

In the data presented here, around 70% of macrolide resistant pathogens were isolated from patients who were on macrolide maintenance therapy, 41% of which were S. pneumoniae. This may suggest that Gram positive bacteria easily acquire resistance to macrolides. In a previous study, macrolide use was associated with macrolide resistance in sputum isolates of S. pneumoniae which persisted for 6 months after administration, this was four times that in the unexposed patients (54 versus 14%). (Desai et al., 2010)

In general, the commensal viridans streptococci expressed varying degrees of AMR to multiple antibiotics in both the antibiotic prophylaxis group and the comparator group. The most frequently encountered AMR was against azithromycin, tetracycline and ampicillin. The multi-drug resistant viridans streptococci were significantly more abundant in patients receiving antibiotic prophylaxis. Macrolide resistance was detected in 95% of the tested viridans
streptococci isolates in the antibiotic prophylaxis group versus 75% in the comparator group. *Streptococcus parasanguinis*, *S. oralis* and *S. mitis* were the most abundant MDR species. Previous studies confirmed that the oropharyngeal carriage of macrolide resistant viridans streptococci is common in populations; it was estimated as 71% in a Belgian cohort of healthy adults and 94% in another Spanish cohort. In the Belgian study, co-resistance to tetracycline was identified in 73% of the isolates where *Streptococcus mitis* was the most abundant macrolide resistant species in this study. (Malhotra-Kumar et al., 2004, Prez-Trallero et al., 2001)

In a previous trial, a direct causal effect was revealed between macrolide use and the carriage of macrolide resistant viridans streptococci in the in nasopharynx, where a single treatment course with either azithromycin (500mg once daily for 3 days), or clarithromycin (500 mg twice daily for seven days) increased the proportion of macrolide-resistant streptococci compared to the placebo arm for up to 180 days later. (Malhotra-Kumar et al., 2007) In the BLESS study, it was reported that macrolide resistant streptococci were enriched in the oropharynx by a median of 28% post one year of erythromycin prophylaxis therapy; this was significantly higher compared to the placebo arm. (Serisier et al., 2013) Furthermore, Taylor *et al.* who studied the acquired resistance in microbiota post a course of intermittent azithromycin therapy for 48 weeks in patients with moderate to severe asthma found that macrolide resistance were enriched among microbiota; specifically among viridans streptococci isolated from sputum; whole genome sequencing revealed that macrolide resistance was caused by transmissible genes which were carried with tetracycline resistance determinants on transposable elements. (Taylor et al., 2019b) Commensal streptococci were previously shown to carry the same macrolide resistance genes as pathogenic streptococci. (Stadler and Teuber, 2002) Therefore, such AMR commensal streptococci may be regarded as a reservoir for AMR in the microbial ecosystem. As a part of the work presented, the resistome of a subset of samples was studied in a pilot study using a metagenomic approach to investigate the underlying
molecular determinants of the observed AMR (data are presented later in Chapter 8).

Airway colonization (based on repeated sputum culture) with *P. aeruginosa* is often seen in advanced stages of various chronic lung disease conditions such as cystic fibrosis, bronchiectasis, COPD and asthma and has always been associated with deterioration in lung function. Once colonization with *P. aeruginosa* is established it is very hard to eradicate. (Yum et al., 2014, Rogers et al., 2014a) Therefore, clinicians opt to treat *P. aeruginosa* infection as soon as the bacteria is detected in sputum with an intensive prolonged ciprofloxacin course in an attempt to eradicate the bacteria and halt colonization. (Hurley and Smyth, 2012) In our cohort, the intensive ciprofloxacin course did not eradicate *P. aeruginosa* in those patients who had received it as a part of their routine health-care. Nevertheless, 56% of the patients who had received the intensive ciprofloxacin course had very low RA <1%, 80% of them were also on prophylactic antibiotics: doxycycline, ciprofloxacin, co-trimoxazole, and azithromycin.

The Australian BLESS study reported that erythromycin 400mg twice daily for 12 months promoted displacement of *H. influenzae* by more macrolide-tolerant pathogen like *P. aeruginosa* in a subset of bronchiectasis patients whose microbiome was not dominated by *Pseudomonas*. (Rogers et al., 2014b) In contrast, the prevalence of both *P. aeruginosa* and *H. influenzae* were similar in both patients' groups who were receiving antibiotic prophylaxis therapy and those not in the data presented. Furthermore, the load of *P. aeruginosa* but not *H. influenzae* was lower in the antibiotic prophylaxis group. In addition, other trials involving azithromycin maintenance therapy in bronchiectasis and other chronic lung diseases did not report similar antibiotic induced bacterial succession by more antimicrobial tolerant bacteria. (Altenburg et al., 2013, Berkhof et al., 2013, Fan et al., 2015, Slater et al., 2014, Taylor et al., 2019b)

CVID is a heterogeneous form of primary immunodeficiency in which the immunoglobulins are under-produced in the body, patients with CVID are prone to recurrent respiratory infection, which can result in progressive bronchiectasis. (Sperlich et al., 2018) Most of the CVID patients included had bronchiectasis (87%)
and were on immunoglobulin replacement therapy. Immunodeficiency was insignificant covariate in the multivariate analysis of weighted UniFrac; however, in unweighted UniFrac p-value by PERMANOVA was close to statistical significance (p=0.06). The airway microbiome profile of the patients with chronic lung disease with an underlying immunodeficiency showed slight differences in the relative abundances of a few minor taxa; particularly, *Finegoldia* was more abundant; whereas, *P. intermedia*, *Bulleidia*, *Sphaerochaeta* and *Filifactor* were less abundant in the immunocompromised patients compared to the immunocompetent patients with chronic lung disease. These are all anaerobic bacteria and all were absent in the negative controls. These taxa have been detected and considered part of airway and oral microbiota in other studies. (Wu et al., 2013) Very little is known about the characteristics, virulence and interactions of these genera in the microbial communities. *Finegoldia* species; particularly *F. magna*, is a common Gram positive anaerobic cocci that normally colonizes the oral cavity and skin surfaces; they are regarded as opportunistic pathogens; especially in chronic wounds and immunocompromised hosts. *Finegoldia* species can induce a pro-inflammatory response and oxidative burse in neutrophils. (Ariane et al., 2020) In the airway microbiome data of the PLW-HIV study (Chapter 7), *Finegoldia* was absent; whereas, the other less abundant taxa in the immunocompromised were all detected in the sequencing data and were present at significantly higher relative abundance in the airway microbiome of the healthy comparator group compared to patients with chronic lung disease. *P. aeruginosa* was significantly more frequently detected in 65% of immunodeficient patients; however, the loads in the immunodeficient patients were significantly low.

Although NTM isolation was reported in the medical records of five participants during the study period, no OTU has been identified as *Mycobacterium*. Therefore, V3-V4 16S rRNA sequencing may be an inappropriate method to capture and study tuberculous and non-tuberculous mycobacterial infections as the short reads of chunks of the variable regions V3-V4 16S rRNA gene lacks sufficient resolution to distinguish *Mycobacterium*. Similar observation was also reported in another study sequencing the V4 variable region of 16S rRNA. (Sulaiman et al., 2018)
Chapter 5: Microbiome and Antimicrobial Treatment

However, another airway microbiome study in tuberculosis study which sequenced V1-V3 could detect Mycobacterium, but the sensitivity was around 50%. (Wu et al., 2013)

It is worth noting that this is an observational study, therefore, participants were not randomly assigned to receive prophylactic antibiotics, and the decision to place a patient on antibiotic prophylaxis therapy was a clinical decision taken by the treating clinicians prior to joining the study. Therefore, some differences between the two study groups exist such as higher proportions of immunocompromised patients, and antibiotic rescue pack prescription were within the antibiotic prophylaxis group. A relatively higher proportion of participants had bronchiectasis compared to COPD; nevertheless, similar proportions of bronchiectasis and COPD patients were present in both study groups. In the cohort, some participants (14%) had overlapping obstructive lung disease and bronchiectasis; however, the predominant condition in each case (defined by Prof. Hurst) has been utilised in the present analysis. Apart from azithromycin, the small number of participants in the other antibiotic prophylaxis regimes did not allow the comparisons between different antibiotics. All these sources of variability might have masked some significant trends in the data.

In addition to being on prophylactic antibiotic treatment, some covariates were found to be significant cofounders in unweighted β-UniFrac but not weighted β-UniFrac. These included the diagnosis of chronic lung disease, SABA, LABA and carbocisteine use. Confounder analysis has been a challenge in microbiome studies; however, stratification analysis has been widely accepted tool to address this issue within the microbiome research community. (Allaband et al., 2019, Kim et al., 2017) As unweighted β-UniFrac accounts on the presence and absence of the taxa regardless of the weight of their relative abundance; therefore, it is more driven by the minor taxa in the microbiome profile unlike the weighted β-UniFrac. Nevertheless, these confounders did not affect the key finding. Apart from Klebsiella and Pseudomonas, the other differentially abundant taxa between the patients who were receiving prophylactic antibiotics and those not were not affected. Klebsiella was more abundant in bronchiectasis patients compared to
COPD; however, it was less abundant in the antibiotic prophylaxis group; even though bronchiectasis patients represented 79% of the patients who were on prophylactic antibiotic treatment. *Pseudomonas* which was significantly lower in both immunocompromised patients and patients on-prophylactic antibiotics. This may be attributed to the fact that 81% of the CVID patients were prescribed antibiotic prophylaxis therapy. Other taxa were associated with different covariates and it is difficult to decide which factor is driving the observed change; for example, *Finegoldia* was more abundant in bronchiectasis (p=0.0004) as well as it was both more abundant and frequent in the immunocompromised patients (p=0.003), *Sphaerochaeta* was less abundant in on-antibiotic prophylaxis group (p=0.01), as well as, it was less abundant in the immunocompromised group (p=0.046), *Prevotella intermedia* was less abundant in the immunocompromised group (p=0.022) but it was significantly higher in short acting β-2 agonists inhaler users.

Carbocisteine is a mucolytic agent which changes the physiochemical properties of sputum such that it could be easily expectorated. It is recommended for patients with excessive sputum secretion as it helps with sputum clearance preventing it from stagnating in the airways which encourages bacterial proliferation. (Wilkinson et al., 2014) Carbocisteine was found to be a significant covariate which may have induced some changes in the microbial composition by suppressing some pathogenic and opportunistic bacteria in the airways.

The main finding in this chapter was that specific taxa that encompass potential bacterial pathogens were selectively suppressed in the antibiotic prophylaxis group without disrupting the homeostasis of the respiratory microbiota. One potential explanation may be attributed to the acquired and/or inherent AMR of the microbiota provides the resilience of the airway microbial community in response to the antimicrobial treatment. On the other side, macrolides are reported to have anti-inflammatory and immune-modulatory properties (as discussed in Ch.1, 1.6.1); therefore, the mechanism of action may be mediated through the extra-pharmacological activities rather than the antimicrobial activity solely. Segal *et al.* have provided a potential explanation; they found that certain inflammatory chemokines: CXCL1, TNF-α and IL 13 and 12p40 declined; whereas, some
bacterial metabolites including glycolic acid and indole-3-acetate were enriched in the BALs samples collected from the ten treated patients who had received intermitted azithromycin therapy (250 mg thrice weekly) for eight weeks compared to the placebo group. The researchers have demonstrated through ex-vivo experiments that the bacterial metabolites rather than azithromycin it-self were responsible for the inhibition of the chemokines production from lipopolysaccharides (LPS)-induced alveolar macrophage. (Segal et al., 2017) This explanation may provide a deeper insight into the mechanism of action of macrolides through modulating the metabolome of the microbiome which in turn mediate the proclaimed anti-inflammatory and immunomodulatory responses. Nevertheless, this hypothesis needs to be validated in future larger study.

In conclusion, antibiotic prophylaxis therapy was associated with reduced phylogenetic alpha diversity of the bacterial communities and lower bacterial density in sputum. It did not induce a definitive compositional shift in the airway microbiota at the cohort level. Nevertheless, the antimicrobial treatment suppressed certain pathogenic bacteria such as 
Pseudomonas, M. catarrhalis and some members of family Enterobacteriaceae but not H. influenzae and S. pneumoniae in the airways of patients who were receiving antibiotic prophylaxis therapy. This observation was confirmed by qPCR and bacteriological culture. This may suggest that prophylactic antibiotics, specifically, azithromycin, are not only acting through their antimicrobial activity which may account for the reductions in the total bacterial density in the airways, but also the current results may support that the proclaimed immune-modulatory effect of macrolides since it selectively suppressed specific taxa that encompass potential respiratory pathogens without disrupting the homeostasis of the respiratory microbiota. In general, macrolide resistance was high in the whole cohort especially among the Gram positive sputum isolates, it was significantly higher in patients receiving macrolide prophylaxis therapy. The detected AMR expressed by pathogenic bacterial isolates in general were associated with the prescribed antibiotics in the rescue packs kept by the patients for self-treatment of exacerbations but not the prophylactic antibiotics.
Chapter 6: Airway Microbiome in Adult Survivors of Extremely Pre-term Birth (EPICure Study)

6.1. Introduction

Pre-term birth has been associated with the development of bronchopulmonary dysplasia (BPD) which is a form of chronic lung disease that develops mostly in pre-term born infants and may lead to a life-long respiratory morbidity in many cases. Survival rates of extremely pre-term born infants; those who are born between 22 to 25 weeks gestation, have increased over the past decade with the current advances in the neonatal care; however, the prevalence of bronchopulmonary dysplasia remains high (around 68%). (Farstad et al., 2011, Costeloe et al., 2012, Jobe, 2011)

Preterm infants have structurally and functionally immature lungs and they require varying degrees of respiratory support. The mechanisms that lead to bronchopulmonary dysplasia are various and multi-factorial; however, lung inflammation seems to be the central mechanism in bronchopulmonary dysplasia pathogenesis. (Kramer et al., 2009, Gien and Kinsella, 2011) Previously, the classical form of bronchopulmonary dysplasia was characterized by marked pulmonary fibrosis and emphysema. However, the new more common form is defined by disrupted distal lung development; this is marked by arrest of alveolarization and interference with proper vascularisation. (Choi, 2010, Kramer et al., 2009, Gien and Kinsella, 2011) The long-term sequelae of bronchopulmonary dysplasia can be presented as a form of obstructive lung disease and reduced aerobic capacity in adulthood which is usually treated similarly as COPD. (Lovering et al., 2014)

The study presented here is a part of the large cohort study EPICure@19. EPICure was a national cohort study, of infants born at or less than 25 completed weeks gestation in the United Kingdom and Ireland between March and December 1995. This unique cohort has been followed up and assessed at 2.5, 6, 9, 19 years old to study the performance of the preterm birth survivors in a wide range of aspects
of growth and health later in life in comparison to their colleagues who were born full-term. (Costeloe et al., 2012, Fawke et al., 2010, Hennessy et al., 2008) At the age of 11 years follow up, the children who were born extremely preterm had significantly more chest deformities, respiratory symptoms, lung function abnormalities with evidence of airway obstruction, ventilation inhomogeneity, gas trapping and airway hyper-responsiveness. The asthma prevalence was double in this cohort compared to their classmates (control group). (Fawke et al., 2010, Lum et al., 2011)

There is a good evidence that bacteria may play a role in the pathogenesis of bronchopulmonary dysplasia in preterm born infants. (Stressmann et al., 2010, Mourani et al., 2011) A recent study revealed a characteristic pattern of airway microbial dysbiosis in the pre-term infants prior to the development of bronchopulmonary dysplasia. This was marked with a decrease in the richness and alpha diversity in the airway microbial community with time; in addition to a shift in the relative abundance of some genera in contrast to the relatively diverse and stable community in the pre-term infants who did not develop bronchopulmonary dysplasia. (Lohmann et al., 2014) However, it is not known to what extent the microbial succession sustains the airway microbial dysbiosis that was associated with the development BPD at infancy into later life stages and how much the airway microbiome of those who were born prematurely differ from the healthy microbiome in the adulthood. The results presented in this chapter are published in the “European Respiratory Journal” (DOI: 10.1183/13993003.01225-2018) (Rofael et al., 2019)

6.2. Methods

6.2.1. Ethics

The EPICure study was approved by the Southampton and South West Hampshire Research Ethics Committee (Reading, UK). (Fawke et al., 2010) All participants had given their consent for tests to be done on their biological samples in relation to the EPICure@19 Study. (Hurst et al., 2020)
6.2.2. Induced sputum samples collection

Induced sputum samples were collected from participants during their visits to University College London Hospital (UCLH), London in the period between January 2014 and March 2015 by the clinical team of the study. Sputum induction was carried out by allowing the participants to inhale nebulised hypertonic saline. Collected samples were held on ice until they were processed for freezing. The samples were split into aliquots and stored at -80°C in the UCL-Royal Free biobank. One aliquot of each sample was unprocessed and these were selected for the microbiome study. Samples were sorted into three groups by the clinical team in EPICure study based on medical history of participants; those who were born prematurely with no history of bronchopulmonary dysplasia; those who were born premature with history of developing neonatal bronchopulmonary dysplasia; and full term born controls. Neonatal BPD was defined as receiving supplemental oxygen or respiratory support at 36 weeks postmenstrual age. Investigators were blinded to the group assignment until analysis was complete. Then, clinical team provided the required metadata of participants for the interpretation of the microbiome data.

6.2.3. Sputum homogenisation and processing

The samples were removed from -80 °C freezers and allowed to thaw at room temperature before 250-500 µL of each sample was transferred to another tube for testing. The samples were treated with an equal volume of freshly diluted Sputasol before they were heated at 95 °C for 30 minutes.

DNA was extracted from 500 µL of Sputasol treated sputum samples using Qiagen DNeasy® Blood and Tissue kit (Qiagen, UK) (Ch.2, 2.6.2.).

Bacterial density estimated by the total number of copies of 16S rna gene, typical respiratory pathogens and P. aeruginosa were detected in sputum using qPCRs (Ch.2, 2.10. Multiplex quantitative polymerase chain reaction (qPCR) for Respiratory pathogens).
Chapter 6: Airway Microbiome of EPICure Study

The V5-V7 regions of the 16S rRNA gene were amplified using the 785 forward and 1175 reverse primer set and amplicons were sequenced on Illumina® MiSeq platform (Ch.2, .2.8.). The samples were processed and sequenced in two runs: the first batch included 40 samples processed by Miss Rachel Troughton as a part of her M.Sc. dissertation (Troughton R. et al., 2015) and the second included 52 samples processed by myself. Out of 92 processed samples, 74 samples successfully amplified V5-V7 variable regions of 16S rRNA gene: 25 samples in the first batch and 49 samples in the second.

The bioinformatic and data analysis of the whole cohort was done by myself, for α and β diversity analyses the biom table was rarefied at the depth of 4000 reads per sample removing four samples with <1000 reads (1 EP with BPD and 3 in the comparator group). All p-values of the differentially abundant taxa were corrected for multiple comparisons using Benjamini-Hochberg False Discovery Rate (FDR) method.

Statement of Contribution: I processed the sputum samples, prepared 16S rRNA sequencing library and performed qPCRs on the second batch of samples. I did the bioinformatic analysis on all samples and generated the results presented in this chapter.

6.3. Results

6.3.1. General Characteristics of the EPICure Study groups

Induced sputum samples were collected from 92 adult participants; mean age 19 year (range: 18 yr: 1 mo to 20 yr: 4 mo). The ex-premature (EP) participants were born extremely pre-term at <26 weeks gestation, while the comparator group included adolescents of the same age who were born full term. The samples were grouped into three groups by EPICure clinical team based on the birth and clinical history of the participants; 37 were extremely preterm born with history of neonatal bronchopulmonary dysplasia (EP+BPD), 14 were extremely preterm born with no history of bronchopulmonary dysplasia (EP no BPD) and 41 were full term born. However, since microbiome analysis was completed on 74 samples: 36 from the
Chapter 6: Airway Microbiome of EPICure Study

comparator group and 38 from the EP group of whom 29 had neonatal BPD. Therefore, the demographics and the clinical characteristics of the whole cohort and the subset involved in the microbiome analysis are presented in Table 6.1.

There were no significant differences between the study groups with respect to the sex of participants, smoking status, passive smoke exposure (>30 mins/week), respiratory events that required treatment or antibiotic administration in the year prior to samples collection, prescribed inhaled medication: SABA, LABA, LAMA, ICS or leukotriene receptor antagonist. Asthma diagnosis was self-reported in 43% participants. The prevalence of self-reported asthma was relatively high in the EP group with BPD (59%) and controls (38%). At the time of sample collection, the mean fractional exhaled nitric oxide (FeNO) concentration (SD) was 16.59 (14.10) and 25.57 (27.71) ppm in the pre-term and control groups respectively (p=0.830). The mean eosinophil counts in blood (SD) were 190 (136) and 232 (156) cells/µL in the preterm group and controls respectively (p=0.399).

Within the sequenced cohort, FEV1 was significantly lower in the EP group with BPD compared to controls and the EP group without BPD where the differences in means were: (-0.91 L, 95% CI: -1.24 L to -0.59 L) and (-0.81 L, 95% CI: -1.31 L to -0.32 L) respectively (p<0.001). Similarly, after adjustment for age, sex, and body size using z-scores, the mean FEV1 z-score of the EP group with BPD was significantly lower compared to the control group and the EP group without BPD, the differences in means (95% CI) were (-1.55 (-2.05 to -1.05)) and (-1.13 (-1.88 to -0.37)) respectively (p<0.001).
Table 6.1: Demographic and clinical data of whole EPICure participants and those with sequenced samples

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Whole Cohort (n=92)</th>
<th></th>
<th>Sequenced Samples (n=74)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EP (n=51)</td>
<td>No BPD</td>
<td>Control</td>
<td>p-value</td>
</tr>
<tr>
<td>N</td>
<td>37</td>
<td>14</td>
<td>41</td>
<td>0.803&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>Age&lt;sup&gt;†&lt;/sup&gt; years</td>
<td>19.02 (0.54)</td>
<td>19.02 (0.41)</td>
<td>19.09 (0.52)</td>
<td>18.93 (0.66)</td>
</tr>
<tr>
<td>Males&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>13 (35%)</td>
<td>6 (43%)</td>
<td>16 (39%)</td>
<td>0.866&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Females</td>
<td>24 (65%)</td>
<td>8 (57%)</td>
<td>25 (61%)</td>
<td>19 (66%)</td>
</tr>
<tr>
<td>Asthma diagnosis&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>19 (57%)</td>
<td>3 (21%)</td>
<td>14 (37%)</td>
<td>0.081&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Current Smoker&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>8 (22%)</td>
<td>4 (29%)</td>
<td>11 (27%)</td>
<td>0.821&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Passive Smoke exposure&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>8 (22%)</td>
<td>4 (29%)</td>
<td>14 (35%)</td>
<td>0.471&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Squamous epithelial cell %&lt;sup&gt;†&lt;/sup&gt;</td>
<td>17.4 (17.7)</td>
<td>17.9 (11.9)</td>
<td>13.8 (11.1)</td>
<td>0.573&lt;sup&gt;4&lt;/sup&gt;</td>
</tr>
<tr>
<td>Prescribed Inhalers&lt;sup&gt;§‡&lt;/sup&gt;</td>
<td>9 (60%)</td>
<td>3 (38%)</td>
<td>7 (64%)</td>
<td>0.557&lt;sup&gt;3&lt;/sup&gt;</td>
</tr>
<tr>
<td>Prescribed ICS&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>4 (27%)</td>
<td>1 (13%)</td>
<td>4 (36%)</td>
<td>0.476&lt;sup&gt;3&lt;/sup&gt;</td>
</tr>
<tr>
<td>Antibiotic treatment in past year&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>1 (7%)</td>
<td>1 (13%)</td>
<td>3 (30%)</td>
<td>0.397&lt;sup&gt;3&lt;/sup&gt;</td>
</tr>
<tr>
<td>Treated for Respiratory problem in the past year&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>7 (20%)</td>
<td>3 (21%)</td>
<td>9 (25%)</td>
<td>0.393&lt;sup&gt;3&lt;/sup&gt;</td>
</tr>
<tr>
<td>FEV1 (L)&lt;sup&gt;†&lt;/sup&gt;</td>
<td>2.66** (0.54)</td>
<td>3.22 (0.76)</td>
<td>3.57 (0.65)</td>
<td>0.000&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>FEV1 z score&lt;sup&gt;†&lt;/sup&gt;</td>
<td>-1.66** (1.09)</td>
<td>-0.911 (0.04)</td>
<td>-0.37 (0.87)</td>
<td>0.000&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>Percent change in FEV1 with 7.93% (6.25)</td>
<td>7.75% (7.81)</td>
<td>5.26% (5.60)</td>
<td>0.077&lt;sup&gt;1&lt;/sup&gt;</td>
<td>9.52% (7.77)</td>
</tr>
</tbody>
</table>
### Chapter 6: Airway Microbiome of EPICure Study

#### 6.3.2. Airway Microbiome Diversity indices

The total bacterial density in sputum was similar across the three study groups where the mean 16S *rRNA* gene copies (SD) was around 5.5 (0.5) log$_{10}$ copies/µL (p= 0.726 by ANOVA) (Figure 6.1).

![Figure 6.1](image)

**Figure 6.1.** The total bacterial density in sputum samples of EPICure cohort indicated by 16S rRNA gene copies measured by qPCR (p=0.726 by ANOVA). EP: extreme preterm birth, BPD: Broncho-pulmonary dysplasia; sample sizes (21, 8, 22) respectively.
Nevertheless, the bacterial community was significantly less rich and less diverse in the sputum of the whole EP group compared to the comparator group where the difference in means of observed number of OTUs (±SEM) was 18 (8.5) OTUs (p=0.041 by t-test). The α-diversity indices: PD_whole tree, Chao_1, and Fisher_alpha were significantly lower in the whole EP group compared to the comparator group where the difference in the means (±SEM) were 0.67 (±0.32, p=0.045 by t-test), 39 (±13, p=0.005 by t-test), 4.84 (±2.28, p=0.037 by t-test) respectively. A similar trend was observed in Shannon index; however, it was not statistically significant (Figure 6.2).

Comparing the richness and alpha diversity indices across the three study groups; a trend was observed in all previously mentioned richness and α-diversity indices in which BPD group had the least diverse and rich microbial communities, while controls had the highest numerical values. This trend was statistically significant with Chao_1 only (p=0.016 by ANOVA); nevertheless, the post hoc comparisons of the differences in the α-diversity indices between the EP participants with history of BPD and the comparator group were statistically significant (Figure 6.2)
E. $p=0.037$ by t-test

F. $p=0.07$ by ANOVA

G. $p=0.045$ by t-test

H. $p=0.125$ by ANOVA

I. $p=0.137$ by MW

J. $p=0.318$ by KW

Figure 6.2. Comparison of the richness and α-diversity indices of bacterial communities in sputum between the pre-term birth survivors (EP), with and without neonatal Bronchopulmonary Dysplasia (BPD), and controls (A, B) Observed number of OTUs (C, D) Chao 1 index, (E, F) Fisher-alpha diversity index (G, H) PD whole tree index (I, J) Shannon index (A, C, E, G, I) comparing the whole EP group (orange, n=37) and Controls (navy blue, n=33), (B, D, F, H, J) comparing the 3 groups: EP+BPD (red, n=28), EP no BPD (yellow, n=9), Controls (navy blue, n=33)

*p-value by ANOVA post-hoc test LSD*
Chapter 6: Airway Microbiome of EPICure Study

In PCoA of weighted UniFrac $\beta$-diversity index, the samples of EP participants clustered regardless the history of developing neonatal bronchopulmonary dysplasia ($p=0.01$ by ANOSIM and $p=0.045$ by PERMANOVA). In unweighted UniFrac neither the samples clustered based on the birth status ($p=0.187$ by ANOSIM) nor preterm birth or neonatal BPD statuses were significant covariates (all $p >0.05$ by PERMANOVA) (Figure 6.3).

**Figure 6.3.** PCoA of weighted UniFrac $\beta$ diversity index (A and B) and unweighted UniFrac $\beta$ diversity index (C and D). (A and C) comparing the whole EP group (orange, $n=37$) and controls (blue, $n=33$) (B and D) comparing the three groups: EP+BPD (red, $n=28$), EP no BPD (yellow, $n=9$), Controls (n=33).
6.3.3. Airway bacterial community composition

At phylum-level, all samples were dominated by Firmicutes (around 52% of the bacterial communities) followed by Bacteroidetes then Proteobacteria. Both EP groups; with and without history of developing BPD, exhibited significantly lower RA of Bacteroidetes in sputum (p=0.01 by KW). This difference was compensated by a non-significant increase in the RA of phylum Firmicutes (p=0.44, Kruskal Wallis test) (Figure 6.4A).

At the genus level (Figure 6.4B), the relative abundance of Prevotella which belongs to phylum Bacteroidetes was significantly lower in both EP groups, with and without BPD, in comparison to the comparator group (p=0.007, KW). Throughout the hierarchy of taxonomy, all the taxa from phylum to the genus level encompassing Prevotella were significantly lower in both EP groups (Figure 6.5). This was compensated by a non-significant and inconsistent increase in RA of other genera such as Streptococcus, Veillonella, Rothia and Neisseria which are normal microbiota in the airways (Figure 6.4B).

Prevotella was not detected in the negative controls. The RA of Prevotella did not correlate with the percentage of squamous epithelial cells in the sputum samples which reflects the level of upper respiratory contamination (r=0.072, p=0.6). (Figure 6.4B). Nevertheless, Prevotella did significantly correlate with the FEV$_1$ z score (rho= 0.272, p=0.02) (Figure 6.6).

Out of 121 OTUs belonging to genus Prevotella; RA of OTU 4458304 accounted for most of the observed difference in the RA of Prevotella between the study groups (Figure 6.6C). By extracting and BLAST searching the 350 bp representative sequence of this OTU against the NCBI 16S ribosomal RNA database using the Nucleotide BLAST tool (NCBI, 1988), the sequence was identified as 100% identical to Prevotella melaninogenica strains (Figure 6.6D).

6.3.4. Asthma

Although the prevalence of a self-reported asthma diagnosis was relatively high in the whole EPICure cohort (43% in the sequenced cohort) (Table 6.1), asthma was not a significant covariate in weighted UniFrac β diversity index (p=0.889 by PERMANOVA) or in unweighted UniFrac (p=0.706 by PERMANOVA); in addition, the samples did not cluster based on the asthma status in both indices (p=0.379 and 0.931 respectively by ANOSIM). Furthermore, no significant differences were detected in the α-diversity of the bacterial community, the RA of genus Prevotella nor the RA of OTU 4458304 (whose RA was significantly lower in EP group relative to the comparator group), between those who were labelled with asthma diagnosis and others in the comparator group and the entire EPICure cohort (Figure 6.7).
Figure 6.4. Comparison of the airway microbiome profiles of individual sputum samples of EPICure cohort at (A) phylum level and (B) genus level.

(A) At Phylum level, the RA of Bacteroidetes was significantly lower in both EP groups compared to controls (p=0.01 by KW), *P<0.05, by MW.

(B) At the Genus level, the RA of Prevotella (sky blue) was significantly lower in both EP groups compared to controls (p=0.007 by KW). NCm: extraction negative control of the saline matrix used for sputum induction, NCr: extraction negative control of the diluted Sputasol and reagents, NCpcr: PCR negative control. Sample size:29 EP+BPD, 9 EP no BPD and 36 Controls.
Figure 6.5. Comparison of relative abundancies of (A) Phylum *Bacteroidetes* \((p=0.01)\), (B) Class: *Bacteroidia* \((p=0.00729)\), (C) Order: *Bacteroidales* \((p=0.00729)\), (D) Family: *Prevotellaceae* \((p=0.00662)\), (E) Genus: *Prevotella* \((p=0.00662)\) across the three study groups. \(p\)-values by KW corrected by BH-FDR, EP: Extremely Preterm birth, BPD: Bronchopulmonary Dyspalsia ; sample size:29, 9 and 36 respectively.
Chapter 6: Airway Microbiome of EPICure Study

Figure 6.6. Comparison between the relative abundance (RA) of *Prevotella* across the three study groups (A) the RA of genus *Prevotella* was significantly lower in both extremely preterm (EP) groups regardless the neonatal broncho-pulmonary dysplasia (BPD) status compared to controls (p=0.007 by KW) (B) the RA of genus *Prevotella* correlated with the FEV1 z-score (Spearman rho=0.272, p=0.02) (C) OTU 4458304 contributed most to the observed difference in the RA of genus *Prevotella* across the study groups (p=0.005, by KW) (D) Phylogenetic Tree of OTU 4458304 was 100% identical to *Prevotella melaninogenica* strains as obtained by BLAST Analysis of the representative sequence against NCBI 16S ribosomal RNA database. * p<0.05; Sample size: 29 EP+BPD, 9 EP no BPD and 36 Controls.
Chapter 6: Airway Microbiome of EPICure Study

Figure 6.7. Comparison of weighted Unifrac $\beta$ diversity index, Chao_1 $\alpha$ diversity index, the RA of genus *Prevotella* and the RA of OTU 4458304 identified as *Prevotella melaninogenica* in sputum samples from participants who were labelled with asthma (pink) and those who were not (blue) within the whole cohort (A, B, C and D respectively) and within the comparator group only (E, F, G and H respectively). No significant differences were detected. 1: $p$-values by T test, 2: $p$-values by MW

PerMANOVA $p=0.889$

PerMANOVA $p=0.395$
6.3.5. Loads and Prevalence of Pathogenic Airway Bacteria

The prevalence and loads of *S. pneumoniae* (Spn), *H. influenzae* (Hi) and *M. catarrhalis* (Mc) were similar within the three study groups (p=0.176 by Chi square, p=0.179 by MANOVA); while *Ps. aeruginosa* was completely undetected by both 16S rRNA sequencing and qPCR in sputum samples from the whole cohort. (Figure 6.8)

*S. pneumoniae* and *H. influenzae* frequently co-existed in sputum samples. In the EP group with neonatal BPD: 14% had *H. influenzae*, 28% had *S. pneumoniae* and 11% had both organisms together. A similar pattern was observed in controls in which 15% had *H. influenzae*, 27% had *S. pneumoniae* and 22% had both organisms together. The bacterial loads of both *S. pneumoniae* and *H. influenzae* were similar in these two group. The mean bacterial load of *S. pneumoniae* was 4.57 (1.23 SD) log$_{10}$ CFU/mL in the BPD group and 4.37 (1.48 SD) in the comparator group. The mean bacterial (SD) load of *H. influenza* was 4.35 (0.91) in BPD group and 4.47 (1.59) log$_{10}$ CFU/mL in the comparator group (Figure 6.8C).

In the EP group without BPD history; *H. influenzae* was the most prevalent organism being detected in 57% of the sputum samples from this group with a mean load (SD) of 4.51 (0.88) log$_{10}$ CFU/mL of sputum. It was found in 36% of the samples alone and in 21% of the samples with *S. pneumoniae*. However, *S. pneumoniae* was the most populous with a mean bacterial load (SD) of 4.96 (1.08) log$_{10}$ CFU/mL (Figure 6.8).

*M. catarrhalis* was the least prevalent and the least populous organism. The mean load of *M. catarrhalis* was 0.7 log$_{10}$ higher in BPD group compared to the other two groups (p=0.681, ANOVA). It is noteworthy that none of the differences in prevalence and load of the three organisms across the study groups achieved statistical significance.

The co-existence of the three bacteria was found in 5% of the comparator group and 3% of BPD group (Figure 6.8B).
Figure 6.8. Prevalence and loads of pathogenic airway bacteria; *H. influenzae* (Hi), *M. catarrhalis* (Mc), *S. pneumoniae* (Spn) (A) Prevalence of the three pathogenic airway bacteria within the three study groups (p=0.176, Chi squared test) (B) Co-existence of the three bacteria in the sputum samples in each group (p=0.88, Fisher Exact test) (C) Mean bacterial load of each of the three bacteria as determined by the multiplex qPCR (p=0.179, MANOVA). Bacterial load was measured in CFU/mL of original sputum sample for each bacterium and the sum of the three (overall) were calculated for each sample. Samples which gave negative results for a given bacteria were excluded from the analysis. Error bars show SD; EP: extreme preterm birth, BPD: Broncho-pulmonary dysplasia; sample sizes (37, 14, 41) respectively.
6.4. Discussion

The airway microbiome in a cohort of survivors of extremely preterm birth (<26 gestations weeks) was studied for the first time. Subjects with and without a history of neonatal bronchopulmonary dysplasia were included in their late adolescence/early adulthood in comparison with a matched group of controls who were born full-term. So far, previous studies had investigated the association between bacteria and microbial dysbiosis and the development of bronchopulmonary dysplasia at infancy. (Wagner et al., 2017, Mourani et al., 2011, Lal et al., 2016, Lal et al., 2018, Beeton et al., 2011, Cordero, 1997, Kotecha et al., 2004, Lohmann et al., 2014, Van Marter et al., 2002)

The respiratory and cardiovascular complications of extreme preterm birth were found to extend to adulthood in the EPICure cohort. The EP group had significantly lower FEV1 z-scores and higher proportions of bronchodilator reversibility, even though they exhibited lower fractional exhaled nitric oxide (FeNO) levels. Respiratory parameters were specifically worse in the subjects with history of neonatal BPD development compared to the controls. (Hurst et al., 2020) Similar trends existed in the cohort subset whose sputum samples were sequenced for airway microbiome characterization. As the respiratory morbidity of premature birth was presented in a form of airflow obstruction in adulthood that may resemble COPD. Investigating this cohort was a privilege to understand the dysbiosis in the airways that may be associated with the early stages of chronic lung disease development.

The overall bacterial density in the airways was similar across the three study groups. Nevertheless, the microbiome profiles of the EP groups were significantly less rich and diverse relative to the comparator group. Although being statistically non-significant, a gradient trend in which many α-diversity indices were lowest in the BPD group and highest in the comparator group, was observed that may reflect clinical significance. Nevertheless, this finding needs to be validated by future larger studies.
In the current study, there was a significant shift away from *Bacteroidetes*; driven particularly by *Prevotella* reduced relative abundance in both EP groups in comparison to the comparator group. This decline was mainly balanced with a relative increase in the *Firmicutes* relative abundance. Although it is not clear how anaerobic bacteria can colonize and survive in aerobic environment like airways and lungs, previous studies have showed a gradient of decreasing partial oxygen pressure along the respiratory tract with the lowest values was in the smallest bronchioles; this was opposite to a gradient of increasing partial pressure of carbon dioxide. (Man et al., 2017) Another study demonstrated that a steep oxygen gradient also existed within a cystic fibrosis mucus plug. (Worlitzsch et al., 2002). Consequently, several authors suggest that, microaerophilic pockets might be created within the stagnant sputum in the airways as result of the reduced mucociliary clearance or within established biofilms that may support the growth of anaerobes within the lower respiratory tract. (Pragman et al., 2016) Nevertheless, *Bacteroidetes* are frequently detected in the microbiome of lower respiratory tract in healthy people; furthermore, a high abundance of *Bacteroidetes* has been linked with a healthy lung microbiome in the literature. (Dickson et al., 2016a)

A shift in community composition away from *Bacteroidetes*; either towards *Proteobacteria* or *Firmicutes*, has been frequently observed in COPD, bronchiectasis, cystic fibrosis and asthmatic patients. (Dickson et al., 2016a, Einarsson et al., 2016, Hilty et al., 2010) A similar shift away from *Bacteroidetes* towards *Firmicutes*; mainly streptococci and a minor increase in *Proteobacteria* has been reported in patients with severe asthma in comparison with non-asthmatics. (Zhang et al., 2016)

In the current data, the relative abundance of *Prevotella* significantly correlated with FEV<sub>1</sub> z-score but there was no association with other clinical parameters such as smoking status, exposure to passive smoking, diagnosis of asthma or prescribed inhalers. On deeper analysis, *Prevotella melaninogenica* was the species that contributed most to the observed reduction in total *Prevotella* relative abundance in the ex-premature groups in comparison to the comparator group. In literature, numerous studies that compared the lung microbiome in health and
Chapter 6: Airway Microbiome of EPICure Study

disease found that *Prevotella* was quickly replaced by members of *Proteobacteria* or *Firmicutes* in various chronic lung conditions. (Hilty et al., 2010, Einarsson et al., 2016, Zhang et al., 2016, Dickson et al., 2016a, Wu et al., 2013)

*Prevotella* species are obligate anaerobes that have been regarded as opportunistic pathogens within the oral microbiota as well as other body sites. They have been associated with periodontitis and have been previously isolated with other bacteria in lower respiratory tract infections before it has been demonstrated that they are part of the normal microbiota in the respiratory system. (Kedzia et al., 2003) *Prevotella* has been a major taxon reported in the microbiome of broncho-alveolar lavage from healthy subjects and even in explanted lung tissues in other studies. (Man et al., 2017, Cabrera-Rubio et al., 2012, Erb-Downward et al., 2011, Charlson et al., 2011, Pragman et al., 2018) Therefore, the *Prevotella* detected cannot be simply regarded as upper respiratory tract contamination. Very little has been published on *Prevotella melaninogenica* so far; it has been most frequently isolated from the sputum in cystic fibrosis by anaerobic culture techniques. It was reported in these studies that the presence of anaerobes was associated with clinical stability. (Council, 2013, Field et al., 2010, Tunney et al., 2008)

Being strictly anaerobic, *Prevotella* is difficult to culture and not usually isolated from specimens in routine diagnostic microbiology laboratories and this may explain why it has received less attention. In a previous study, the potential pathogenic role of *P. melaninogenica* in cystic fibrosis patients was investigated, the lipopolysaccharides of *P. melaninogenica* were found to elicit less inflammatory and interleukin response compared to that of *Ps. aeruginosa* within the same cell lines. Contrary to expectation, this observation indicated that *P. melaninogenica* minimally contributed to the overall inflammatory response in cystic fibrosis. Furthermore, this study revealed that antibody response against *P. melaninogenica* was higher in cystic fibrosis patients compared to non-disease controls, such immune response may explain the reduced relative abundance of *Prevotella* associated with chronic lung disease. (Council, 2013) Further research is required to understand the role of *Prevotella* species in the context of the lung microbial population, its interactions with other microbes and the host immunity.
and to characterize the species and strains that are associated with respiratory health both genotypically and phenotypically.

In the current study, no significant differences in the microbiota composition between the two ex-premature groups was found. The samples from both ex-premature groups clustered together in principle component analysis. No significant differences in the richness and alpha diversities were observed either.

Nevertheless, the airway microbiome of the ex-premature adolescents in the studied cohort was compared with what had been published in literature about the microbiology of the premature born infants. Lal et al. studied the airway microbiome at birth in both preterm and full term born infants, in their results *Prevotella* was detected in tracheal aspirates at day one of life. Looking closely at their published microbiome profiles a trend was noticed in which the relative abundance of *Prevotella* was relatively higher in the full-term born infants compared to extremely low birth weight preterm infants; especially among those who were predisposed to BPD. This supports the current findings and suggests that the observed bacterial dysbiosis may have originated during infancy and was sustained through to adulthood. (Lal et al., 2016)

Lohmann et al. reported notable changes in airway microbiome of preterm infants who subsequently developed BPD compared to those who did not. Immediately after birth; most of the preterm infants’ airway microbiome was mainly dominated with *Proteobacteria*; particularly *Acinetobacter*. Over time, the relative abundance of *Firmicutes* increased driven mainly by *Staphylococcus* and *Proteobacteria* decreased in those infants who developed BPD in contrast to the relatively diverse and stable community in the non-BPD group. (Lohmann et al., 2014) *Staphylococcus* was also associated with BPD development in another study. (Wagner et al., 2017) In the current results, the airway bacterial community in the preterm birth adolescent survivors was mainly dominated by *Firmicutes*; with two genera *Streptococcus* and *Veillonella* having the highest contributions. *Bacteroidetes*, *Proteobacteria*, and *Actinobacteria* respectively contributed relatively less to the microbiome profile at phylum level. This was similar to the
microbiome profiles of the EPICure controls as well as the control group of relatively healthy subjects (HIV negative control group) presented in the following chapter (Ch.7, Figure 7.6). *Acinetobacter* and *Staphylococcus* were rarely identified in the airway microbiome of our cohort; their relative abundances were <0.1%. Although not being statistically significant it is noteworthy to highlight that *Staphylococcus* when detected its relative abundance was relatively higher in the BPD group compared to the other two groups.

Numerous studies found an important role for respiratory colonization with *Ureaplasma*; a genital mycoplasma in the development BPD. (Beeton et al., 2011, Van Marter et al., 2002, Cordero, 1997, Kotecha et al., 2004) In the current results, *Ureaplasma* was identified in the airways of six participants at RA<0.01% from EP groups as well as the control group. In another study, *Corynebacterium* was associated with the development of severe BPD in premature infants. (Imamura et al., 2017) Again, *Corynebacterium* was present at similar relative abundance in the three studied groups at around 1%.

In the EPICure Cohort, the prevalence and the abundance of the three respiratory pathogens: *S. pneumoniae, H. influenzae*, and *M. catarrhalis* were similar across the three study groups. Compared to the healthy comparator group in Chapter 7, the detected loads and prevalence of the three respiratory pathogens were also similar between the EPICure study groups and the HIV negative group (all p-values >0.05). *S. pneumoniae* mean load was 4.45 (SD 1.20) CFU/mL in both the BPD and control groups, this was comparable to what Wu *et al.* found (4.24 ± 1.78 log₁₀ copies/ml) in the sputum samples from healthy individuals using another qPCR assay. (Wu et al., 2014) *S. pneumoniae* frequently co-existed with *H. influenzae*. *M. catarrhalis* was the least prevalent among the tested pathogenic bacteria in the three groups. Whilst, *P. aeruginosa* was completely absent in the sputum samples from the three study groups.

This study has a small sample size; especially the EP group without BPD, which could not power small differences to statistical significance. However, this is an inherent and unavoidable limitation as neonatal BPD is strongly related with the
The incidence rate of bronchopulmonary dysplasia was 68% among the EP survivors in the EPICure cohort. (Costeloe et al., 2012) Asthma prevalence was relatively high, even within the comparator group in which asthma prevalence was 37%, compared to the national levels for general population within the same age group; according to the British Lung Foundation the prevalence in this age group in the UK was 20% (BLF, 2012). It is likely that there was significant over-diagnosis of ‘asthma’ in this longitudinal cohort. Asthma diagnosis was self-reported. EPICure and other preterm cohort studies have previously reported that BPD survivors often have airflow obstruction later in life which can be mis-labelled as asthma. (Fawke et al., 2010, Northway et al., 1990, Vrijlandt et al., 2005) Nevertheless, asthma does not interfere with the findings of the current study, asthma status was not found to be a significant covariate is (p>0.05 by both adonis and PERMANOVA). None of the microbiome describing parameters (α and β indices) nor the *Prevotella* relative abundance were different between subjects labelled as asthmatics and non-asthmatics, either in the whole cohort or the control group.

To conclude, extremely preterm birth results in a significant dysbiosis in the airway microbiome which is maintained into early adult life. This is characterised by less rich and diverse bacterial community in the airways. The microbiome composition was shifted away from the *Bacteroidetes*; as manifest by decline in the relative abundance of *Prevotella*; which was correlated with FEV1 z-score. Particularly, *Prevotella melaninogenica* was the species showing most variation within this genus. None of the reported airway dysbiosis or specific bacterial colonization associated with the development of BPD in preterm infants in literature was identified in the current adult cohort of preterm survivors. There were no significant differences in the microbiota composition between extreme pre-term survivors with and without neonatal BPD history.
Chapter 7: Airway Microbiome of People Living with HIV (PLW-HIV) in comparison to HIV-uninfected Individuals

7.1. Introduction

Highly active antiretroviral therapy (ART) has greatly influenced the morbidity and mortality of HIV infection. People living with HIV (PLW-HIV) have a better chance to live a near normal life expectancy in the era of ART. (van Sighem et al., 2010) HIV complications have also been shifted from the traditional AIDS opportunistic infections and neoplasms into non-AIDS related comorbidities. (George et al., 2016) However, respiratory illness continues to be a common comorbidity in this population. HIV infected individuals remain at higher risk for developing pneumonia as well as chronic lung disease including COPD. (Grubb et al., 2006, Presti et al., 2017)

As HIV infection is known to impact innate and adaptive immunity in the lungs; therefore, it can alter the lung microenviroment in a way that can affect the homeostasis of the microbial communities present. (Shenoy and Lynch, 2018) Current evidence, mostly from USA cohort studies, suggests that HIV infection alters the airway microbiome in different ways, generally it reduces the richness and alpha diversity of the microbial communities in the airways, enriches signature bacteria such as *Tropheryma whippelii* and some taxa such as *Prevotella* and *Veillonella* which were claimed to contribute to lung inflammation. (Lozupone et al., 2013, Lawani and Morris, 2016, Twigg et al., 2017, Presti et al., 2017) However, with effective ART these deviations in the microbiome are alleviated. Data from Lung HIV Microbiome Project (LHMP) suggests that the lung microbiome in healthy HIV-infected individuals on effective ART with preserved CD4 counts is similar to uninfected population. (Beck et al., 2015)

The aim in this chapter is to explore the deviations in the airway microbiome of PLW-HIV in a cohort of closely monitored healthy HIV infected subjects (PLW-HIV) on effective ART, with preserved peripheral CD4 count and most with undetectable viral load in comparison to HIV uninfected matched controls. The results presented
Chapter 7: Airway Microbiome of PLW-HIV


7.2. Methods

7.2.1. Study design and population

The work presented was a part of larger prospective observational longitudinal cohort study conducted in the Royal Free London NHS Foundation Trust (RFL) comparing the incidence of acute respiratory illness in PLW-HIV and HIV uninfected individuals. (Brown et al., 2020) The study was approved by the London Hampstead Research Ethics Committee (14/LO/1409) and registered with the ISRCTN registry (ISRCTN38386321). Participants were recruited between November 2015 and January 2017. PLW-HIV were invited to participate when they attended ambulatory HIV care appointments. The only eligibility criteria were age over 18 years, consent to participate and absence of symptoms of acute respiratory illness at study entry. HIV uninfected participants were recruited from Sexual Health Clinics at Barnet Hospital (the Clare Simpson clinic) and the Royal Free Hospital (Marlborough clinic), as well as, primary care settings; primary-care records were used to invite potential participants with similar age, gender and smoking status to the expected characteristics of the HIV positive population sampled. HIV uninfected participants had their HIV status confirmed by blood test at recruitment. All participants provided written informed consent. (Brown, 2019)

Sputum samples collected from participants who could expectorate at study entry were submitted to the present analysis. Sputum samples were stored unprocessed, within 3 hours at -70°C.

The Methods are explained in detail in Chapter 2. Briefly, sputum samples were homogenised and metagenomic DNA was then extracted on the automated LIAISON ® Ixt extraction platform. The V3-V4 regions of the 16S rRNA gene were amplified using 341 forward and 805 reverse primer set. The library was sequenced on Illumina Miseq platform. In bioinformatic analysis, the biom table was rarefied at the depth of 20,000 reads. Total bacterial load was quantified using
16S rRNA qPCR in the sputum samples (Ch.2, 2.8.) Typical respiratory pathogens and *Ps. aeruginosa* were detected in sputum using an qPCR (Ch. 2, 2.10.) for respiratory pathogens.

**Statement of Contribution:** The clinical study was designed and conducted by Prof. Lipman and Dr. Brown. The sputum samples and the clinical metadata were collected by Dr. Brown. Dr. Brown and myself processed the sputum samples and performed qPCRs. I did the 16S rRNA sequencing, bioinformatic and data analysis and generated the results presented in this chapter.

### 7.3. Results

#### 7.3.1. Clinical and demographic cohort characteristics

Sputum was collected from 38 HIV uninfected individuals (HIV negative group) and 64 PLW-HIV (HIV positive group). No significant differences were found in the demographic characteristics between the two study groups regarding age (range: 74-30 years old), body mass index (BMI), ethnicity, educational level, sex, and self-reported co-morbidities; e.g. COPD, asthma, heart disease, diabetes and cancer (Table 7.1). However, significantly more PLW-HIV were smoking at the time of sample collection and had been using recreational drugs where the prevalence was 3.5 and 3 times higher than that in the HIV negative comparator group respectively.

In the HIV positive group, all participants were on ART, 53%, had been taking ART for over 10 years. Eighty four percent had undetectable HIV viral load (VL) (<40 copies/mL) in plasma, while 16% (n=10) had detectable VL (median 122 copies/mL, IQR 71-418). The median blood CD4 count was 676 cells/µL (IQR 467-880 cells/µL). The median length of time between HIV diagnosis and starting ART was 1342 days (IQR 448-3131 days) and the median nadir CD4 (the lowest CD4 count detected) was 225 cells/µL (IQR 107-371 cells/µL) of which 44% had nadir CD4 <200 cells/µL.
Table 7.1: Demographics and clinical characteristics of the HIV study cohort

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>HIV Positive (n=64)</th>
<th>HIV Negative (n=38)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>52 (10.47)</td>
<td>53 (7.76)</td>
<td>0.594³</td>
</tr>
<tr>
<td>Sex</td>
<td>Males 55 (86%)</td>
<td>29 (76%)</td>
<td>0.218²</td>
</tr>
<tr>
<td></td>
<td>Females 9 (14%)</td>
<td>9 (24%)</td>
<td></td>
</tr>
<tr>
<td>Sexuality</td>
<td>Heterosexual 13 (20%)</td>
<td>29 (76%)</td>
<td>&lt;0.001²</td>
</tr>
<tr>
<td></td>
<td>Homosexual 42 (66%)</td>
<td>9 (24%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bisexual 9 (14%)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Caucasian 51 (80%)</td>
<td>36 (95%)</td>
<td>0.108³</td>
</tr>
<tr>
<td></td>
<td>Black 7 (11%)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Asian 2 (3%)</td>
<td>1 (3%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Others 4 (6%)</td>
<td>1 (3%)</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>University Degree 28 (46%)</td>
<td>20 (54%)</td>
<td>0.912³</td>
</tr>
<tr>
<td></td>
<td>Secondary 21(34%)</td>
<td>10(28%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Education or equivalent No Qualification 4 (7%)</td>
<td>2 (5%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Others 8 (13%)</td>
<td>5 (14%)</td>
<td></td>
</tr>
<tr>
<td>BMI (Kg/m²)</td>
<td>26.18 (4.93)</td>
<td>25.04 (2.91)</td>
<td>0.789¹</td>
</tr>
<tr>
<td>Self-reported Comorbidities</td>
<td>Asthma 8 (21%)</td>
<td>5 (36%)</td>
<td>0.292³</td>
</tr>
<tr>
<td></td>
<td>COPD 2 (5%)</td>
<td>1 (7%)</td>
<td>1.000³</td>
</tr>
<tr>
<td></td>
<td>Cancer 7 (18%)</td>
<td>1 (7%)</td>
<td>0.665³</td>
</tr>
<tr>
<td></td>
<td>Heart Disease 5 (13%)</td>
<td>0 (0%)</td>
<td>0.309³</td>
</tr>
<tr>
<td></td>
<td>Diabetes 4 (10%)</td>
<td>0 (0%)</td>
<td>0.563³</td>
</tr>
<tr>
<td>Smoking</td>
<td>Current Smoker 25 (39%)</td>
<td>4 (11%)</td>
<td>0.009²</td>
</tr>
<tr>
<td></td>
<td>Ex-smoker 21 (33%)</td>
<td>16 (43%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Never smoked 18 (28%)</td>
<td>17 (46%)</td>
<td></td>
</tr>
<tr>
<td>Recreational drug use (ever)</td>
<td>48 (75%)</td>
<td>20 (57%)</td>
<td>0.067²</td>
</tr>
<tr>
<td>Recreational drug use (last 3 mo)</td>
<td>25 (39%)</td>
<td>5 (14%)</td>
<td>0.007²</td>
</tr>
<tr>
<td>CD4 count (cells/µL)</td>
<td>676 (467-880)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nadir CD4 (cells/µL)</td>
<td>225 (107-371)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD8 count (cells/µL)</td>
<td>1030 (723-1377)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD4/CD8 Ratio</td>
<td>0.67 (0.43-0.92)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjects with Blood HIV Viral load &lt;40 copies/mL</td>
<td>54 (84%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blood HIV Viral load &gt;40 copies/mL</td>
<td>10 (16%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Viral load (copies/mL)</td>
<td>122 (71-418)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prescribed Antiretroviral therapy</td>
<td>63 (100%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prescribed Antiretroviral therapy over 10 years</td>
<td>10.5 (5-15)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹Mean (SD)                                    ³Median (IQR)                                    ³n (%)
²p-value by Mann-Whitney Test                  ¹p-value by Chi square
³p-value by Fisher Exact Test

BMI: Body Mass Index
7.3.2. Respiratory status of participants

The participants were free from the following acute respiratory infection symptoms: cough, sore throat, blocked or runny nose with or without a sensation of facial pain or pressure, breathlessness or pain on breathing at time of samples’ collection; therefore, the participants were considered at baseline respiratory status; i.e. free from acute respiratory illness.

Lung function reflected by the spirometry results (Table 7.2) was normal for most participants, no significant differences were detected between the two study groups in FEV1, FVC, FEV1 and FVC z-scores, FEV1 and FVC percentage predicted or in FEV1/FVC ratio. No statistically significant differences were detected in prevalence of obstructive airflow (defined as FEV1/FVC <0.7) or restrictive airflow (defined as FEV1/FVC >0.7 and FVC< 80% predicted); however, it is important to note that the airflow obstruction was three times higher in the HIV positive group (Figure 7.1).

Table 7.2: Lung Function parameters and Inhalers use

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>HIV Positive (n=64)</th>
<th>HIV Negative (n=38)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEV1 (L)§</td>
<td>3.28 (2.77-3.79)</td>
<td>3.62 (2.84-4.09)</td>
<td>0.132¹</td>
</tr>
<tr>
<td>FEV1 % predicted§</td>
<td>90% (72-104%)</td>
<td>90% (80-107%)</td>
<td>0.316¹</td>
</tr>
<tr>
<td>FEV1 z-score§</td>
<td>-0.53 (-1.18-0.07)</td>
<td>-0.35 (-0.85-0.32)</td>
<td>0.115¹</td>
</tr>
<tr>
<td>FVC (L)§</td>
<td>4.30 (3.37-4.74)</td>
<td>4.61 (3.78-5.27)</td>
<td>0.061¹</td>
</tr>
<tr>
<td>FVC % predicted§</td>
<td>91% (76-104%)</td>
<td>96% (84-114%)</td>
<td>0.134¹</td>
</tr>
<tr>
<td>FVC z-score</td>
<td>-0.27 (-0.95-0.25)</td>
<td>-0.13 (-0.76-0.46)</td>
<td>0.326</td>
</tr>
<tr>
<td>FEV1/FVC</td>
<td>0.81 (0.75-0.86)</td>
<td>0.78 (0.73-0.84)</td>
<td>0.32¹</td>
</tr>
<tr>
<td>(FEV1/FVC) % predicted</td>
<td>79.3% (78.5-80.2%)</td>
<td>79.1% (78.2-80.3%)</td>
<td>0.994¹</td>
</tr>
<tr>
<td>FEV1/FVC z-score</td>
<td>0.68 (0.66-0.70)</td>
<td>0.68 (0.67-0.69)</td>
<td>0.708¹</td>
</tr>
<tr>
<td>Airflow Obstruction*‡</td>
<td>10 (16%)</td>
<td>2 (5%)</td>
<td>0.202³</td>
</tr>
<tr>
<td>Restrictive Airflow**‡</td>
<td>14 (22%)</td>
<td>3 (8%)</td>
<td>0.062²</td>
</tr>
<tr>
<td>Inhaled respiratory medication‡</td>
<td>10 (16%)</td>
<td>4 (11%)</td>
<td>0.500²</td>
</tr>
<tr>
<td>β2 Agonist‡</td>
<td>5 (8%)</td>
<td>2 (6%)</td>
<td>1.00³</td>
</tr>
<tr>
<td>ICS‡</td>
<td>4 (7%)</td>
<td>0 (0%)</td>
<td>0.293³</td>
</tr>
</tbody>
</table>

§Median (IQR)  †n (%)  ICS: Inhaled Corticosteroids
¹p-value by Mann-Whitney Test  ³p-value by Fisher Exact Test
²p-value by Chi square
FEV1: Forced Expiratory volume in 1 sec  FVC: Forced Vital Capacity
* FEV1/FVC <0·7  **FEV1/FVC ≥0·7 and FVC< 80% predicted
Chapter 7: Airway Microbiome of PLW-HIV

Figure 7.1. Comparison between the rates of air flow obstruction (FEV1/FVC) in the HIV positive (Orange, n=63) and HIV negative comparator group (Blue, n=38)

7.3.3. Richness and alpha Diversity of bacterial communities

The total bacterial load, quantified by 16S *rrra* qPCR, was significantly lower in the sputum of PLW-HIV compared to the HIV negative comparator group where the mean (SD) of log total 16S *rrna* copies per µL was 5.72 (0.13) in the HIV positive group compared to 6.21 (0.12) in the HIV negative group, p= 0.026 by Mann Whitney (Figure 7.2).

Figure 7.2. Total bacterial density was significantly lower in the HIV positive group (orange, n=55) compared to the HIV negative group (blue, n=38); p=0.026 by MW; Total bacterial load was quantified by measuring the total 16S rRNA (V3-V4) gene copies/µL of reaction mixture.
In 16S rRNA sequencing data, a trend was observed in which all the richness and α-diversity indices, were slightly lower in the HIV positive group compared to the HIV negative comparator group; however, none of the observed differences was statistically significant (Table 7.3).

**Table 7.3. Comparison between the α-diversity indices in HIV positive and negative groups**

<table>
<thead>
<tr>
<th>α-diversity indices</th>
<th>HIV Positive (n=64)</th>
<th>HIV Negative (n=38)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Observed OTUs</td>
<td>300 (9)†</td>
<td>314 (8)</td>
<td>0.247†</td>
</tr>
<tr>
<td>Chao 1</td>
<td>367.2 (10.9)</td>
<td>385.2 (9.5)</td>
<td>0.216†</td>
</tr>
<tr>
<td>Fisher alpha</td>
<td>50.5 (1.8)</td>
<td>53.04 (1.5)</td>
<td>0.292†</td>
</tr>
<tr>
<td>PD whole tree</td>
<td>16.42 (0.42)</td>
<td>16.99 (0.35)</td>
<td>0.301†</td>
</tr>
<tr>
<td>Shannon</td>
<td>5.11 (0.08)</td>
<td>5.29 (0.09)</td>
<td>0.126²</td>
</tr>
<tr>
<td>Simpson</td>
<td>0.926 (0.005)</td>
<td>0.934 (0.006)</td>
<td>0.251²</td>
</tr>
</tbody>
</table>

†Mean (SD)

¹p-value by t-test

²p-value by Mann-Whitney Test

Within the HIV positive group, the following richness and α-diversity indices were significantly lower in participants whose blood CD4 count was <500 cells/µL (n=22) where the differences in means of: the number of observed OTUs, Chao1, PD whole tree and Fisher alpha were -42 (p=0.026), -56.3 (p=0.013), -1.88 (p=0.033) and -8.52 (p=0.026) respectively (Figures 7.3). The blood CD4 count also significantly correlated with Chao1 (Spearman’s rho 0.257, p=0.04) and PD whole tree (Spearman’s rho 0.249, p=0.047).
A. \( p=0.026 \)  
B. \( p=0.013 \)  
C. \( p=0.033 \)  
D. \( p=0.026 \)

**Figure 7.3.** Richness and alpha diversity of airway bacterial communities in PLW-HIV with respect to the blood CD4 count; < 500 cells/µL burgundy (n=22), ≥500 cells/µL, green, (n=42), p-values by t-test

### 7.3.4. Beta Diversity of the bacterial community

The samples were distinguished based on the HIV infection status in principal co-ordinate analysis (PCoA) of unweighted Unifrac β-diversity index (PERMANOVA \( p=0.002 \) by PERMANOVA, \( p=0.031 \) by ANOSIM), but not in weighted Unifrac β-diversity index (\( p=0.283 \) by PERMANOVA, \( p=0.273 \) by ANOSIM). (Figure 7.4). In PCoA of unweighted Unifrac, the control comparator group formed a cluster while the HIV positive group samples were more dispersed. This could be explained by the increased abundance or prevalence of some minor taxa e.g. *Staphylococcus aureus*, *Pseudomonas*, *Klebsiella*, and *Lactobacillus* as shown in Figure 7.5.

Smoking (adonis \( p=0.001 \)) and recreational drug use (adonis \( p=0.001 \)) were significant covariates in principle component analysis of both weighted and unweighted Unifrac β-diversity index.
Figure 7.4. Comparing the microbiome profiles through β-diversity with respect to the HIV status (A) PCoA weighted Unifrac β-diversity index showing no separation with respect to the HIV status (p=0.283 by PERMANOVA, p=0.273 by ANOSIM) (B) PCoA plot of unweighted Unifrac β-diversity in which samples clustered based on the HIV status (p=0.002 by PERMANOVA, p=0.031 by ANOSIM). Group size: 38 HIV-ve group (blue) and 64 HIV +ve group (orange).
Figure 7.5. PCoA plot of unweighted Unifrac β-diversity (p=0.002 by PERMANOVA, p=0.031 by ANOSIM) demonstrating that the presence of some minor taxa may explain the dispersion of PLW-HIV samples; Group size: 38 HIV-ve group and 64 HIV +ve group.
Chapter 7: Airway Microbiome of PLW-HIV

Figure 7.6. Airway microbiome profiles of HIV negative and HIV positive groups (A) Mean RA of bacterial phyla (all p-values >0.05 by White non-parametric t-test) (B) Individual microbiome profiles at the genus level. NCext: extraction negative control, NCpcr: no-template PCR negative control, Group size: 38 HIV-ve group and 64 HIV +ve group
7.3.5. Composition of Bacterial Communities

At the phylum level, the microbiome profiles were similar and no significant differences in the RA abundances of the bacterial phyla were detected between the HIV positive and HIV negative groups. At genus level, the core taxa were very similar between the HIV positive and negative groups (Figures 7.6). Nevertheless, some low abundance minor taxa such as *S. aureus* (p=0.0016), *Pseudomonas* (p=0.0093), *Klebsiella* (p=0.024), genus *Lactobacillus* (p=0.029), *Bilophila* (p<0.0001), and some OTUs in the *Enterobacteriaceae* (p=0.04) were significantly more prevalent and/or more abundant in the HIV positive samples compared to the HIV negative comparator group (Figure 7.6). *S. aureus* (OTU 1040220) and *Pseudomonas* (OTU 144452), each was present in 8% of PLW-HIV, and *Klebsiella* (OTU 144814) in 9%. In contrast, these OTUs were completely absent in controls. *Bilophila* (OTU 359872) was detected in 13% of HIV positive and 3% of HIV negative group. Both genus *Lactobacillus* and family *Enterobacteriaceae* taxa were significantly more frequent and their RA was higher in the HIV positive group – being present in 66% and 30% respectively of PLW-HIV samples, while they were present at 45% and 18% respectively in the HIV negative group (Figure 7.7A and E).

The differential abundances of the previously reported taxa, apart from *Lactobacillus* which was also differentially abundant with the smoking and recreational drug use (Figure 7.8), were only explained by the HIV status and not by any other variable in the metadata. None of the previously reported taxa were detected in the negative controls.

To get a deeper resolution and better insight about the identity of the differentially abundant OTU, the representative sequences were extracted and aligned against multiple databases: Silva (Quast C. et al., 2013) and Ribosomal Database Project (RDP) (Maidak et al., 1996), Green genes and NCBI 16S rRNA type strains databases. With the first three databases, the resolution of taxonomy was down to the genus level; however, in case of *Staphylococcus* taxon, alignment against Green genes gave resolution down to species level (*S. aureus*). In the NCBI BLAST analysis (NCBI, 1988), the top hits (with the highest max and total scores and the lowest E-values) for the following OTUs: 1040220, 144452, and 144814 were 99% identical to *S. aureus, Pseudomonas aeruginosa*, and *Klebsiella pneumoniae* strains respectively (Appendix 3.8). The presence of *Ps. aeruginosa* in the HIV positive group was confirmed using qPCR method; while it was completely absent in the HIV negative group. However, the detected loads were below the method’s sensitivity for quantification.
Figure 7.7. Comparison between the relative abundance of (A) genus *Lactobacillus* (p=0.029) (B) *S. aureus* (p=0.0016) (C) genus *Pseudomonas* (p=0.0093) (D) genus *Klebsiella* (p=0.024) (E) Family *Enterobacteriaceae* (p=0.04) (F) genus *Bilophila* (p<0.0001) in the individual sputum samples of PLW-HIV and the comparator group; p-values by White non-parametric t-test, corrected by Benjamini-Hochberg FDR.
Chapter 7: Airway Microbiome of PLW-HIV

A.

B.

C.

Figure 7.8. Differential abundance of genus *Lactobacillus* with respect to (A) HIV status (p=0.029 by White t-test), (B) recreational drug use (red: drug users, n=30 while blue: non drug users, n=71) (p=0.023 by White t-test) and (C) smoking status: Ex-smokers (lilac) n=37, Non-smokers (green) n=35, Smokers (navy blue) n=29, (p=0.000834, by KW). All p-values are corrected by Benjamini-Hochberg FDR method.
Chapter 7: Airway Microbiome of PLWHIV

7.3.6. Relation of the microbiology findings and the clinical status

The presence of one or more of the pathogenic bacterial taxa; *S. aureus*, *Pseudomonas*, *Klebsiella*, and *Bilophila* and the differentially abundant OTUs belonging to the *Enterobacteriaceae* family was associated with longer time between HIV diagnosis and starting ART (Spearman-rho=0.279, p=0.028). There was no significant association found between presence of these taxa and the value of the nadir CD4 count.

7.3.7. Covariates in the microbiome data

Recreational drug use (adonis p=0.001) and smoking (adonis p=0.001) were significant covariates in principle component analysis of both weighted and unweighted Unifrac β-diversity index (Figure 7.4). But neither were explanatory covariates for the differential abundance of the previously reported taxa apart from *Lactobacillus* (Figure 7.8).

Recreational drug use resulted in a significant reduction of the microbial community α-diversity when measured by Shannon α-diversity index (p=0.0001, by Mann-Whitney). At the genus level, RA of *Lactobacillus* was enriched in drug users’ samples (p=0.023); whereas the relative abundancies of the following taxa: *Mogibacterium* (p=0.011), *Leptotrichia* (p=0.017), *Fusobacterium* (p=0.022) and *Catonella* (p=0.046) were significantly lower (Figure 7.9).

With smoking α-diversity is significantly reduced where Shannon α-diversity index was lowest in smokers and highest in non-smokers (p=0.002, by KW). Smoking was the explanatory covariate for enriching the following taxa in the samples: *Streptococcus* (p=0.01), *Lactobacillus* (p=0.0008), and *Scardovia* (p=0.045) and inhibiting *Capnocytophaga* (p=0.013), *Cardiobacterium* (p=0.006), *Catonella* (p=0.008), *Fusobacterium* (p=0.014), *Lautropia* (p=0.012), *Leptotrichia* (p=0.037), *Mogibacterium* (p=0.006), *Neisseria* (p=0.001), *Oribacterium* (p=0.027), *Peptococcus* (p=0.034), *Peptostreptococcus* (p=0.0006), *Tannerella* (p=0.022) (Figure 7.10).
Figure 7.9. Differences in the airway microbiota associated with recreational drug use in the PLW-HIV cohort and the HIV negative comparator group (A) Shannon α-diversity index was significantly lower in drug users (p=0.0001, by MW) (B) Weighted Unifrac β-diversity index (p<0.001, adonis) (C) RA of genus Lactobacillus (p=0.023) was higher whereas the (D) RA of Mogibacterium (p=0.011), (E) Leptotrichia (p=0.017), (F) Fusobacterium (p=0.022) (G) Catonella (p=0.046) were significantly reduced; p: White non-parametric t-test+BH-FDR. Group size: 30 Drug users (red) and 71 non-drug users (blue).
Figure 7.10. Differences in the airway microbiota associated with the smoking status in the PLW-HIV cohort and the HIV negative comparator group (A) Shannon \( \alpha \)-diversity index was significantly reduced in smokers (\( p=0.002, \) KW) *p*-value by MW (B) Weighted Unifrac \( \beta \)-diversity index (\( p<0.001, \) adonis). The RA of genera (C) Streptococcus (\( p=0.01 \)), (D) Lactobacillus (\( p=0.0008 \)), and (E) Scardovia (\( p=0.045 \)) were significantly enriched; \( p \)-values by KW, corrected by BH-FDR. Group size: 29 Smokers (navy blue), 37 Ex-smokers (lilac) and 35 Non-smokers.
Chapter 7: Airway Microbiome of PLW-HIV

Differences in the airway microbiota associated with the smoking status in the PLW-HIV cohort and the HIV negative comparator group, the RA of genera (F) Capnocytophaga (p=0.013), (G) Cardiobacterium (p=0.006), (H) Catonella (p=0.008), (I) Fusobacterium (p=0.014), (J) Lautropia (p=0.012), (K) Leptotrichia (p=0.037), (L) Mogibacterium (p=0.006), (M) Neisseria (p=0.001), (N) Oribacterium (p=0.027), (O) Peptococcus (p=0.034), (P) Peptostreptococcus (p=0.0006), (Q) Tannerella (p=0.022) were significantly reduced; p-values by KW t-test, corrected by BH-FDR. Group size: 29 Smokers (navy blue), 37 Ex-smokers (lilac) and 35 Non-smokers (green).
7.3.8. Load and Prevalence of Airway Bacteria using qPCR

The prevalence of the three airway bacteria: *S. pneumoniae* (Spn), *H. influenzae* (Hi) and *M. catarrhalis* (Mc) were similar within the two study groups (Figure 7.11A). *M. catarrhalis* was the least prevalent in both groups being detected in 8% of HIV positive and 14% in HIV negative group. In the HIV negative group, 18% had *S. pneumoniae*, 13% had *H. influenzae* and 16% had both organisms together. In the HIV positive group, *S. pneumoniae* was slightly but not significantly more prevalent (p=0.101) being detected in 27% of samples alone and in 25% with *H. influenzae*. The co-existence of the three respiratory pathogens were similar in both groups <5% (Figure 7.11B).

*S. pneumoniae* was the most populous in both groups with a mean bacterial load (SD) of 5 (1.3) log₁₀ CFU/mL; this was similar in the two groups. The mean bacterial load of *M. catarrhalis* was slightly lower by 1.3 log₁₀ in the HIV positive group compared to the HIV negative group; however, this difference did not achieve statistical significance (Figure 7.11C). The mean bacterial load of *H. influenzae* was lower in the HIV positive group compared to controls; the difference between the means of the two groups was -1.04 log₁₀ (p=0.008). Within the HIV positive group, the mean bacterial load of *H. influenzae* in participants with CD4 blood count <500 cells/µL was significantly lower by 0.95 log₁₀ compared those whose CD4 blood count ≥500 cells/µL (p=0.035) (Figure 7.11D).
Figure 7.11. Prevalence and loads of pathogenic airway bacteria; *H. influenzae* (Hi), *M. catarrhalis* (Mc), *S. pneumoniae* (Spn) within the PLW-HIV compared to the comparator group (A) Prevalence of the three airway bacteria within the two study groups (p=0.488, Chi squared test) (B) Co-existence of the three bacteria in the sputum samples of each group (p=0.469, Fisher Exact test) (C) Mean bacterial load of each of the three bacteria as determined by the multiplex qPCR (p=0.867, MANOVA). (D) Comparison of *H. influenzae* load between HIV negative group and PLW-HIV with CD4 count ≥500 cells/µL and those with CD4 count <500 cells/µL (p=0.035, t-test) (p=0.005 by ANOVA) *LSD post hoc test. Loads were measured in (CFU/mL) of original sputum sample for each bacterium. Samples which gave negative results for a given bacteria were excluded from the analysis. Error bars show ±1 SD; Group size: 38 HIV-ve group and 64 HIV +ve group
Chapter 7: Airway Microbiome of PLW-HIV

7.4. Discussion

In this chapter, the airway microbiome was studied in a cohort of 64 PLW-HIV with undetectable or low viral load, preserved peripheral CD4 blood count and on effective long-term ART therapy for the first time in the UK, in comparison with a matched HIV negative comparator group recruited from primary care setting.

The results were consistent with earlier findings of similar USA cohort studies involving healthy HIV infected population, (Beck et al., 2015, Cribbs et al., 2016) where the core microbiome was similar in both HIV positive and negative groups, and richness and α-diversity of the microbial communities were not significantly different. This was also confirmed here, in the PCoA of weighted Unifrac β-diversity index in which samples did not separate with respect to the HIV status.

Nevertheless, in PCoA of unweighted Unifrac β-diversity index which unlike the weighted Unifrac is more sensitive to the presence or absence of minor taxa, the HIV negative control group were more similar and clustered together whereas the samples from the HIV positive group were more heterogeneous and dispersed.

Looking more closely at the individual microbiome profiles at the genus level the prevalence and/or relative abundance of some minor taxa, whose relative abundance were <1%, were significantly higher in HIV positive group. The enrichment of the respiratory microbial communities in subsets of HIV positive group with these taxa could explain the dispersal of HIV positive samples in the PCoA of unweighted Unifrac β-diversity as shown in Figure 7.5.

Some of the reported enriched taxa are known to be normal respiratory microbiota such as *Lactobacillus* which was significantly more abundant in PLW-HIV samples. However, *Lactobacillus* was also significantly more abundant in smokers and recreational drug users; therefore, HIV infection may not be the only explanatory covariate (Figure 7.8). It is noteworthy to highlight that both smoking and recreational drug use were significantly higher in the HIV positive group. However, in a previously published study, *Lactobacillus* was not differentially abundant in
BAL and oral washes from smokers and non-smokers healthy subject. (Morris et al., 2013) On the other hand, increased abundance of Lactobacillus was previously reported by Twigg et al. and Yang et al. in HIV infected individuals. (Twigg et al., 2016, Twigg et al., 2017, Yang et al., 2019) Lactobacillus has been associated with favourable outcomes in other microbiome studies. Studies on gut microbiome demonstrated that Lactobacillus exhibited anti-inflammatory properties in murine models for asthma and intestinal mucositis. (Sagar et al., 2014, Justino et al., 2015) Intra-nasally administered Lactobacillus species promoted alveolarization in germ free mice. (Justino et al., 2015) In newly born infants, Lactobacillus was detected in the tracheal aspirates and its abundance significantly declined in the preterm infants who developed bronchopulmonary dysplasia. (Lal et al., 2016)

Most of the reported enriched taxa in the airway microbiome of PLW-HIV were potential pathogens that may have a role in pathogenesis of the HIV associated respiratory comorbidities such as S. aureus, Pseudomonas and Klebsiella; these were significantly more frequently detected in a subset of HIV positive group (around 8.5%) and almost absent in the HIV negative group. These mostly pathogenic taxa have been implicated in the pathogenesis of chronic lung disease and acute pneumonia infections in literature. (O'Dwyer et al., 2016, Gordin et al., 2008, Dickson et al., 2016a)

Other taxa which were enriched in the sputum of some PLW-HIV are regarded as gut bacteria such as Bilophila and OTUs belonging to family Enterobacteriaceae. In the lung HIV microbiome project, Lozupone et al. reported a similar finding where bronchoalveolar lavage from HIV infected individuals was enriched with Tropheryma whipplei which is a gut bacterium and the etiological agent of Whipple’s disease. Moreover, they found that the relative abundance of this bacterium was reduced following effective ART. (Lozupone et al., 2013) The enrichment of lung microbiome with gut bacteria may be explained by the breakdown of the gut mucosal barrier in HIV infection leading to increased translocation of gut microbes and microbial products into systemic circulation through which they can eventually lodge and find a favourable ecological niche in HIV infected lungs. (Brenchley et al., 2006, Lawani and Morris, 2016, Twigg et al.,
In a previous study Dickson et al., demonstrated evidence on the translocation of bacteria through the gut-lung axis in patients with acute respiratory syndrome whose lung microbiome was found to be enriched with gut bacteria. Using murine model of abdominal sepsis which is the most common cause of acute respiratory syndrome, the translocation of viable bacteria from the lower gastrointestinal tract to the lungs was confirmed. (Dickson et al., 2016b)

A positive correlation was found between the presence of the potential pathogenic and/or the gut taxa in sputum of HIV positive participants and the time between HIV diagnosis and starting ART in HIV infected participants. This may suggest a delayed start of ART may contribute to this phenomenon. However, this observation needs to be confirmed by a properly powered future study.

Although previous studies have associated between the enrichment of respiratory microbiome with *Prevotella* and *Veillonella*; and the enhanced alveolar inflammation which may contribute to chronic lung pathology in HIV infected population. (Segal et al., 2013, Twigg et al., 2016, Dickson et al., 2016b) In our cohort, the relative-abundances of *Prevotella* and *Veillonella* were similar in the sputum samples of HIV positive and negative groups. In a Canadian cohort, these two genera were even less abundant in bronchoscopic specimens from PLW-HIV compared to HIV negative controls. (Xu et al., 2018)

Recent epidemiological data showed that although the incidence of pneumonia has been greatly reduced in the ART era, PLW-HIV are still at higher risk of developing bacterial pneumonia infections which are a major cause of mortality in this population. (Sogaard et al., 2009) *S. pneumoniae* followed by *S. aureus* and *H. influenzae* were the most identified etiologic agents of pneumonia in HIV-infected population. (Gordin et al., 2008) Morris et al. have also found a permanent decline in the lung functions: FEV1, FVC, and FEV1/FVC following pneumonia infection in HIV infected population. (Morris et al., 2000) Therefore, it was important to specifically detect the prevalence and loads of typical respiratory bacterial pathogens that are the common causes of community acquired pneumonia in our cohort using specific qPCR method. Quantitative PCR results demonstrated that
S. pneumoniae followed by H. influenzae were the most prevalent in the HIV positive group and this is consistent with Nimmo et al. findings using the same method in a similar cohort. (Nimmo et al., 2015) There were no significant differences in the loads and prevalence of M. catarrhalis and S. pneumoniae between HIV positive and negative groups; however, S. pneumoniae was slightly more prevalent in the HIV positive group but this study lacked the power to prove a significant difference. H. influenzae was equally prevalent in both groups however a trend in H. influenzae load has been observed in which it was highest in the HIV negative control group and lowest in the HIV positive group with CD4 count <500 cells/µL (Figure 7.11D). Compared to the results of the EPICure cohort, although the prevalence of H. influenzae was similar, the load of H. influenzae was significantly lower in the HIV positive group (mean load 2.89 (SD 1.1) CFU/mL) compared to the EPICure cohort (mean load 4.38 (SD 1.3) CFU/mL), (p<0.001) (Ch.6, Figure 6.8) The loads and prevalence of the other two pathogens were similar to what observed in the EPICure cohort.

In this cohort, there were significantly more participants with obstructive and/or restrictive airflow, identified by spirometry parameters, in the HIV positive group (38%) compared with the HIV negative comparator group (13%), (p=0.007), which might be a confounder in our results. Nevertheless, this is an inherent problem in similar cohorts because PLW-HIV are at higher risk of developing chronic lung disease (Grubb et al., 2006); therefore, differences in the participant groups reflect differences in the populations being sampled. However, the degrees of airflow obstructions were small as shown in Figure 7.1. Excluding those participants with obstructive and restrictive airflow from α-diversity analysis did not change the results. Furthermore, both factors were statistically insignificant covariates in weighted and unweighted UniFrac β- diversity indices (adonis p-values >0.05) and none of them was explanatory covariate for the differential abundances of the reported potentially pathogenic taxa.

Another consideration is that the ability of the participants to expectorate sputum might not be random and it may represent a source of selection bias which means that this cohort may not fully represent the wider HIV infected population. Also,
current tobacco smoking and recreational drug use that were significantly higher in the HIV positive group compared to the comparator group and they have been significant covariates in β-diversity indices. This is an inherent limitation due to the nature of the cohort and it is a common in similar cohort study. Nevertheless, apart from *Lactobacillus*, the taxa contributing to the observed significant differences with respect to the smoking status and recreational drug use were all oral taxa and did not involve the main finding which relates to the detection of some potentially pathogenic and gut bacteria taxa that are not normally oral taxa, in samples from HIV positive group while these taxa were rare or completely absent in the HIV negative samples and the experimental negative controls. This agrees with Morris *et al.* finding that smoking mainly impacts the oral microbiome but has minimal effect on lung microbiome. (Morris *et al.*, 2013)

In conclusion, the core microbiome in the airways of people living with HIV who are on effective ART and clinically well managed is indistinguishable from the general population. However, HIV infection may alter the local lung microenvironment to be more permissive to respiratory pathogenic bacteria like *S. aureus*, *P. aeruginosa* and *Klebsiella sp.* and gut bacteria like *Bilophila* and members of family *Enterobacteriaceae* to occupy ecological niches in the lower respiratory tracts of HIV infected population. The carriage of these pathogenic bacteria may contribute to respiratory comorbidities. The prompt start of effective ART for HIV infected patients may reduce the carriage of the pathogenic bacteria in the respiratory tract.
Chapter 8: The Respiratory Resistome in Chronic Lung Disease

8.1. Introduction

The global burden of antimicrobial resistance (AMR) is growing at an alarming pace. WHO has recognised the problem as one of the greatest threats to public health worldwide and called for improving global AMR surveillance. (WHO, 2014)

Management of chronic lung disease is often necessarily managed with frequent and prolonged courses of antibiotics, either for prophylaxis and/or treatment of acute exacerbations. It is easy to imagine that the microbiome could play a major role in the acquisition and spread of antibiotic resistance to pathogenic bacteria. Therefore, the microbiota can act a reservoir for antibiotic resistance. (Sommer et al., 2009, Penders et al., 2013)

Molecular techniques provide promising tools such as metagenomic sequencing for exploring the microbiome and resistome. The resistome is the collective sum of all the antimicrobial resistance determinants in a microbial ecosystem. (Wright, 2007) Unlike the metataxonomics that relies on the amplification and sequencing of the taxonomic marker gene; 16S rRNA, metagenomics is the collection of genes and genomes from all the microbial members in an ecosystem. It is performed by sequencing the DNA extracted directly from specimens without the need for a previous cultivation or amplification steps. Therefore, metagenomics go beyond the metataxonomic analysis and can provide deeper insights into the microbial functional genomics, virulence and AMR molecular determinants at the population level. (Marchesi and Ravel, 2015)

The Oxford Nanopore MinION is one of the latest advances in the sequencing technologies (3rd generation sequencing) that offers deep sequencing and long sequence reads. It has the advantage of being small size, portable, and does not necessarily require sophisticated laboratory facilities. (Lavezzo et al., 2016)
In this chapter, the hypothesis that extensive use of antimicrobials in the context of chronic lung disease may select for a higher prevalence of AMR genes in the airways has been investigated through an optimized metagenomic sequencing workflow for sputum samples using the state of art Oxford Nanopore MinION sequencing technology and a newly developed bioinformatic pipeline for the analysis of long read sequencing data.

8.2. Methods

In a pilot study in collaboration between UCL CCM (Sylvia Rofael and Prof. Tim McHugh) and UCL institute of Genetics (Dr Lucy Van Drop and Prof. Francois Balloux) funded by UCL doctoral students’ small grant scheme (Ref: 156425), metagenomic sequencing was carried out on 25 sputum samples using MinION system (Oxford Nanopore Technologies, ONT, UK). Briefly, aliquots of sputum samples homogenised with Sputasol which had not been heat killed, were treated with saponin 2.5% and HL-SAN endonuclease (Arcticzymes®, Norway) to deplete human DNA as per the published method (Charalampous et al., 2019). Then, metagenomic DNA was extracted on the automated LIAISON® Ixt extraction platform (Ch.2, 2.6.1.). The bacterial load was quantified by the 16S rRNA V3-V4 qPCR method (Ch2., 2.10.3.) and for human DNA, a human RNA polymerase II qPCR method was used; the details of the primers and probe sequences are presented in Table 8.1 and its thermo-cycling condition is similar to that of 16S rRNA method. (Mwaigwisya, 2018)

Table 8.1: The sequence of primers and probe of the Human RNA polymerase II method.

<table>
<thead>
<tr>
<th>Primer/Probe ID</th>
<th>Sequence 5'-'3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hum_RNApol2_Fw Pr.</td>
<td>TGAAGCCGTGCGGAAGG</td>
</tr>
<tr>
<td>Hum_RNApol2_Rv Pr.</td>
<td>ACAAGAGAGGCAATGTGCG</td>
</tr>
<tr>
<td>Hum_RNApol2_probe</td>
<td>[6FAM] TAC CAC GTC ATC TCC TTT GAT GGC TCC TAT [OQA]</td>
</tr>
</tbody>
</table>

The extracted DNA was purified using magnetic AMPure XP beads (Beckman Coulter, UK) and washed twice with ethanol 80%. The sequencing library was prepared by multiplexing six samples in addition to the negative control per run.
using the Rapid Barcoding Kit (Ref: SQK-RLB004) which involved a tagmentation step followed by a non-specific PCR amplification step of 25 cycles of: 15 sec denaturation at 95°C, 15 sec annealing at 56 °C and 6 min extension at 65 °C, and a final 6 min extension step at 65 °C which enriches and tags all DNA fragments from each sample with unique barcodes. The DNA from differently tagged samples was pooled at equal concentrations at a maximum of 500 ng each; then the pooled library was purified using magnetic AMPure XP beads (Beckman Coulter, UK) and double ethanol 80% washes. The sizes of the DNA fragments were checked on Tape Station Automated Electrophoresis (Agilent, UK) and they were around 3500 bp (a sample report is in Appendix 3.9). The library was adjusted to 10 µL of 200 fmol final DNA concentration (20 nM) prior it to adding the rapid adaptors and other reagents from the kit (as per manufacturer’s directions). Quality control checks were carried out on the flow cells and the number of active pores was recorded prior to sequencing run. The sequencing was run for 48 hrs (the run was restarted after 20 to 24 hrs to replenish the sequencing pores) or until all the sequencing pores have been exhausted (a sample report is in Appendix 3.9).

ONT Albacore Sequencing Pipeline Software (version 2.3.4) was used to carry out base-calling, de-multiplex the samples per run and convert the resultant Fast5 files into fastq files which were then transferred to Dr. Lucy van Drop who run the bioinformatic analysis. Briefly, the genomes were assembled using the miniasm/minimap pipeline (Li, 2016) and a BLAST search against the ResFinder database performed (Zankari et al., 2012) to detect the antimicrobial resistance (AMR) genes. The Comprehensive Antibiotic Resistance Database (CARD) database was used to identify the mechanisms of resistance of the detected AMR genes. (Alcock et al., 2020) The alignments with accuracy less than 90% have been excluded. The prevalence of AMR genes within the samples was measured in part per million reads (ppm) i.e. the number of sequence reads identified as AMR genes relative to the total number of reads representing the sample.

Statement of Contribution: I processed the samples, ran the metagenomic sequencing, performed basecalling and generated the fastq files. Dr. Lucy Van
Chapter 8: Resistome in Chronic Lung Disease

Drop processed the sequence reads and ran the bioinformatic analysis. Both of us participated in the data analysis.

8.3. Results

8.3.1. Cohort Characteristics

Twenty-five sputum samples were analysed through metagenomic sequencing: 17 samples were from patients with bronchiectasis at stable respiratory state (a subset of the cohort described in Ch. 4, 4.2): eight of them were on-prophylactic antibiotics: seven were on 250 mg azithromycin thrice weekly and one on ciprofloxacin (250 mg twice daily), nine were not receiving antibiotic prophylaxis therapy. In addition, eight samples from HIV negative control group (Ch. 7, 7.2) were included as a comparator group; these samples were collected from healthy participants who did not have significant known comorbidities and had normal lung function (normal spirometry parameters) recruited from primary care settings (Table 8.2).

For the bronchiectasis patients, the most common antibiotics prescribed for the treatment of acute exacerbations were β-lactams (53%); amoxicillin, co-amoxiclav and cephalexin, doxycycline (29%), and ciprofloxacin (14%). In most of the cases, these were prolonged antibiotic courses that extended for 10 days to two weeks. The patients who were on antibiotic prophylaxis therapy had significantly less exacerbations that needed treatment with antibiotics in the year prior to samples' collection. Data on antibiotic consumption was not available for the comparator group; however, it was assumed that the healthy participants had been much less exposed to antibiotic prescription compared to patients with chronic lung disease. The clinical and demographic characteristics of the patients whose samples took part in the presented metagenomic analysis are presented in Table 8.2.
Table 8.2: Clinical and demographic characteristics of participants in resistome analysis

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Bronchiectasis patients</th>
<th>Healthy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prophylactic Antibiotic (n=8)</td>
<td>No Prophylactic Antibiotic (n=9)</td>
</tr>
<tr>
<td>Age (years)†</td>
<td>62 (7)</td>
<td>71 (9)</td>
</tr>
<tr>
<td>Sex‡</td>
<td>Males</td>
<td>Females</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Immunocompromised‡</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Acute exacerbations in previous year§</td>
<td>2 (1-3)</td>
<td>4 (2-5)</td>
</tr>
<tr>
<td>Prophylactic Antibiotics‡</td>
<td>Azithromycin</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Ciprofloxacin</td>
<td>1</td>
</tr>
<tr>
<td>Rescue Pack Antibiotics§</td>
<td>Amoxicillin</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Amoxicillin/clavulanic</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Cephalexin</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Doxycycline</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Ciprofloxacin</td>
<td>2</td>
</tr>
<tr>
<td>FEV1 (L)§</td>
<td>1.44 (1.04-1.68)</td>
<td>1.49 (1.14-2.10)</td>
</tr>
<tr>
<td>FEV1 % Predicted§</td>
<td>49% (32-92%)</td>
<td>77% (58-92%)</td>
</tr>
<tr>
<td>FVC (L)§</td>
<td>2.17 (1.63-3.05)</td>
<td>2.27 (1.60-2.82)</td>
</tr>
<tr>
<td>FVC % Predicted§</td>
<td>80% (45-106%)</td>
<td>122% (70-123%)</td>
</tr>
<tr>
<td>FEV1/FVC§</td>
<td>0.68 (0.49-0.75)</td>
<td>0.60 (0.54-0.66)</td>
</tr>
<tr>
<td>(FEV1/FVC) % predicted§</td>
<td>77% (61-103%)</td>
<td>77% (70-88%)</td>
</tr>
</tbody>
</table>

† Mean (SD) § Median (IQR) n
8.3.2. Human DNA Depletion

The efficiency of the human DNA depletion step was evaluated by measuring the total bacterial and human DNA loads in samples before and after the depletion process using qPCR. For the human DNA the Ct values shifted by a mean of +7 units in the treated samples compared to the corresponding untreated samples (Figure 8.1) which is equivalent to a 100 fold reduction in human DNA load. On the other hand, Ct values of 16S rRNA gene copies increased by a mean of 1.5 units in the treated samples which is equivalent to a 5-fold reduction in the bacterial DNA load. This confirmed that the depletion process worked well in most cases to reduce the human DNA load to an acceptable level that permits metagenomic sequencing with a minimal loss in the bacterial DNA. The samples with human DNA Ct value >30 were selected.

After sequencing and bioinformatic analysis, the mean percentage of sequences (SD) that were classified as microbial reads were 77% (14%); whereas the mean human unclassified reads (SD) were 17% (9%) of the total number of reads per sample.

![Figure 8.1](image.png)

**Figure 8.1.** Reduction in the human DNA load in the treated sample manifested by a shift of a mean of +7 in the Ct values of real-time PCR targeting human RNA polymerase II gene, dashed lines: the treated samples to deplete the human DNA, solid line: untreated samples
8.3.3. Prevalence of AMR genes

Although not being statistically significant due to the small sample sizes, there was a trend in which AMR genes were more frequently detected in the bronchiectasis patients (92% of the total number of AMR genes detected) compared to the controls (8% of the total number of AMR genes detected). The median (IQR) prevalence of AMR genes was 26 ppm sequence reads (11–79 ppm, n=17) in the bronchiectasis patients versus 11 ppm (6–30 ppm, n=6) in the controls (p=0.155 by MW). Specifically, there was a trend to a higher prevalence of AMR genes among patients on prophylactic antibiotics with median 28 ppm (IQR: 14–43 ppm, n=8) compared to controls (p=0.093 by MW) (Figure 8.2). Within the bronchiectasis patient group, those not receiving antibiotic prophylaxis therapy exhibited a broader range of AMR prevalence where median (IQR) was 21 ppm (6–121 ppm, n=9) and a greater diversity of AMR genes which confer resistance to multiple antibiotic classes as shown in Figure 8.3.

Figure 8.2. Comparison between the relative abundance (expressed in part per million reads) of the total AMR genes detected in sputum samples from bronchiectasis patients on antibiotic prophylaxis (blue, n=8), patients having frequent exacerbations treated with antibiotics but not receiving antibiotic prophylaxis therapy (yellow, n=9) and relatively healthy participants in the comparator group (green, n=8) (p-value=0.331 by KW). BTX: Bronchiectasis, ppm: part per million.
The most frequently detected AMR gene families were *tet* (32% of the total detected AMR genes) particularly 25% were *tet* genes which encode ribosomal protection proteins and confer resistance to tetracyclines by target protection such as *tet*(O) and *tet*(M) and 5% were mosaic *tet* genes such as *tet*(O/32/O), *tet*(S/M), *tet*(W/32/O), *tet*(O/W/32/O), *tet*(O/W/32/O/W/O). Whereas, the *tet* genes which confer resistance to tetracyclines through efflux pumps such as *tet*(A), *tet*(B) represented 2% of detected AMR genes.

Macrolide resistance genes were the second most frequent, 14% of the detected AMR genes were *erm*; particularly *erm*(B) and *erm*(F) which encodes 23S ribosomal methyl-transferase and confers Macrolide–Lincosamide–StreptograminB (MLSb) resistant phenotype. Both the major facilitator efflux pump encoding gene *mef*(A) and the MDR ATP binding cassette (ABC-F) encoding gene *msr*(D) which confer resistance to macrolides and multiple antimicrobials that target the protein synthesis process such as tetracyclines, phenicols, lincosamides, and oxazolidinones, each of these gene families represented 5% of the detected AMR genes. Although there were no statistically significant differences in the prevalence of the above described AMR genes between the three study groups, it is noteworthy to highlight that the prevalence of macrolide resistance genes was slightly higher in the group of bronchiectasis patients not receiving antibiotic prophylaxis therapy compared to the other two groups (Figure 8.3).

Both the MDR ABC-F ribosomal protection protein gene family *lsr*(C) and ciprofloxacin phosphotransferase encoding gene *crpP* were equally detected in both bronchiectasis patients’ groups. The detected *crpP* gene which confer resistance to fluoroquinolones through antibiotic inactivation was not detected in the patient who was on ciprofloxacin prophylaxis therapy but in a patient on azithromycin prophylaxis having chronic *P. aeruginosa* colonization and was prescribed ciprofloxacin in the rescue pack for the treatment of exacerbations.
Figure 8.3. The prevalence of AMR genes in bronchiectasis patients on antibiotic prophylaxis (group 1, blue, n=8), bronchiectasis patients having frequent exacerbations treated with antibiotic rescue packs but not receiving antibiotic prophylaxis therapy (group 2, blue, n=9) and healthy participants in the comparator group (group 3, green, n=8) (All p-values >0.05 by KW); Rp: ribosomal protection proteins, AG: Aminoglycosides resistance, FQ: Fluoroquinolones resistance, RF: Rifamycin resistance, Sulf/tri: Sulfonamides/Trimethoprim.
The following AMR genes were detected only in the group of bronchiectasis patients not receiving antibiotic prophylaxis. First AMR genes which encode antibiotic inactivating enzymes such as class A β-lactamase enzymes encoding genes: \textit{blaBRO}, \textit{blaZ} and \textit{cfxA}, macrolide phosphotransferase encoding genes \textit{mphA}, aminoglycosides inactivating enzymes encoding genes: \textit{aac(3)} and \textit{ant(3”)}, rifampin ADP-ribosyltransferase encoding gene \textit{arr-3}. Second, the AMR genes which confer resistance through antibiotic target modification: sulfonamide resistant dihydropteroate synthase encoding genes \textit{sul1} and trimethoprim resistant dihydrofolate reductase encoding gene \textit{dfrA17} (\textbf{Figure 8.3}).

Forty-three percent of the AMR genes detected such as \textit{msr(D)}, \textit{mef(A)}, \textit{tet(O)}, \textit{crpP}, \textit{ACC (3)}, and \textit{ANT(3”)} have been known to be linked to mobile genetic elements such as plasmids, transposons and integrons. (Alcock et al., 2020)

\textbf{8.4. Discussion}

In this chapter, the airway resistome is studied in sputum of 17 patients with bronchiectasis; eight of whom were on-prophylactic antibiotics in comparison to eight healthy participants with normal lung function and without known chronic lung disease applying a novel metagenomics approach involving the state of art sequencing technology; the ONT MinION system.

To date, antimicrobial resistance patterns have principally been studied using culture based techniques mainly disc diffusion and minimum inhibitory concentration (MIC) techniques which have the advantage of demonstrating the expressed phenotypic changes in the susceptibility of bacteria towards antimicrobial agents; however, most of the established antimicrobial susceptibility testing protocols and the published breakpoints were developed for conventional pathogenic bacteria. These techniques do not provide insights into the carriage of resistance determinants by members of the wider bacterial community; especially, the unculturable bacteria. In addition, culture based techniques do not reveal the underlying molecular basis of resistance or the possibility of transmissibility of resistance determinants between bacteria.
Metagenomic sequencing provides a comprehensive technique to study AMR in microbial communities. It provides a holistic view on the diversity and the prevalence of AMR determinants at the microbial population level without introducing biases from cultures or a priori selection of a panel of resistance determinants as in PCR-based methods. Metagenomics has also the potential for discovering novel resistance determinants. (Taylor et al., 2018) In addition it can provide information on different aspects of the microbial community such meta-taxome including information about bacteria and DNA viruses, microbial genomics and functions. Nevertheless, the cost per sample is much higher compared to other techniques such as; culture, qPCRs and amplicon sequencing, because of the high sequencing depth required to be achieved to reliably characterise the microbial genes in specimens.

Applying metagenomic sequencing on sputum samples was a significant challenge given the high ratio of human DNA to microbial DNA in sputum. Efficient human DNA depletion or microbial DNA enrichment techniques are crucial for the success of metagenomics, without it the sequencing capacity would be wasted in sequencing human DNA while the microbial DNA would represent 2-6% of the sequenced reads. (Feigelman et al., 2017) The applied human deletion procedure was successful in reducing the human DNA load in the treated samples by an order of 100-fold such that the microbial reads in the sequencing results represented a mean of 77% of the total reads per sample.

MinION technology was proved to be a promising tool to be applied in metagenomic studies as it provides deep sequencing and generates exceptionally long sequencing reads. The depth of sequencing would depend on the number of multiplexed samples per run and length of sequencing run. The obtained lengths of sequence reads were in the range of 2000-7000 bp long. This length was good enough to characterise the diverse and repeat rich regions and helped with the assemblage of genomes, which would be challenging to assemble using short read sequencing technology like Illumina whose read length is typically 2x150 bp long. Nevertheless, Nanopore sequencing currently has relatively high error rates (Lavezzo et al., 2016, Chandak et al., 2020, Bowden et al., 2019) Consequently, it
was unreliable to account for the exact variants of the AMR genes; instead analysis was done at the level of AMR gene family. For reliable determination of the exact AMR gene variant, the sequencing libraries would need to be sequenced to greater depth. Since the technology is novel in resistome analyses, the results would need to be validated against the current standard using the Illumina short read sequencing technology which currently provides the best accuracy and the lowest error rate. (Schirmer et al., 2016, Almeida, 2017)

In Chapter 5, AMR was studied in sputum isolates and demonstrated the association between the repeated use of antibiotics and the high prevalence of phenotypic resistance in the pathogenic bacteria isolated from the sputum of chronic lung disease patients. In this chapter, the prevalence of AMR molecular determinants in a subset of the previously studied cohort: eight bronchiectasis patients on-prophylactic antibiotics (seven on azithromycin and one on ciprofloxacin), nine bronchiectasis patients not receiving antibiotic prophylaxis therapy were compared to eight healthy participants from the HIV negative comparator group (Chapter 7). It is assumed that the healthy participants had received fewer antibiotic courses compared to the bronchiectasis patients. The patients not receiving antibiotic prophylaxis therapy had more exacerbation frequency in the previous year which were treated with prolonged antibiotic courses. The most frequently prescribed antibiotics in the rescue pack were co-amoxiclav, doxycycline and ciprofloxacin.

Resistome analysis revealed a trend in which the AMR genes were richer (in terms of frequency of detection) in the samples from bronchiectasis patients on-prophylactic antibiotics, however, a broader diversity of AMR genes was observed in the group of bronchiectasis patients who had frequent exacerbations events treated with antibiotics rescue packs but were not receiving antibiotic prophylaxis therapy. Nevertheless, due to the small sample sizes, the study lacked the power to prove any of the observed trends to be statistically significant.

The \textit{tet} gene family which confer resistance to tetracyclines by ribosomal target protection and to a lesser extent the \textit{tet} genes encoding for efflux pumps were the most abundant resistance determinants in sputum samples in the three study
groups. The *erm* gene family which encodes 23S ribosomal methyltransferase enzyme and confers resistance to macrolides, lincosamide and streptogramin B (MLS-B phenotype) was generally highly abundant in the bronchiectasis patients, their prevalence was even higher in patients not receiving macrolide prophylaxis therapy (Figure 8.3). It is known that *erm*(B) gene and the tetracycline resistance determinant *tet*(M) are both present on the same mobile genetic element mainly in Gram positive bacteria but also in some Gram negative bacteria. (Clewell et al., 1995) In a previous study, resistance determinants which induce various macrolide resistance mechanisms were detected in sputum of cystic fibrosis patients who did not take macrolides in the previous year to the study. (Lim et al., 2014)

In two recent published studies, similar resistance determinants were found in the respiratory resistome of COPD and asthma patients. (Ramsheh et al., 2019, Taylor et al., 2019a) In the COPD study, the airway resistome in sputum and bronchial brush samples from a cohort of COPD patients and healthy subjects was studied using a panel of qPCR targeting 279 AMR determinant target. They found that resistance determinants to macrolides, tetracycline and β-lactams were the most abundant in both types of respiratory samples, with no significant differences in the prevalence between the study groups. An interesting observation was reported in which the AMR gene prevalence declined significantly through exacerbation and recovery in the group of COPD patients whose microbiome was dominated by *Gamma-Proteobacteria* but not in the patients whose microbiome was dominated by *Firmicutes*. Given that during exacerbations, the diversity of airway microbiome decline significantly and the members of *Proteobacteria* predominates the microbiome as demonstrated in Chapter 4 (Figure 4.5). Therefore, this may suggest that a major source of the resistance determinants detected in metagenomics may be the commensal microbiota belonging to phylum *Firmicutes*. (Ramsheh et al., 2019)

The AMAZES study was the first to study the resistome before and after the administration of azithromycin 500 mg thrice weekly in patients with persistent uncontrolled asthma in the context of double blind randomised placebo controlled trial using metagenomic sequencing approach. Out of 89 AMR genes which were
common in all samples, seven transmissible genes were enriched after azithromycin therapy but not placebo: five genes (erm(B), erm(F), msr(E), mef(A), and mel) confer resistance to macrolides, while two (tet(W) and tet(M)) confer resistance to tetracyclines. (Taylor et al., 2019a)

As shown in Chapter 5 (Figure 5.19), the commensal viridans streptococci isolates were demonstrated to express high resistance levels to multiple antibiotics; including macrolides and tetracyclines in the sputum of chronic lung disease patients using culture based antimicrobial suitability testing. Similarly, Taylor et al. in their study found that the total proportion of azithromycin-resistant isolates increased significantly after azithromycin therapy, of which viridans streptococci were the only significant subgroup. The whole genome sequencing of the resistant viridans streptococci isolates, detected the macrolide resistance genes erm(B), mel, and mef(A), as well as tet(W) and tet(M). Furthermore, both erm(B) and tet(M) genes were carried on a conjugative transposon (Tn916) in six isolates. Sequenced Prevotella and Porphyromonas isolates, which are common microbiota in the airways, had regions that map to erm(F), while, sequenced Rothia and Pseudomonas isolates had regions that map to msr(E). (Taylor et al., 2019a) In another previous study, a single treatment course with clarithromycin selected the highly macrolide resistant viridans streptococci strains coding for erm(B) in the oropharynx and the observed disruption persisted up to 180 days. (Malhotra-Kumar et al., 2007) All these results provide evidence that microbiota within the airways may act as a reservoir for transmissible resistance determinants.

The mechanism of resistance of some detected resistance determinants such as bla, cfxA, mphA, aac(3), ant(3’), and arr-3 is through inactivating β-lactams, macrolides, aminoglycosides and rifampin respectively. The expression of these inactivating enzymes by commensals bacteria in the ecosystem may risk reducing the bioavailability of the antibiotics at the target site and therefore its action on the pathogenic bacteria. Future metabolomic studies would be required to investigate this hypothesis.
Many AMR genes detected in the current study and reported by others are multi-drug resistance determinants and are known to be linked to mobile genetic elements. (Alcock et al., 2020) Therefore, there is a risk of horizontal transfer of resistance determinants from the commensal to pathogenic bacteria within the ecosystem. (Sherrard et al., 2014) In an in-vitro study, macrolide resistant *H. influenzae* isolates from children with cystic fibrosis were found to carry *erm*(B) and *erm*(F) which were transferred through conjugation experiments to other susceptible *H. influenzae* strains as well as to Gram positive bacterium *E. faecalis*. The resulted trans-conjugates exhibited increased MICs to macrolide antibiotics azithromycin and erythromycin. (Roberts et al., 2011) Further, studies are required to evaluate the possibility of horizontal gene transfer between the microbiota and common respiratory pathogens.

Dissemination of antimicrobial resistance in the wider population is another major concern. A recent study had detected AMR determinants such as (*erm*(B), *mef*(A) and *tet*(M)) in face-mask collected exhalation samples from COPD patients which may suggest airborne dissemination into the wider community and may explain the high prevalence of AMR determinants in healthy cohorts who were not exposed to these antimicrobial agents before. (Kennedy et al., 2018) In a Belgium cohort of healthy adults, 65% of the macrolide resistant viridans streptococci from the oropharynx carried either *erm*(B) either alone or with the macrolide efflux pump encoding gene *mef*(A). Co-resistance to tetracycline was identified in 73% isolates, of which *tet*(M) was present in 67% of the resistant isolates. (Malhotra-Kumar et al., 2004) Another example, in Albert’s *et al.* azithromycin prophylaxis trial in COPD patients, at the time of enrolment, the researchers reported that around 55% of participants exhibited macrolide resistance in the bacterial isolates from the nasopharynx before starting the therapy. (Albert et al., 2011)

To conclude, the airway microbiota can be regarded as a reservoir for AMR genes. In chronic lung disease, the resistome was enriched with diverse resistance determinants; especially, against macrolides and tetracyclines. Many of the detected AMR genes are known to be linked to mobile genetic elements that risks horizontal gene transfer between bacteria within the microbial ecosystem and risks
dissemination in the wider community. This pilot study demonstrated the potential of metagenomic sequencing approaches using MinION technology as a promising tool to explore different aspects of the microbial community directly from specimens and can provide the basis for future studies exploring microbiome and resistome.
Chapter 9: General Discussion and Conclusions

Chronic lung disease is one of the main causes of morbidity and mortality worldwide. COPD is already the third leading cause of death globally. Moreover, the diagnosis of bronchiectasis has been rising in different parts of the globe. The health and economic burden imposed by chronic lung disease on individuals and health-care systems is great. Acute exacerbations in chronic lung disease substantially contribute to morbidity, disability and increased health-care use and cost. Chronic infections with respiratory pathogens is a common feature among patients with chronic lung disease and can result in a poor prognosis and faster decline in lung function (as it was discussed in Chapter 1, 1.4.3.).

The recent discovery of the lung microbiota has transformed our understanding of respiratory infections and the pathogenesis of chronic lung disease. Previously, it was believed that the healthy lung was normally sterile and a respiratory infection happened when an inoculum of pathogenic micro-organism was contracted and gained access to the lower respiratory tract where it could overwhelm the immune defences; consequently, the pathogen proliferates unrestrictedly establishing infection.

Over the past decade, our knowledge about the lung microbiome has greatly expanded. Traditionally pathogenic bacteria such as *S. pneumoniae, H. influenzae* and *M. catarrhalis* were detected at relatively low abundances in the respiratory microbiome of healthy individuals as it was shown in the work presented here in the sputum of the healthy comparator groups (Figures 6.8 and 7.11) as well as in other lower respiratory specimens and lung tissue biopsies in other published studies. (Pragman et al., 2018, Yu et al., 2016) Therefore, a more appropriate hypothesis of respiratory infection from the microbiome perspective might be that the potential pathogenic microorganisms are normal opportunistic members of the microbial community that inhabit the airways. A disturbance in the normal homeostasis of the microbial ecosystem, for example due to contracting a new bacterial strain or viral infection or inflammation, may select for specific pathogens leading to uncontrolled proliferation that in turn induces a stronger host
inflammatory response; forming a positive feedback loop in which inflammation propels the bacterial proliferation that induces more inflammation until infection is established. (Dickson et al., 2016a, Kitsios et al., 2018) Further research is needed to investigate this hypothesis and identify the potential factors and stimuli that may elicit the disturbances of the microbial ecosystem causing infections.

In chronic lung disease, microbial dysbiosis rather than presence or load of individual pathogens has therefore been demonstrated here and in similar studies to be more critical in the pathogenesis of chronic respiratory conditions. Since the microbiota and pathogenic bacteria are living together in balance within one ecosystem, they are competing together for resources and nutrients under the control of the immune system of the host. Therefore, it is not just the presence and the load of pathogenic bacteria which matter in the pathophysiology of chronic lung conditions, biodiversity within the microbial community is also crucial. The decline in bacterial community diversity or disturbance in their microbial functions and metabolome may prepare the stage for potential opportunistic pathogens to take over and this can contribute to the pathogenesis of chronic lung disease. Microbiome studies have revealed that the α-diversity has significantly and consistently declined with the development of many chronic lung diseases; in addition, further decline has been observed during acute exacerbation events; that the bacterial community has been dominated by one or two taxa representing over 50% of the microbial reads.

Classical microbiology has focused on pathogenic micro-organisms. Well-established conventional microbiology laboratory protocols were designed to isolate and identify pathogens then characterise their antimicrobial susceptibility profiles. Therefore, limitations of culture based techniques in studying bacterial communities are related to being selective to a panel of pathogenic bacteria; while less sensitive to growing other microbiota, highly dependent on the operator’s experience, and time consuming due to the extended incubation periods required for some bacteria to grow. Also, different classes of bacteria would need various nutritional requirements in culture media and a wide spectrum of incubation conditions that can be expensive and laborious. In addition, low abundant bacteria
can be easily obscured by predominant ones and be missed. Nevertheless, culture based techniques have some advantages, such as identifying bacteria to species and strain level, characterising the phenotypes of bacteria; their expressed functions, virulence and AMR traits, and obtaining pure cultures of bacteria which can be employed in subsequent experiments to study interactions between the microorganisms and host immune components and to perform whole genome sequencing for full genomic characterisation of the bacterium.

On the other hand, 16S rRNA amplicon sequencing which has been used extensively to characterise microbiome profiles, and is a more comprehensive approach as it provides a holistic view over the composition of all bacterial inhabitants that normally reside in one habitat (meta-taxome) directly from specimens without the need for growing bacteria. The current main technical limitations and challenges of the 16S rRNA amplicon sequencing using NGS (discussed in detail in Chapter 3, 3.4) are related to the short read length generated by most of the available NGS technologies; therefore, only a subset of the 16S rRNA gene can be practically sequenced imposing a challenge over the choice of the universal primers set which provide the broadest universality and maximum resolution, and making it hard to identify most bacteria beyond the genus level in most cases. As it was shown in Chapter 3 (Figure 3.3 and 3.6), V3-V4 hypervariable regions of 16S rRNA sequencing were found to generate more accurate and robust representation of the microbiome profile at different level of taxonomy and was highly reproducible, while V5-V7 hypervariable regions offered more diversity at the lower taxonomic levels (genus level). Amplification steps in most protocols may introduce some biases in amplifying 16S rRNA sequences from different sources which may be reflected in a biased representation from the true structure of bacterial community. The lack of universal standardisation in bioinformatic analysis makes it difficult to meta-analyse data from various sources.

In my opinion, culture dependant and independent techniques are not alternatives; they complement each other in which 16S rRNA sequencing can be used for screening and identifying important microbial signatures which in turn can be targeted by the specific culture techniques to isolate and characterise the microbial
signatures of interest both phenotypically and genotypically, and study the bacterium’s interactions with other microorganisms and host defences through in-vitro and ex-vivo studies.

In the presented work, the airway microbiome was studied in sputum samples from a cohort of patients with COPD, and non-cystic fibrosis bronchiectasis; more than half of the cohort (56%) was receiving antibiotic prophylaxis therapy (Chapters 5). The airway microbiome in two further cohorts was studied; the first was young adult survivors of preterm birth (19 years old) with and without a history of neonatal bronchopulmonary dysplasia (BPD) compared to matched controls, of the same age, who were born full term (Chapter 6). The fourth cohort was healthy PLW-HIV compared to a group of healthy HIV uninfected adults who did not have defined respiratory morbidities (Chapter 7). The latter two cohorts were studied in the context of the presented thesis as these patient groups are known to be at higher risk of development of chronic respiratory morbidities compared to the general population, therefore, it was of interest to study whether an early bacterial dysbiosis could be detected that may indicate early development of chronic lung disease.

As it was shown in Chapters 6 and 7 (Figure 6.4 and Figure 7.6), the microbial community in sputum samples from the healthy comparator groups was composed mainly of the following five bacterial phyla in descending order: *Firmicutes, Actinobacteria, Bacteroidetes, Proteobacteria* and *Fusobacteria*. At the genus level, around 150 taxa were detected. The most frequently detected taxa whose relative abundance was above 1 % in the studied sputum samples were *Streptococcus, Rothia, Prevotella, Veillonella, Actinomyces, Haemophilus, Neisseria, Fusobacterium, Granulicatella, Gemella, Leptotrichia* and *Porphyromonas*. These may represent the core microbiota of the respiratory system. Many of these taxa are known to be of oral origin. Nevertheless, as discussed in Chapter 1, 1.3 previous studies have demonstrated that the microbial population in the oral cavity is the most probable source community for lung microbiome. (Venkataraman et al., 2015, Bassis et al., 2015) In Chapter three, 3.3.4, the microbiome profiles of paired sputum and saliva samples of ten patients demonstrated that despite the potential risk of upper respiratory contamination, a
meaningful microbial signature from the lower respiratory tract was not obscured in sputum samples. This is important since sputum is an easy non-invasive respiratory sample (in contrast, say to protected brush or BAL) and would promote large microbiome studies. Therefore, sputum was used in the presented work to study the respiratory microbiome with confidence in disease conditions as it can reflect perturbations in the microbial community composition in the lower airways.

In Chapter 4, as it was shown in Ch4, 4.3.4, the airway bacterial community was seen to be highly individualised and stable over the one year follow up period. Even though there were some temporal changes in response to factors such as exacerbations and use of antibiotics which caused disturbances in the phylogenetic make-up of the bacterial communities in the airways (Ch.4, 4.3.5), the bacterial communities were resilient as it reverted to its previous base-line structure as soon as the influencer ceased (Ch.4, 4.3.6).

The various forms of chronic lung disease studied in the context of this thesis were compared through a meta-analysis performed on all the studied cohorts. However, as there were significant differences between V3-V4 and V5-V7 16S rRNA sequencing as it was shown in Chapter 3, 3.3.4., Figure 3.9; and due to differences in the age, type of sputum (induced or spontaneous) and the method of sputum processing, DNA extraction and sequencing. Therefore, in the context of meta-analysis, patients with COPD and bronchiectasis were compared to the HIV negative healthy comparator group (Chapter 7, 7.2); whereas, the extremely preterm born groups: with and without history of neonatal BPD, were compared to their matched comparator group (EPICure controls). Nevertheless, in all forms of the studied chronic lung diseases, common trends were found.
Chapter 9: Discussion and Conclusions

First, a significant decline in the bacterial community α-diversity was observed in all forms of chronic lung disease in comparison to the corresponding comparator groups (Figure 9.1)

![Figure 9.1. Decline of α-diversity in chronic lung diseases (A) Comparison of Chao 1 α-diversity metric in COPD and bronchiectasis patients compared to the HIV negative comparator group, p<0.001 by ANOVA, V3-V4 16S rRNA sequencing (B) Comparison of Chao 1 α-diversity metric in the extremely preterm born (EP) groups with and without BPD compared to the EPICure controls, p=0.016 by ANOVA, V5-V7 16S rRNA sequencing. **p<0.01 by LSD ad-hoc test.](image)

The airway microbiome profiles of patients with COPD and bronchiectasis were distinct from each other and from that in healthy airways; where sputum samples from the groups of patients with COPD, bronchiectasis and healthy comparator groups, were separated in the PCoA of both weighted and unweighted UniFrac β-diversity based on the diagnosed respiratory condition (Figure 9.2). Therefore, all normal airway microbiome was similar and clustered together; while in the microbiome in chronic lung diseases, each was distinguished in some way.
Figure 9.2. Airway microbiome is distinguished by the respiratory condition: COPD (red, n=23), bronchiectasis (blue, n=61), healthy controls (green, n=32) in the PCoA of (A) Unweighted UniFrac $\beta$-diversity (p=0.001 by PERMANOVA, p=0.001 by ANOSIM) (B) Weighted UniFrac $\beta$-diversity (p=0.001 by PERMANOVA, p=0.002 by ANOSIM)

Second, a specific pattern of bacterial dysbiosis characterised by a shift in the bacterial community composition away from the bacterial phylum Bacteroidetes which includes most anaerobic bacteria, towards either Firmicutes or Proteobacteria was observed in different forms of chronic lung disease. In the extreme preterm birth survivors and COPD patients, the shift was mainly towards the Firmicutes; which encompass Gram positive bacteria such as Streptococcus. In COPD and bronchiectasis patients; a decline in the relative abundance of Actinobacteria and Fusobacteria as well as Bacteroidetes was also observed. Greater decline in the relative abundance of Bacteroidetes was associated with the presence of airflow obstruction. Proteobacteria which encompass common Gram negative bacterial pathogens was significantly enriched in the airways of bronchiectasis patients. (Figure 9.3).

Figure 9.3. Bacterial dysbiosis at the phylum level in the airways of patients with COPD, bronchiectasis and adult survivors of extreme preterm birth with history of developing neonatal BPD.

*Significant difference p<0.05.
Chapter 9: Discussion and Conclusions

Third, genus *Prevotella* specifically was suppressed in comparison to the corresponding healthy comparator groups in the extreme preterm birth survivors (Chapter 6) as well as in both COPD and bronchiectasis patients as shown in Figure 9.4 A and B. Furthermore, the relative abundance of genus *Prevotella* correlated with the FEV1 % predicted in all studied cohorts. (Figure 9.4 C and D)

As it was shown in Chapter 6, Figure 6.6, *Prevotella melaninogenica* was found to be the species which was driving the observed decline in genus *Prevotella* and phylum *Bacteroidetes*.

![Figure 9.4](image)

**Figure 9.4.** The relative abundance of genus *Prevotella* significantly declined in the presence of chronic respiratory disease and correlated with FEV1% predicted.

*Prevotella* species are strictly anaerobic and require specific culture and incubation conditions for growing them in the laboratory. In previous studies, *Prevotella melaninogenica* was the most frequently isolated species from the sputum of cystic
fibrosis patients in anaerobic cultures. (Tunney et al., 2008, Field et al., 2010, Council, 2013) Before the discovery of lung microbiome, these species were regarded as strictly pathogenic bacteria; nevertheless, the application of 16S rRNA sequencing on various lower respiratory specimens has expanded our knowledge about the respiratory microbiota and *Prevotella* is a genus that is present on almost all microbiome lists of lower respiratory samples in patients as well as healthy subjects in the published literature. It is not completely clear; how the anaerobic bacteria survive in the aerobic conditions of the respiratory tract. However, it is known from previous studies that there is a gradient of decreasing partial oxygen pressure along the respiratory tract with the lowest values in the smallest bronchioles; this is opposite to a gradient of increasing partial pressure of carbon dioxide. (Man et al., 2017)

The presence of *Prevotella* species and its relative abundance has been associated with health and clinical stability of chronic respiratory conditions, an observation that has been repeatedly reported in various studies (as discussed in Chapter 6, 6.4). Nevertheless, the direction of the causality is impossible to define in the context of observational studies and would require future research involving animal microbiome models to identify whether *Prevotella* was a cause or consequence. A potential explanation can be attributed to the developing inflammatory process and rising oxidative stress that may interfere with the survival of anaerobic bacteria in airways, therefore, leading to decline in their relative abundance. Nevertheless, in this case the decline in *Prevotella* may be used as indicator on an evolving respiratory pathology. The *Prevotella* species and strains that are associated with respiratory health, the role of *Prevotella* species in the context of the lung microbial population, its interactions with other microbes and the host immunity are all points for future research.

It is noteworthy that 16S rRNA sequencing is a semi-quantitative method in which the abundance of the taxa is expressed in terms of relative abundance; i.e. the proportion of sequences classified into one taxon relative to the total number of sequences representing the whole microbiome of a sample. Microbiome data is more focused on representing the balance between bacterial constituents in a
microbial population. Therefore, an increase or decrease in relative sequence
abundances does not necessarily mean that the corresponding bacterial group
occurs at higher or lower cell densities; so the results of microbiome needs to be
augmented by qPCR methods. As it was as shown in Chapter 3.3.5-3.3.7, qPCR
methods provide unparalleled specificity and sensitivity to detect and measure the
quantitative loads of specific pathogenic bacteria. qPCR demonstrated that the
total bacterial density in sputum was significantly higher in the bronchiectasis
patients compared to the healthy comparator group; however, it did not differ
significantly in COPD patients or in extreme preterm birth survivors as shown in
Figure 9.5.

![Figure 9.5](image)

**Figure 9.5.** Comparison between the total bacterial densities in sputum in chronic
lung diseases expressed in the total 16S rRNA gene (copies/µL), EP: Extreme Pre-
term birth survivors.

The prevalence and loads of *S. pneumoniae, H. influenzae* and *M. catarrhalis* were
comparable in all studied groups: COPD, bronchiectasis, PLW-HIV, and HIV
negative comparator group as well as in the EPICure study groups. On the
contrary, *P. aeruginosa* was more prevalent and present at much higher bacterial
load in patients with COPD and bronchiectasis compared to all other participants.
*P. aeruginosa* was more frequently detected in PLW-HIV (Ch7, **Figure 7.7C**) and
CVID patients (Ch.5, 5.3.5) compared to the corresponding comparator groups;
even though its load was very low compared to that in established chronic infections in more advanced chronic lung disease, which means that *P. aeruginosa* could occupy niches in the airway in the presence of an element of immunodeficiency (acquired in the former and primary in the latter cohort), despite the antiretroviral therapy and immunoglobulin replacement therapy in the two cohorts respectively. In contrast, *P. aeruginosa* was completely absent in the extreme preterm birth survivors as well as the two comparator groups.

In PLW-HIV, the microbiota composition was similar to that of the healthy comparator group (as shown in Chapter 7, Figure 7.6). Nevertheless, some taxa which are classified as potential respiratory pathogens and gut bacteria such as *S. aureus*, *P. aeruginosa*, *Klebsiella*, *Enterobacteriaceae* and *Bilophila* were enriched in the airways of PLW-HIV. The enrichment of lung microbiome with gut bacteria may be explained by the breakdown of the gut mucosal barrier in HIV infection leading to increased translocation of gut microbes and microbial products which they can eventually lodge and find a favourable ecological niche in HIV infected lungs. The carriage of these pathogenic bacteria may represent a nucleus that could contribute to respiratory morbidities later in these populations. As it was shown in Ch7, 7.3.6, the prompt start of effective antiretroviral therapy (ART) for HIV infected patients may reduce the carriage of the pathogenic bacteria in respiratory tract.

Chronic respiratory diseases are not curable; however, various forms of treatment such as inhaled bronchodilators and corticosteroids, pulmonary rehabilitation programmes and antimicrobial treatment can help to manage symptoms, improve patients’ quality of life, and perhaps decelerate the decline in the lung function. Patients with chronic lung diseases often receive frequent and prolonged courses of antimicrobial agents in current clinical practice in the UK; especially during periods of acute exacerbations. β-lactams, tetracyclines, macrolides and fluoroquinolones are among the first and second line treatments of infective exacerbations. Nevertheless, the benefit of antibiotic in the treatment of acute exacerbation; especially in COPD has remained a matter of controversy for many years and the evidence supporting the universal treatment of exacerbation with
antibiotics is limited. (GOLD, 2020) This may be attributed to the difficulty in proving the infectious aetiology in acute exacerbation of COPD and clinically identifying acute exacerbations of bacterial origin. As it was shown in Chapter 4, 4.3.5.1, microbiome data available from acute exacerbation events in a few patients revealed a decline in the \( \alpha \)-diversity and elevation in the relative abundance of gamma-Proteobacteria. This needs to be confirmed by future larger study looking specifically at microbiome changes around acute exacerbation events. Nevertheless, this observation highlights the potential of microbiome signatures in defining exacerbation and may help in identifying the patients who may benefit from antibiotic treatment.

Antibiotic prophylaxis therapy has been one of the proposed strategies to manage COPD and bronchiectasis; especially, in advanced cases. Substantial scientific evidence derived from randomised controlled clinical trials supports the use of macrolide prophylaxis therapy; especially erythromycin and azithromycin, as they were shown to reduce the rate of exacerbations and improve quality of life. Studying the adverse consequences of the prolonged antimicrobial therapy in terms of its impact on the homeostasis of the bacterial communities in the body and its contribution to the AMR was the aim of the clinical study presented in Chapters 5 and 8.

In Chapter 5, the airway microbiome of patients with COPD and non-cystic fibrosis bronchiectasis receiving antibiotic prophylaxis therapy, most of them (66%) on macrolides, was compared to those not. As it was shown in Ch5, 5.3.2, antibiotic prophylaxis therapy was associated with reduced phylogenetic alpha diversity (PD-whole tree) of the bacterial communities and lower bacterial density in sputum samples (as shown in Ch.5, Figure 5.9C). However, antimicrobial treatment suppressed certain pathogenic bacteria such as \( P. \) aeruginosa, \( M. \) catarrhalis and some members of family \( Enterobacteriaceae \) in the airways of patients receiving antibiotic prophylaxis therapy, this observation was also confirmed through standard bacteriological cultures and qPCRs for specific respiratory pathogens (as shown in Ch.5, Figures 5.7, 5.8, 5.9, 5.15). This may suggest that macrolides, especially, azithromycin, are not only acting through their antimicrobial activity.
which may account for the reductions in the total bacterial load in the airways, but also the current results may support the proclaimed immune-modulatory effect of macrolides (as discussed in Ch.1, 1.6.1.) since it selectively suppressed specific taxa that encompass potential respiratory pathogens without disrupting the homeostasis of the respiratory microbiota (as it was shown in Ch.5, 5.3.4). Segal et al. provided interesting findings that may explain the immune-modulatory activity of azithromycin in which some bacterial metabolites were found to be upregulated in COPD patients who received azithromycin maintenance therapy; whereas, pro-inflammatory chemokines were suppressed. In an ex-vivo study, they demonstrated that the upregulated microbial metabolites, but not azithromycin, down regulated the secretion of the same chemokines in LPS-induced alveolar macrophages. (Segal et al., 2017) Therefore, these findings suggest that the activity of azithromycin may be mediated through the modulation of the metabolome of the microbiota which in turn mediates the observed anti-inflammatory effect.

The risk of emergence and dissemination of AMR associated with the extensive use of antibiotics in the sector of chronic lung disease is of great concern; especially with the current global situation of an evolving AMR crisis. In Chapter 5, using the standard antibiotic disc diffusion technique, high resistance to penicillins and fluoroquinolones was detected among the pathogenic Gram negative sputum isolates while resistance to macrolides and tetracyclines was more among Gram positives; including commensals viridans streptococci. In general, the detected AMR were more related to the prescribed antibiotics in rescue packs kept by the patients for the prompt initiation of self-treatment of exacerbations rather than to the prophylactic antibiotic. Nevertheless, multi-drug resistant viridans streptococci were significantly more abundant in patients receiving antibiotic prophylaxis therapy. In Chapter 8, although the pilot study was underpowered, a metagenomics approach on a smaller sample size demonstrated that the airway resistome was enriched with diverse resistance determinants, especially against macrolides and tetracyclines in chronic lung disease patients compared to the healthy controls. A trend was observed in which the AMR genes were richer (in terms of frequency of detection) in the samples from bronchiectasis patients on-
Chapter 9: Discussion and Conclusions

prophylactic antibiotics (as shown in Figure 8.2), however, a broader diversity of AMR genes was observed in the group of bronchiectasis patients who had frequent exacerbations events treated with antibiotics but were not receiving antibiotic prophylaxis therapy (as shown in Figure 8.3). Many of the detected AMR genes in resistome analysis are known to be linked to mobile genetic elements that may indicate horizontal gene transfer between bacteria within the microbial ecosystem with risk of dissemination in the wider community.

That said, the practice of prescribing antibiotic rescue packs might be more concerning in this sector rather than antibiotic prophylaxis therapy; especially in patients who suffers frequent exacerbations. This may justify the added benefit of the antibiotic prophylaxis therapy in those patients. Patients’ education on the rational use of antibiotics during exacerbations is crucial. Also, rotating the antibiotic prescriptions between different classes of antibiotics may help to tackle the AMR problem in this sector, although sometimes the available options might be limited. The development and clinical implementation of an efficient microbiome based diagnostic tool would be invaluable; as it would promote personalised medicine by tailoring antimicrobial treatment based on the individual microbiome, as well as improving chemotherapeutic antimicrobial prescribing practices and slowing down the development of AMR in this sector.

The presented work has mainly studied the changes in the bacterial community composition in sputum that may be associated with chronic respiratory disease development and prognosis using 16S rRNA profiling. Nevertheless, other microbial inhabitants such as viruses, fungi and protozoa, the host immune and inflammatory responses, and environmental factors such as the exposure to allergens may also play important roles and contribute to the pathogenesis of the chronic respiratory disease; therefore, all these factors need to be considered in future studies. Metagenomic sequencing provides deeper insights into metatransomes, resistome and microbial functional analysis of the microbial communities over the 16S rRNA sequencing. Since it can avoid the initial amplification step and it captures other microbes such as DNA viruses and eukaryotes, it therefore provides a more accurate representation and broader view over the microbial
community composition. The main challenges in the metagenomic sequencing was the low microbial to human DNA ratio that would need an efficient human DNA depletion technique without much loss in the microbial DNA, and complicated bioinformatic analysis tools that can be challenging to implement. Using the MinION system was a great privilege because it generated long sequence reads which facilitated the assemblage of microbial genomes due to the relatively long scaffold reads. The workflow and bioinformatic pipeline developed and optimised by Dr. Lucy van Drop and myself for metagenomic sequencing using the MinION system in the context of the pilot study presented in Chapter 8 would be a useful resource for future studies studying microbiome and resistome in sputum. This would improve the phylogenetic and taxonomic resolution in future studies, and may enable the identification of the source of AMR genes and their carriage on mobile genetic elements.

Finally, microbiome research has the capacity to discover novel therapeutics or identify innovative markers to be used as diagnostic tools that would promote personalised medicine in the future. There are existing examples of novel microbiome based therapeutics derived from gut microbiome research to treat challenging gut conditions. Fecal Microbiota transplantation (FMT) has been an innovative microbiome based therapeutic for the treatment of *Clostridium difficile* superinfection and some reports have shown that it was also successful in knocking down multidrug resistant bacteria in the gut. Probiotics and prebiotics are another approach to modulate the microbial community composition towards a healthier balanced microbiome structure. The difference between the probiotics and prebiotics is that the former is formulated as a pharmaceutical therapeutic delivering beneficial bacteria identified from gut microbiome research; while the latter is mostly a dietary supplement delivering nutritional factors such as certain fibres that may promote the growth of ‘good’ bacteria. However, manipulating the lung microbiome is still at its infancy, in the future, we may have respiratory inhaled probiotics to modulate the microbiome structure towards more healthy composition or a microbiome based diagnostic tools which have the potential to predict exacerbations and response to chemotherapeutic antimicrobial treatment.
Chapter 9: Discussion and Conclusions

To conclude, the work presented has demonstrated that the airway microbiome was stable over one year and it was highly resilient. Despite some inter-individual differences, all normal airway microbiomes were similar and clustered together; while for the microbiome in chronic lung diseases, each was distinct in some way. Nevertheless, all forms of chronic lung disease were associated with a decline in the biodiversity of bacterial community and a significant dysbiosis manifested as a shift in the bacterial community composition away from phylum Bacteroidetes; particularly genus Prevotella whose relative abundance correlated with important lung function parameters. In PLW-HIV, potential respiratory pathogens and gut bacteria were enriched in the airway microbiome which may explain why this population is at higher risk of respiratory morbidities and pneumonia. The airway of chronic lung disease patients harboured resistant microbiota and various AMR determinants; especially for macrolides and tetracyclines. The AMR determinants were more related to the antibiotics used as rescue packs for prompt initiation of self-treatment of exacerbations than the prophylactic antibiotics. Antibiotic prophylaxis therapy was associated with lower total bacterial load and it suppressed some recognised pathogenic bacteria in the airways of chronic lung disease patients. Nevertheless, it had minimal effect on the homeostasis of the respiratory microbiota; as no definite directional shift in the microbiome composition was identified at the group level of patients on-prophylactic antibiotics.
10. References


References


References


Faintuch, J. and Faintuch, S. 2019. Microbiome and Metabolome in Diagnosis, Therapy, and other Strategic Applications. Elsevier Inc.


dysplasia - prevalence, severity and predictive factors in a national cohort

Fawke, J., Lum, S., Kirkby, J., Hennessy, E., Marlow, N., Rowell, V., Thomas, S.
and Stocks, J. 2010. Lung function and respiratory symptoms at 11 years in
children born extremely preterm: the EPI Cure study. Am J Respir Crit Care

Feigelman, R., Kahlert, C., Baty, F., Rassouli, F., Kleiner, R., Kohler, P. and
Brutsche, M. 2017. Sputum DNA sequencing in cystic fibrosis: non-invasive
access to the lung microbiome and to pathogen details. Microbiome, 5(1).


Flume, P. A., Chalmers, J. D. and Olivier, K. N. 2018. Advances in bronchiectasis:
endotyping, genetics, microbiome, and disease heterogeneity. The Lancet

Fuks, G., Elgart, M., Amir, A., Zeisel, A., Turnbaugh, P. J., Soen, Y. and Shental,
N. 2018. Combining 16S rRNA gene variable regions enables high-
resolution microbial community profiling. Microbiome, 6(1), pp. 17. doi:
10.1186/s40168-017-0396-x.

Gao, Y.-H., Guan, W.-J., Xu, G., Lin, Z.-Y., Tang, Y., Lin, Z.-M., Gao, Y., Li, H.-M.,
Zhong, N.-S., Zhang, G.-J. and Chen, R.-C. 2015. The role of viral infection
in pulmonary exacerbations of bronchiectasis in adults: a prospective study.

Garcha, D. 2014. Prevalence and Load of Airway Bacteria in Chronic Obstructive
Pulmonary Disease. Thesis (Ph.D.), University College London.

Garcha, D. S., Thurston, S. J., Patel, A. R., Mackay, A. J., Goldring, J. J.,
prevalence and load of airway bacteria using quantitative PCR in stable and

of local variations in antibiotic consumption and geographical differences of
erithromycin and penicillin resistance in Streptococcus pneumoniae.
Journal of Clinical Microbiology, 40(1), pp. 159-164. doi:

George, M. P., Singh, V. and Gladwin, M. T. 2016. Noninfectious and
Nonneoplastic Conditions Associated with Human Immunodeficiency Virus


References


References


References


References


Rogers, G. B., Daniels, T. W., Tuck, A., Carroll, M. P., Connett, G. J., David, G. J. and Bruce, K. D. 2009. Studying bacteria in respiratory specimens by using


Snell, N., Gibson, J., Jarrold, I. and Quint, J. K. 2019. Epidemiology of bronchiectasis in the UK: Findings from the British lung foundation's


Taylor, S., Leong, L., Mobegi, F., Choo, J., Wesselingh, S., Yang, I., Upham, J., Reynolds, P., Hodge, S., James, A., Jenkins, C., Peters, M., Baraket, M.,


References


References


Appendices

Appendix 1: Approved Ethics Documents of “Changes in Microbiome with Treatment in Chronic Lung Diseases”
- Appendix 1.1. HRA Approval Letter
- Appendix 1.2. Research Protocol V.1.5.3
- Appendix 1.3. Participant Information Sheet
- Appendix 1.3. Consent Form
- Appendix 1.4. Approved IRAS Form
- Appendix 1.5. Symptom Detailing Diary Card

Appendix 2. Study Protocol of CVID Patients

Appendix 3 Supplementary Methods
- Appendix 3.1: Patient’s detailed advice sheet on Sputum Samples’ Collection
- Appendix 3.2: Preparation of Artificial Sputum Base
- Appendix 3.3: UCL Genomics 16S rRNA Metagenomic Profiling Assay Protocol
- Appendix 3.4: Internal Standard Curve of P. aeruginosa and total bacterial load in each qPCR run
- Appendix 3.5. Sample of MALDI-TOF Report
- Appendix 3.6. Coverage of 16S rRNA qPCR (A) primers: Bact340F and Bact806R and (B) Probe
- Appendix 3.7. Specificity of the typical respiratory qPCR primers and probes to the corresponding targets through BLAST Analysis
  A. H. influenzae qPCR forward and reverse primers’ specificity through BLAST Analysis
  B. H. influenzae qPCR probe (reverse complement) and reverse primer Specificity through BLAST Analysis
  C. S. pneumoniae qPCR forward and reverse primers’ specificity through BLAST Analysis
  D. S. pneumoniae qPCR forward primer and probe specificity through BLAST Analysis
  E. M. catarrhalis qPCR forward and reverse primers’ specificity through BLAST Analysis
  F. M. catarrhalis qPCR probe and reverse primers’ specificity through BLAST Analysis
  G. P. aeruginosa qPCR forward and reverse primers’ specificity through BLAST Analysis
  H. P. aeruginosa qPCR probe and reverse primer specificity through BLAST Analysis
- Appendix 3.8. BLAST Analysis of the differentially abundant taxa in PLW-HIV (Chapter 7)
  3.8.A. Pseudomonas
  3.8.B. Klebsiella
  3.8.C. Bilophila
  3.8.D.i. S. aureus NCBI BLAST
  3.8.D.ii. S. aureus SILVA ACT
- Appendix 3.9. MinION Metagenomic Sequencing Library Tape Station Report
- Appendix 3.10. MinION Sequencing Report

Appendix 4. Publications
Appendix 1: Approved Ethics Documents of “Changes in Microbiome with Treatment in Chronic Lung Diseases”

Appendix 1.1. HRA Approval Letter

NHS
Health Research Authority

Prof. Timothy D. McHugh
Professor of Medical Microbiology, Div of Infection & Immunity
Faculty of Medical Sciences, University College London and
Director of UCL Centre for Clinical Microbiology
University College London
UCL Centre for Clinical Microbiology, Royal Free Campus,
Rowland Hill Street,
London, UK
NW3 2PF

14 November 2016

Dear Professor McHugh

Study title: Changes of Microbiome in Response to Treatment in Patients with Chronic Lung Diseases
IRAS project ID: 202883
REC reference: 16/LO/1490
Sponsor: University College of London

I am pleased to confirm that HRA Approval has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications noted in this letter.

Participation of NHS Organisations in England
The sponsor should now provide a copy of this letter to all participating NHS organisations in England.

Appendix B provides important information for sponsors and participating NHS organisations in England for arranging and confirming capacity and capability. Please read Appendix B carefully, in particular the following sections:

- Participating NHS organisations in England – this clarifies the types of participating organisations in the study and whether or not all organisations will be undertaking the same activities.
- Confirmation of capacity and capability - this confirms whether or not each type of participating NHS organisation in England is expected to give formal confirmation of capacity and capability. Where formal confirmation is not expected, the section also provides details on the time limit given to participating organisations to opt out of the study, or request additional time, before their participation is assumed.
- Allocation of responsibilities and rights are agreed and documented (4.1 of HRA assessment criteria) - this provides detail on the form of agreement to be used in the study to confirm capacity and capability, where applicable.
Further information on funding, HR processes, and compliance with HRA criteria and standards is also provided.

It is critical that you involve both the research management function (e.g. R&D office) supporting each organisation and the local research team (where there is one) in setting up your study. Contact details and further information about working with the research management function for each organisation can be accessed from www.hra.nhs.uk/hra-approval.

Appendices
The HRA Approval letter contains the following appendices:
- A – List of documents reviewed during HRA assessment
- B – Summary of HRA assessment

After HRA Approval
The document “After Ethical Review – guidance for sponsors and investigators”, issued with your REC favourable opinion, gives detailed guidance on reporting expectations for studies, including:
- Registration of research
- Notifying amendments
- Notifying the end of the study
The HRA website also provides guidance on these topics, and is updated in the light of changes in reporting expectations or procedures.

In addition to the guidance in the above, please note the following:
- HRA Approval applies for the duration of your REC favourable opinion, unless otherwise notified in writing by the HRA.
- Substantial amendments should be submitted directly to the Research Ethics Committee, as detailed in the After Ethical Review document. Non-substantial amendments should be submitted for review by the HRA using the form provided on the HRA website, and emailed to hra.amendments@nhs.net.
- The HRA will categorise amendments (substantial and non-substantial) and issue confirmation of continued HRA Approval. Further details can be found on the HRA website.

Scope
HRA Approval provides an approval for research involving patients or staff in NHS organisations in England.

If your study involves NHS organisations in other countries in the UK, please contact the relevant national coordinating functions for support and advice. Further information can be found at http://www.hra.nhs.uk/resources/applying-for-reviews/nhs-hsc-nd-review/.

If there are participating non-NHS organisations, local agreement should be obtained in accordance with the procedures of the local participating non-NHS organisation.
User Feedback
The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please email the HRA at hra.approval@nhs.net. Additionally, one of our staff would be happy to call and discuss your experience of HRA Approval.

HRA Training
We are pleased to welcome researchers and research management staff at our training days – see details at http://www.hra.nhs.uk/hra-training/

Your IRAS project ID is 202883. Please quote this on all correspondence.

Yours sincerely

Sharon Northe
Senior Asessor

Email: [redacted]

Copy to: Miss Jenise Davidson, Portfolio co-ordinator/ Joint Research Office – Sponsor and R&D contact

NIHR CRN Portfolio Applications Team
### Appendix A - List of Documents

The final document set assessed and approved by HRA Approval is listed below.

<table>
<thead>
<tr>
<th>Document</th>
<th>Version</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Costing template [Cost and Finance Spread Sheet]</td>
<td></td>
<td>04 July 2016</td>
</tr>
<tr>
<td>Covering letter on headed paper [Cover Letter ]</td>
<td></td>
<td>26 July 2016</td>
</tr>
<tr>
<td>Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [UCL Insurance certificate]</td>
<td></td>
<td>22 July 2016</td>
</tr>
<tr>
<td>IRAS Application Form [IRAS_Form_14102016]</td>
<td></td>
<td>14 October 2016</td>
</tr>
<tr>
<td>Other [Risk Assessment by Sponsor]</td>
<td>1</td>
<td>17 August 2015</td>
</tr>
<tr>
<td>Other [Sponsorship Checklist]</td>
<td>2</td>
<td>16 June 2016</td>
</tr>
<tr>
<td>Other [Statement of Activities]</td>
<td>1.0</td>
<td>11 November 2016</td>
</tr>
<tr>
<td>Other [Schedule of Events]</td>
<td>1.0</td>
<td>11 November 2016</td>
</tr>
<tr>
<td>Other [PI letter in response to the provisional REC letter]</td>
<td></td>
<td>07 October 2016</td>
</tr>
<tr>
<td>Other [Sylvia Rofael eOCP course certificate]</td>
<td></td>
<td>03 June 2016</td>
</tr>
<tr>
<td>Participant consent form [Patient Consent 18/0178_v_1.5_25.10.16]</td>
<td>1.5</td>
<td>25 October 2016</td>
</tr>
<tr>
<td>Participant consent form [Patient Consent 18/0178_v_1.5_25.10.16] with track changes</td>
<td>1.5</td>
<td>25 October 2016</td>
</tr>
<tr>
<td>Participant information sheet (PIS) [PIS 18/178_v_1.6_25.10.16] with track changes</td>
<td>1.6</td>
<td>25 October 2016</td>
</tr>
<tr>
<td>Participant information sheet (PIS) [PIS 18/0178_v_1.6_25.10.2016] with track changes</td>
<td>1.6</td>
<td>25 October 2016</td>
</tr>
<tr>
<td>Referee’s report or other scientific critique report [Dr. David Spratt]</td>
<td></td>
<td>05 July 2016</td>
</tr>
<tr>
<td>Referee’s report or other scientific critique report [Dr. Marc Lipman]</td>
<td></td>
<td>01 July 2016</td>
</tr>
<tr>
<td>Research protocol or project proposal [Research Protocol 16/0178 signed]</td>
<td>1.5.3</td>
<td>07 July 2016</td>
</tr>
<tr>
<td>Sample diary card/patient card [Sample Diary Card October ]</td>
<td>2</td>
<td>27 September 2016</td>
</tr>
<tr>
<td>Summary CV for Chief Investigator (CI) [Prof. Timothy McDHugh CV]</td>
<td></td>
<td>28 March 2016</td>
</tr>
<tr>
<td>Summary CV for student [Sylvia Rofael C.V]</td>
<td></td>
<td>20 May 2016</td>
</tr>
<tr>
<td>Summary CV for supervisor (student research) [Dr. John Hurst C.V]</td>
<td></td>
<td>15 March 2016</td>
</tr>
<tr>
<td>Summary, synopsis or diagram (flowchart) of protocol in non technical language [Study flow chart and summary]</td>
<td>1.5.3</td>
<td>07 July 2016</td>
</tr>
<tr>
<td>Validated questionnaire [CAT Questionnaire]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Validated questionnaire [St. George Respiratory Questionnaire (SGRO)]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B - Summary of HRA Assessment

This appendix provides assurance to you, the sponsor and the NHS in England that the study, as reviewed for HRA Approval, is compliant with relevant standards. It also provides information and clarification, where appropriate, to participating NHS organisations in England to assist in assessing and arranging capacity and capability.

For information on how the sponsor should be working with participating NHS organisations in England, please refer to the, participating NHS organisations, capacity and capability and Allocation of responsibilities and rights are agreed and documented (4.1 of HRA assessment criteria) sections in this appendix.

The following person is the sponsor contact for the purpose of addressing participating organisation questions relating to the study:

Name: [blank]
Tel: [blank]
Email: [blank]

### HRA assessment criteria

<table>
<thead>
<tr>
<th>Section</th>
<th>HRA Assessment Criteria</th>
<th>Compliant with Standards</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>IRAS application completed correctly</td>
<td>Yes</td>
<td>No comments</td>
</tr>
<tr>
<td>2.1</td>
<td>Participant information/consent documents and consent process</td>
<td>Yes</td>
<td>No comments</td>
</tr>
<tr>
<td>3.1</td>
<td>Protocol assessment</td>
<td>Yes</td>
<td>No comments</td>
</tr>
<tr>
<td>4.1</td>
<td>Allocation of responsibilities and rights are agreed and documented</td>
<td>Yes</td>
<td>The sponsor is using the Statement of Activities as the agreement between the sponsor and the NHS organisation to participate.</td>
</tr>
<tr>
<td>4.2</td>
<td>Insurance/indemnity arrangements assessed</td>
<td>Yes</td>
<td>Where applicable, independent contractors (e.g. General Practitioners) should ensure that the professional indemnity provided by their medical defence organisation covers the activities expected of them for this research study</td>
</tr>
<tr>
<td>Section</td>
<td>HRA Assessment Criteria</td>
<td>Compliant with Standards</td>
<td>Comments</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------</td>
<td>--------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>4.3</td>
<td>Financial arrangements assessed</td>
<td>Yes</td>
<td>No funding is provided to NHS organisations as detailed in the Statement of Activities. The submitted schedule of events does not have the visit or cost distribution completed by the sponsor. Further information is provided in the submitted cost and finance spreadsheet.</td>
</tr>
<tr>
<td>5.1</td>
<td>Compliance with the Data Protection Act and data security issues assessed</td>
<td>Yes</td>
<td>No comments</td>
</tr>
<tr>
<td>5.2</td>
<td>CTIMPS – Arrangements for compliance with the Clinical Trials Regulations assessed</td>
<td>Not Applicable</td>
<td>No comments</td>
</tr>
<tr>
<td>5.3</td>
<td>Compliance with any applicable laws or regulations</td>
<td>Yes</td>
<td>No comments</td>
</tr>
<tr>
<td>6.1</td>
<td>NHS Research Ethics Committee favourable opinion received for applicable studies</td>
<td>Yes</td>
<td>No comments</td>
</tr>
<tr>
<td>6.2</td>
<td>CTIMPS – Clinical Trials Authorisation (CTA) letter received</td>
<td>Not Applicable</td>
<td>No comments</td>
</tr>
<tr>
<td>6.3</td>
<td>Devices – MHRA notice of no objection received</td>
<td>Not Applicable</td>
<td>No comments</td>
</tr>
<tr>
<td>6.4</td>
<td>Other regulatory approvals and authorisations received</td>
<td>Not Applicable</td>
<td>No comments</td>
</tr>
</tbody>
</table>

**Participating NHS Organisations in England**

This provides detail on the types of participating NHS organisations in the study and a statement as to whether the activities at all organisations are the same or different.

There is one NHS organisation undertaking the study. Therefore there is only one study site ‘type’ involved in the research.

The Chief Investigator or sponsor should share relevant study documents with participating NHS
organisations in England in order to put arrangements in place to deliver the study. The documents should be sent to both the local study team, where applicable, and the office providing the research management function at the participating organisation. For NIHR CRN Portfolio studies, the Local CRN contact should also be copied into this correspondence. For further guidance on working with participating NHS organisations please see the HRA website.

If chief investigators, sponsors or principal investigators are asked to complete site level forms for participating NHS organisations in England which are not provided in IRAS or on the HRA website, the chief investigator, sponsor or principal investigator should notify the HRA immediately at hra.approval@nhs.net. The HRA will work with these organisations to achieve a consistent approach to information provision.

**Confirmation of Capacity and Capability**

This describes whether formal confirmation of capacity and capability is expected from participating NHS organisations in England.

NHS organisations in England that are participating in the study will be expected to formally **confirm their capacity and capability** to host this research.

- Following issue of this letter, participating NHS organisations in England may now confirm to the sponsor their capacity and capability to host this research, when ready to do so. How capacity and capacity will be confirmed is detailed in the Allocation of responsibilities and rights are agreed and documented (4.1 of HRA assessment criteria) section of this appendix.

The Assessing, Arranging, and Confirming document on the HRA website provides further information for the sponsor and NHS organisations on assessing, arranging and confirming capacity and capability.

**Principal Investigator Suitability**

This confirms whether the sponsor position on whether a PI, LC or neither should be in place is correct for each type of participating NHS organisation in England and the minimum expectations for education, training and experience that PIs should meet (where applicable).

A Principal Investigator is expected to be in place at the NHS organisation.

GCP training is **not** a generic training expectation, in line with the HRA statement on training expectations.

**HR Good Practice Resource Pack Expectations**

This confirms the HR Good Practice Resource Pack expectations for the study and the pre-engagement checks that should and should not be undertaken.

As a non-commercial study undertaken by local staff, it is unlikely that letters of access or honorary research contracts will be applicable, except where local network staff employed by another Trust (or University) are involved (and then it is likely that arrangements are already in place). Where arrangements are not already in place, network staff (or similar) undertaking any of the research activities listed in A18 or A19 of the IRAS form (except for administration of questionnaires), would be
expected to obtain an honorary research contract from one NHS organisation (if university
employed), followed by Letters of Access for subsequent organisations. This would be on the basis
of a Research Passport (if university employed) or an NHS to NHS confirmation of pre-engagement
checks letter (if NHS employed). These should confirm enhanced DBS checks, including appropriate
barred list checks, and occupational health clearance. For research team members only
administering questionnaires, a Letter of Access based on standard DBS checks and occupational
health clearance would be appropriate.

Other Information to Aid Study Set-up

This details any other information that may be helpful to sponsors and participating NHS organisations in
England to aid study set-up.

- The applicant has indicated that they intend to apply for inclusion on the NIHR CRN Portfolio.
- HRA Assessment is only applicable to the NHS organisation
Appendix 1.2. Research Protocol V.1.5.3

Changes in Microbiome with Treatment in Chronic Lung Diseases, Protocol version 1.5.3; 7/7/2016
PI: Tim McHugh

Changes of Microbiome in Response to Treatment in Patients with Chronic Lung Diseases
(Student Study)

Short Title:
Changes in Microbiome with Treatment in Chronic Lung Diseases

Chief Investigator: Prof. Timothy D. McHugh, PhD
Professor of Medical Microbiology
Director of UCL Centre for Clinical Microbiology,
Faculty Graduate Tutor, Medical Sciences,
Department of Infection and Immunity, UCL.

Sponsored by: University College London (UCL)

R&D / Sponsor Reference Number: 16/0178

Study Protocol

Version 1.5.3;
7th July 2016
### PROTOCOL VERSIONS

<table>
<thead>
<tr>
<th>Version Stage</th>
<th>Versions No</th>
<th>Version Date</th>
<th>Protocol updated &amp; finalised by</th>
<th>Reason for the protocol update</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current</td>
<td>1.5.3</td>
<td>7/7/2016</td>
<td>Sylvia Rofael</td>
<td>Following JRO fourth review by Miss. Jenise Davidson</td>
</tr>
<tr>
<td>Previous</td>
<td>1.5.2</td>
<td>12/6/2016</td>
<td>Sylvia Rofael</td>
<td>Dr. Hurst modifications excluding the extra patients’ visits and Following JRO third review by Miss. Jenise Davidson</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>10/6/2016</td>
<td>Sylvia Rofael</td>
<td>Following JRO second review by Miss. Jenise Davidson</td>
</tr>
<tr>
<td></td>
<td>1.4</td>
<td>15/5/2016</td>
<td>Sylvia Rofael</td>
<td>Following JRO Review by Miss. Jenise Davidson</td>
</tr>
<tr>
<td></td>
<td>1.3.4</td>
<td>18/4/2016</td>
<td>Sylvia Rofael</td>
<td>Following Thesis committee review in 6 mo supervisory meeting</td>
</tr>
<tr>
<td></td>
<td>1.3.3</td>
<td>1/4/2016</td>
<td>Sylvia Rofael</td>
<td>Tim McHugh Modifications</td>
</tr>
<tr>
<td></td>
<td>1.3.2</td>
<td>20/3/2016</td>
<td>Sylvia Rofael</td>
<td>John Hurst second revision</td>
</tr>
<tr>
<td></td>
<td>1.3.1</td>
<td>20/2/2016</td>
<td>Sylvia Rofael</td>
<td>Following Peer review in UCL Respiratory Academic meeting</td>
</tr>
<tr>
<td></td>
<td>1.2</td>
<td>1/2/2016</td>
<td>Sylvia Rofael</td>
<td>John Hurst modifications</td>
</tr>
<tr>
<td></td>
<td>1.1</td>
<td>25/1/2016</td>
<td>Sylvia Rofael</td>
<td>Early Draft</td>
</tr>
</tbody>
</table>
DECLARATIONS

The undersigned confirm that the following protocol has been agreed and accepted and that the investigator agrees to conduct the study in compliance with the approved protocol and will adhere to the Research Governance Framework 2005 (as amended thereafter), the Trust Data & Information policy, Sponsor and other relevant SOPs and applicable Trust policies and legal frameworks.

I (investigator) agree to ensure that the confidential information contained in this document will not be used for any other purposes other than the evaluation or conduct of the clinical investigation without the prior written consent of the Sponsor.

I (investigator) also confirm that an honest accurate and transparent account of the study will be given; and that any deviations from the study as planned in this protocol will be explained and reported accordingly.

Chief Investigator: 
Signature:............................ Date: 19th July, 2016

Print Name (in full): Timothy D. McHugh

Position: Professor of Medical Microbiology, Director of UCL Centre for Clinical Microbiology, Faculty Graduate Tutor, Medical Sciences, Department of Infection and Immunity, UCL.

On behalf of the Study Sponsor:

Signature:... Date.15JULY./2016

Print Name (in full): Miss Jenise Davidson

Position: 

<table>
<thead>
<tr>
<th>Identifiers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IRAS Number</td>
<td>202883</td>
</tr>
<tr>
<td>REC Reference No</td>
<td></td>
</tr>
<tr>
<td>Sponsor Reference No</td>
<td>16/0178</td>
</tr>
<tr>
<td>Other research reference number(s) (if applicable)</td>
<td></td>
</tr>
</tbody>
</table>

| Full (Scientific) title                         | Changes of Microbiome in Response to Treatment in Patients with Chronic Lung Diseases |
| Health condition(s) or problem(s) studied       | Chronic Lung Diseases; Chronic Obstructive Pulmonary Disease (COPD) and Bronchiectasis |
| Study Type                                      | Prospective Observational Comparative Cohort Study |
| Target sample size                              | A pilot study aims to recruit 50 patients which will inform the power calculations for development of a full study. |

**STUDY TIMELINES**

| Study Duration/length                           | 3 years            |
| Expected Start Date                             | 15/9/2016          |
| End of Study definition and anticipated date    | last visit of the last subject undergoing research 1/8/2019 |
| Key Study milestones                            | study submission, study approvals, first patient recruitment, last patient recruitment, Data Analysis, Thesis submission |

**FUNDING & Other**

| Funding                                         | Newton Mosharafa scholarship offered by the Egyptian Government, British Council Egypt and British Embassy Egypt to the PhD Student Sylvia Rofael; in addition to, UCL research funds held by Prof. McHugh |
| Other support                                   | Joint Research Office, 1st Floor Maple House (Suite B) |

**STORAGE of SAMPLES**

| Human tissue samples                            | UCL Centre for Clinical Microbiology (CCM) Royal Free Campus of UCL |
| Data collected / Storage                        | UCL Respiratory and UCL Centre for Clinical Microbiology Royal Free Campus of UCL |

**KEY STUDY CONTACTS**

| Chief Investigator                              | Prof. Tim McHugh |

---

IRAS Number: 202883  Sponsor Reference Number: 16/0178
KEY WORDS
Microbiome, Microbiota, Resistome, Antimicrobial Resistance, Azithromycin, COPD, Bronchiectasis

LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AE</td>
<td>Acute Exacerbations</td>
</tr>
<tr>
<td>AZI</td>
<td>Azithromycin</td>
</tr>
<tr>
<td>CAT</td>
<td>COPD Assessment Test</td>
</tr>
<tr>
<td>COPD</td>
<td>Chronic Obstructive Pulmonary Disease</td>
</tr>
<tr>
<td>CT</td>
<td>Computed Tomography</td>
</tr>
<tr>
<td>Diff</td>
<td>Difference</td>
</tr>
<tr>
<td>Exc</td>
<td>Exacerbation</td>
</tr>
<tr>
<td>FEV1</td>
<td>Forced Expiratory Volume in the first second</td>
</tr>
<tr>
<td>GOLD</td>
<td>Global initiative for Chronic Obstructive Lung Disease</td>
</tr>
<tr>
<td>Hr</td>
<td>Hour(s)</td>
</tr>
<tr>
<td>Inc</td>
<td>Increase</td>
</tr>
<tr>
<td>Mo</td>
<td>Month(s)</td>
</tr>
<tr>
<td>OTU</td>
<td>Operational Taxonomic Units</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>QIIME</td>
<td>Quantitative Insights In to Microbial Ecology</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal RiboNucleic Acid</td>
</tr>
<tr>
<td>SGRQ</td>
<td>St Georges Respiratory Questionnaire</td>
</tr>
<tr>
<td>Sig.</td>
<td>Significant</td>
</tr>
<tr>
<td>R</td>
<td>Resistance</td>
</tr>
<tr>
<td>VC</td>
<td>Vital Capacity</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organisation</td>
</tr>
<tr>
<td>Yr</td>
<td>Year</td>
</tr>
</tbody>
</table>
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>8</td>
</tr>
<tr>
<td>Figure 1: Overview of Study</td>
<td>9</td>
</tr>
<tr>
<td>BACKGROUND AND RATIONALE</td>
<td>10</td>
</tr>
<tr>
<td>Other Studies Underway in This Field</td>
<td>15</td>
</tr>
<tr>
<td>STUDY OBJECTIVES</td>
<td>16</td>
</tr>
<tr>
<td>Hypothesis</td>
<td>16</td>
</tr>
<tr>
<td>Goal of the Study</td>
<td>16</td>
</tr>
<tr>
<td>Objectives</td>
<td>16</td>
</tr>
<tr>
<td>OUTCOMES</td>
<td>17</td>
</tr>
<tr>
<td>Primary Study Outcome</td>
<td>17</td>
</tr>
<tr>
<td>Secondary Study Outcome</td>
<td>17</td>
</tr>
<tr>
<td>METHODS</td>
<td>18</td>
</tr>
<tr>
<td>Systematic Review</td>
<td>18</td>
</tr>
<tr>
<td>Study Design</td>
<td>19</td>
</tr>
<tr>
<td>First Study</td>
<td>19</td>
</tr>
<tr>
<td>Patient Recruitment</td>
<td>20</td>
</tr>
<tr>
<td>Eligibility Criteria:</td>
<td>20</td>
</tr>
<tr>
<td>Consent</td>
<td>21</td>
</tr>
<tr>
<td>DATA COLLECTION</td>
<td>22</td>
</tr>
<tr>
<td>DATA MANAGEMENT AND HANDLING</td>
<td>23</td>
</tr>
<tr>
<td>STUDY SCHEDULE</td>
<td>23</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>--------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>STUDY PROCEDURES</td>
<td>25</td>
</tr>
<tr>
<td>BIOLOGICAL MATERIAL AND SAMPLES STORAGE</td>
<td>26</td>
</tr>
<tr>
<td>PROSPECTIVE FOLLOW UP AND USE OF DIARY CARDS</td>
<td>27</td>
</tr>
<tr>
<td>RISKS AND RECORDING OF ADVERSE EVENTS</td>
<td>27</td>
</tr>
<tr>
<td>DATA AND STATISTICAL ANALYSIS</td>
<td>27</td>
</tr>
<tr>
<td>STUDY SIZE</td>
<td>28</td>
</tr>
<tr>
<td>CONFIDENTIALITY</td>
<td>28</td>
</tr>
<tr>
<td>STUDY CONDUCT</td>
<td>28</td>
</tr>
<tr>
<td>PEER AND REGULATORY REVIEW</td>
<td>29</td>
</tr>
<tr>
<td>FUNDING</td>
<td>29</td>
</tr>
<tr>
<td>ARCHIVING</td>
<td>29</td>
</tr>
<tr>
<td>INDEMNITY ARRANGEMENTS</td>
<td>29</td>
</tr>
<tr>
<td>PATIENT AND PUBLIC INVOLVEMENT</td>
<td>30</td>
</tr>
<tr>
<td>TRANSLATION INTO CLINICAL BENEFIT</td>
<td>30</td>
</tr>
<tr>
<td>PROJECT PERSONNEL</td>
<td>31</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>33</td>
</tr>
</tbody>
</table>
Introduction

The socioeconomic burden of chronic lung diseases as well as the symptom burden on patients is considerable worldwide. Patients with chronic lung diseases periodically suffer of acute episodes of sustained worsening in their respiratory symptoms known as acute exacerbations (AEs) that may need hospitalization in severe cases. Those patients are often prescribed frequent and prolonged courses of antibiotics either as prophylaxis or treatment of AEs. However, the clinical evidence supporting the usefulness of antibiotics in the treatment of AEs is surprisingly limited.

Since the excessive and inappropriate use of antimicrobials is the main reason for development and wide spread of antimicrobial resistance in population. We hypothesise that the frequently prescribed antibiotics in this patient cohort may enrich the prevalence of antimicrobial resistance genes (resistome) in the body’s microflora. This may act as a potential reservoir for antimicrobial resistance genes in the body. In addition, the inappropriate use of antimicrobials may exert selection pressure on the normal microbiota resulting in collateral damage in the homeostasis of the microbial ecology in human body (microbiome).

Patients with chronic lung diseases who are either hospitalized or attending the respiratory clinics or pulmonary rehabilitation program in Royal Free London NHS Trust will be invited to participate in the study. The participating patients will be closely followed for one year. Repeated biological samples (spontaneously expectorated sputum samples, nasopharyngeal swabs, and saliva, stool and blood samples) will be collected from the same patients five times over one year. The changes in the microbiome and the resistome in the repeated samples will be analyzed and correlated with the different therapies given to patients. This will be an observational study which does not involve active intervention. We will not be changing the routine care that patients receive or any of the regular patients’ management prescribed by their physicians. So we do not anticipate any risk on the participants.

This pilot study will inform the design of future studies but the evidence collected will begin to reveal the impact of various therapeutic interventions on the individual microbiome and resistome in patients with chronic lung diseases. It will provide a better understanding and deeper insights into how antibiotics affect the airway, oral and gut microbiome and resistome. This study may help in the development of individualized medicine based on tailoring therapy according to the individual microbiome in a way that may limit excessive and inappropriate antibiotics utilization and control antimicrobial resistance development and spread.
Recruitment of Patients
Collection of demographic data, history, baseline data, biological samples and spirometry

Patients are clinically assessed quarterly in their regular follow up appointments by the health care team
Collection of samples and CAT questionnaires

Patients fill diary cards daily and record any change in the baseline symptoms that may denote sign of acute exacerbation

After 1 year study complete.
Analysis of results

Figure 1: Overview of study
Background and Rationale

Not long ago, it was believed that the lower respiratory tract and the lungs of a healthy subject were sterile and the lower airways became colonized with potentially pathogenic microorganisms only during respiratory infections. Until recently, this was a well-established concept that has been taught to all medical students. However, the use of advanced molecular techniques such as 16S rRNA pyrosequencing have showed that the lower respiratory tracts and lungs; like many parts of the body, harbour an unprecedented diversity of microorganisms, collectively known as the airway microbiota. In 2010, Hilty et al.(1) was the first to demonstrate that the bronchial tree was not sterile, and contained a mean of 2,000 bacterial genomes per centimetre square. Since then, there was a growing interest in lung microbiome. Many studies were conducted trying to characterize airway microbiome in health and disease conditions. The core pulmonary bacterial microbiome was found to include genera such as Streptococcus, Haemophilus, Enterobacteriaceae, Prevotella, Veillonella, Actinomyces, Rothia and Gemella.(2-5) Currently, “microbial dysbiosis” represents the basis of a new conceptualized model in which alterations in the healthy microbiome; e.g. in richness, bacterial load or diversity, could be associated with the disease development and progression in a wide range of chronic respiratory conditions.(6)

The global burden of diseases is shifting from communicable to non-communicable diseases, with chronic conditions. According to WHO, these are now the chief causes of morbidity and mortality worldwide.(7, 8) Recently, there has been a growing concern in two progressive chronic lung diseases; chronic obstructive pulmonary disease (COPD) and non-cystic fibrosis bronchiectasis. COPD is characterised by chronic inflammation of the airways and lung in response to noxious particles and gases with consequent respiratory impairment and distress. It is estimated that COPD will be the third leading cause of death by 2020.(9) Bronchiectasis is another chronic lung disease that has been neglected for a long time. It is characterised by irreversibly damaged and dilated airways with poor mucus clearance and persistent bacterial colonization.(10, 11)

The impact of obstructive lung diseases on health and quality of life is considerable, particularly in advanced cases. It has been suggested that symptom burden in this patient cohort is comparable to that of patients with cancer, although patients with obstructive lung diseases have a greater live expectancy.(12) Obstructive lung diseases interfere with the patients’ productivity, some patients experience difficulties in normal physical exertion, social and family activities especially those with severe disease.(13)

Acute Exacerbations in chronic lung diseases can be defined as “a sustained worsening of the patient’s condition, from the stable state and beyond normal day-to-day variations, that is acute in onset and necessitates a change in regular medication in
a patient with underlying chronic lung disease". (14) Exacerbations in 50–70% of cases are due to respiratory infections (including bacteria, atypical organisms and respiratory viruses), in 10% are due to environmental pollution (depending on season and geographical placement), and up to 30% are of unknown etiology – likely also infection, reflecting the difficulty of standard tests to detect all organisms. (15, 16) Acute exacerbations cause significant morbidity and may accelerate disease progression as a result of the vicious cycle of: airway obstruction, bacterial colonization, more inflammation, and progressive tissue destruction. In addition, exacerbations represent a great economic burden on the health care system since it is a major cause for primary care visits and may require hospitalization in severe cases. (17, 18)

Consequently, recent clinical trials concentrating on preventing or even reducing exacerbations were conducted aiming at slowing the progression of these inflammatory pulmonary diseases. Currently, there is a growing body of scientific evidence on the clinical usefulness of the long-term prophylactic use of antibiotics; especially the macrolides, to reduce the frequency of exacerbations in COPD (19, 20) bronchiectasis (21-23) (Table 1). It has not been proved yet whether the observed benefit of the macrolides may be due to the antibacterial activity or another pharmacological action such as anti-inflammatory or pro-kinetic effects. Nevertheless, the long term use of antibiotics in such chronic cases may exert a selection pressure on the airway microbiota. In addition, a great concern has been raised regarding the emergence and spread of antimicrobial resistance in the general population. (24)
### Table 1: Clinical trials studying the usefulness of macrolides long term prophylaxis in chronic lung diseases

<table>
<thead>
<tr>
<th>First Author</th>
<th>Year</th>
<th>Place</th>
<th>Study Design</th>
<th>Participants</th>
<th>Condition</th>
<th>Antibiotic</th>
<th>Dose</th>
<th>Duration</th>
<th>Point of comparison</th>
<th>Main Findings</th>
<th>P-value</th>
<th>Main Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seemungal (25)</td>
<td>2008</td>
<td>UK</td>
<td>Randomized double blind placebo controlled study</td>
<td>109</td>
<td>COPD</td>
<td>Erythromycin</td>
<td>250 mg twice daily</td>
<td>12 mo.</td>
<td>Frequency of AE</td>
<td>Control 125, Treated 81</td>
<td>0.003</td>
<td>• Rate ratio for AE for the macrolides-treated compared with placebo was 0.648 • Patients had shorter duration exacerbations compared with placebo • No effect on airway or systemic inflammatory markers</td>
</tr>
<tr>
<td>Albert (19) Han (20)</td>
<td>2011</td>
<td>12 academic health centres in USA</td>
<td>Randomized prospective, parallel-group, Placebo Controlled Clinical Trial</td>
<td>1142 (1:1 ratio)</td>
<td>COPD</td>
<td>Azithromycin (Oral)</td>
<td>250 mg Once daily</td>
<td>1 yr</td>
<td>median time to 1st AE / (AE/ patient-yr Patients% with sig. quality of life improvement / (Macrolides R))</td>
<td>174 days 266 days 1.83 1.48 36% 43% 41% 81%</td>
<td>&lt;0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Wong (EMBRA CE Study) (21)</td>
<td>2012</td>
<td>Auckland, New Zealand</td>
<td>Randomized Double blind Parallel group Placebo controlled clinical trial</td>
<td>141 (1:1 ratio)</td>
<td>Bronchiectasis</td>
<td>Azithromycin (Oral)</td>
<td>500 mg 3 times weekly</td>
<td>6 mo.</td>
<td>Rate (AE/patient) Δ Prebronchodilator FEV1 SGRQ total score</td>
<td>1.57 0.59 -0.004L 0 -1.92 -5.17</td>
<td>10^-14</td>
<td>0.25</td>
</tr>
</tbody>
</table>
Changes in Microbiome with Treatment in Chronic Lung Disease, Protocol version 1.5.3; 7/7/2016
Pl. Tim McHugh

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Country</th>
<th>Randomise d Double blind Group</th>
<th>Placebo controlled clinical trial</th>
<th>143 (1:1 ratio)</th>
<th>Bronchiectasis</th>
<th>Azithromycin (Oral)</th>
<th>250 mg daily</th>
<th>12 mo.</th>
<th>Median number of AE</th>
<th>No of individual having at least 1 exc/yr</th>
<th>A Predicted FEV1/3 mo</th>
<th>Gastrointestinal adverse effects</th>
<th>Macrolides Resistance Rate</th>
<th>2</th>
<th>0</th>
<th>Median diff.</th>
<th>(AE)/patient-yr</th>
<th>12 hr sputum production</th>
<th>Attenuated lung function decline</th>
<th>Median inc in Macrolides R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altenburg</td>
<td>2013</td>
<td>Netherlands</td>
<td>Randomised Double blind Parallel group Placebo controlled clinical trial</td>
<td>143 (1:1 ratio)</td>
<td>Bronchiectasis</td>
<td>Azithromycin (Oral)</td>
<td>250 mg daily</td>
<td>12 mo.</td>
<td>Median number of AE</td>
<td>No of individual having at least 1 exc/yr</td>
<td>A Predicted FEV1/3 mo</td>
<td>Gastrointestinal adverse effects</td>
<td>Macrolides Resistance Rate</td>
<td>2</td>
<td>0</td>
<td>Median diff.</td>
<td>(AE)/patient-yr</td>
<td>12 hr sputum production</td>
<td>Attenuated lung function decline</td>
<td>Median inc in Macrolides R</td>
<td></td>
</tr>
<tr>
<td>Serierier</td>
<td>2013</td>
<td>Queensland</td>
<td>Randomised Double blind Parallel group Placebo controlled clinical trial</td>
<td>117 (1:1 ratio)</td>
<td>Bronchiectasis</td>
<td>Erythromycin ethylsuccinate (Oral)</td>
<td>400 mg Twice daily</td>
<td>12 mo.</td>
<td>Median number of AE</td>
<td>No of individual having at least 1 exc/yr</td>
<td>A Predicted FEV1/3 mo</td>
<td>Gastrointestinal adverse effects</td>
<td>Macrolides Resistance Rate</td>
<td>2</td>
<td>0</td>
<td>Median diff.</td>
<td>g</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AE: Acute Exacerbation  exc: exacerbation  AZI: azithromycin  R: resistance  Δ: change... f: Frequency ...diff.: difference... SGRQ: St Georges Respiratory Questionnaire  FEV1: Forced Expiratory Volume in the first second  mo.: months  yr: year

• Lower AE rate  • Improved quality of life  • No change in the microbiological profile  • Higher resistance levels  • Higher gastrointestinal adverse effects

• Reduced AE Frequency  • Reduced 24 hr sputum production  • Attenuated lung function decline  • Increased proportion of macrolides-resistant oropharyngeal streptococci
Patients with chronic lung diseases receive frequent and prolonged courses of antimicrobial agents in current clinical practice in the UK; especially during periods of acute exacerbations. β-lactams, tetracyclines, macrolides and fluoroquinolones are the most prescribed classes of antibiotics to treat exacerbations. Nevertheless, the evidence of antibiotic usefulness in the treatment of acute exacerbation; especially in COPD has remained a matter of controversy for many years. In the GOLD guidelines, the evidence for antibiotic use in COPD exacerbations is classed as Category B.(9) This may be attributed to the difficulty in proving the aetiology of bacterial infection in acute exacerbation of COPD and clinically identifying patients with acute exacerbation of bacterial origin.(26) Therefore, in spite of the widespread of antibiotics use in exacerbations, the evidence for this usage is surprisingly weak. In addition the excessive antibiotic use has a number of adverse effects in terms of direct side effects for patients, adds to the healthcare cost and contributes to antimicrobial resistance.

A direct correlation was reported between antimicrobial use and the extent of antimicrobial resistance.(27) Antibiotic misuse and/or overuse is a major contributing factor in developing resistance not only in the pathogenic bacteria but also the commensals.(28) It is easy to imagine that uncultivable bacteria in the microbiome could play a major role in the acquisition and spread of antibiotic resistance to pathogenic bacteria. Therefore, microbiota can act a reservoir for antibiotic resistance.(29, 30) In many studies, the human gut microbiota was shown to act as a reservoir of antibiotic resistance genes. The resistant bacteria can persist for years in the gut, even under short-term antibiotic administration and can transfer the resistance gene to other bacteria, even if the bacteria are just passing through the intestine.(31-34) However, to the best of our knowledge, little is known about the lung’s resistome (collection of all the antibiotic resistance genes and their precursors in pathogenic and non-pathogenic bacteria in a certain ecosystem).(35)

Many studies on the respiratory tract have suggested links between the microbiome composition and disease progression, severity and exacerbations across wide spectrum of chronic pulmonary disorders.(36-39) Not much work has been conducted on the influence of the therapy on the individual microbiome. It is not clear till now the actual benefit of antimicrobials or their impact on the microbiome and the resistome when administered as prophylactic treatment in controlling the chronic respiratory conditions.(40) Cross-sectional studies have informed our understanding of the airway microbiome in chronic lung diseases, revealing a complex and highly variable microbiome in a constantly perturbed ecosystem.(2, 37, 41-47) However, in order to better understand the impact of therapeutic interventions on the lung microbiome, it is important to investigate dynamic changes in the bacterial communities over time within the same patient.
Other studies underway in this field

A search of clinicaltrials.gov reveals two studies examining the impact of antimicrobial therapy on the airway microbiome.

a. Evaluation of Inhaled Antibiotics on Bacterial Diversity and Richness in the Cystic Fibrosis Lung

This study is taking place in Dartmouth-Hitchcock Medical Centre (U.S). The purpose of this study is to characterize bacterial diversity and richness in the sputum of cystic fibrosis patients treated with every-other-month TOBI™ Podhaler™ (tobramycin inhalation powder) and continuous alternating therapy with TOBI™ Podhaler and colistimethate (colistin).

b. Cystic Fibrosis Microbiome-determined Antibiotic Therapy Trial in Exacerbations: Results Stratified (CFMATTERS)

CFMATTERS will provide a randomized multi-centre controlled trial of microbiome derived antimicrobial treatments versus current empirical therapy in cystic fibrosis. The intervention they will be looking at; tobramycin and ceftazidime antibiotic therapy. This study is sponsored by University College Cork and will take place in a multi-centre in Europe and the United States.


Changes in Microbiome with Treatment in Chronic Lung Diseases, Protocol version 1.5.3; 7/7/2016
PI, Tom McHugh

Study objectives

Hypothesis

We hypothesise that frequent and prolonged courses of chemotherapeutic antimicrobial agents; in addition to, corticosteroids, which are prescribed to patients with chronic respiratory disorders either as a part of the lifelong management therapy or for the treatment of acute exacerbations, may affect the homeostasis of the individual microbiome and may contribute to the enrichment of the resistome.

Goal of the study

This study aims to reveal the impact of therapeutic interventions on the individual microbiome and resistome in patients with chronic lung diseases with the general goal of tailoring individualized treatment regimes according to the individual microbiome.

Objectives

1. To closely follow a prospective cohort of patients with chronic lung disease who are on long term azithromycin prophylaxis therapy and describe the changes in their microbiome compared to a comparative control group which are not on azithromycin prophylaxis therapy.
2. To elucidate the impact of the different therapeutic interventions to treat acute exacerbations on the individual airway microbiome.
3. To characterize the antimicrobial resistance patterns both phenotypically and genotypically in patients with chronic lung disease administering antibiotics through adopting a metagenomic approach.
4. To compare between the effect of different routes of administration of the medication on the microbiome.
5. To elucidate the role of respiratory viral agents in the airway microbiome during stable state and acute exacerbations in chronic respiratory conditions.
6. To investigate the impact of different antimicrobial treatments on the resistome of the airways, gut and oral cavity
7. To investigate the relation between the inflammatory and immunological biomarkers of infection in the plasma and response to treatment.
8. To report the influence of long term administration of azithromycin as prophylaxis on the frequency of exacerbations in patients with chronic lung disease.
9. To assess the impact of long term azithromycin prophylaxis on the health associated quality of life in patients with chronic lung disease.
10. To measure the associations between the microbiome parameters and patients' characteristics: spirometry, body mass index, smoking history, diet, occupation.

IRAS Number: 202883   Sponsor Reference Number: 16/0178
Outcomes

Primary study outcome

1. Define the changes in the microbiome parameters (community richness, evenness, alpha and beta diversity indices, bacterial load) with antimicrobial treatment in patients with chronic lung diseases.

2. Measure the prevalence of antimicrobial resistance genes against commonly used broad spectrum antibiotics including: macrolides, β-lactams, tetracyclines and fluoroquinolones among the microbiota in patients with chronic lung diseases receiving frequent courses of antimicrobial chemotherapeutic agents.

Secondary study outcome

1. Report the prevalence and the load of respiratory viral agents such as rhinovirus, human respiratory syncytial viruses (RSVs), influenza virus and adeno virus in the airways during stable state and acute exacerbations in chronic respiratory conditions.

2. Measure the inflammatory and immunological biomarkers related to infections in plasma.

3. Report the frequency (total and time to next pulmonary exacerbation), duration, severity of exacerbations and change in FEV1 in patients receiving long term azithromycin prophylaxis.

4. Measure the health related quality of life in patients receiving long term azithromycin prophylaxis.
Methods

Systematic Review

A systematic review of the literature will be undertaken looking at the impact of therapeutic interventions in the clinical management of chronic lung diseases on the individual microbiome and resistome; and the contribution of antimicrobial therapy on the prevalence of antimicrobial resistance.

Systematic reviews of clinical and laboratory studies relevant to these questions will be undertaken. As outlined in the Cochrane Handbook and the PRISMA statement,(48, 49) review preparation will follow six steps:

1. Formulation of the problem
2. Location and selection of studies
3. Critical appraisal of the studies
4. Collection of the data
5. Analysis and presentation of the results
6. Interpretation of the results.

Search strategies will include both computerised literature search of a number of databases including Pub Med/Medline/PreMedline, Scopus, Embase, Google Scholar, Web of Science, and Index to Theses. Search terms will be defined in discussion with a reference librarian with extensive experience in literature searches to ensure capture of all relevant studies. Relevant references from articles/reports identified will also be obtained. We will also conduct hand-searches of key journals, reports and bulletins and website searches of relevant studies. Completing these reviews will enable a summary of the current evidence base to be produced.
Study Design

This study will be a prospective observational comparative cohort study comprising two linked studies;

First Study

A group of patients with severe chronic lung disease who will be starting long term azithromycin prophylaxis prescribed by their respiratory consultant, as a part of their management regime, will be recruited. A spontaneously coughed sputum sample, nasopharyngeal swabs, saliva, stool and blood samples will be collected from these patients at baseline just before starting the course, after a month from the start date and by the end of the study. Further sputum samples and nasopharyngeal swabs will be collected at three months’ time intervals during their regular follow up appointments in the clinic. The baseline airway microbiome will be defined by next generation sequencing techniques and compared over a period of one year within the same patient to assess the influence of long term azithromycin prophylaxis on the individual microbiome and on the development or acquisition of macrolides’ resistance.

Another group of patients who are already on long term azithromycin prophylaxis for at least six months will also be recruited. This group will be of great value in assessing the effect of longer treatment periods on the levels of antimicrobial resistance and the more persistent changes in the airway and gut microbiome. The results will be compared with a control group of patients with similar clinical conditions but not on long term azithromycin prophylaxis.

Acute Exacerbations will be defined as the worsening of respiratory symptoms for two or more days (the first day defined the start of the exacerbation) with at least one major symptom and another major or minor symptom that necessitates healthcare utilization to evaluate the clinical condition by a clinician and to prescribe appropriate therapy. Major symptoms are sputum purulence, dyspnoea and sputum volume, and minor symptoms are colds, cough, sore throat and wheeze. In this way we have combined both the symptom based and event based definitions of exacerbations to exclude the co morbid events that could be wrongly included in the symptom based definition alone.(50, 51) Patients that experienced two or more exacerbations in the year preceding the date of the first sputum sample collection will be defined as frequent exacerbators. Patients will be considered stable 4 weeks after (measured from the end of the exacerbation when symptoms have returned to baseline levels), and two weeks before, the onset of any exacerbation.
Patient Recruitment

Patients attending pulmonary rehabilitation programs or the out-patient respiratory clinics at Royal Free London NHS Foundation Trust as well as hospitalized patients due to pulmonary exacerbation events will be invited to participate in the study.

Eligibility Criteria:

BRONCHIECTASIS INCLUSION: Patients with primary diagnosis of bronchiectasis confirmed by their treating clinicians and documented by high-resolution computed tomographic scan.

COPD INCLUSION: Patients with primary diagnosis of COPD confirmed by their treating clinicians will be included if the forced expiratory volume in one second (FEV1) was ≤ 80% predicted and FEV1/ vital capacity (VC) ratio was <0.7, in keeping with GOLD stages II to IV.(52) A history of chronic symptoms (dyspnoea, sputum production, wheeze and cough), and smoking history (number of pack-years smoked, current smoking status) will be taken in to consideration.

Table 2: Inclusion and exclusion criteria

<table>
<thead>
<tr>
<th>Inclusion Criteria</th>
<th>Exclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Confirmed primary diagnosis of COPD or Bronchiectasis</td>
<td>1. Risk of non-compliance or inability to comply with the study procedures</td>
</tr>
<tr>
<td>2. History of frequent exacerbations ≥2/year</td>
<td>2. Unable to produce sputum</td>
</tr>
<tr>
<td>3. Able to expectorate and produce sputum sample</td>
<td>3. Patients receiving Immunosuppressant drug not including steroids</td>
</tr>
<tr>
<td>4. Able to perform spirometry/maneuvers reproducibly</td>
<td>4. Patients with confirmed diagnosis of asthma, cystic fibrosis, current pneumonia or other respiratory disorders (e.g.; tuberculosis, lung cancer)</td>
</tr>
<tr>
<td>5. Male or female 20 years and above</td>
<td>5. Presence of other conditions that the principal investigator judges may interfere with the study findings</td>
</tr>
<tr>
<td>6. Willing to participate in study for the whole duration</td>
<td>6. Known HIV infection</td>
</tr>
<tr>
<td>7. Able to provide written informed consent</td>
<td>7. Patients who cannot comprehend spoken English or Arabic</td>
</tr>
</tbody>
</table>

IRAS Number: 202883  Sponsor Reference Number: 16/0178
Consent

Patients will be approached by their usual clinical care team (Dr. Hurst and COPD team) during routine encounters in their routine clinical environment. Patients will be given an information sheet and any questions can be answered by the researchers. For those who cannot read English the patient information sheet and consent can be translated to their language using online translation services; however the participants are required to comprehend spoken English or Arabic languages. Recruitment and initial review of the study with participants on joining the cohort can happen after routine hospital visit or at another date if more convenient for the patients.

The participants will be given at least 24 hours to decide whether it is convenient for them to participate in the study or not. Those who agree to participate in the study will be asked to fill out the consent form and sign it in arranged appointment according to the patient and the study team convenience. All participants will require the ability to consent for themselves to participate in the study. No children or vulnerable individuals will be recruited. The original consent forms signed by the participants and the patient notes will be kept in a locked office at the Royal Free Campus of UCL.
Data collection

Participants in the study will be interviewed at recruitment and then reviewed regularly every three months during the study. Data will be recorded in a proforma and entered into a database under allocated study number. This data will only be accessible by the study investigators. The personal information and link will be stored in a protected encrypted database.

Data collection for all study participants will consist of:

- Demographic details
- Medical history and lung function data
- Smoking History
- History of immunisation including pneumococcal and influenza vaccines
- Medication history
- Social history: employment, occupation, mode of delivery, diet, early life feeding mode.
- The burden of respiratory symptoms will be assessed by use of the St George’s Respiratory Questionnaire (SGRQ)(53) and COPD Assessment Test (CAT)(54) for COPD patients
- Spontaneously coughed sputum collection for detection of respiratory bacterial pathogens by means of culture and molecular diagnostic techniques
- Nasopharyngeal swabs for detection of respiratory viruses using molecular diagnostic techniques
- Sputum samples will stored for airway microbiome and resistome analyses
- Stool and saliva samples collection for gut and oral microbiome and resistome analyses respectively.
- Blood samples collection for inflammatory and immunological biomarkers investigation
- Patients will be asked to complete daily diary cards of their on-going respiratory symptoms and medication to allow capturing all exacerbation events and to document antibiotic usage and any change in the treatment regimes.

Although many have criticized sputum samples as not being the most ideal sample for studying the lung microbiome due to possible contamination from the oropharynx; it is the most practical and pragmatic sample. In addition, recent published studies have demonstrated that sputum is predictive sample that reflects the microbiology of the lower respiratory compartments and is suitable for the airway microbiome studies. The contamination from upper airway was found to be insignificant.(55-57)

Patients in this study, as part of their routine healthcare, should have their regular follow up appointments every three months with their consultants; we match the visits
for the study with the patients’ scheduled appointments to reduce the burden on the participants. During the extra visit the patients will be seen in the Grove Clinic in the Royal Free Hospital.

The health related quality of life will be assessed by asking the patients to fill two questionnaires; CAT(54) and SGRQ(53) which was originally developed to measure quality of life in COPD, but has subsequently been used in many different respiratory conditions. The SGRQ takes about 15 minutes to compete and the CAT takes around 5 minutes to complete.

Patients may also have tests as required by their clinical care team; such tests may include Chest X-Ray, CT chest, lung function tests and echocardiography. Although these tests are routine care and as such fall outside of the research study protocol, we will inform participants of our intention to collect and use these data in the study and secure consent for the use of this information from such tests will be included in the participant consent form.

Data Management and Handling

Source documentation paper files, original paper questionnaires and consent forms will be stored in separate files in a locked cupboard in a code accessed office at the Royal Free Campus of UCL. Electronic files will contain only pseudonymised data and will be saved as password protected files on password accessed UCL computers. Email may be used to transfer non-identifiable anonymised data. Data for analysis on laptops will be anonymised and no personally identifiable data will be kept on lap tops. This data will only be accessible by members of the study team. Dr. John Hurst will be the custodian of the collected data.

Study Schedule

This study is expected to be running for three years. Recruitment is estimated to take up to 2 years. Each participant will take part for up to one year. We define the end of the study as last visit of the last subject undergoing research. However the final data set will be available at approximately five years to allow further analysis of the data.
Table 3: Timing of data collection

<table>
<thead>
<tr>
<th></th>
<th>First visit</th>
<th>1+ mo</th>
<th>3 mo</th>
<th>6 mo</th>
<th>9 mo</th>
<th>12 mo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic details</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Collection of details regarding past medical history</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>History of immunisation</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Medication History</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Smoking history</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>St George's Respiratory Questionnaire (SGRQ)</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>COPD Assessment TEST (CAT)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Spirometry</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Sputum for detection of bacteria</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Sputum for microbiome analysis</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Sputum for resistome analysis</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Nasopharyngeal swabs for respiratory virus detection</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Stool for microbiome and resistome analysis</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Saliva for microbiome and resistome analysis</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Blood for inflammatory and immunological biomarkers</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

*One month review visit will be only for patients starting long term azithromycin prophylactic treatment.
Study Procedures

Spirometry

This will be performed at recruitment to the study to assess the lung function and confirm the primary diagnosis of obstructive lung disease, and at the end of the study to reveal the impact of the therapeutic intervention on the lung function. Subjects will undergo verbal instruction in the technique, including what the test entails and how they may feel during and after the test. Spirometry will be performed a minimum of three times, whilst seated, with appropriate single use one-way filter. Forced expiratory volume during the first second (FEV1), vital capacity (VC) and peak flow, plus quality of flow loop will be recorded at each attempt. All attempts will be recorded and the best used as study measure. The FEV1 and VC at two attempts should not differ by ±5%. Coughing during an attempt will render the attempt invalid. The results (including lung age) will be conveyed to the subject at their request. Contraindications to spirometry consist of: recent pneumothorax or thoracic surgery (within last 3 months), recent myocardial infarction or stroke (within last 1 month), recent abdominal, eye or neurosurgery (within last 2 months), recent perforated tympanic membrane (within last 3 months) Unstable angina, haemoptysis of unknown origin, systolic blood pressure >190mmHg.
Biological Material and Samples Storage

In this study, spontaneously coughed sputum, nasopharyngeal swabs, stool, saliva and blood samples will be collected from patients during their visits to the Grove Clinic in the Royal Free Hospital in accordance with the patient consent form and patient information sheet and shall include all tissue samples or other biological materials and any derivatives, portions, progeny or improvements as well as all patient information and documentation supplied in relation to them.

The biological samples will be sent to the UCL Centre for Clinical Microbiology for storage and microbiological testing. Sputum samples will be processed by conventional culture techniques and molecular diagnostic techniques such as multiplex quantitative PCR for identification of potential bacterial respiratory pathogens including, but not limited to, non-typeable Haemophilus influenzae, Moraxella catarrhalis, Streptococcus pneumoniae, Pseudomonas aeruginosa, and Staphylococcus aureus. Nasopharyngeal swabs will be tested for potential viral respiratory pathogens including Human Respiratory Syncytial Viruses, Rhinovirus, Influenza and Corona virus using molecular diagnostic technique. All biological samples are processed (if applicable), frozen and stored for the microbiome and resistome analysis in accordance with the analytical plan agreed with the chief investigator. These samples will be stored in a secure location in the UCL Centre for Clinical Microbiology, located at the Royal Free Campus. Samples will be stored in pseudo-anonymised form by study number and access to the study numbers will be restricted to study investigators. Culture isolates will be tested for antimicrobial sensitivity and then stored in microbiology department freezers for possible further disease-related testing.

The UCL Centre for Clinical Microbiology, located at the Royal Free Campus will process, store and dispose of the sputum samples in accordance with all applicable legal and regulatory requirements, including the Human Tissue Act 2004 and any amendments thereafter.

Samples taken during this study will not be transferred to any party not identified in this protocol and are not to be processed and/or transferred other than in accordance with the patients’ consent. After ethics approval for the study has expired, the biological material will be disposed of in accordance with the Human Tissue Act 2004, and any amendments thereto, or transferred to a licensed tissue bank.
Prospective follow up and use of diary cards

Study subjects will be followed for 1 year. Over this time they will be routinely reviewed every three months by the clinical care team, thus providing five routine reviews during the study.

In addition to these scheduled visits, patients will be requested to complete daily diary cards reporting any changes in their respiratory symptoms and their medication throughout the study period. In the diary cards patients will be asked if they experience any symptoms suggestive of acute exacerbation/respiratory infection such as; change in sputum volume, purulence and viscosity, dyspnoea, worse cough, chest wheezes, sore throat, fever, nasal congestion, running nose, increased breathing rate or heart rate; whether they are experiencing any adverse event and whether there is any change in their medication. To optimise compliance, we will provide diary cards in printed form. The intention of the diary cards is to ensure capture and documentation of all episodes of acute exacerbation/respiratory infection and any change in the medication. This will reduce recall bias that would be inevitable with retrospective documentation of this information. The majority of the time, most participants will simply record “no new symptoms” and “no change in the medication”. This should take less than 5 minutes in the morning. Self-reported diary cards have been used successfully in respiratory cohort studies of chronic obstructive lung disease.(14)

Risks and Recording of Adverse Events

This is an observational study which does not involve active intervention. We will not be changing the routine care that patients receive or any of the regular patients’ management prescribed by their physicians. All subjects will be encouraged to seek treatment early for a suspected exacerbation. So we do not anticipate any risk on the participants due to the study. Adverse Events will be reported by the patients in the daily diary and serious events may be recorded in the patient’s medical records.

Data and Statistical Analysis

Principal components analysis (PCA) will be applied to the combined clinical and microbiological parameters to identify those characteristics that are driving the pathogenesis of chronic lung disease.
For microbiome analysis, the 16S rRNA amplicon raw sequencing data will be analyzed using Quantitative Insights Into Microbial Ecology (QIIME) software (version 1.5.0) which allows further read processing and clustering into relevant Operational Taxonomic Units (OTUs). Community richness (number of taxa detected), evenness (relative distribution of taxa in a community), and alpha diversity (which consider both the number of taxa and their relative abundance) statistics will be performed on the obtained phylogenetic data of the various samples using the R statistical software package (version 3.2.2)(58). The taxa will be partitioned into core and rare taxa groups such that the persistent and abundant core taxa will be defined as those in more than 75% of all samples, while all other species falling outside of the upper quartile will be considered rare. Beta-diversity measures will be calculated to examine the differences between the microbiome profiles and the taxa turnover between the consecutive samples. Other flexible approaches to statistical analysis may also be used.

Study Size

This is an observational study that will assess the associations between disease state, treatment success and the lung microbiome/resistome. The primary outcome in this study will be defining the changes in the microbiome/resistome parameters with various short and long term antibiotic interventions in patients with chronic lung diseases. The sample size is established on a pragmatic recruitment rate of one patient per week until twenty five participants are allocated in each study group (on and off azithromycin prophylactic treatment) then we will use the generated information to inform the power calculations for a larger study.

Confidentiality

Full patient confidentiality will be maintained throughout the study, with access to samples and data restricted to the study investigators and their deputies.

Study Conduct

The chief investigator and study supervisors will be responsible for study oversight. We will have support from The Joint Research Office which is a partnership between UCL, UCL Hospitals NHS Foundation Trust and Royal Free London NHS Foundation Trust.
Peer and Regulatory Review

The study has been peer reviewed in accordance with the requirements outlined by UCL. This study has been developed in the context of Mrs Rofael’s PhD studies. The protocol has been reviewed by Dr. David Spratt and Dr. Marc Lipman upon. The protocol has also been presented and discussed in UCL respiratory academic meeting and UCL Centre for Clinical Microbiology (CCM) meeting. The Sponsor has verified that the supervisors of the project have undertaken sufficient review of the protocol in line with the requirements of their departments.

The study was deemed to require regulatory approval from Health Research Authority (HRA). Approval will be obtained before the study commences.

Funding

The study funding has been reviewed by the UCL Joint Research Office, and deemed sufficient to cover the requirements of the study. NHS costs will be supported via Royal Free London NHS Trust. This project is a part of a PhD program supported by Newton Mosharafa scholarship offered by the Egyptian Government, British Council, Egypt and British Embassy, Egypt to the PhD Student Sylvia Rofael; in addition to: UCL research funds held by Prof. Timothy McHugh.

Archiving

UCL and each participating site recognise that there is an obligation to archive study-related documents at the end of the study (as such end is defined within this protocol). The Chief Investigator confirms that he will archive the study master file at UCL Royal Free campus for the period stipulated in the protocol (five years) and in line with all relevant legal and statutory requirements.

Indemnity Arrangements

University College London holds insurance against claims from participants for harm caused by their participation in this clinical study. Participants may be able to claim compensation if they can prove that UCL has been negligent. However, if this clinical study is being carried out in a hospital, the hospital continues to have a duty of care to the participant of the clinical study. University College London does not accept liability for any breach in the hospital’s duty of care, or any negligence on the part of hospital employees. This applies whether the hospital is an NHS Trust or otherwise.
Patient and Public Involvement

The study needs to be acceptable to participants to have adequate retention in the study so that it can provide satisfying answers to the proposed research questions. This study addresses key issues to patients with chronic lung diseases and public which is the pros and cons of the frequent antibiotic courses on patients’ microbiome and quality of life; on the other hand, how the excessive consumption of antibiotics in this patient group can influence the prevalence of antimicrobial resistance in the population. We discussed our proposed research protocol with a sample of patients in pulmonary rehabilitation program held in Peckwater Centre, London adjusting the design in the light of their comments. During the study, the patients will be asked to monitor their respiratory symptoms daily through filling daily diary cards; the patients will be seen during their regular routine reviews in the clinic. We will be keen to feed back the patients with any relevant results who will help us to disseminate the findings of our study among the population. Therefore, involvement of patients is essential to complete the study. The study will be performed in line with INVOLVE principles.(59)

Translation into Clinical Benefit

Patients with chronic lung diseases are often prescribed frequent and prolonged courses of antibiotics either as a prophylaxis or treatment to acute exacerbations in UK. However, extensive debate around the usefulness of antibiotics in the treatment of exacerbations has not been concluded as the clinical evidence supporting this practice is surprisingly low. Excessive and inappropriate use of antimicrobials often results in the emergence of multidrug-resistant bacterial strains known as superbugs which is associated with treatment failure, higher morbidity and mortality, and increased health care costs. In addition, there might be significant collateral damage in the human microbiome. So far, studies that have assessed the ecological impact of antibiotics on the human microbiome are scarce and have focused mainly on the gut. To the best of our knowledge, this study will be one of the early studies in the UK to investigate the impact of different therapeutic interventions on the airway microbiome and resistome in patients with chronic lung diseases. The results of this study will provide a better understanding and deeper insights into how antibiotics affect the microbial ecology in airways, oral cavity and gut. This may help in the development of individualized medicine based on tailoring therapy according to the individual microbiome.
Project Personnel

Principal Investigator
Prof. Timothy D. McHugh,
Professor of Medical Microbiology
Director, UCL Centre for Clinical Microbiology,
Faculty Graduate Tutor, Medical Sciences,
Department of Infection and Immunity.

Primary Researcher
Mrs. Sylvia Rofael,
PhD student
Division of Infection and Immunity
University College London.

Project supervisors
Prof. Timothy D. McHugh,
Professor of Medical Microbiology
Director, UCL Centre for Clinical Microbiology,
Faculty Graduate Tutor, Medical Sciences,
Department of Infection and Immunity.

As detailed above

Dr. John Hurst,
Reader in Respiratory Medicine,
Consultant Respiratory Physician,
University College London,
Royal Free London NHS Foundation Trust,
**Collaborators**

**Dr. David Spratt,**

Reader in Microbial Ecology,  
Division of Microbial Diseases  
Eastman Dental Institute  
University College London.

**Dr. Marc Lipman,**

Senior Clinical Lecturer,  
Consultant Respiratory Physician,  
University College London.

**Dr. Rama Vancheeswaran,**

Consultant respiratory and general physician,  
Royal Free London NHS Foundation Trust.

**Ms. Elisha Pickett,**

Clinical Trials Practitioner
References


Appendix 1.3. Participant Information Sheet

Changes in Microbiome with Treatment in Chronic Respiratory Diseases. PI: Tim McHugh
Sponsor Reference Number: 16/0178

Participant Information Sheet (version 1.6)

Changes of Microbiome in Response to Treatment in Patients with Chronic Lung Diseases
(Student Study)

We would like to invite you to participate in a research study. This study aims to look at the impact of different treatments (mainly antibiotics and other medicines) on the bacteria that normally reside inside the human body in people with lung diseases. Please take time to read the following information carefully and discuss it openly. Please feel free to ask us if anything is not clear or if you would like any further information.

Names of Investigators: Mrs. Sylvia Rafael (PhD Student, Pharmaceutical Microbiologist)
Dr. John Hurst (Consultant and Reader in Respiratory Medicine)
Prof. Timothy D. McHugh (Professor of Medical Microbiology, PI)

What is the study about?

People with long-term lung problems such as Chronic Obstructive Lung disease (COPD) and bronchiectasis can suffer from episodes of worsening in their chest symptoms known as acute exacerbations (AEs). Exacerbations are usually due to infections (caused by bacteria, atypical organisms and viruses). Frequent and prolonged courses of antibiotics may need to be prescribed either to treat or prevent these events. However, the evidence supporting the usefulness of antibiotics is surprisingly low. It may be due to the fact that only half of these events are associated with bacteria (antibiotics are effective only in bacterial infections). Recently, there has been evidence from some clinical trials on the clinical usefulness of the long term Azithromycin preventive therapy (one of the macrodide antibiotics) in managing certain cases of COPD and bronchiectasis.

Excess use of antibiotics has a number of potential problems both in terms of side effects as well as antibiotic resistance. The frequent antibiotic courses for patients may increase the presence of antibiotic resistance genes (resistome) among the bacteria that normally reside inside the human body (microbiome). This may act as a reservoir for resistance genes in the body. Our research is designed to look at this. We will closely follow up a group of patients with chronic lung conditions for one year. We will investigate how different treatments impact the levels of resistance and types of bacteria present in the lung, gut and mouth.

What will happen in the study And How long will it last for?

We will have two groups of participants with chronic lung disease; a group who are commencing or already on Azithromycin preventive treatment and a comparable control group who are not. In the control group, we will also be looking at the impact of other antibiotics and medication prescribed to treat exacerbations on the body's normal flora. We
plan to follow up the people for one year during which each patient will be seen five times at their regular scheduled follow up appointments in the Grove clinic in Royal Free Hospital. Each participant will have a lung assessment test (spirometry) at the beginning and end of the study. Sputum sample (phlegm), saliva, stool, blood samples and a swab from the nose and the throat (nasopharyngeal swabs) will be collected in all visits and sent to UCL Centre for Clinical Microbiology for analysis. Volunteers will be asked to monitor their respiratory symptoms with a daily diary.

Why have I been invited?

You have been invited to join this study because you have been diagnosed with COPD or bronchiectasis and you have a history of frequent exacerbations in the previous year. This study is based on two groups of participants; so you may join the Azithromycin group if you are receiving or have been recommended to start long term Azithromycin course by your consultant; or you may join the control group if you are not receiving Azithromycin. The clinical decision for you to be on or off Azithromycin prophylactic therapy is taken by your clinician before you are invited to participate in the study and taking part in this study will not change this decision.

Who is sponsoring the research?

This study is being sponsored by University College London (UCL).

Do I have to take part?

It is up to you to decide if you want to take part. We will describe the study and go through this information sheet with you. If you agree to take part we will then ask you to sign a consent form. You are free to withdraw at any time, without giving a reason. This will not affect any care or treatment you may be receiving.

What will happen if I decide to participate in the study?

1. We will arrange an initial meeting to complete some documentation for the study. This will take at most 1 hour. At this initial appointment we will ask you to sign a consent form. We will then collect some details regarding your medical history and ask you to complete two questionnaires concerning your current respiratory symptoms.
2. We will then take a swab from your nose and throat (nasopharyngeal swab) and ask you to provide sputum (phlegm), saliva and stool samples. These samples will be tested for bacteria, viruses and antimicrobial resistance genes.
3. For convenience, you will be given a collection kit to collect the stool sample at home and bring it with you in the following visit.
4. We will also take blood samples (10 mL) to test for markers of infection and the body's response to infection.
5. We will perform a simple test of lung function (spirometry): this simply involves blowing into a hand-held machine that will measure the amount of air that you are able to blow out.
6. You will be seen four times over one year at your regular scheduled follow up appointments in the hospital; however, you may have to spend up to 30 minutes extra time during your hospital visit to allow time to collect the extra biological samples for the study.
7. If you are starting azithromycin treatment we will need to see you after one month from starting the course of antibiotics.
8. We would like you to fill in daily diary cards which will be provided in a printed form to report any changes in your respiratory symptoms and/or medication throughout the study period. Most of the time, you will simply record “no new symptoms” and “no change in the medication” which should take you less than 5 minutes; however, it may take you a longer time to detail the experienced symptoms and the new medication received during periods of exacerbations.
9. If you have any tests (Routine Microbiology tests, Chest X-Rays or CT scans, echocardiograms or lung function tests) performed as part of your routine medical care whilst you are in this study then we may use information from these tests (with your permission), but no scans or X-Rays will be performed for research purposes.

Why should I fill a daily diary cards?

You will be asked to record any changes in sputum volume, purulence and viscosity, dyspnoea, worse cough, chest wheezes, sore throat, fever, nasal congestion, running nose; and whether there is any change in your routine medication. It is important to monitor your baseline respiratory symptoms to capture all episodes of exacerbations and to document any change in your medication; as this may impact the results of the research. Filling the diary card on daily basis would reduce recall bias. However if you find this may be a daunting task, please discuss it with one of the study team.

Are there any possible disadvantages or risks of taking part?

This is an observational study; it does not include anything other than monitoring the changes in the bacteria in the lung and other parts of the body. **We are not changing any of your routine treatment.** Therefore, we don’t anticipate any harm or risk apart from a mild discomfort that you may experience during taking the nasopharyngeal swabs. You may find bruising after we take the blood samples. Blood will be taken by a suitably qualified person using procedures approved by NHS Trust to ensure your safety and comfort.
Will information about me be passed on to anyone else?

All information collected about you during this study will be confidential. Your personal information will be stored with a study code and number in a secure place in UCL Respiratory Medicine department. Your name will not be used in the study at all. No information that could identify you will be passed on to anyone outside of the research group.

Will my results be shared with the clinical team?

The results of the extra lung function tests will be available to your treating clinician. The rest of the study data will not be fed back to your clinicians except in exceptional circumstances in which the results need the attention of the clinical team. The standard care you receive should not be influenced by the current study data.

What will happen to the results of the research study?

We hope that the results of this study will help to improve care in the future. Some results may be presented at medical conferences or published in medical journals. You will not be identifiable from any published information. Participants who wish to be notified of the response will receive a lay summary of our findings by mail or e-mail as they wish by the end of the study. If you are interested to know more about the results please let us know and we can provide you access to the final study report and all the scientific publications that are produced from this study.

What will happen to the samples taken during this study?

This study will involve taking sputum (phlegm), saliva, stool and blood samples; in addition to, swabs from your nose and throat. Some of these samples will be tested immediately and some will be stored for later analysis. The stored samples will be pseudoanonymised (labeled with the study number only). These samples will be frozen and stored in UCL Centre for Clinical Microbiology at Royal Free Hospital. If you give your consent any remaining biological samples collected in this study may be used for future ethically approved research.

What if things went wrong?

If you wish to complain about your treatment by members of staff due to your participation in the research, National Health Service or UCL complaints mechanisms are available to you. Please ask your researcher if you would like more information on this. In the unlikely event that you are harmed by taking part in this study, or if you have concern about any aspect of this study, you should ask to speak to a member of the study.
team who will do their best to answer your questions. If you remain unhappy or wish to complain formally, you can do this through the NHS Complaints Procedure. Details can be obtained from the hospital Patient Advice and Liaison Service.

What will happen if I don’t want to carry on with the study?

If you no longer want us to use your data in this way, please contact Dr. Hurst and we will delete your files. In this case, you will have the right to request destruction of the biological samples collected from you. If, for some reason, you lose the ability to consent to the study (for instance if you are too unwell to consent) then we would withdraw you from the study although in this circumstance we would still use the information obtained whilst you retained the capacity to consent.

Finally, it is up to you to decide whether you want to be part of this study. If you do not want to be in this study, that is okay and your usual NHS care will continue as normal. You can stop during the study at any time without giving a reason. Your doctor will still look after you as normal. We hope, though, that you will tell us why you wish to stop the study.

Thank you for taking time to consider this study. Please ask any questions and let us know if there are things that you do not understand, or would like more information. Please address any further questions to:

Dr. John Hurst, Respiratory Consultant at the Royal Free Hospital.

Mrs. Sylvia Rofael (PhD Student, Div. of Infection and Immunity, UCL)

Email: 
Telephone: 
30/9/2016
Appendix 1.3. Consent Form

Changes in Microbiome with Treatment in Chronic Respiratory Diseases. PI: Tim McHugh
Sponsor Reference Number: 16/9178

Royal Free Hospital  Principal Investigator: Prof. Timothy McHugh
Study Number: 0000  Clinical Supervisor: Dr. John Hurst
Participant Identification:  Researcher name: Sylvia Rafael
Participant Initials:
Participant Date of Birth:

CONSENT FORM

Changes of Microbiome in Response to Treatment in Patients with Chronic Lung Diseases
(Student Study)

1. I confirm that I have read the information sheet dated 25/10/2016 (version 1.6) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. □

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected. □

3. I understand that relevant sections of my medical notes and data collected during the study may be looked at by individuals from the UCL Centre for Clinical Microbiology, respiratory research team from the Royal Free Hospital, from regulatory authorities or from the NHS Trust, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records. □

4. I understand that biological samples taken from me as part of this research project will be stored for analysis by UCL Centre for Clinical Microbiology. □

5. I agree that the information and biological samples collected from me in the context of this study may be used to support other ethically approved research in the future, and may be shared anonymously with other researchers. (Optional) □ Yes □ No

6. I agree that if I have other tests such as (routine microbiology tests, Chest X-Rays, CT scans, echocardiograms or lung function tests) whilst I am in this study the research team may use this information, but that scans or X-Rays will not be performed for research purposes. □

8. I agree to take part in the above study. □

____________________________  __________________________  __________________________
Name of Participant  Date  Signature

____________________________  __________________________  __________________________
Name of Person taking consent  Date  Signature

IRAS Number: 202883  Participant Consent Form version 1.5 (25/10/2016)
Appendix 1.4. Approved IRAS Form

Welcome to the Integrated Research Application System

**IRAS Project Filter**

The integrated dataset required for your project will be created from the answers you give to the following questions. The system will generate only those questions and sections which (a) apply to your study type and (b) are required by the bodies reviewing your study. Please ensure you answer all the questions before proceeding with your applications.

Please complete the questions in order. If you change the response to a question, please select 'Save' and review all the questions as your change may have affected subsequent questions.

---

Please enter a short title for this project (maximum 70 characters)
Changes in Microbiome with Treatment in Chronic Lung Diseases

1. Is your project research?
   - [ ] Yes
   - [ ] No

2. Select one category from the list below:
   - [ ] Clinical trial of an investigational medicinal product
   - [ ] Clinical investigation or other study of a medical device
   - [ ] Combined trial of an investigational medicinal product and an investigational medical device
   - [ ] Other clinical trial to study a novel intervention or randomised clinical trial to compare interventions in clinical practice
   - [ ] Basic science study involving procedures with human participants
   - [ ] Study administering questionnaires/interviews for quantitative analysis, or using mixed quantitative/qualitative methodology
   - [ ] Study involving qualitative methods only
   - [ ] Study limited to working with human tissue samples (or other human biological samples) and data (specific project only)
   - [ ] Study limited to working with data (specific project only)
   - [ ] Research tissue bank
   - [ ] Research database

If your work does not fit any of these categories, select the option below:
   - [ ] Other study

2a. Will the study involve the use of any medical device without a CE Mark, or a CE marked device which has been modified or will be used outside its intended purposes?
   - [ ] Yes
   - [ ] No

2b. Please answer the following question(s):
   - a) Does the study involve the use of any ionising radiation?
     - [ ] Yes
     - [ ] No
   - b) Will you be taking new human tissue samples (or other human biological samples)?
     - [ ] Yes
     - [ ] No
   - c) Will you be using existing human tissue samples (or other human biological samples)?
     - [ ] Yes
     - [ ] No

Date: 13/09/2016
d) Will the study involve any other clinical procedures with participants (e.g. MRI, ultrasound, physical examination)?

☐ Yes  ☑ No

3. In which countries of the UK will the research sites be located? (Tick all that apply)

☑ England
☐ Scotland
☐ Wales
☐ Northern Ireland

3a. In which country of the UK will the lead NHS R&D office be located:

☑ England
☐ Scotland
☐ Wales
☐ Northern Ireland
☐ This study does not involve the NHS

4. Which applications do you require?

IMPORTANT: If your project is taking place in the NHS and is led from England select ‘IRAS Form’. If your project is led from Northern Ireland, Scotland or Wales select ‘NHS/HSC Research and Development Offices’ and/or relevant Research Ethics Committee applications, as appropriate.

☑ IRAS Form
☐ Confidentiality Advisory Group (CAG)
☐ National Offender Management Service (NOMS) (Prisons & Probation)

For NHS/HSC R&D Offices in Northern Ireland, Scotland and Wales the CI must create NHS/HSC Site Specific Information forms, for each site, in addition to the study wide forms, and transfer them to the PIs or local collaborators.

For participating NHS organisations in England different arrangements apply for the provision of site specific information. Refer to IRAS Help for more information.

Most research projects require review by a REC within the UK Health Departments’ Research Ethics Service. Is your study exempt from REC review?

☐ Yes  ☐ No

5. Will any research sites in this study be NHS organisations?

☑ Yes  ☐ No

5a. Are all the research costs and infrastructure costs (funding for the support and facilities needed to carry out research e.g. NHS Support costs) for this study provided by a NIHR Biomedical Research Centre, NIHR Biomedical Research Unit, NIHR Collaboration for Leadership in Health Research and Care (CLAHRC), NIHR Patient Safety Translational Research Centre or a Diagnostic Evidence Co-operative in all study sites?

Please see information button for further details.

☐ Yes  ☐ No

Date: 13/09/2016
5b. Do you wish to make an application for the study to be considered for NIHR Clinical Research Network (CRN) Support and inclusion in the NIHR Clinical Research Network Portfolio?

Please see information button for further details.

- Yes
- No

The NIHR Clinical Research Network provides researchers with the practical support they need to make clinical studies happen in the NHS e.g. by providing access to the people and facilities needed to carry out research "on the ground".

If you select yes to this question, you must complete a NIHR Clinical Research Network (CRN) Portfolio Application Form (PAF) immediately after completing this project filter question and before submitting other applications. Failing to complete the PAF ahead of other applications e.g. HRA Approval, may mean that you will be unable to access NIHR CRN Support for your study.

6. Do you plan to include any participants who are children?

- Yes
- No

7. Do you plan at any stage of the project to undertake intrusive research involving adults lacking capacity to consent for themselves?

- Yes
- No

Answer Yes if you plan to recruit living participants aged 16 or over who lack capacity, or to retain them in the study following loss of capacity. Intrusive research means any research with the living requiring consent in law. This includes use of identifiable tissue samples or personal information, except where application is being made to the Confidentiality Advisory Group to set aside the common law duty of confidentiality in England and Wales. Please consult the guidance notes for further information on the legal frameworks for research involving adults lacking capacity in the UK.

8. Do you plan to include any participants who are prisoners or young offenders in the custody of HM Prison Service or who are offenders supervised by the probation service in England or Wales?

- Yes
- No

9. Is the study or any part of it being undertaken as an educational project?

- Yes
- No

Please describe briefly the involvement of the student(s):
PhD student will lead the laboratory analysis, and project funded as a PhD

9a. Is the project being undertaken in part fulfillment of a PhD or other doctorate?

- Yes
- No

10. Will this research be financially supported by the United States Department of Health and Human Services or any of its divisions, agencies or programs?

- Yes
- No

11. Will identifiable patient data be accessed outside the care team without prior consent at any stage of the project (including identification of potential participants)?

Date: 13/09/2016
☐ Yes  ☐ No
Integrated Research Application System
Application Form for Basic science study involving procedures with human participants

IRAS Form (project information)

Please refer to the E-Submission and Checklist tabs for instructions on submitting this application.

The Chief Investigator should complete this form. Guidance on the questions is available wherever you see this symbol displayed. We recommend reading the guidance first. The complete guidance and a glossary are available by selecting Help.

Please define any terms or acronyms that might not be familiar to lay reviewers of the application.

Short title and version number: (maximum 70 characters - this will be inserted as header on all forms)
Changes in Microbiome with Treatment in Chronic Lung Diseases

Please complete these details after you have booked the REC application for review.

REC Name:
London - Harrow Research Ethics Committee

REC Reference Number: 16LO/1490
Submission date: 13/09/2016

PART A: Core study information

1. ADMINISTRATIVE DETAILS

A1. Full title of the research:
Changes of Microbiome in Response to Treatment in Patients with Chronic Lung Diseases

A2. Educational projects

Name and contact details of student(s):

<table>
<thead>
<tr>
<th>Student 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title: Forename/Initials Surname</td>
</tr>
<tr>
<td>Mrs. Sylvia Rafael</td>
</tr>
<tr>
<td>Address</td>
</tr>
<tr>
<td>Post Code</td>
</tr>
<tr>
<td>E-mail</td>
</tr>
<tr>
<td>Telephone</td>
</tr>
<tr>
<td>Fax</td>
</tr>
</tbody>
</table>

Give details of the educational course or degree for which this research is being undertaken:

Name and level of course/ degree:
MPhil/PhD Division of Infection and Immunity

Date: 13/09/2016
Name of educational establishment:
University College London

Name and contact details of academic supervisor(s):

**Academic supervisor 1**

Title: Prof.
Forename/Initials: Timothy D.
Surname: McHugh

Address

Post Code
E-mail
Telephone
Fax

**Academic supervisor 2**

Title: Dr.
Forename/Initials: John
Surname: Hurst

Address

Post Code
E-mail
Telephone
Fax

Please state which academic supervisor(s) has responsibility for which student(s):
Please click “Save now” before completing this table. This will ensure that all of the student and academic supervisor details are shown correctly.

<table>
<thead>
<tr>
<th>Student(s)</th>
<th>Academic supervisor(s)</th>
</tr>
</thead>
</table>
| Student 1 Mrs. Sylvia Rofael | ✔️ Prof. Timothy D. McHugh  
| | ✔️ Dr. John Hurst |

A copy of a current CV for the student and the academic supervisor (maximum 2 pages of A4) must be submitted with the application.

A2-2. Who will act as Chief Investigator for this study?

- ○ Student
- ● Academic supervisor
- ○ Other

A3-1. Chief Investigator:

Date: 13/09/2016
Title: Prof. Timothy D. McHugh

Post: Professor of Medical Microbiology, Div of Infection & Immunity Faculty of Medical Sciences, University College London and Director of UCL Centre for Clinical Microbiology

Qualifications: B. Sc.(Hons), PhD

Employer: University College London

* This information is optional. It will not be placed in the public domain or disclosed to any other third party without prior consent. A copy of a current CV (maximum 2 pages of A4) for the Chief Investigator must be submitted with the application.

A4. Who is the contact on behalf of the sponsor for all correspondence relating to applications for this project?
This contact will receive copies of all correspondence from REC and HRA/R&D reviewers that is sent to the CI.

Title: Miss Jenise Davidson

Address: 

Post Code: 

E-mail: 

Telephone: 

Fax: 

A5-1. Research reference numbers. Please give any relevant references for your study:

 Applicant's/organisation's own reference number, e.g. R & D (if available): 16/0178

Sponsor's/protocol number:

Protocol Version: 1.5.3

Protocol Date: 07/07/2016

Funder's reference number:

Project website:

Registry reference number(s):

The Department of Health's Research Governance Framework for Health and Social Care and the research governance frameworks for Wales, Scotland and Northern Ireland set out the requirement for registration of trials. Furthermore, Article 19 of the World Medical Association Declaration of Helsinki adopted in 2008 states that "every clinical trial must be registered on a publicly accessible database before recruitment of the first subject", and the International Committee of Medical Journal Editors (ICMJE) will consider a clinical trial for publication only if it has been registered in an appropriate registry. Please see guidance for more information.
A5-2. Is this application linked to a previous study or another current application?

☐ Yes  ☐ No

Please give brief details and reference numbers.

2. OVERVIEW OF THE RESEARCH

To provide all the information required by review bodies and research information systems, we ask a number of specific questions. This section invites you to give an overview using language comprehensible to lay reviewers and members of the public. Please read the guidance notes for advice on this section.

A6-1. Summary of the study. Please provide a brief summary of the research (maximum 300 words) using language easily understood by lay reviewers and members of the public. Where the research is reviewed by a REC within the UK Health Departments’ Research Ethics Service, this summary will be published on the Health Research Authority (HRA) website following the ethical review. Please refer to the question specific guidance for this question.

Frequent courses of antibiotics and corticosteroids are often prescribed to patients with chronic lung diseases to treat chronic lung infections and manage acute exacerbations (AE). Currently, there is scientific evidence on the clinical usefulness of prolonged Azithromycin prophylactic therapy in reducing the frequency of exacerbations in COPD and bronchiectasis. There is a growing concern that the excessive administration of antibiotics to this patient cohort may enrich the prevalence of antimicrobial resistance genes among the body’s microflora (resistome). This may act as a potential reservoir for antimicrobial resistance genes in the body and the whole population. In addition, the excessive inappropriate use of antimicrobials may result in collateral damage in the homeostasis of the microbial ecology in different parts of the body (microbiome).

This study is an observational study that addresses the following question: What is the impact of therapeutic interventions on the airway, oral and gut microbiome and resistome in patients with chronic lung diseases? COPD and bronchiectasis patients from Royal Free NHS Trust will be recruited and allocated into two groups; a group who are commencing or already on Azithromycin prophylactic therapy and a comparable control group who are not. The participants will be followed up for one year. Repeated biological samples (sputum, nasopharyngeal swabs, saliva, stool, blood) will be collected five times over one year. The changes in the microbiome and resistome in the repeated samples within the same individual will be analyzed and correlated to therapies given to patients.

This study will provide deeper insights into how antibiotics affect the airway, gut and oral microbiome and resistome. This may help in the development of individualized medicine based on tailoring therapy according to the individual microbiome in a way that may limit excessive inappropriate antibiotics utilization and control the development and spread of antimicrobial resistance in population.

A6-2. Summary of main issues. Please summarise the main ethical, legal, or management issues arising from your study and say how you have addressed them.

Not all studies raise significant issues. Some studies may have straightforward ethical or other issues that can be identified and managed routinely. Others may present significant issues requiring further consideration by a REC, HRA, or other review body (as appropriate to the issue). Studies that present a minimal risk to participants may raise complex organisational or legal issues. You should try to consider all the types of issues that the different reviewers may need to consider.

This study aims at investigating the impact of the therapeutic interventions on the individual microbiome and resistome in patients with chronic lung diseases. This is an observational study which does not involve any active interventions. We will not be changing the routine clinical care that patients receive or any of the regular patients’ management planned by their clinicians. The study data will not be fed back to the clinician during the study except in exceptional circumstances in which the results indicate a clinical review is required. That is because the analysis will be done in batches so it will not be available in real time for patient management. The results of the extra spirometry may be available to the clinicians but this is unlikely to influence the standard care because the lung function data of
this patient cohort would be known to the clinical team from previous spirometry tests. So we do not anticipate any risk on the participants. We do not expect that the study data would influence the current treatment or the standard care of the patients. We do not expect that this study may raise any ethical or legal issues either.

Purpose of the Study
Our study is designed to explore a new avenue in this field which is investigating the impact of the therapeutic interventions on the individual microbiome and resistome in patients with chronic lung diseases. At the beginning, we were concerned with the impact airway microbiome and resistome but after discussing the protocol UCL respiratory academic meeting, with Dr. Marc Lipman, and with a group of patients with chronic respiratory conditions in a pulmonary rehabilitation program; we decided it is worth to think about the impact of the therapeutic interventions on the microbiome and resistome of other parts of the body such as the oral cavity and the gut as these medicines are administered systemically and reach all body compartments. We aim also to compare the different routes of drug administration as well in terms of their impact on the microbiome. By studying the airway, oral and gut microbiome within the same patient this will allow us to investigate whether there is a link between the microbiome in these physically linked tracts.

Design
This study will be a prospective observational comparative cohort study. Most of the studies that have been done so far are cross-sectional studies which have informed our understanding of the airway microbiome in chronic lung diseases, revealing a complex and highly variable microbiome in a constantly perturbed ecosystem. However, in order to better understand the impact of therapeutic interventions on the lung microbiome, we find it important to design our study as a longitudinal study which investigates dynamic changes in the bacterial communities within the same patient over time.

The biological samples which will be analyzed have been discussed in supervisory meetings. We agreed on spontaneously expectorated sputum samples to reflect the lower respiratory airways and lungs. Although many have criticized sputum samples as not being the most ideal sample for studying the lung microbiome due to possible contamination from the oropharynx; yet it is the most practical, pragmatic and least invasive lower respiratory sample. In addition, recent published studies have demonstrated that sputum is predictive sample that reflects the microbiology of the lower respiratory compartments and is suitable for the airway microbiome studies. The contamination from upper airway was found to be insignificant. Saliva was selected to represent the oral cavity and stool samples to represent the gut. Blood samples will be collected to test inflammatory and immunological biomarkers to investigate the host response to treatment. Nasopharyngeal swabs will be taken to test for respiratory viruses and investigate their role in the microbiome.

Patient Recruitment and consent
Patients attending pulmonary rehabilitation programs or the out-patient respiratory clinics at Royal Free London NHS Foundation Trust as well as hospitalized patients due to pulmonary exacerbation events will be invited to participate in the study. Patients will be approached by their usual clinical care team during routine encounters. Patients will be given an information sheet and any questions can be answered by the researchers. Patients will be given at least 24 hours to decide if it is convenient for them to participate in the study. No vulnerable participants or children will be recruited. Patient decision will not affect the routine care he/she receives. Those who show interest need to sign an informed consent before participation. After taking the consent, we will collect; demographic data (to ensure that there is no bias in our sample), medical history and spirometry (to confirm clinical condition being investigated), medication, immunization, smoking history (which may influence some microbiome parameters) and social history such as employment; mode of delivery, diet, early life feeding mode (which can be correlated with some microbiome parameters).

Inclusion Criteria;
Inclusion Criteria are set to confirm the medical condition being studied and ensure that the participant’s clinical condition is relevant to the study (frequent exacerbators and able to expectorate sputum).

Exclusion Criteria
Exclusion Criteria are set exclude patients who are immune-compromised or have a medical condition that may influence the microbiome.

Risks, Burden and Benefits

Risks
We do not anticipate major risks on participants because our study is observational study. We will not be changing any of the regular patients’ management therapy prescribed by their consultant. All subjects will be encouraged to seek treatment early for a suspected exacerbation. Blood will be taken by a suitably qualified person using procedures approved by NHS Trust to ensure the participant’s safety and to keep the risks of bruising and needle stick injury associated with blood tests minimal.

Burdens
Study subjects should be committed to participate in the study for 1 year. Over this time they will be routinely reviewed every three months by the clinical care team in their regular follow up appointments, thus providing five routine reviews.

Date: 13/09/2016
during the study. Patients who will be starting azithromycin prophylactic treatment will be asked to pay one more extra study visit after one month from starting the antibiotic course. Nasopharyngeal swabs may cause mild discomfort. In addition to these scheduled visits, patients will be requested to complete daily diary cards reporting any changes in their respiratory symptoms; any adverse effects and their medication throughout the study period. The majority of the time, most participants will simply record “no new symptoms” and “no change in the medication”; this should take less than 5 minutes once daily. However during periods of exacerbations it may take longer to detail the chest symptoms and the medications received in the diary cards.

Benefits
Developing new knowledge that may improve the future care of chronic lung diseases.
Patients taking part will learn more about their condition and they will have access to specialist advice during the study.

Confidentiality
Full patient confidentiality will be maintained throughout the study with access to samples and data restricted to the study investigators. Data will be stored pseudo-anonymously in password protected and encrypted files.

Conflict of interest
We do not anticipate conflict of interest between our research and patient care. By the end of the study, the results will be disseminated through scientific conferences and peer reviewed journals.

Storage and future use of samples
Where there is sufficient clinical sample to allow subsequent testing, samples will be stored in UCL Centre for Clinical Microbiology for analysis. Participants in the study will be asked for consent for retaining their samples for ethically approved research in the future. The samples will only be used by members of the research team to help answer questions relevant to this study and will not be made available to third parties. Participants will have the opportunity to decline to have their samples retained and in this event the samples will be destroyed after testing for this study.

3. PURPOSE AND DESIGN OF THE RESEARCH

A7. Select the appropriate methodology description for this research. Please tick all that apply:

☐ Case series/ case note review
☐ Case control
☐ Cohort observation
☐ Controlled trial without randomisation
☐ Cross-sectional study
☐ Database analysis
☐ Epidemiology
☐ Feasibility/ pilot study
☐ Laboratory study
☐ Metaanalysis
☐ Qualitative research
☐ Questionnaire, interview or observation study
☐ Randomised controlled trial
☐ Other (please specify)

A19. What is the principal research question/objective? Please put this in language comprehensible to a lay person.

1. What is the impact of therapeutic interventions on airway microbiome and resistome in patients with chronic lung diseases receiving frequent courses of antibiotics and corticosteroids?

2. What is the impact of therapeutic interventions on gut and oral resistome in patients with chronic lung diseases receiving frequent courses of antibiotics and corticosteroids?

Date: 13/09/2016
A11. What are the secondary research questions/objectives if applicable? Please put this in language comprehensible to a lay person.

1. In what way do the airway microbiome and resistome parameters change in patients with chronic lung diseases receiving long term azithromycin prophylaxis compared to a control group?
2. What is the influence of long term azithromycin prophylaxis on the frequency of exacerbations and health associated quality of life in patients with chronic lung disease?
3. What are the phenotypic and genotypic antimicrobial resistance patterns of the respiratory bacterial pathogens isolated from patients with chronic lung diseases receiving frequent courses of antibiotics?
4. How does the route of administration of the drugs (whether being oral or inhaled) affect the airway microbiome parameters?
5. How do the respiratory viruses contribute to airway microbiome in chronic lung diseases?
6. How can the inflammatory and immunological biomarkers in the plasma be correlated to treatment response?
7. Is there any association between the frequency of exacerbations, individual microbiome parameters and the cohort characteristics (smoking history, diet, body mass index, mode of delivery, early feeding mode)?

A12. What is the scientific justification for the research? Please put this in language comprehensible to a lay person.

Until recently, lower respiratory tracts and the lungs were believed to be sterile and that lower airways became colonized with potentially pathogenic microorganisms only during respiratory infections. In 2010, Filtyl et al. through using advanced molecular techniques was the first to demonstrate that the bronchial tree was not sterile and it is inhabited by normal microflora but most of these bacteria are not cultivable. Since then, there was a growing interest in lung microbiome. Many studies were conducted trying to characterize airway microbiome in health and disease conditions. Currently, "microbial dysbiosis" represents the basis of a new conceptualized model in which alterations in the healthy microbiome; e.g. in richness, bacterial load or diversity, could be associated with the disease development and progression in a wide range of chronic respiratory conditions. Many studies on the respiratory tract have suggested links between the microbiome composition and disease progression, severity and exacerbations across wide spectrum of chronic pulmonary disorders. Not much work has been conducted on the influence of the therapy on the individual microbiome.

Patients with chronic lung diseases are often prescribed frequent and prolonged courses of antibiotics and corticosteroids in UK. However, the extensive debate around the usefulness of antibiotics in the treatment of exacerbations has not been concluded yet as the clinical evidence supporting this practice is surprisingly low. A number of recent clinical trials have demonstrated the usefulness of the long-term prophylactic use of antibiotics; especially the macrolides, to reduce the frequency of exacerbations in COPD and bronchiectasis. For example, in BLESS study (a randomised double blind placebo controlled clinical trial in Australia), Erythromycin (400mg twice daily for one year) successfully reduced exacerbation frequency, reduced 24 hr sputum production and attenuated lung function decline in bronchiectasis patients; but the proportion of macrolide-resistant oropharyngeal streptococci was doubled in the treated group. Albert et al.'s randomised double blind placebo controlled clinical trial on 1142 patients in USA showed that 250 mg of Azithromycin once daily for 12 months decreased the frequency of exacerbations; improved quality of life; however it increased hearing decrements by 5% and doubled the frequency of macrolide resistance in treated group. It is not clear till now the actual benefit of antimicrobials in the treatment of acute exacerbations or their role when administered as prophylactic treatment in controlling the chronic respiratory conditions and their impact on the microbiome and the resistome.

Cross-sectional studies have informed our understanding of the airway microbiome in chronic lung diseases, revealing a complex and highly variable microbiome in a constantly perturbed ecosystem. However, in order to better understand the impact of therapeutic interventions on the lung microbiome, it is important to investigate dynamic changes in the bacterial communities within the same patient over time. Excessive and inappropriate use of antimicrobials often results in the emergence of multidrug-resistant bacterial strains known as superbugs which is associated with treatment failure, higher morbidity and mortality, and increased health care costs. In addition, there might be significant collateral damage in the human microbiome due to excessive antibiotics administration. So far, studies that have assessed the ecological impact of antibiotics on the human microbiome are scarce and have focused mainly on the gut. Studies on gut microbiome and resistome studies have showed that uncultivable bacteria in the microbiome could play a major role in the acquisition and spread of antibiotic resistance to pathogenic bacteria even under short-term antibiotic administration. Therefore, microbiota can act a reservoir for antibiotic resistance. But little is known about the lung's resistome. To the best of our knowledge, this study will be one of the early studies in the UK to study the impact of therapeutic interventions mainly antibiotics on the airway microbiome and resistome in Chronic respiratory disorders. The results of this study will provide a better understanding and deeper insights into how antibiotics affect the microbial ecology in airways. This may help in the development of individualized medicine based on tailoring therapy according to the individual microbiome.

A13. Please summarise your design and methodology. It should be clear exactly what will happen to the research participant, how many times and in what order. Please complete this section in language comprehensible to the lay person.

Date: 13/09/2016

11

202883/1017635/37/744
Hypothesis:
We hypothesise that frequent and prolonged courses of chemotherapeutic antimicrobial agents; in addition to, corticosteroids, which are prescribed to patients with chronic respiratory disorders either as a part of the lifelong management therapy or for the treatment of acute exacerbations, may affect the homeostasis of the individual microbiome and may contribute to the enrichment of the resistantome.

Study Design:
This study will be a prospective observational comparative cohort study. The proposed research protocol was discussed with a sample of patients in pulmonary rehabilitation program held in Peckwater Centre, London. The design was adjusted in the light of their comments.

A group of patients with severe chronic lung disease who will be starting long term azithromycin prophylaxis prescribed by their respiratory consultant, as a part of their management regime, will be recruited. A spontaneously coughed sputum sample, nasopharyngeal swabs, saliva, stool and blood samples will be collected from these patients at baseline just before starting the course, after a month from the start date and by the end of the study.

Further sputum samples and nasopharyngeal swabs will be collected at three months’ time intervals during their regular follow up appointments in the clinic. The baseline airway microbiome will be defined by next generation sequencing techniques and compared over a period of one year within the same patient to assess the influence of long term azithromycin prophylaxis on the individual microbiome and on the development or acquisition of macrolides’ resistance.

Another group of patients who are already on long term azithromycin prophylaxis for at least six months will also be recruited. This group will be of great value in assessing the effect of longer treatment periods on the levels of antimicrobial resistance and the more persistent changes in the airway and gut microbiome. The results will be compared with a control group of patients with similar clinical conditions but not on long term azithromycin prophylaxis. So each patient will be seen five times at their regular scheduled follow up reviews through out the year. The patients who will be starting azithromycin prophylaxis course will have an extra review appointment after one month of starting the antibiotic course. Data will be recorded in a proforma and entered into a database under allocated study number. This data will only be accessible by the study investigators. The personal information and link will be stored in a protected encrypted database.

Data collection for all study participants will consist of (Please refer to Table 3 in the protocol version 3.4):
- Demographic details (1st visit)
- Medical history and lung function data (1st visit)
- Smoking History (in all visits)
- History of immunization including pneumococcal and influenza vaccines (All visits)
- Medication history (All visits)
- Social history: employment; mode of delivery, diet, early life feeding mode. (1st visit)
- The burden of respiratory symptoms will be assessed by use of the St George’s Respiratory Questionnaire (SGRQ) (the first and last visit)
- COPD Assessment Test (CAT) for COPD patients (All visits)
- Spontaneously coughed sputum collection for detection of respiratory bacterial pathogens by means of culture and molecular diagnostic techniques (All visits)
- Nasopharyngeal swabs for detection of respiratory viruses using molecular diagnostic techniques (all visits)
- Sputum samples will stored for airway microbiome and resistome analysis
- Stool and saliva samples collection for gut and oral microbiome and resistome analyses respectively.(first and last visits)
- Blood samples collection for inflammatory and immunological biomarkers investigation. (first and last visits)
- Patients will be asked to complete daily diary cards of their on-going respiratory symptoms and medication to allow capturing all exacerbation events and to document antibiotic usage, any adverse events and any change in the treatment regimes. (daily)

The sample size is established on a pragmatic recruitment rate of one patient per week until twenty five participants are allocated in each study group (on and off azithromycin prophylactic treatment) then we will use the generated information to inform the power calculations for a larger future study.

A14.1. In which aspects of the research process have you actively involved, or will you involve, patients, service users, and/or their carers, or members of the public?

- [ ] Design of the research
- [ ] Management of the research
- [x] Undertaking the research
- [ ] Analysis of results

Date: 13/09/2016

202883/1017635/37/744
Give details of involvement. or if none please justify the absence of involvement.
The study needs to be acceptable to participants to have adequate retention in the study so that it can provide satisfying answers to the proposed research questions. This study addresses key issues to patients with chronic lung diseases as well as public which is the pros and cons of the frequent antibiotic courses on patients' normal microbiota (good bacteria) and quality of life; on the other hand, how the excessive consumption of antibiotics in this patient group can influence the prevalence of antimicrobial resistance in the population. We discussed our proposed research with a sample of patients, adjusting the design in the light of their comments. The patients showed an interest in the project and provided constructive reviews. During the study, the patients will monitor their respiratory symptoms daily through filling daily diary cards and to contact the study team as soon as they feel deterioration in their baseline respiratory symptoms that may suggest acute exacerbations. We will be keen to feedback the patients who will be interested with any relevant results. These patients will help us to disseminate the findings of our study among the population. Therefore, involvement of patients is essential to complete the study. The study will be performed in line with INVOLVE principles.

4. RISKS AND ETHICAL ISSUES

RESEARCH PARTICIPANTS

A15. What is the sample group or cohort to be studied in this research?

Select all that apply:

- Blood
- Cancer
- Cardiovascular
- Congenital Disorders
- Dementias and Neurodegenerative Diseases
- Diabetes
- Ear
- Eye
- Generic Health Relevance
- Infection
- Inflammatory and Immune System
- Injuries and Accidents
- Mental Health
- Metabolic and Endocrine
- Musculoskeletal
- Neurological
- Oral and Gastrointestinal
- Paediatrics
- Renal and Urogenital
- Reproductive Health and Childbirth
- Respiratory
- Skin
- Stroke
A17-1. Please list the principal inclusion criteria (list the most important, max 5000 characters).

1. Confirmed primary diagnosis of COPD or Bronchiectasis
   BRONCHIECTASIS INCLUSION: Patients with primary diagnosis of bronchiectasis confirmed by their treating clinicians and documented by high-resolution computed tomographic scan.
   COPD INCLUSION: Patients with primary diagnosis of COPD confirmed by their treating clinicians will be included if the forced expiratory volume in one second (FEV1) was ≥ 80% predicted and FEV1/FVC ratio was < 0.7, in keeping with GOLD stages II to IV. A history of chronic symptoms (dyspnoea, sputum production, wheeze and cough), and smoking history (number of pack-years smoked, current smoking status) will be taken in to consideration.

2. History of frequent exacerbations two or more in the year preceding recruitment
3. Male or female 20 years and above
4. Patients should be able to expectorate and produce sputum sample
5. Patient should be able to perform spirometry maneuvers reproducibly.
6. Patients should be able to provide a written informed consent.
7. Willing to participate in study for the whole duration.

A17-2. Please list the principal exclusion criteria (list the most important, max 5000 characters).

1. Risk of non-compliance or inability to comply with the study procedures
2. Unable to produce sputum
3. Patients receiving immunosuppressant drug not including steroids
4. Patients with confirmed diagnosis of asthma, cystic fibrosis, current pneumonia or other respiratory disorders (e.g.: tuberculosis, lung cancer).
5. Known HIV infection.
6. Presence of other conditions that the principal investigator judges may interfere with the study findings.
7. Patients who cannot communicate in English.

RESEARCH PROCEDURES, RISKS AND BENEFITS

A18. Give details of all non-clinical intervention(s) or procedure(s) that will be received by participants as part of the research protocol. These include seeking consent, interviews, non-clinical observations and use of questionnaires.

Please complete the columns for each intervention/procedure as follows:

1. Total number of interventions/procedures to be received by each participant as part of the research protocol.
2. If this intervention/procedure would be routinely given to participants as part of their care outside the research, how many of the total would be routine?
3. Average time taken per intervention/procedure (minutes, hours or days)
4. Details of who will conduct the intervention/procedure, and where it will take place.

<table>
<thead>
<tr>
<th>Intervention or procedure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invitation and information sheet</td>
<td>1</td>
<td>0</td>
<td>10 mins</td>
<td>Staff at Royal Free Hospital</td>
</tr>
<tr>
<td>Consent and answering participant questions about the study</td>
<td>1</td>
<td>0</td>
<td>15 min</td>
<td>Research nurse or researcher at Royal Free Hospital</td>
</tr>
<tr>
<td>Collection of baseline demographic and medical history</td>
<td>1</td>
<td>0</td>
<td>15 min</td>
<td>Research doctor or researcher at Royal Free Hospital</td>
</tr>
</tbody>
</table>

Date: 13/09/2016
A19. Give details of any clinical intervention(s) or procedure(s) to be received by participants as part of the research protocol. These include uses of medicinal products or devices, other medical treatments or assessments, mental health interventions, imaging investigations and taking samples of human biological material. Include procedures which might be received as routine clinical care outside of the research.

Please complete the columns for each intervention/procedure as follows:

1. Total number of interventions/procedures to be received by each participant as part of the research protocol.
2. If this intervention/procedure would be routinely given to participants as part of their care outside the research, how many of the total would be routine?
3. Average time taken per intervention/procedure (minutes, hours or days).
4. Details of who will conduct the intervention/procedure, and where it will take place.

<table>
<thead>
<tr>
<th>Intervention or procedure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collection of spontaneous sputum samples</td>
<td>6</td>
<td>0</td>
<td>3 min</td>
<td>Research doctor or research nurse at Royal Free Hospital</td>
</tr>
<tr>
<td>Spirometry</td>
<td>2</td>
<td>1</td>
<td>15 mins</td>
<td>Research doctor at Royal Free Hospital</td>
</tr>
<tr>
<td>Collection of Nasopharyngeal swabs for detection of respiratory viruses</td>
<td>6</td>
<td>0</td>
<td>5 min</td>
<td>Research doctor or research nurse at Royal Free Hospital</td>
</tr>
<tr>
<td>Collection of Saliva samples for oral microbiome and resistome analysis</td>
<td>3</td>
<td>0</td>
<td>3 min</td>
<td>Research doctor or research nurse at Royal Free Hospital</td>
</tr>
<tr>
<td>Collection of Stool samples for gut microbiome and resistome analysis</td>
<td>3</td>
<td>0</td>
<td>5 mins</td>
<td>Research doctor or research nurse at Royal Free Hospital</td>
</tr>
<tr>
<td>Collection of Blood Samples for immunological analysis</td>
<td>3</td>
<td>0</td>
<td>5 mins</td>
<td>Research doctor or research nurse at Royal Free Hospital</td>
</tr>
</tbody>
</table>

A21. How long do you expect each participant to be in the study in total?
For One year

A22. What are the potential risks and burdens for research participants and how will you minimise them?
For all studies, describe any potential adverse effects, pain, discomfort, distress, intrusion, inconvenience or changes to lifestyle. Only describe risks or burdens that could occur as a result of participation in the research. Say what steps would be taken to minimise risks and burdens as far as possible.

This is an observational study which does not involve active intervention. We will not be changing the routine care that patients receive or any of the regular patients' management prescribed by their physicians. All subjects will be encouraged to seek treatment early for a suspected exacerbation. So we do not anticipate any risk on the participants apart from the study burden due to their commitment to be on the study for one year and filling the daily diary. Filling the diary cards everyday will take in average three minutes every day in most of the times; however during periods of exacerbations it may take longer. To make it easier we will provide the diary cards in a printed form. Also the patients may experience mild discomfort while taking the nasopharyngeal swab. Blood will be taken by a suitably qualified person using procedures approved by NHS Trust to ensure the participant's safety and to keep the risks of bruising and needle stick injury associated with blood tests minimal.

A24. What is the potential for benefit to research participants?
Developing new knowledge that may improve the future care of chronic lung diseases. Patients taking part will learn more about their condition and they will have access to specialist advice during the study.

RECRUITMENT AND INFORMED CONSENT

In this section we ask you to describe the recruitment procedures for the study. Please give separate details for different study groups where appropriate.

A27-1. How will potential participants, records or samples be identified? Who will carry this out and what resources will be used? For example, identification may involve a disease register, computerised search of GP records, or review of medical records. Indicate whether this will be done by the direct healthcare team or by researchers acting under arrangements with the responsible care organisation(s).

Patients will be approached and invited to take part in our studies at COPD clinics and Pulmonary Rehabilitation classes across Camden, by Dr Hurst and the COPD team. Patients willing to take part will be given a study information sheet and an appointment to discuss taking part arranged with the researcher. If the patient is willing to take part, written informed consent will be completed prior to any study involvement.

A27-2. Will the identification of potential participants involve reviewing or screening the identifiable personal information of patients, service users or any other person?
- Yes
- No

Please give details below:

A28. Will any participants be recruited by publicity through posters, leaflets, adverts or websites?
- Yes
- No

A29. How and by whom will potential participants first be approached?

Patients will be approached by their usual clinical care team (Dr. Hurst and COPD team) during routine encounters in their routine clinical environment. Patients will be given an information sheet and any questions can be answered by the researchers.

A30-1. Will you obtain informed consent from or on behalf of research participants?
- Yes
- No

If you will be obtaining consent from adult participants, please give details of who will take consent and how it will be done, with details of any steps to provide information (a written information sheet, videos, or interactive material). Arrangements for adults unable to consent for themselves should be described separately in Part B Section 6, and for
children in Part B Section 7.

If you plan to seek informed consent from vulnerable groups, say how you will ensure that consent is voluntary and fully informed.

After delivering a full study description to the participant, those who agree to participate in the study will be asked to fill out the consent form and sign it. All participants will require the ability to consent for themselves to participate in the study.

If you are not obtaining consent, please explain why not.

Please enclose a copy of the information sheet(s) and consent form(s).

A30-2. Will you record informed consent (or advice from consultees) in writing?

- Yes  
- No

A31. How long will you allow potential participants to decide whether or not to take part?

A minimum of 24 hours

A32. Will you recruit any participants who are involved in current research or have recently been involved in any research prior to recruitment?

- Yes  
- No  
- Not Known

If Yes, please give details and justify their inclusion. If Not Known, what steps will you take to find out?

Participants will be asked about this at the recruitment visit. Prior studies would not be a barrier to participation. There will be no reason to exclude individuals from this study due to participation in other research trials unless this trial would directly impact on the primary outcome (e.g. if the participant is blinded to administration of an antibiotic or corticosteroids).

A33-1. What arrangements have been made for persons who might not adequately understand verbal explanations or written information given in English, or who have special communication needs? (e.g. translation, use of interpreters)

All the information will be delivered in English. Patients who can not comprehend English will not be recruited.

A35. What steps would you take if a participant, who has given informed consent, loses capacity to consent during the study? Tick one option only.

- The participant and all identifiable data or tissue collected would be withdrawn from the study. Data or tissue which is not identifiable to the research team may be retained.
- The participant would be withdrawn from the study. Identifiable data or tissue already collected with consent would be retained and used in the study. No further data or tissue would be collected or any other research procedures carried out on or in relation to the participant.
- The participant would continue to be included in the study.
- Not applicable – informed consent will not be sought from any participants in this research.
- Not applicable – it is not practicable for the research team to monitor capacity and continued capacity will be assumed.

Further details:

If you plan to retain and make further use of identifiable data/tissue following loss of capacity, you should inform
**CONFIDENTIALITY**

In this section, personal data means any data relating to a participant who could potentially be identified. It includes pseudonymised data capable of being linked to a participant through a unique code number.

**Storage and use of personal data during the study**

A36. Will you be undertaking any of the following activities at any stage (including in the identification of potential participants)? *(Tick as appropriate)*

- [ ] Access to medical records by those outside the direct healthcare team
- [ ] Access to social care records by those outside the direct social care team
- [X] Electronic transfer by magnetic or optical media, email or computer networks
- [ ] Sharing of personal data with other organisations
- [ ] Export of personal data outside the EEA
- [ ] Use of personal addresses, postcodes, faxes, emails or telephone numbers
- [ ] Publication of direct quotations from respondents
- [ ] Publication of data that might allow identification of individuals
- [ ] Use of audio/visual recording devices
- [ ] Storage of personal data on any of the following:
  - [X] Manual files (includes paper or film)
  - [ ] NHS computers
  - [ ] Social Care Service computers
  - [ ] Home or other personal computers
  - [X] University computers
  - [ ] Private company computers
  - [ ] Laptop computers

**Further details:**

We will need to keep a record of patients' telephone number and address to enable contact. Patients will be given a study number. Source documentation paper files, original paper questionnaires and consent forms will be kept in a locked office at the Royal Free Campus of UCL. Electronic files will contain only pseudonymised data and will be saved as password protected files on password accessed UCL computers. Email may be used to transfer non-identifiable anonymised data. Data for analysis on laptops will be anonymised and no personally identifiable data will be kept on lap tops.

A37. Please describe the physical security arrangements for storage of personal data during the study?

The Source documentation paper files which contain personal data, original paper questionnaires, participants' contacts and the original signed consents will be kept in a file in a locked cupboard in a room locked by a coded access in UCL centre for Clinical Microbiology or UCL respiratory department, in Royal Free Campus. Electronic data will contain only pseudonymised data and will be saved as password protected files on password accessed UCL computers with a limited access for the central research team and the chief investigator. Security badge access is required at the Royal Free campus to access UCL Respiratory and UCL Centre for Clinical Microbiology.

A38. How will you ensure the confidentiality of personal data? *Please provide a general statement of the policy and procedures for ensuring confidentiality, e.g. anonymisation or pseudonymisation of data.*

All the identification data will be saved in a password protected file on a password accessed UCL computer and kept

Date: 13/09/2016
separately from research data. Research data will be pseudonymised using a study number.

A40. Who will have access to participants' personal data during the study? Where access is by individuals outside the direct care team, please justify and say whether consent will be sought.

Only the primary researcher (Mrs. Sylvia Roaf)} and the supervisors (Prof. Tim McHugh and Dr. John Hurst), and sponsor as necessary.

Storage and use of data after the end of the study

A41. Where will the data generated by the study be analysed and by whom?

Generated data will be analyzed by the research team and UCL statistics team in UCL. Any identifiable information will not be included at this stage.

A42. Who will have control of and act as the custodian for the data generated by the study?

<table>
<thead>
<tr>
<th>Title</th>
<th>Forename/Initials</th>
<th>Surname</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr.</td>
<td>John</td>
<td>Hurst</td>
</tr>
</tbody>
</table>

Reader in Respiratory Medicine/ Honorary Consultant Physician, Royal Free London NHS Foundation Trust

Qualifications: PHD, FRCP, FRCPE

Work Address

Post Code

Work Email

Work Telephone

Fax

A43. How long will personal data be stored or accessed after the study has ended?

- [ ] Less than 3 months
- [ ] 3 – 6 months
- [ ] 6 – 12 months
- [ ] 12 months – 3 years
- [x] Over 3 years

If longer than 12 months, please justify:

Recruitment is estimated to take up to 2 years. Patients will take part for up to one year, but the final data set will be available at approximately five years to allow further analysis of the data.

A44. For how long will you store research data generated by the study?

Years: 20

Months:

A45. Please give details of the long term arrangements for storage of research data after the study has ended. Say where data will be stored, who will have access and the arrangements to ensure security.

Date: 13/09/2016
The research team will keep the research data for 20 years, held securely at UCL.

**INCENTIVES AND PAYMENTS**

A46. Will research participants receive any payments, reimbursement of expenses or any other benefits or incentives for taking part in this research?

- [ ] Yes  [ ] No

A47. Will individual researchers receive any personal payment over and above normal salary, or any other benefits or incentives, for taking part in this research?

- [ ] Yes  [ ] No

A48. Does the Chief Investigator or any other investigator/collaborator have any direct personal involvement (e.g. financial, share holding, personal relationship etc.) in the organisations sponsoring or funding the research that may give rise to a possible conflict of interest?

- [ ] Yes  [ ] No

**NOTIFICATION OF OTHER PROFESSIONALS**

A49-1. Will you inform the participants’ General Practitioners (and/or any other health or care professional responsible for their care) that they are taking part in the study?

- [ ] Yes  [ ] No

*If Yes, please enclose a copy of the information sheet/letter for the GP/health professional with a version number and date.*

**PUBLICATION AND DISSEMINATION**

A50. Will the research be registered on a public database?

*The Department of Health's Research Governance Framework for Health and Social Care and the research governance frameworks for Wales, Scotland and Northern Ireland set out the requirement for registration of trials. Furthermore: Article 19 of the World Medical Association Declaration of Helsinki adopted in 2008 states that “every clinical trial must be registered on a publicly accessible database before recruitment of the first subject”; and the International Committee of Medical Journal Editors (ICMJE) will consider a clinical trial for publication only if it has been registered in an appropriate registry. Please see guidance for more information.*

- [ ] Yes  [ ] No

*Please give details, or justify if not registering the research. This is an observational study that focuses on laboratory methods.*

*Please ensure that you have entered registry reference number(s) in question A5-1.*

A51. How do you intend to report and disseminate the results of the study? Tick as appropriate:

- [x] Peer reviewed scientific journals
- [x] Internal report
- [x] Conference presentation

**Date:** 13/09/2016
A52. If you will be using identifiable personal data, how will you ensure that anonymity will be maintained when publishing the results?

Identifiable personal data will not be used when publishing the results.

A53. Will you inform participants of the results?

☐ Yes  ☐ No

Please give details of how you will inform participants or justify if not doing so. Participants who wish to be notified of the results will receive a lay summary of our findings by mail or e-mail as they wish by the end of the study. The interested participants can be given access to the final study report and the scientific publication that are produced from this study upon their request.

5. Scientific and Statistical Review

A54. How has the scientific quality of the research been assessed? Tick as appropriate:

☐ Independent external review
☐ Review within a company
☐ Review within a multi-centre research group
☑ Review within the Chief Investigator's institution or host organisation
☑ Review within the research team
☐ Review by educational supervisor
☐ Other

Justify and describe the review process and outcome. If the review has been undertaken but not seen by the researcher, give details of the body which has undertaken the review. The protocol has also been reviewed by other researchers in host institution; it has been discussed in UCL respiratory academic meeting and UCL Centre for Clinical Microbiology lab meeting. It has been also reviewed by experts including Dr Marc Lipman and Dr David Spratt. The former is highly experienced clinical researcher and he has advised regarding collection of appropriate patient samples. The latter is highly experienced in microbial ecology and microbiome studies and he has given supportive reviews of the study protocol. The study has been also approved by the PhD student sponsor, and by UCL as suitable for a PhD.

For all studies except non-doctoral student research, please enclose a copy of any available scientific critique reports, together with any related correspondence.

For non-doctoral student research, please enclose a copy of the assessment from your educational supervisor/ institute.

A56. How have the statistical aspects of the research been reviewed? Tick as appropriate:

☐ Review by independent statistician commissioned by funder or sponsor
☐ Other review by independent statistician
A57. What is the primary outcome measure for the study?

1. Investigate the changes in the microbiome parameters (community richness, evenness, alpha and beta diversity indices, bacterial load) with antibiotic treatment in patients with chronic lung disease.

2. Assess the prevalence of antimicrobial resistance genes against commonly used broad spectrum antibiotics including; macrolides, β-lactams, tetracyclines and fluoroquinolones among the microbiota in patients with chronic lung diseases receiving frequent courses of antimicrobial chemotherapeutic agents.

A58. What are the secondary outcome measures? (If any)

1. Report the prevalence and the load of respiratory viral agents such as rhino virus, human respiratory syncytial virus (RSVs), influenza virus and adeno virus in the airways during stable state and acute exacerbations in chronic respiratory conditions.

2. Measure the inflammatory and immunological biomarkers related to infections in plasma.

3. Report the frequency (total and time to next pulmonary exacerbation), duration, severity of exacerbations and change in FEV1 in patients receiving long term azithromycin prophylaxis.

4. Measure the health related quality of life in patients receiving long term azithromycin prophylaxis.

A59. What is the sample size for the research? How many participants/samples/data records do you plan to study in total? If there is more than one group, please give further details below.

Total UK sample size: 50
Total international sample size (including UK): 0

Date: 13/09/2016
Further details:
This is a pilot study that will assess the associations between disease state, treatment success and the lung microbiome/resistome. The primary outcome in this study will be defining the changes in the microbiome/resistome parameters with various short and long term antibiotic treatment in patients with chronic lung disease. The sample size is established on a pragmatic recruitment rate of one patient per week until twenty five participants are recruited in each study group (on and off azithromycin prophylactic treatment) then we will use the generated information to inform power calculations for a larger study. In this study we are not changing the treatment regime of any of the participants; it is just an observational study.

A60. How was the sample size decided upon? If a formal sample size calculation was used, indicate how this was done, giving sufficient information to justify and reproduce the calculation.

This is an observational study that will assess the associations between disease state, treatment success and the lung microbiome/resistome. The sample size is established on a pragmatic recruitment rate of one patient per week until twenty five participants are recruited in each study group (on and off azithromycin prophylactic treatment) then we will use the generated information to inform the power calculations.

A61. Will participants be allocated to groups at random?

☐ Yes  ☐ No

A62. Please describe the methods of analysis (statistical or other appropriate methods, e.g. for qualitative research) by which the data will be evaluated to meet the study objectives.

Principal components analysis (PCA) will be applied to the combined clinical and microbiological parameters to identify those characteristics that are driving the pathogenesis of chronic lung disease. For microbiome analysis, the 16S rRNA amplicon raw sequencing data will be analyzed using Quantitative Insights In to Microbial Ecology (QIIME) software (version 1.5.0), which allows further read processing and clustering into relevant Operational Taxonomic Units (OTUs). Community richness (number of taxa detected), evenness (relative distribution of taxa in a community), and alpha diversity (which consider both the number of taxa and their relative abundance) statistics will be performed on the obtained phylogenetic data of the various samples using the R statistical software package (version 3.2.2). The taxa will be partitioned into core and rare taxa groups such that the persistent and abundant core taxa will be defined as those in more than 75% of all samples, while all other species falling outside of the upper quartile will be considered rare. Beta-diversity measures will be calculated to examine the differences between the microbiome profiles and the taxa turnover between the consecutive samples. Other flexible approaches to statistical analysis may also be used.

6. MANAGEMENT OF THE RESEARCH

A63. Other key Investigators/collaborators. Please include all grant co-applicants, protocol co-authors and other key members of the Chief Investigator’s team, including non-doctoral student researchers.

Title: For name/Initials Surname
Prof. Timothy D. McHugh

Post: Professor of Medical Microbiology, Div. of Infection & Immunity Faculty of Medical Sciences,
University College London and Director of UCL Centre for Clinical Microbiology

Qualifications: B. Sc. (Hons), PhD

Employer: University College London

Work Address:

Post Code

Date: 13/09/2016

202883/1017635/37/744
Title: Forename/Initials Surname
Dr. John Hurst
Post: Reader in Respiratory Medicine/ Honorary Consultant Physician, Royal Free London NHS Foundation Trust
Qualifications: PHD, FRCP, FRCPE
Employer: University College London, Royal Free London NHS Foundation Trust

Date: 13/09/2016
#### IRAS Form

<table>
<thead>
<tr>
<th>Post</th>
<th>Senior Lecturer, Division of Medicine, UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualifications</td>
<td>BA (Hons), MB BS, MA, MRCP, MD, FRCP</td>
</tr>
<tr>
<td>Employer</td>
<td>University College of London</td>
</tr>
<tr>
<td>Work Address</td>
<td></td>
</tr>
<tr>
<td>Post Code</td>
<td></td>
</tr>
<tr>
<td>Telephone</td>
<td></td>
</tr>
<tr>
<td>Fax</td>
<td></td>
</tr>
<tr>
<td>Mobile</td>
<td></td>
</tr>
<tr>
<td>Work Email</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Title</th>
<th>Forename/Initials</th>
<th>Surname</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr.</td>
<td>Rama</td>
<td>Vancheeswaran</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Post</th>
<th>Consultant respiratory and general physician</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualifications</td>
<td>FRCP, MSc Immunology, PhD</td>
</tr>
<tr>
<td>Employer</td>
<td>Royal Free London NHS Foundation Trust</td>
</tr>
<tr>
<td>Work Address</td>
<td></td>
</tr>
<tr>
<td>Post Code</td>
<td></td>
</tr>
<tr>
<td>Telephone</td>
<td></td>
</tr>
<tr>
<td>Fax</td>
<td></td>
</tr>
<tr>
<td>Mobile</td>
<td></td>
</tr>
<tr>
<td>Work Email</td>
<td></td>
</tr>
</tbody>
</table>

### A64. Details of research sponsor(s)

#### A64.1. Sponsor

**Lead Sponsor**

<table>
<thead>
<tr>
<th>Status:</th>
<th>NHS or HSC care organisation</th>
<th>Academic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pharmaceutical industry</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medical device industry</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Local Authority</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Other social care provider (including voluntary sector or private organisation)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Other</td>
</tr>
</tbody>
</table>

**Commercial status:**

*If Other, please specify:*

**Contact person**

- Name of organisation: University College of London
- Given name: Tabitha

Date: 13/09/2016
Family name: Kavoli

Is the sponsor based outside the UK?
☐ Yes ☒ No

Under the Research Governance Framework for Health and Social Care, a sponsor outside the UK must appoint a legal representative established in the UK. Please consult the guidance notes.

A65. Has external funding for the research been secured?
☐ Funding secured from one or more funders
☐ External funding application to one or more funders in progress
☒ No application for external funding will be made

What type of research project is this?
☐ Stand-alone project
☐ Project that is part of a programme grant
☐ Project that is part of a Centre grant
☐ Project that is part of a fellowship/ personal award/ research training award
☐ Other

Other – please state: PhD student Project

A66. Has responsibility for any specific research activities or procedures been delegated to a subcontractor (other than a co-sponsor listed in A64-1)? Please give details of subcontractors if applicable.

☐ Yes ☐ No

A67. Has this or a similar application been previously rejected by a Research Ethics Committee in the UK or another country?

☐ Yes ☐ No

Please provide a copy of the unfavourable opinion letter(s). You should explain in your answer to question A6-2 how the reasons for the unfavourable opinion have been addressed in this application.

A68-1. Give details of the lead NHS R&D contact for this research:

Date: 13/09/2016
A68-2. Select Local Clinical Research Network for NHS Organisation Identified in A68-1:

North Thames

For more information, please refer to the question specific guidance.

A69-1. How long do you expect the study to last in the UK?

Planned start date: 15/09/2016
Planned end date: 01/08/2019
Total duration:
Years: 2  Months: 10  Days: 17

A70. Definition of the end of trial, and justification in the case where it is not the last visit of the last subject undergoing the trial (1)

We define the end of the trial as last visit of the last subject undergoing research. We anticipate we may take 2 years to recruit patients.

A71-1. Is this study?

☐ Single centre
☐ Multicentre

A71-2. Where will the research take place? (Tick as appropriate)

☑ England
☐ Scotland
☐ Wales
☐ Northern Ireland
☐ Other countries in European Economic Area

Total UK sites in study 2

Does this trial involve countries outside the EU?

☐ Yes  ☐ No

Date: 13/09/2016
A72. Which organisations in the UK will host the research? Please indicate the type of organisation by ticking the box and give approximate numbers if known:

- NHS organisations in England 1
- NHS organisations in Wales
- NHS organisations in Scotland
- HSC organisations in Northern Ireland
- GP practices in England
- GP practices in Wales
- GP practices in Scotland
- GP practices in Northern Ireland
- Joint health and social care agencies (e.g., community mental health teams)
- Local authorities
- Phase 1 trial units
- Prison establishments
- Probation areas
- Independent (private or voluntary sector) organisations
  - Educational establishments 1
  - Independent research units
  - Other (give details)

Total UK sites in study: 2

A73-1. Will potential participants be identified through any organisations other than the research sites listed above?

- Yes
- No

A76. Insurance/indemnity to meet potential legal liabilities

Note: In this question to NHS indemnity schemes include equivalent schemes provided by Health and Social Care (HSC) in Northern Ireland

A78-1. What arrangements will be made for insurance and/or indemnity to meet the potential legal liability of the sponsor(s) for harm to participants arising from the management of the research? Please tick box(es) as applicable.

- NHS indemnity scheme will apply (NHS sponsors only)
- Other insurance or indemnity arrangements will apply (give details below)

UCL Sponsorship

Please enclose a copy of relevant documents.

Date: 13/09/2016
A76.2. What arrangements will be made for insurance and/or indemnity to meet the potential legal liability of the sponsor(s) or employer(s) for harm to participants arising from the design of the research? Please tick box(es) as applicable.

Note: Where researchers with substantive NHS employment contracts have designed the research, indemnity is provided through NHS schemes. Indicate if this applies (there is no need to provide documentary evidence). For other protocol authors (e.g., company employees, university members), please describe the arrangements and provide evidence.

☐ NHS indemnity scheme will apply (protocol authors with NHS contracts only)
☐ Other insurance or indemnity arrangements will apply (give details below)

UCL Sponsorship

Please enclose a copy of relevant documents.

A76.3. What arrangements will be made for insurance and/or indemnity to meet the potential legal liability of investigators/collaborators arising from harm to participants in the conduct of the research?

Note: Where the participants are NHS patients, indemnity is provided through the NHS schemes or through professional indemnity. Indicate if this applies to the whole study (there is no need to provide documentary evidence). Where non-NHS sites are to be included in the research, including private practices, please describe the arrangements which will be made at these sites and provide evidence.

☐ NHS indemnity scheme or professional indemnity will apply (participants recruited at NHS sites only)
☐ Research includes non-NHS sites (give details of insurance/indemnity arrangements for these sites below)

The sponsorship and insurance are obtained through UCL

Please enclose a copy of relevant documents.

A78. Could the research lead to the development of a new product/process or the generation of intellectual property?

☐ Yes  ☐ No  ☐ Not sure

Part B: Section 5 – Use of newly obtained human tissue (or other human biological materials) for research purposes

1. What types of human tissue or other biological material will be included in the study?

Spontaneously expectorated sputum samples, nasopharyngeal swabs, stool, saliva and blood samples

2. Who will collect the samples?

Research doctor or research nurse at Royal Free Hospital

3. Who will the samples be removed from?

☐ Living donors
☐ The deceased

4. Will informed consent be obtained from living donors for use of the samples? Please tick as appropriate

☐ Yes  ☐ No

Date: 13/09/2016 29 202883/1017635/37/744
In future research?
- Yes  
- No  
- Not applicable

6. Will any tissues or cells be used for human application or to carry out testing for human application in this research?
- Yes  
- No

8. Will the samples be stored: [Tick as appropriate]
   - In fully anonymised form? (link to donor broken)
     - Yes  
     - No
   - In linked anonymised form? (linked to stored tissue but donor not identifiable to researchers)
     - Yes  
     - No

   If Yes, say who will have access to the code and personal information about the donor.
   Samples will be stored in linked anonymised form using sample number alone. Only the investigators of this study will have the original code for patient identifiable data. The personal information and link will be stored in a protected encrypted database.

   In a form in which the donor could be identifiable to researchers?
   - Yes  
   - No

9. What types of test or analysis will be carried out on the samples?
   - Bacteriological culture.
   - Antimicrobial susceptibility testing on the bacterial isolates.
   - Molecular diagnostic testing for viral respiratory pathogens
   - Molecular diagnostic testing for bacterial respiratory pathogens
   - Molecular diagnostic testing for antimicrobial resistance marker genes
   - Bacterial Whole Genome sequencing using next generation sequencing technology
   - 16S rRNA next generation sequencing
   - Biochemical tests on the plasma for inflammatory and immunological biomarkers

10. Will the research involve the analysis or use of human DNA in the samples?
    - Yes  
    - No

11. Is it possible that the research could produce findings of clinical significance for donors or their relatives?
    - Yes  
    - No

12. If so, will arrangements be made to notify the individuals concerned?
    - Yes  
    - No  
    - Not applicable

13. Give details of where the samples will be stored, who will have access and the custodial arrangements.
   Samples will be stored in the Centre For Clinical Microbiology located at the Royal Free Hospital. The department is locked and only accessible by electronic access. The stored samples will in pseudo-anonymised form by study.

Date: 13/09/2016
14. What will happen to the samples at the end of the research? Please tick all that apply and give further details.

- Transfer to research tissue bank

(if the bank is in England, Wales or Northern Ireland the institution will require a licence from the Human Tissue Authority to store relevant material for possible further research.)

- Storage by research team pending ethical approval for use in another project

(Unless the researcher's institution holds a storage licence from the Human Tissue Authority, or the tissue is stored in Scotland, or it is not relevant material, a further application for ethical review should be submitted before the end of this project)

- Storage by research team as part of a new research tissue bank

(The institution will require a licence from the Human Tissue Authority if the bank will be storing relevant material in England, Wales or Northern Ireland. A separate application for ethical review of the tissue bank may also be submitted.)

- Storage by research team of biological material which is not “relevant material” for the purposes of the Human Tissue Act

- Disposal in accordance with the Human Tissue Authority's Code of Practice

- Other

- Not yet known

Please give further details of the proposed arrangements:
PART C: Overview of research sites

Please enter details of the host organisations (Local Authority, NHS or other) in the UK that will be responsible for the research sites. For NHS sites, the host organisation is the Trust or Health Board. Where the research site is a primary care site, e.g. GP practice, please insert the host organisation (PCT or Health Board) in the Institution row and insert the research site (e.g. GP practice) in the Department row.

<table>
<thead>
<tr>
<th>Investigator Identifier</th>
<th>Research site</th>
<th>Investigator Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>IN1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NHS site</td>
<td>John</td>
</tr>
<tr>
<td></td>
<td>Non-NHS site</td>
<td>Hurst</td>
</tr>
<tr>
<td>Country: England</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organisation name</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Address</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post Code</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IN2</td>
<td></td>
<td>Timothy D. McHugh</td>
</tr>
<tr>
<td></td>
<td>NHS site</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-NHS site</td>
<td></td>
</tr>
<tr>
<td>Institution name</td>
<td>University College London, Royal Free campus</td>
<td></td>
</tr>
<tr>
<td>Department name</td>
<td>Centre for Clinical Microbiology, Div. of Infection and Immunity</td>
<td></td>
</tr>
<tr>
<td>Street address</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Town/city</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post Code</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Date: 13/09/2016
PART D: Declarations

D1. Declaration by Chief Investigator

1. The information in this form is accurate to the best of my knowledge and belief and I take full responsibility for it.

2. I undertake to abide by the ethical principles underlying the Declaration of Helsinki and good practice guidelines on the proper conduct of research.

3. If the research is approved I undertake to adhere to the study protocol, the terms of the full application as approved and any conditions set out by review bodies in giving approval.

4. I undertake to notify review bodies of substantial amendments to the protocol or the terms of the approved application, and to seek a favourable opinion from the main REC before implementing the amendment.

5. I undertake to submit annual progress reports setting out the progress of the research, as required by review bodies.

6. I am aware of my responsibility to be up to date and comply with the requirements of the law and relevant guidelines relating to security and confidentiality of patient or other personal data, including the need to register when necessary with the appropriate Data Protection Officer. I understand that I am not permitted to disclose identifiable data to third parties unless the disclosure has the consent of the data subject or, in the case of patient data in England and Wales, the disclosure is covered by the terms of an approval under Section 281 of the NHS Act 2006.

7. I understand that research records/data may be subject to inspection by review bodies for audit purposes if required.

8. I understand that any personal data in this application will be held by review bodies and their operational managers and that this will be managed according to the principles established in the Data Protection Act 1998.

9. I understand that the information contained in this application, any supporting documentation and all correspondence with review bodies or their operational managers relating to the application:

   - Will be held by the REC (where applicable) until at least 3 years after the end of the study; and by NHS R&D offices (where the research requires NHS management permission) in accordance with the NHS Code of Practice on Records Management.
   - May be disclosed to the operational managers of review bodies, or the appointing authority for the REC (where applicable), in order to check that the application has been processed correctly or to investigate any complaint.
   - May be seen by auditors appointed to undertake accreditation of RECs (where applicable).
   - Will be subject to the provisions of the Freedom of Information Acts and may be disclosed in response to requests made under the Acts except where statutory exemptions apply.
   - May be sent by email to REC members.

10. I understand that information relating to this research, including the contact details on this application, may be held on national research information systems, and that this will be managed according to the principles established in the Data Protection Act 1998.

11. Where the research is reviewed by a REC within the UK Health Departments Research Ethics Service, I understand that the summary of this study will be published on the website of the National Research Ethics Service (NRES), together with the contact point for enquiries named below. Publication will take place no earlier than 3 months after issue of the ethics committee’s final opinion or the withdrawal of the application.

Contact point for publication (Not applicable for R&D Forms)

NRES would like to include a contact point with the published summary of the study for those wishing to seek further information. We would be grateful if you would indicate one of the contact points below.

☐ Chief Investigator

Date: 13/09/2016
Access to application for training purposes (Not applicable for R&D Forms)
Optional – please tick as appropriate:

☐ I would be content for members of other RECs to have access to the information in the application in confidence for training purposes. All personal identifiers and references to sponsors, funders and research units would be removed.

This section was signed electronically by Dr Timothy Daniel McHugh on 07/10/2016 12:40.

Job Title/Post: Professor
Organisation: UCL
Email: [redacted]
D2. Declaration by the sponsor’s representative

If there is more than one sponsor, this declaration should be signed on behalf of the co-sponsors by a representative of the lead sponsor named at A64-1.

I confirm that:

1. This research proposal has been discussed with the Chief Investigator and agreement in principle to sponsor the research is in place.

2. An appropriate process of scientific critique has demonstrated that this research proposal is worthwhile and of high scientific quality.

3. Any necessary indemnity or insurance arrangements, as described in question A76, will be in place before this research starts. Insurance and indemnity policies will be renewed for the duration of the study where necessary.

4. Arrangements will be in place before the study starts for the research team to access resources and support to deliver the research as proposed.

5. Arrangements to allocate responsibilities for the management, monitoring and reporting of the research will be in place before the research starts.

6. The duties of sponsors set out in the Research Governance Framework for Health and Social Care will be undertaken in relation to this research.

   Please note: The declarations below do not form part of the application for approval above. They will not be considered by the Research Ethics Committee.

7. Where the research is reviewed by a REC within the UK Health Departments Research Ethics Service. I understand that the summary of this study will be published on the website of the National Research Ethics Service (NRES), together with the contact point for enquiries named in this application. Publication will take place no earlier than 3 months after issue of the ethics committee’s final opinion or the withdrawal of the application.

8. Specifically, for submissions to the Research Ethics Committees (RECs) I declare that any and all clinical trials approved by the HRA since 30th September 2013 (as defined on IRAS categories as clinical trials of medicines, devices, combination of medicines and devices or other clinical trials) have been registered on a publically accessible register in compliance with the HRA registration requirements for the UK, or that any deferral granted by the HRA still applies.

This section was signed electronically by Miss Tabitha Kavoli on 12/10/2016 09:51.

Job Title/Post: Research Management and Governance Manager

Organisation: University College London

Email:

Date: 13/09/2016
D3. Declaration for student projects by academic supervisor(s)

1. I have read and approved both the research proposal and this application. I am satisfied that the scientific content of the research is satisfactory for an educational qualification at this level.

2. I undertake to fulfill the responsibilities of the supervisor for this study as set out in the Research Governance Framework for Health and Social Care.

3. I take responsibility for ensuring that this study is conducted in accordance with the ethical principles underlying the Declaration of Helsinki and good practice guidelines on the proper conduct of research, in conjunction with clinical supervisors as appropriate.

4. I take responsibility for ensuring that the applicant is up to date and complies with the requirements of the law and relevant guidelines relating to security and confidentiality of patient and other personal data, in conjunction with clinical supervisors as appropriate.

Academic supervisor 1
This section was signed electronically by Mr John Hurst on 10/10/2016 17:21.

Job Title/Post: Reader / Consultant
Organisation: UCL
Email: 

Academic supervisor 2
This section was signed electronically by Dr Timothy Daniel McHugh on 07/10/2016 12:39.

Job Title/Post: Professor
Organisation: UCL
Email: 

Date: 13/09/2016
Appendix 1.5. Symptom Detailing Diary Card

<table>
<thead>
<tr>
<th>DATE</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHANGE in Symptoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHANGE in Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seek Medical advice</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adverse Events</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DATE</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHANGE in Symptoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHANGE in Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seek Medical advice</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adverse Events</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Please state any change in; (1) sputum volume, (2) sputum purulence (3) sputum viscosity, (4) dyspnoea, (5) worse cough, (6) chest wheezes, (7) sore throat, (8) fever, (9) nasal congestion, or (10) running nose by just recording the number given to each respiratory symptom in CHANGE in Symptoms cells every day and for the whole period when you are feeling unwell. If there is no change in your baseline respiratory symptoms just record “NO new symptoms”.
- Please state any Extra or New medicine taken every day; if it is just your routine treatment record “NO change in medication” in the CHANGE in Treatment cells.
- Please record if you seek any medical advice for your new respiratory symptoms from hospital clinic, a doctor or a nurse in the GP practice, a pharmacist or hospital emergency department in “Seek Medical advice” cells every day otherwise just record “NO”.
- Please record any Adverse event (Side effect) you may think it is related to your medication.
Changes in Microbiome with Treatment in Chronic Lung Diseases Study

<table>
<thead>
<tr>
<th>DATE</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHANGE in Symptoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHANGE in Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seek Medical advice</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adverse Events</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DATE</th>
<th>25</th>
<th>26</th>
<th>27</th>
<th>28</th>
<th>29</th>
<th>30</th>
<th>31</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHANGE in Symptoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHANGE in Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seek Medical advice</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adverse Events</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Please state any Change in; (1) sputum volume, (2) sputum purulence (3) sputum viscosity, (4) dyspnoea, (5) worse cough, (6) chest wheezes, (7) sore throat, (8) fever, (9) nasal congestion, or (10) running nose by just recording the number given to each respiratory symptom in CHANGE in Symptoms cells every day and for the whole period when you are feeling unwell. If there is no change in your baseline respiratory symptoms just record “NO new symptoms”.
- Please state any Extra or New medicine taken every day; if it is just your routine treatment record “NO change in medication” in the CHANGE in Treatment cells.
- Please record if you seek any medical advice for your new respiratory symptoms from hospital clinic, a doctor or a nurse in the GP practice, a pharmacist or hospital emergency department in “Seek Medical advice” cells every day otherwise just record “NO”.
- Please record any Adverse event (Side effect) you may think it is related to your medication.
Appendix 2. Study Protocol of CVID Patients

Studying the microbiome of antibody deficient patients

Background
Microbial antibiotic resistance is one of the most urgent issues threatening the provision of medical care in the twenty-first century [1, 2]. Underlying factors include the inappropriate, widespread use of antibiotics together with transmission of resistant organisms between vulnerable patients in clinical settings and horizontal transmission of resistance genes between bacteria in colonised hosts [2, 3].

The Clinical Immunology Department at the Royal Free Hospital, based within the Institute for Immunity and Transplantation, London, is a national referral centre for patients with primary (genetic) and secondary (often iatrogenic) immunodeficiency. The department manages 395 patients on immunoglobulin replacement, including 170 people who attend the department for intravenous infusions (IVIG) every three or four weeks in an open bay. It has previously been demonstrated that transmission of organisms (specifically Mycoplasma amphilforme) occurs between immunodeficient patients within this clinical service [4].

Antibody-deficient patients remain vulnerable to upper and lower respiratory tract infection despite immunoglobulin replacement [5]. The majority of patients receive treatment courses of antibiotics frequently, usually on the basis of clinical symptoms alone and often as self-medication with ‘rescue packs’ of medication. A recent daily diary card exercise in 69 patients with Common Variable Immunodeficiency (CVID) identified 76 antibiotic-treated respiratory exacerbations over a three month period. Despite this intensive antibiotic usage, in a parallel study we identified a pathogenic virus in 56% of symptomatic respiratory exacerbations while pathogenic bacteria (including colonising organisms) were identified in only 33%.

Furthermore, a high percentage (currently 59% of Common Variable Immunodeficiency patients) receive prophylactic antibiotics of various classes (predominantly macrolides, tetracyclines, quinolones and co-trimoxazole).

It has recently been demonstrated that immunocompromised patients are permissive hosts for pathogen evolution [6] and this may be expected to occur more rapidly in response to antimicrobial challenge. This could have considerable public health importance if resistant pathogens are then transmitted within the department and onwards to the community.

The respiratory bacterial flora in antibody-deficient patients has not, to our knowledge, been systematically studied before. We plan to explore the related issues of microbial antibacterial resistance, inter-patient pathogen transmission and appropriateness of therapy within this antibiotic-experienced population.

We believe that the data generated will be relevant to many other intensive clinical settings where immunologically vulnerable patients regularly attend hospital. If there is evidence of widespread pathogen resistance and transmission, this may inform novel strategies for treating or preventing infection in these patients, including nebulised immunoglobulin and host-directed strategies to prevent bronchiectasis. We also hope to be able to identify changes in microbiome which predict imminent infection.
Aims and Hypotheses

Aim 1: To investigate the background microbial flora in antibody-deficient patients and the presence of antibiotic resistance in resident organisms.

Hypothesis 1: The widespread use of antibiotics for prophylaxis and treatment in patients with hypogammaglobulinemia is associated with resistance in potentially pathogenic organisms.

Aim 2: To investigate within-host bacterial evolution in the absence of an effective immune response, especially in response to antibiotic therapy.

Hypothesis 2: Antibiotic usage directly leads to bacterial evolution, evident at genome level, within immunocompromised hosts.

Aim 3: To investigate the transmission of pathogenic organisms between immunocompromised hosts in regular casual contact and to close contacts in the community.

Hypothesis 3: Pathogenic and resistant organisms are transmitted between patients who regularly attend the Royal Free Hospital Clinical Immunology department and to household contacts.

Aim 4: To investigate the ability to predict imminent symptomatic infection from changes in microbiome.

Hypothesis 4: Changes in the respiratory tract microbial flora in antibody-deficient patients predict the imminent development of infection.

Methods

Inclusion criteria:

- Age over 18 years
- Regularly infuse IV/G at the Royal Free Hospital
- Consent to research (V7 or V8 protocol)
- Agree to provide samples as requested

Participating patients will be given an information sheet (Appendix A) and asked to:

- Complete a baseline questionnaire (Appendix B)
- At each visit provide saliva, stool and – if possible – sputum
- At each visit complete a brief questionnaires (Appendix C)
- If they have symptomatic infection between infusions to contact the department and provide additional samples: saliva, sputum and naso-pharyngeal swab for viral PCR

An aliquot of sputum will be processed for culture and sensitivity as per routine diagnostic protocol.

A further aliquot of sputum and all other samples will be stored at -80 C for future analysis of microbiome. If a pathogenic organism is recurrently isolated from a single individual then we will store it for future DNA sequencing of each isolate over time.
References

1.) Khabbaz RF, Moseley RR, Steiner RJ, Levitt AM, Bell BP. Challenges of infectious diseases in the USA. Lancet 2014; 384(9937):53-63

Appendix A – Information Sheet

Studying the microbiome of antibody-deficient patients

Many thanks for considering taking part in this study.

Why are we doing this study?
We are interested in the ‘microbiome’ (i.e. the bacteria which normally live in our bodies) in patients with antibody deficiency. In particular, we want to see whether the bacteria – especially those which can cause infection – are affected by the antibiotics you take and whether bacteria are passed between patients having infusions at the hospital.

We also want to see whether bacteria ‘evolve’ within your bodies because your immune system is unable to eliminate them effectively. Finally we want to see whether we can predict imminent respiratory infection from changes in the microbiome beforehand.

What will happen if I decide to take part?
If you decide to take part you will be asked to complete Part A of the short questionnaire attached (Part B is for the doctor). We will then ask you to provide saliva, stool and, if possible, sputum at each infusion. You can produce the samples at the hospital or beforehand at home and bring them with you. You will also be asked to complete a brief questionnaire at each visit.

If you develop a respiratory infection we will ask you to get in touch with us and to provide more samples. You can come in to the hospital to do this or send them from home.

Are there any risks?
No, there should not be any risks from giving the samples.

Will I see the results?
If you want to see the results of the study, please let the clinical team know and they will discuss the details with you once available (note that this could take a year or more). The study will only generate results for the patient population as a whole. If we find any results of clinical significance for you as an individual we will let you know straight away.
Appendix B – Baseline questionnaire

Part A – for the patient to complete

1. How often do you come for infusions? (Please circle)

Every 2 weeks          Every 3 weeks          Every 4 weeks          Other (specify)............

2. How often do you have respiratory infections? (These include colds, sinusitis, sore throat and bronchitis. Please tick one box)

At least once a month       □
Less that once a month but at least 6 times per year □
About 4 or 5 times per year □
About 2 or 3 times per year □
About once a year          □
Less than once a year      □

3. How often do you have diarrhoea? (Please tick one box)

Always
Not always but at least 15 days per month □
Less than 15 days but more than 7 days per month □
At least once per month but less than 7 days per month □
Less than once per month but several times per year □
Rarely or never              □

Please provide additional details if necessary: ..........................................................
........................................................................................................................................

4. Do you frequently suffer from other types of infection? (Please circle)

Yes                   No

If yes, what sorts of infection? ..............................................................
........................................................................................................................................

5. Are you on a long-term (prophylactic) antibiotic? (Please circle)

Yes                   No
If yes, which antibiotic and how often? .................................................................

6. How often do you take courses of antibiotics in a year (in addition to the long-term antibiotic if you are taking one)? (Please tick one box)

At least once a month ☐
Less than once a month but at least 6 times per year ☐
About 4 or 5 times per year ☐
About 2 or 3 times per year ☐
About once a year ☐
Less than once a year ☐

7. Approximately when did you last take a course of antibiotics (apart from the long-term antibiotic, if you are taking one)?

....................................................................................................................................

....................................................................................................................................

8. Do you usually have a cough? (Please circle)

Yes ☐ No ☐

If yes, do you usually produce sputum (phlegm)? (Please circle)

Yes ☐ No ☐

9. Do you live with other people? (Please circle)

Yes ☐ No ☐

If yes, please specify how many adults? ...........

And how many children? ...........
Part B – to be completed by the doctor

Immunodeficiency diagnosis

Other diagnoses

Bronchiectasis? Yes  No  Unknown

Positive sputum microbiology in last year?

Positive respiratory virus PCR in last year?

Enteric pathogen identified in last year?

Immunoglobulin product and dose

Year started immunoglobulin

Number of years attending RFH for infusions

Current medications

Confirm consented for research  □
Appendix C – interim questionnaire

Since your last visit:

1. Have you had any respiratory infections? (Please circle)
   
   Yes  No

   If yes, which of the following symptoms have you had? (Please tick all that apply, using more than one column if you have had more than one infection. Please do not record long-term or ‘usual’ symptoms, only symptoms which were new or different with the infection)

<table>
<thead>
<tr>
<th></th>
<th>First infection</th>
<th>Second infection</th>
<th>Third infection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cough</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sputum (phlegm) production</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shortness of breath</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fever</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blocked or runny nose</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Sore throat</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

   About how long did the infection last or has it lasted so far? …… Days …… Days …… Days

   Are the infection symptoms still present? (Please circle)
   
   Yes  No

   Please give further details if required ………………………………………………………………………………………………………………..

   ………………………………………………………………………………………………………………..

   ………………………………………………………………………………………………………………..

   Did you contact the hospital team? Yes  No

   Did you provide additional samples? Yes  No
2. Have you had any new diarrhoea or a change in diarrhoea symptoms? (Please circle)
   Yes   No

   If yes, please give further details .................................................................
   .......................................................................................................................
   .......................................................................................................................
   .......................................................................................................................

3. Have you started, stopped or changed a long-term (prophylactic) antibiotic? (Please circle)
   Yes   No

   If yes, please give further details .................................................................
   .......................................................................................................................

4. Have you taken any additional or ‘breakthrough’ antibiotics? (Please circle)
   Yes   No

   If yes, which antibiotic(s)? ...........................................................................

   For what reason(s)? ......................................................................................

   For how many days? ......................................................................................

   Are you still taking additional or ‘breakthrough’ antibiotics today? (Please circle)
   Yes   No

   If yes, which antibiotic(s)? ...........................................................................

   How many days do you plan to take them for in total? .................................

For the nurse / doctor:

Please tick if collected

   Saliva sample   
   Stool sample   
   Sputum sample
Appendix 3 Supplementary Methods

Appendix 3.1: Patient’s detailed advice sheet on Sputum Samples’ Collection

How to collect a sputum sample

Preparation
- If you are able, please collect a sputum sample on the same day of your appointment; preferably the first thing in the morning; and bring it with you to the clinic.
- **Sputum is phlegm that comes from the lungs.** It is produced by a deep cough from the chest.
- **Sputum is not the same as saliva.** For our purpose here, we want to collect sputum.
- Do not eat, drink, smoke or brush your teeth right before you collect the sample.

Collection
- Open the two plastic pots without touching the inside.
- Breathe in deeply three times.
- Next, cough deeply from the chest to bring up sputum.
- **Try to cough up at least one teaspoon full of sputum for each pot.**
- Spit the sputum into the pots.
- Close their caps tightly.
- Place the pot in the provided plastic bag
- Keep the specimen in the fridge (at 2-8°C) until being transferred to the lab.
Appendix 3.2: Preparation of Artificial Sputum Base

Dr. Isobella Honeyborne (UCL CCM internal document)

Modified protocol from the method developed by Dinesh et al (Nature Protocol Exchange).

<table>
<thead>
<tr>
<th>Material</th>
<th>Amount</th>
<th>Cat #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distilled water</td>
<td>400 mL</td>
<td></td>
</tr>
<tr>
<td>Mucin from porcine stomach</td>
<td>2.5 g</td>
<td></td>
</tr>
<tr>
<td>DTPA*</td>
<td>3.0 g</td>
<td>D6518-10G</td>
</tr>
<tr>
<td>Sodium chloride</td>
<td>2.5 g</td>
<td>S/3160/63</td>
</tr>
<tr>
<td>Potassium chloride</td>
<td>1.1 g</td>
<td>P9541-500G</td>
</tr>
<tr>
<td>20 amino acids**</td>
<td>125 mg</td>
<td>09416-1E</td>
</tr>
</tbody>
</table>

Adjust pH to 7.0 using 2 M Tris base.

After autoclaving and leaving to cool, add

| Egg yolk emulsion         | 5 mL     | 536262U        |

After preparation, store the ASM in the fridge

* DTPA diethylenetriaminepentaacetic acid

** add all amino acids except for trans-4-hydroxy-L-proline and L-cystine
Appendix 3.3: UCL Genomics 16S rRNA Metagenomic Profiling

Assay Protocol

UCL Genomics 16S Metagenomic Profiling Assay

Introduction
The following protocol details how to prepare and MiSeq sequence up to 96 16S amplicon libraries for the purpose of metagenomic profiling. The sequences used here target V5-V7, however, it should be possible to redesign them to target whichever region best suits your needs. This protocol is fully compatible with the Illumina MiSeq 16S workflow envoked through MiSeq Reporter v2.3 Libraries are created using a single (non-nested) PCR and are dual indexed with sequences identical to those used in standard Illumina dual indexed assays. A double-bind Ampure XP high stringency clean up is used to remove primer dimer before samples are quantitative, pooled, re-quantified and sequenced.

Materials Required
Primers (all HPLC purified) – full list of primer sequences including indexes can be found in the appendix

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>785F</td>
<td>AATGATACGGCGACACACCTAGTACAC(15 Index)ACGTACGTAGCTAGATACCCCCBGGTTAGTTC</td>
</tr>
<tr>
<td>1175R</td>
<td>CAGGCAAGGAACCCGCGATA(17 Index)AGTCAGTCAGCCACGTCRTGCCDCCTTGCCTC</td>
</tr>
<tr>
<td>Read1</td>
<td>ACGTACGTAGCTAGATACCCCCBGGTTAGTTC</td>
</tr>
<tr>
<td>i7Index</td>
<td>GAAAGAGGGCGGCAAGCAGGTGAAGTACGT</td>
</tr>
<tr>
<td>Read2</td>
<td>AGTCAGTCAGCCACGTCRTGCCDCCTTGCCTC</td>
</tr>
</tbody>
</table>

PCR
MolTag 16S Basic Master Mix, (VHBIO #0550400100 (Molzym))
PCR grade water (Various)
Ampure XP beads, (Beckman Coulter #)

High Stringency Cleanup
Elution buffer (10mM Tris-HCl pH 7.8 / 0.1% Tween 20) or Qiagen EB
PEG-8000 (Sigma #89510)
2.5M NaCl (Life Technologies #AM9760G)
80% Ethanol (Various)

Quality Control
Bioanalyzer DNA 1000 Assay (Agilent #5067-4626)

Before you begin
If not already done so, resuspend every primer to 100µM in Elution Buffer. Aliquot each out amplification primer into 2µL aliquots (PCR tube strip) to limit freeze-thaws. Aliquot each read primer into 5µL aliquots to limit freeze-thaws.

Program a thermo-cyler as follows, saving the protocol as 16S_AMP

1. 95°C for 5 minutes
2. 95°C for 30 seconds
3. 55°C for 40 seconds
4. 72°C for 60 seconds
5. Goto Step 2 an additional 29 times
6. 72°C for 10 mins
7. 4°C forever

Step 1 - Prepare and perform PCR reactions

1. Dilute required R&F amplification primers to 2.5µL by adding 78µL of PCR grade water to each 2µL aliquot
2. Aliquot 5ng of each sample into a well of a thin-walled PCR plate and bring the volume up to 11µL with PCR grade water

3. Align the R primer strip (12 tubes) on the top row of a plate rack. Use a multichannel to transfer 2µL into each row of samples on the PCR plate. Change tips each time

4. Align the F primer strip (8 tubes) on the left column of a plate rack. Use a multichannel to transfer 2µL into each column of samples on the PCR plate. Change tips each time

5. Use a multichannel to add 10µL of 2.5X MoIaq Basic Mastermix to each well. Mix by gentle pipetting after each addition and discard tips. The total volume of each well should now be 25µL.

6. Seal the plate with a PCR seal and spin the plate briefly to collect the PCR reaction to the bottom of the wells.

7. Place the plate on a preheated thermo-cycler running the 16S_AMP program.

8. After the PCR has completed, remove PCR from the block. Samples can be stored at -20°C or proceed to clean up.

9. Optionally, run 2µL of a selection of samples are run on a 1% agarose gel for 45 minutes at 120V to ensure the PCR was successful. You should see a band around 500bp.

**Step 2 - High Stringency Binding Buffer Preparation**

1. Dilute 15mL of 5M NaCl solution with 15mL DNase free water to create 30mL of 2.5M NaCl

2. Zero a balance containing a 50mL measuring cylinder

3. Weigh 5g of PEG-8000 directly into the measuring cylinder

4. Add 2.5M NaCl solution to approximately the 20mL mark on the cylinder, seal with parafilm and vortex gently until all the PEG has dissolved.

5. Top up the solution to 25mL with 2.5M NaCl and mix by inversion

6. Decant the binding buffer into a fresh DNase free 50mL Falcon tube

7. Store at 4°C along with the Ampure XP beads.

**Step 3 - High Stringency Clean up of PCR products**

1. 30 minutes before clean up, remove the binding buffer and Ampure XP beads from the fridge and allow to equilibrate to room temperature

2. Add 25µL elution buffer to each sample for a total volume of 50µL

3. Add 40µL of Ampure XP beads and mix by pipetting (>6 times)

4. Incubate at room temperature for 5 minutes

5. Pellet beads on the magnet (~2 minutes)

6. Remove and discard supernatant

7. Keep the tube/plate on the magnet and add 200µL freshly prepared 80% EtOH.

8. Incubate for 30 seconds to allow all the beads to re-pellet.
Appendix 3.4. Internal Standard Curve of *P. aeruginosa* and total bacterial load in each qPCR run

**P. aeruginosa**

A. Amplification plot

B. Standard curve

**16S rRNA**

C. Amplification plot

D. Standard curve

Figure 5: Amplification plots (A,C) and standard curves (B,D) of qPCR for *P. aeruginosa* and total bacterial load
# Appendix 3.5. Sample of MALDI-TOF Report

## Bruker MALDI Biotyper Identification Results

**Run Info:**
- **Run Identifier:** 170913-1248-100
- **Comment:**
- **Operator:** mic1@FLEX-PC
- **Run Creation Date/Time:** 2017-09-13T13:12:57.085
- **Number of Tests:** 25
- **Type:** Standard
- **BTS-QC:** not present
- **Instrument ID:** 254472.00140
- **Server Version:** 4.1.70 (PYTH) 48 2016-10-26_15-05-35

## Result Overview

<table>
<thead>
<tr>
<th>Sample Name</th>
<th>Sample ID</th>
<th>Organism (best match)</th>
<th>Score Value</th>
<th>Organism (second-best match)</th>
<th>Score Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>D3 (+++) (A)</td>
<td>782 hi</td>
<td>Streptococcus vestibularis</td>
<td>7.81</td>
<td>Streptococcus salivarius</td>
<td>1.95</td>
</tr>
<tr>
<td>D2 (+++) (A)</td>
<td>TD PS</td>
<td>Pseudomonas aeruginosa</td>
<td>7.31</td>
<td>Pseudomonas aeruginosa</td>
<td>7.31</td>
</tr>
<tr>
<td>D7 (+++) (A)</td>
<td>HP CLED 1</td>
<td>Escherichia coli</td>
<td>7.42</td>
<td>Escherichia coli</td>
<td>7.18</td>
</tr>
<tr>
<td>D9 (+++) (A)</td>
<td>HP CLED 2</td>
<td>Escherichia coli</td>
<td>7.42</td>
<td>Escherichia coli</td>
<td>7.41</td>
</tr>
<tr>
<td>D11 (+++) (B)</td>
<td>789 HI</td>
<td>Neisseria perflava</td>
<td>7.71</td>
<td>Neisseria subflava</td>
<td>7.18</td>
</tr>
<tr>
<td>E1 (-) (C)</td>
<td>790 HI</td>
<td>No Organism Identification Possible</td>
<td>1.87</td>
<td>No Organism Identification Possible</td>
<td>1.57</td>
</tr>
<tr>
<td>E3 (+++) (A)</td>
<td>781 PR</td>
<td>Proteus mirabilis</td>
<td>7.27</td>
<td>Proteus mirabilis</td>
<td>7.09</td>
</tr>
</tbody>
</table>

*Result overview table—continued on next page*
<table>
<thead>
<tr>
<th>Sample Name</th>
<th>Sample ID</th>
<th>Organism (best match)</th>
<th>Score Value</th>
<th>Organism (second-best match)</th>
<th>Score Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>F6 (++) (A)</td>
<td>781 CLED GR (standard)</td>
<td>Escherichia coli</td>
<td>1.98</td>
<td>Escherichia coli</td>
<td>1.42</td>
</tr>
<tr>
<td>F7 ++ (A)</td>
<td>782 B PS (standard)</td>
<td>Pseudomonas aeruginosa</td>
<td>1.21</td>
<td>Pseudomonas aeruginosa</td>
<td>2.77</td>
</tr>
<tr>
<td>F9 (-) (C)</td>
<td>782B CLED (standard)</td>
<td>No Organism Identification Possible</td>
<td>1.00</td>
<td>No Organism Identification Possible</td>
<td>1.00</td>
</tr>
<tr>
<td>F11 (++) (A)</td>
<td>782B CLED YEL (standard)</td>
<td>Rothia mucilaginosa</td>
<td>1.26</td>
<td>Rothia mucilaginosa</td>
<td>2.21</td>
</tr>
<tr>
<td>F7 (++) (A)</td>
<td>782B CLED GREY (standard)</td>
<td>Rothia mucilaginosa</td>
<td>1.15</td>
<td>Rothia mucilaginosa</td>
<td>2.62</td>
</tr>
<tr>
<td>F4 (++) (A)</td>
<td>792 CLED WH (standard)</td>
<td>Candida albicans</td>
<td>2.13</td>
<td>Candida albicans</td>
<td>2.61</td>
</tr>
<tr>
<td>F6 (-) (C)</td>
<td>791 CLED (standard)</td>
<td>No peaks found</td>
<td>0.00</td>
<td>No peaks found</td>
<td>0.00</td>
</tr>
<tr>
<td>F7 (--) (C)</td>
<td>GB CLED WH BIG (standard)</td>
<td>No peaks found</td>
<td>0.00</td>
<td>No peaks found</td>
<td>0.00</td>
</tr>
<tr>
<td>F9 (++) (A)</td>
<td>GB YEL 5T (standard)</td>
<td>Rothia mucilaginosa</td>
<td>1.32</td>
<td>Rothia mucilaginosa</td>
<td>2.04</td>
</tr>
<tr>
<td>F11 (--) (C)</td>
<td>GB CLED YEL SM (standard)</td>
<td>No peaks found</td>
<td>0.00</td>
<td>No peaks found</td>
<td>0.00</td>
</tr>
<tr>
<td>G1 (+++) (A)</td>
<td>782 B CH HI (standard)</td>
<td>Haemophilus influenzae</td>
<td>1.72</td>
<td>Haemophilus influenzae</td>
<td>2.31</td>
</tr>
<tr>
<td>G3 (+++) (B)</td>
<td>782B 3AS (standard)</td>
<td>Neisseria subflava</td>
<td>1.12</td>
<td>Neisseria perflava</td>
<td>1.12</td>
</tr>
<tr>
<td>G5 (+) (A)</td>
<td>782B P5 (standard)</td>
<td>Pseudomonas aeruginosa</td>
<td>1.45</td>
<td>Pseudomonas aeruginosa</td>
<td>2.45</td>
</tr>
<tr>
<td>G7 (+) (B)</td>
<td>782B (standard)</td>
<td>Rothia mucilaginosa</td>
<td>1.72</td>
<td>Rothia mucilaginosa</td>
<td>1.72</td>
</tr>
<tr>
<td>G9 (-) (C)</td>
<td>CNA WH (standard)</td>
<td>No peaks found</td>
<td>0.00</td>
<td>No peaks found</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Result overview table—continued on next page.
### Result overview table—continued from previous page

<table>
<thead>
<tr>
<th>Sample Name</th>
<th>Sample ID</th>
<th>Organism (best match)</th>
<th>Score Value</th>
<th>Organism (second-best match)</th>
<th>Score Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1(L) (C)</td>
<td>GB CH TRANSGB CH (standard)</td>
<td>No Organism Identification Possible</td>
<td>1.00</td>
<td>No Organism Identification Possible</td>
<td>1.00</td>
</tr>
<tr>
<td>H1 (C)</td>
<td>784 PS PDR (standard)</td>
<td>no peaks found</td>
<td>0.00</td>
<td>no peaks found</td>
<td>0.00</td>
</tr>
<tr>
<td>H3 (++++) (A)</td>
<td>HP ORIG PR. (standard)</td>
<td>Proteus mirabilis</td>
<td>2.17</td>
<td>Proteus mirabilis</td>
<td>2.17</td>
</tr>
</tbody>
</table>

### Meaning of Score Values

<table>
<thead>
<tr>
<th>Range</th>
<th>Interpretation</th>
<th>Symbols</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.00 - 3.00</td>
<td>High-confidence identification</td>
<td>(++++)</td>
<td>green</td>
</tr>
<tr>
<td>1.70 - 1.99</td>
<td>Low-confidence identification</td>
<td>(+)</td>
<td>yellow</td>
</tr>
<tr>
<td>0.00 - 1.69</td>
<td>No Organism Identification Possible</td>
<td>(−)</td>
<td>red</td>
</tr>
</tbody>
</table>

### Meaning of Consistency Categories (A - C)

<table>
<thead>
<tr>
<th>Category</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A)</td>
<td><strong>High consistency</strong>: The best match is a high-confidence identification. The second-best match is (1) a high-confidence identification in which the species is identical to the best match, (2) a low-confidence identification in which the species or genus is identical to the best match, or (3) a non-identification.</td>
</tr>
<tr>
<td>(B)</td>
<td><strong>Low consistency</strong>: The requirements for high consistency are not met. The best match is a high- or low-confidence identification. The second-best match is (1) a high- or low-confidence identification in which the genus is identical to the best match or (2) a non-identification.</td>
</tr>
<tr>
<td>(C)</td>
<td><strong>No consistency</strong>: The requirements for high or low consistency are not met.</td>
</tr>
</tbody>
</table>
Appendix 3.6. Coverage of 16S rRNA qPCR (A) primers: Bact340F and Bact806R and (B) Probe

A.
B. 16S rRNA qPCR Probe
Appendix 3.7. Specificity of the typical respiratory qPCR primers and probes to the corresponding targets through BLAST Analysis

A. *H. influenzae* qPCR forward and reverse primers’ specificity through BLAST Analysis

Detailed primer reports

**Primer pair 1**

<table>
<thead>
<tr>
<th></th>
<th>Sequence (5'→3')</th>
<th>Length</th>
<th>Tm</th>
<th>GC%</th>
<th>Self complementarity</th>
<th>Self 3' complementarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward primer</td>
<td>CCGGTGTCGGTAGGATTAAAT</td>
<td>23</td>
<td>58.87</td>
<td>43.48</td>
<td>5.00</td>
<td>2.00</td>
</tr>
<tr>
<td>Reverse primer</td>
<td>CTGATTTCAGTCTGTTCCTTTGC</td>
<td>24</td>
<td>59.56</td>
<td>41.67</td>
<td>4.00</td>
<td>2.00</td>
</tr>
</tbody>
</table>

Products on intended target

Products on allowed transcript variants

Products on potentially unintended templates

Products on target templates

>NC_009071 Haemophilus influenzae Rd KW20 chromosome, complete genome

product length = 91
Forward primer 1 CCGGTGTCGGTAGGATTAAAT 23
Template 737791 737812
Reverse primer 1 CTGATTTCAGTCTGTTCCTTTGC 24
Template 737881 737858
B. *H. influenzae* qPCR probe (reverse complement) and reverse primer Specificity through BLAST Analysis

<table>
<thead>
<tr>
<th>Primer-BLAST+ JOB ID: AQyewu14Z3Go_w9bfl1zdqahxW7oA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primer-BLAST Results</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>Input PCR template</td>
</tr>
<tr>
<td>none</td>
</tr>
<tr>
<td>Specificity of primers</td>
</tr>
<tr>
<td>Target templates were found in selected database: RefSeq Representative Genome Database</td>
</tr>
<tr>
<td>Other reports</td>
</tr>
<tr>
<td>Search Summary</td>
</tr>
</tbody>
</table>

## Detailed primer reports

### Primer pair 1

<table>
<thead>
<tr>
<th></th>
<th>Sequence (5'-&gt;3')</th>
<th>Length</th>
<th>Tm</th>
<th>GC%</th>
<th>Self complementarity</th>
<th>Self 3' complementarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward primer</td>
<td>ACAGGCCACAACGGTAAAGTGTTCCTACGT</td>
<td>28</td>
<td>66.57</td>
<td>46.43</td>
<td>4.00</td>
<td></td>
</tr>
<tr>
<td>Reverse primer</td>
<td>CTGATTTTTCAGTGCTGTCTTTGC</td>
<td>24</td>
<td>59.56</td>
<td>41.67</td>
<td>2.00</td>
<td></td>
</tr>
</tbody>
</table>

- Products on intended target
- Products on allowed transcript variants
- Products on potentially unintended templates
- Products on target templates

> NC_000907.1 Haemophilus influenzae Rd KW20 chromosome, complete genome

<table>
<thead>
<tr>
<th>product length</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward primer 1</td>
<td>ACAGGCCACAACGGTAAAGTGTTCCTACGT</td>
</tr>
<tr>
<td>Template</td>
<td>737822</td>
</tr>
<tr>
<td>Reverse primer 1</td>
<td>CTGATTTTTCAGTGCTGTCTTTGC</td>
</tr>
<tr>
<td>Template</td>
<td>737861</td>
</tr>
</tbody>
</table>

BLAST is a registered trademark of the National Library of Medicine

[Support center Mailing list YouTube](https://www.ncbi.nlm.nih.gov/books/blastprimerblastprimertool.cgi?bc_emes=1547164510&job_key=AQyewu14Z3Go_w9bfl1zdqahxW7oA)
C. S. pneumoniae qPCR forward and reverse primers’ specificity through BLAST Analysis

Detailed primer reports

Primer pair 1

<table>
<thead>
<tr>
<th>Sequence (5'-&gt;3')</th>
<th>Length</th>
<th>Tm</th>
<th>GC%</th>
<th>Self complementarity</th>
<th>Self 3' complementarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward primer</td>
<td>AGTCGTTCAAGGAACAAAGCTT</td>
<td>23</td>
<td>59.61</td>
<td>43.48</td>
<td>4.00</td>
</tr>
<tr>
<td>Reverse primer</td>
<td>ACCAAGCTGACGACCTCTTT</td>
<td>20</td>
<td>58.87</td>
<td>50.00</td>
<td>4.00</td>
</tr>
</tbody>
</table>

Products on intended target

Products on allowed transcript variants

Products on potentially unintended templates

Products on target templates

>NC_003098.1 Streptococcus pneumoniae R6 chromosome, complete genome

<table>
<thead>
<tr>
<th>product length</th>
<th>= 156</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward primer 1</td>
<td>AGTCGTTCAAGGAACAAAGCTT</td>
</tr>
<tr>
<td>Template</td>
<td>558025</td>
</tr>
<tr>
<td>Reverse primer 1</td>
<td>ACCAAGCTGACGACCTCTTT</td>
</tr>
<tr>
<td>Template</td>
<td>558160</td>
</tr>
</tbody>
</table>
D. S. pneumoniae qPCR forward primer and probe specificity through BLAST Analysis

Detailed primer reports

Primer pair 1

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5'-&gt;3')</th>
<th>Length</th>
<th>Tm</th>
<th>GC%</th>
<th>Self complementarity</th>
<th>Self 3' complementarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward</td>
<td>AGTCGGTCCAAGTAAAGGTCT</td>
<td>23</td>
<td>59.61</td>
<td>43.48</td>
<td>4.00</td>
<td>3.00</td>
</tr>
<tr>
<td>Reverse</td>
<td>TACATGTAGGAACACTTTTCTCACAAA</td>
<td>29</td>
<td>59.37</td>
<td>31.03</td>
<td>14.00</td>
<td>10.00</td>
</tr>
</tbody>
</table>

Products on intended target

Products on allowed transcript variants

Products on potentially unintended templates

>NC_003098.1 Streptococcus pneumoniae R6 chromosome, complete genome

product length = 134
Forward primer 1: AGTCGGTCCAAGTAAAGGTCT 23
Template: 558025
Reverse primer 1: TACATGTAGGAACACTTTTCTCACAAA 29
Template: 558150

BLAST is a registered trademark of the National Library of Medicine
Support center  Mailing list  YouTube

E. *M. catarrhalis* qPCR forward and reverse primers’ specificity through BLAST Analysis

- NCBI Home
- Sign in to NCBI
- Skip to Main Content
- Skip to Navigation
- About NCBI Accesskeys

**Primer-BLAST** > JOB ID:SUOXbC--1hYFKDgtNU0cH09WDS1iRRYwYw
Primer-BLAST Results

[?] Input PCR template
none
Specificity of primers
Target templates were found in selected database: RefSeq Representative Genome Database (Organism limited to Bacteria)
Other reports
Search Summary

**Detailed primer reports**

You can re-search for specific primers by accepting some of the unintended targets, check the box(es) next to the ones you accept and try again to re-search for specific primers.

[2] Primer pair 1

<table>
<thead>
<tr>
<th></th>
<th>Sequence (5'-&gt;3')</th>
<th>Length</th>
<th>Tm</th>
<th>GC%</th>
<th>Self complementarity</th>
<th>Self 3' complementarity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Forward</strong></td>
<td>GTGAGTGGCGGCTTTTACAACC</td>
<td>21</td>
<td>60.33</td>
<td>62.38</td>
<td>4.00</td>
<td></td>
</tr>
<tr>
<td><strong>Reverse</strong></td>
<td>TGTATCGCGCTGCCAGACAA</td>
<td>20</td>
<td>59.68</td>
<td>50.00</td>
<td>5.00</td>
<td>3.00</td>
</tr>
</tbody>
</table>

Products on intended targets

Products on allowed targets

Products on allowed transcript variants

Products on potentially unintended templates

Products on target templates

>NC_014147.1 Moraxella catarrhalis BBH18, complete genome

<table>
<thead>
<tr>
<th>product length = 72</th>
<th>Forward primer</th>
<th>1</th>
<th>CTGAGTGGCGGCTTTTACAACC</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template</td>
<td>511328</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reverse primer</td>
<td>1</td>
<td></td>
<td>TGTATCGCGCTGCCAGACAA</td>
<td>20</td>
</tr>
<tr>
<td>Template</td>
<td>511399</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
F. M. catarrhalis qPCR probe and reverse primers’ specificity through BLAST Analysis

- NCBI Home
- Sign in to NCBI
- Skip to Main Content
- Skip to Navigation
- About NCBI Accessions

U.S. National Library of Medicine
NCBI National Center for Biotechnology Information

- My NCBI
- Sign in to NCBI
- Register
- Sign Out

Primer-BLAST® JOB ID: vbdjmN/91RxK88awkZrHLhV-s6VRuEz1A
Primer-BLAST Results
[2]

Input PCR template
none
Specificity of primers
Target templates were found in selected database: RefSeq Representative Genome Database (Organism limited to Bacteria)
Other reports Search Summary

Detailed primer reports

You can re-search for specific primers by accepting some of the unintended targets, check the box(es) next to the ones you accept and try again to re-search for specific primers [Submit]
[2]

Primer pair 1

<table>
<thead>
<tr>
<th>Sequence (5'→3')</th>
<th>Length</th>
<th>Tm</th>
<th>GC%</th>
<th>Self complementarity</th>
<th>Self 3' complementarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward primer</td>
<td>TGCTTTTGCACTGTGGATGCAGGCTAA</td>
<td>27</td>
<td>67.38</td>
<td>48.15</td>
<td>7.00</td>
</tr>
<tr>
<td>Reverse primer</td>
<td>TGTATCGCCTGCGCAAGACAA</td>
<td>20</td>
<td>59.68</td>
<td>50.00</td>
<td>5.00</td>
</tr>
</tbody>
</table>

Products on intended targets

Products on allowed targets

Products on allowed transcript variants

Products on potentially unintended templates

Products on target templates

>NC_014147.1 Moraxella catarrhalis BBH18, complete genome

product length - 49
Forward primer 1 TGCTTTTGCACTGTGGATGCAGGCTAA 27
Template 511351 .......................... 511377
Reverse primer 1 TGTATCGCCTGCGCAAGACAA 20
Template 511399 ...........A............ 511380
**G. P. aeruginosa** qPCR forward and reverse primers’ specificity through BLAST Analysis

- Input PCR template: none
- Specificity of primers: Target templates were found in selected database: RefSeq Representative Genome Database
- Other reports: Search Summary

## Detailed primer reports

### Primer pair 1

<table>
<thead>
<tr>
<th>Sequence (5′-3′)</th>
<th>Length</th>
<th>Tm</th>
<th>GC%</th>
<th>Self complementarity</th>
<th>Self 3′ complementarity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Forward primer</strong></td>
<td>TGCTGGTGCCACAGGACAT</td>
<td>19</td>
<td>61.83</td>
<td>57.89</td>
<td>5.00</td>
</tr>
<tr>
<td><strong>Reverse primer</strong></td>
<td>TTTGGTTGGCAGTTCTCTATTG</td>
<td>22</td>
<td>59.64</td>
<td>45.45</td>
<td>4.00</td>
</tr>
</tbody>
</table>

Products on intended target

Products on allowed transcript variants

Products on potentially unintended templates

Products on target templates

>**NC_002516.2** Pseudomonas aeruginosa PA01 chromosome, complete genome

**Product length = 65**

Features associated with this product:

- **Transcriptional regulator ToolR**

Forward primer 1

<table>
<thead>
<tr>
<th>Sequence (5′-3′)</th>
<th>Length</th>
<th>Tm</th>
<th>GC%</th>
<th>Self complementarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template</td>
<td>TGCTGGTGCCACAGGACAT</td>
<td>19</td>
<td>61.83</td>
<td>57.89</td>
</tr>
<tr>
<td>Reverse primer 1</td>
<td>TTTGGTTGGCAGTTCTCTATTG</td>
<td>22</td>
<td>59.64</td>
<td>45.45</td>
</tr>
</tbody>
</table>

Template

>**NW_006916612.1** Endocarpon pusillum Z07020 unplaced genomic scaffold scaffold8762, whole genome shotgun sequence

**Product length = 144**

Features associated with this product:

Forward primer 1

<table>
<thead>
<tr>
<th>Sequence (5′-3′)</th>
<th>Length</th>
<th>Tm</th>
<th>GC%</th>
<th>Self complementarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template</td>
<td>TGCTGGTGCCACAGGACAT</td>
<td>19</td>
<td>61.83</td>
<td>57.89</td>
</tr>
</tbody>
</table>

Template

BLAST is a registered trademark of the National Library of Medicine

**Support center Mailing list YouTube**
H. P. aeruginosa qPCR probe and reverse primer specificity through BLAST Analysis

- NCBI Home
- Sign in to NCBI
- Skip to Main Content
- Skip to Navigation
- About NCBI Accesskeys

NCBI National Center for Biotechnology Information

- My NCBI
- Sign in to NCBI
- Register
- Sign Out

Primer-BLAST Job ID: T0WRamu8JhQBKnwvMU8YHUtUCS9mmRx1yZw
Primer-BLAST Results [2]

Input PCR template
none

Specificity of primers
Target templates were found in selected database: RelSeq Representative Genome Database (Organism limited to Bacteria)

Other reports
Search Summary

Detailed primer reports

You can re-search for specific primers by accepting some of the unintended targets, check the box(es) next to the ones you accept and try again to re-search for specific primers [2]

Primer pair 1

<table>
<thead>
<tr>
<th>Sequence (5'-&gt;3')</th>
<th>Length</th>
<th>Tm</th>
<th>GC%</th>
<th>Self complementarity</th>
<th>Self 3' complementarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward primer</td>
<td>CAGATGCTTTGCTCAA</td>
<td>17</td>
<td>52.08</td>
<td>47.06</td>
<td>3.00</td>
</tr>
<tr>
<td>Reverse primer</td>
<td>TTGTGGTGCGATTCCTCGATTG 22</td>
<td>59.64</td>
<td>45.45</td>
<td>4.00</td>
<td>3.00</td>
</tr>
</tbody>
</table>

Products on intended targets

Products on allowed targets

Products on allowed transcript variants

Products on potentially unintended templates

Products on target templates

>NC_002516.2 Pseudomonas aeruginosa PA01, complete genome

product length  =  45
Forward primer 1  CAGATGCTTTGCTCAA  17
Template 780325  780309
Reverse primer 1  TTGTGGTGCGATTCCTCGATTG  22
Template 780281  780302
Appendix 3.8. BLAST Analysis of the differentially abundant taxa in PLW-HIV (Chapter 7)

3.8.A. Pseudomonas

03/12/2018

BLAST Results

Job title: Nucleotide Sequence (427 letters)

<table>
<thead>
<tr>
<th>RID</th>
<th>ID</th>
<th>Query ID</th>
<th>Database Name</th>
<th>Description</th>
<th>Program</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA6E9NUG14</td>
<td>1</td>
<td>query_16390</td>
<td>rRNA_16S_prokaryotic</td>
<td>16S ribosomal RNA (Bacteria and Archaea)</td>
<td>BLASTN 2.8.1+</td>
</tr>
</tbody>
</table>

Query Length: 427

Graphic Summary

Distribution of the top 100 Blast Hits on 100 subject sequences

Color key for alignment scores

<table>
<thead>
<tr>
<th>Query Length</th>
<th>&lt;40</th>
<th>40-50</th>
<th>50-80</th>
<th>80-200</th>
<th>&gt;200</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>80</td>
<td>160</td>
<td>240</td>
<td>320</td>
<td>400</td>
</tr>
</tbody>
</table>

### Descriptions

Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>Description</th>
<th>Max score</th>
<th>Total score</th>
<th>Query cover</th>
<th>E value</th>
<th>Ident</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pseudomonas aeruginosa strain DSM 50071 16S ribosomal RNA gene, partial sequence</td>
<td>764</td>
<td>784</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_117678.1</td>
</tr>
<tr>
<td>Pseudomonas aeruginosa strain NBRC 12689 16S ribosomal RNA gene, partial sequence</td>
<td>784</td>
<td>784</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_113599.1</td>
</tr>
<tr>
<td>Pseudomonas aeruginosa strain ATCC 10145 16S ribosomal RNA gene, partial sequence</td>
<td>764</td>
<td>784</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_114471.1</td>
</tr>
<tr>
<td>Pseudomonas guerinei strain RA26 16S ribosomal RNA, partial sequence</td>
<td>773</td>
<td>773</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_114957.1</td>
</tr>
<tr>
<td>Pseudomonas gualdiogonospora strain Stx2-Z 16S ribosomal RNA gene, partial sequence</td>
<td>767</td>
<td>767</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_118458.1</td>
</tr>
<tr>
<td>Pseudomonas officinalis strain MCC10330 16S ribosomal RNA, partial sequence</td>
<td>767</td>
<td>767</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_043289.1</td>
</tr>
<tr>
<td>Pseudomonas resinovorans strain ATCC 14035 16S ribosomal RNA, partial sequence</td>
<td>767</td>
<td>767</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_112962.1</td>
</tr>
<tr>
<td>Pseudomonas alcicolum strain NBRC 14159 16S ribosomal RNA gene, partial sequence</td>
<td>761</td>
<td>761</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_113646.1</td>
</tr>
<tr>
<td>Pseudomonas alcicolum 16S ribosomal RNA gene, partial sequence</td>
<td>761</td>
<td>761</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_117827.1</td>
</tr>
<tr>
<td>Pseudomonas alcicolum strain ATCC 14009 16S ribosomal RNA gene, partial sequence</td>
<td>758</td>
<td>758</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_114472.1</td>
</tr>
<tr>
<td>Pseudomonas indica strain NBRC 103045 16S ribosomal RNA gene, partial sequence</td>
<td>756</td>
<td>756</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_114196.1</td>
</tr>
<tr>
<td>Pseudomonas indica strain IMT37 16S ribosomal RNA gene, partial sequence</td>
<td>756</td>
<td>756</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_028801.1</td>
</tr>
<tr>
<td>Pseudomonas lynyrgonensis strain LYBRD3-7 16S ribosomal RNA gene, partial sequence</td>
<td>750</td>
<td>750</td>
<td>100%</td>
<td>0.0</td>
<td>96%</td>
<td>NR_117838.1</td>
</tr>
<tr>
<td>Pseudomonas agitissata strain CC-OPY-1 16S ribosomal RNA gene, partial sequence</td>
<td>741</td>
<td>741</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_118347.1</td>
</tr>
<tr>
<td>Pseudomonas kraschkussii strain 813 16S ribosomal RNA, partial sequence</td>
<td>739</td>
<td>739</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_121733.1</td>
</tr>
<tr>
<td>Pseudomonas citronellicola strain NBRC 100343 16S ribosomal RNA gene, partial sequence</td>
<td>739</td>
<td>739</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_114194.1</td>
</tr>
<tr>
<td>Pseudomonas nitratireducens strain NBRC 12684 16S ribosomal RNA gene, partial sequence</td>
<td>739</td>
<td>739</td>
<td>100%</td>
<td>0.0</td>
<td>96%</td>
<td>NR_113681.1</td>
</tr>
<tr>
<td>Pseudomonas delhissii strain RLD-1 16S ribosomal RNA gene, partial sequence</td>
<td>739</td>
<td>739</td>
<td>100%</td>
<td>0.0</td>
<td>96%</td>
<td>NR_043731.1</td>
</tr>
<tr>
<td>Pseudomonas nitratireducens strain DSM 14389 16S ribosomal RNA gene, partial sequence</td>
<td>739</td>
<td>739</td>
<td>100%</td>
<td>0.0</td>
<td>96%</td>
<td>NR_114975.1</td>
</tr>
<tr>
<td>Pseudomonas nitratireducens strain IAM 1439 16S ribosomal RNA gene, partial sequence</td>
<td>739</td>
<td>739</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_042435.1</td>
</tr>
<tr>
<td>Pseudomonas alcicolum strain IAM 12411 16S ribosomal RNA, partial sequence</td>
<td>739</td>
<td>739</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_043419.1</td>
</tr>
<tr>
<td>Description</td>
<td>Max score</td>
<td>Total score</td>
<td>Query cover</td>
<td>E value</td>
<td>Ident</td>
<td>Accession</td>
</tr>
<tr>
<td>-----------------------------------------------------------------------------</td>
<td>-----------</td>
<td>-------------</td>
<td>-------------</td>
<td>---------</td>
<td>-------</td>
<td>------------</td>
</tr>
<tr>
<td>Pseudomonas chlororhabdus strain ATCC 13674 16S ribosomal RNA, partial sequence</td>
<td>739</td>
<td>739</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_112009.1</td>
</tr>
<tr>
<td>Pseudomonas multiresistovora strain ATCC 700690 16S ribosomal RNA, partial sequence</td>
<td>739</td>
<td>739</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_119225.1</td>
</tr>
<tr>
<td>Pseudomonas knaomussali strain B13 16S ribosomal RNA, partial sequence</td>
<td>739</td>
<td>739</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_841702.1</td>
</tr>
<tr>
<td>Pseudomonas chlororhabdus strain DSM 50332 16S ribosomal RNA gene, partial sequence</td>
<td>739</td>
<td>739</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_926533.1</td>
</tr>
<tr>
<td>Pseudomonas onyxiae strain NM-3 16S ribosomal RNA, partial sequence</td>
<td>734</td>
<td>734</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_133023.1</td>
</tr>
<tr>
<td>Pseudomonas nitritireducens strain WCBFD-3-5A2 16S ribosomal RNA, partial sequence</td>
<td>734</td>
<td>734</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_133020.1</td>
</tr>
<tr>
<td>Pseudomonas pseudocaligenes strain NBR1C 141457 16S ribosomal RNA gene, partial sequence</td>
<td>734</td>
<td>734</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_113653.1</td>
</tr>
<tr>
<td>Pseudomonas penicillinii strain Spec 1 16S ribosomal RNA gene, partial sequence</td>
<td>734</td>
<td>734</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_844209.1</td>
</tr>
<tr>
<td>Pseudomonas indoxylsulfuricola strain IPL1 16S ribosomal RNA gene, partial sequence</td>
<td>734</td>
<td>734</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_115522.1</td>
</tr>
<tr>
<td>Pseudomonas pseudocaligenes strain JCM 5998 16S ribosomal RNA, partial sequence</td>
<td>734</td>
<td>734</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_112065.1</td>
</tr>
<tr>
<td>Pseudomonas pseudocaligenes strain Starkey 03 16S ribosomal RNA gene, partial sequence</td>
<td>734</td>
<td>734</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_907000.1</td>
</tr>
<tr>
<td>Pseudomonas aeruginosa strain DSM 50370 16S ribosomal RNA, complete sequence</td>
<td>732</td>
<td>732</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_926078.1</td>
</tr>
<tr>
<td>Pseudomonas oleovorans strain ATCC 8092 16S ribosomal RNA gene, partial sequence</td>
<td>730</td>
<td>730</td>
<td>100%</td>
<td>0.0</td>
<td>97%</td>
<td>NR_114478.1</td>
</tr>
<tr>
<td>Pseudomonas fluvialis strain ASS-1 16S ribosomal RNA, partial sequence</td>
<td>728</td>
<td>728</td>
<td>100%</td>
<td>0.0</td>
<td>97%</td>
<td>NR_159318.1</td>
</tr>
<tr>
<td>Azotobacter beijerinckii strain ICM® 8671 16S ribosomal RNA gene, partial sequence</td>
<td>728</td>
<td>728</td>
<td>100%</td>
<td>0.0</td>
<td>97%</td>
<td>NR_842071.1</td>
</tr>
<tr>
<td>Pseudomonas mendocina strain NCIB 10541 16S ribosomal RNA, partial sequence</td>
<td>728</td>
<td>728</td>
<td>100%</td>
<td>0.0</td>
<td>97%</td>
<td>NR_843421.1</td>
</tr>
<tr>
<td>Pseudomonas stutzeri strain ATCC 17588 16S ribosomal RNA, complete sequence</td>
<td>723</td>
<td>723</td>
<td>100%</td>
<td>0.0</td>
<td>97%</td>
<td>NR_103934.2</td>
</tr>
<tr>
<td>Pseudomonas stutzeri strain NBR1C 141455 16S ribosomal RNA gene, partial sequence</td>
<td>723</td>
<td>723</td>
<td>100%</td>
<td>0.0</td>
<td>97%</td>
<td>NR_113652.1</td>
</tr>
<tr>
<td>Pseudomonas oleovorans strain NBR1C 13583 16S ribosomal RNA gene, partial sequence</td>
<td>723</td>
<td>723</td>
<td>100%</td>
<td>0.0</td>
<td>97%</td>
<td>NR_113617.1</td>
</tr>
<tr>
<td>Pseudomonas anguilliseptica strain S 1 16S ribosomal RNA gene, partial sequence</td>
<td>723</td>
<td>723</td>
<td>100%</td>
<td>0.0</td>
<td>97%</td>
<td>NR_929319.1</td>
</tr>
<tr>
<td>Pseudomonas stutzeri strain VSM B-975 16S ribosomal RNA gene, partial sequence</td>
<td>723</td>
<td>723</td>
<td>100%</td>
<td>0.0</td>
<td>97%</td>
<td>NR_116489.1</td>
</tr>
<tr>
<td>Pseudomonas peli strain R-20805 16S ribosomal RNA gene, partial sequence</td>
<td>723</td>
<td>723</td>
<td>100%</td>
<td>0.0</td>
<td>97%</td>
<td>NR_842451.1</td>
</tr>
<tr>
<td>Pseudomonas stutzeri strain ATCC 17588 16S ribosomal RNA, partial sequence</td>
<td>723</td>
<td>723</td>
<td>100%</td>
<td>0.0</td>
<td>97%</td>
<td>NR_841715.1</td>
</tr>
</tbody>
</table>

**Alignments**

**Pseudomonas aeruginosa strain DSM 50071 16S ribosomal RNA gene, partial sequence**

Sequence ID: NR_117878.1  Length: 1527  Number of Matches: 1
Range 1: 343 to 788

<table>
<thead>
<tr>
<th>Score</th>
<th>Expect</th>
<th>Identities</th>
<th>Gaps</th>
<th>Strand</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>764 (424)</td>
<td>0.0</td>
<td>426/427 (99%)</td>
<td>0/427 (0%)</td>
<td>Plus/Plus</td>
<td></td>
</tr>
</tbody>
</table>

Features:

Query 1

<table>
<thead>
<tr>
<th>Score</th>
<th>Expect</th>
<th>Identities</th>
<th>Gaps</th>
<th>Strand</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>402</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Subject 343

<table>
<thead>
<tr>
<th>Score</th>
<th>Expect</th>
<th>Identities</th>
<th>Gaps</th>
<th>Strand</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>120</td>
<td>463</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Subject 463

<table>
<thead>
<tr>
<th>Score</th>
<th>Expect</th>
<th>Identities</th>
<th>Gaps</th>
<th>Strand</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>180</td>
<td>522</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Query 121

<table>
<thead>
<tr>
<th>Score</th>
<th>Expect</th>
<th>Identities</th>
<th>Gaps</th>
<th>Strand</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>240</td>
<td>582</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Subject 181

<table>
<thead>
<tr>
<th>Score</th>
<th>Expect</th>
<th>Identities</th>
<th>Gaps</th>
<th>Strand</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>240</td>
<td>582</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Query 21 |

<table>
<thead>
<tr>
<th>Score</th>
<th>Expect</th>
<th>Identities</th>
<th>Gaps</th>
<th>Strand</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>624</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Subject 241

<table>
<thead>
<tr>
<th>Score</th>
<th>Expect</th>
<th>Identities</th>
<th>Gaps</th>
<th>Strand</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>360</td>
<td>703</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Query 361

<table>
<thead>
<tr>
<th>Score</th>
<th>Expect</th>
<th>Identities</th>
<th>Gaps</th>
<th>Strand</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>426</td>
<td>762</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Subject 361

<table>
<thead>
<tr>
<th>Score</th>
<th>Expect</th>
<th>Identities</th>
<th>Gaps</th>
<th>Strand</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>426</td>
<td>762</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Query 421

<table>
<thead>
<tr>
<th>Score</th>
<th>Expect</th>
<th>Identities</th>
<th>Gaps</th>
<th>Strand</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>427</td>
<td>769</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Subject 421

<table>
<thead>
<tr>
<th>Score</th>
<th>Expect</th>
<th>Identities</th>
<th>Gaps</th>
<th>Strand</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>427</td>
<td>769</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Pseudomonas aeruginosa strain NBRC 12609 16S ribosomal RNA gene, partial sequence**

Sequence ID: NR_113556.1  Length: 1461  Number of Matches: 1
Range 1: 325 to 751

3.8.B. *Klebsiella*

**BLAST Results**

**Job title:** Nucleotide Sequence (427 letters)

<table>
<thead>
<tr>
<th>Query ID</th>
<th>IciQuery_117149</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecule type</td>
<td>Nucleic acid</td>
</tr>
<tr>
<td>Query Length</td>
<td>437</td>
</tr>
</tbody>
</table>

**Database Name**

- RNA_typestrains/prokaryotic_165_ribosc

**Description**

- 16S ribosomal RNA (Bacteria and Archaea)

**Program**

BLASTN 2.8.1+

**Graphic Summary**

Distribution of the top 100 Blast Hits on 100 subject sequences

<table>
<thead>
<tr>
<th>Query</th>
<th>Color key for alignment scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>80 40-50 50-80 80-200 &gt;=200</td>
</tr>
<tr>
<td>160</td>
<td></td>
</tr>
<tr>
<td>240</td>
<td></td>
</tr>
<tr>
<td>320</td>
<td></td>
</tr>
<tr>
<td>400</td>
<td></td>
</tr>
</tbody>
</table>
# Descriptions

Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>Description</th>
<th>Max score</th>
<th>Total score</th>
<th>Query cover</th>
<th>E value</th>
<th>Ident</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>Klebsiella pneumoniae subsp. ozzaeae strain ATCC 11296 16S ribosomal RNA gene, partial sequence</td>
<td>778</td>
<td>778</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_041759.1</td>
</tr>
<tr>
<td>Klebsiella quasipneumoniae subsp. quasipneumoniae strain 01A030 16S ribosomal RNA, partial sequence</td>
<td>773</td>
<td>773</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_114652.1</td>
</tr>
<tr>
<td>Klebsiella pneumoniae strain DSM 30104 16S ribosomal RNA gene, partial sequence</td>
<td>773</td>
<td>773</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_117633.1</td>
</tr>
<tr>
<td>Klebsiella pneumoniae subsp. rhinoscleromatis strain ATCC 13884 16S ribosomal RNA gene, partial sequence</td>
<td>773</td>
<td>773</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_114507.1</td>
</tr>
<tr>
<td>Klebsiella pneumoniae subsp. rhinoscleromatis strain R-70 16S ribosomal RNA gene, partial sequence</td>
<td>773</td>
<td>773</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_037064.1</td>
</tr>
<tr>
<td>Serratia liquefaciens strain JCM1245 16S ribosomal RNA gene, partial sequence</td>
<td>773</td>
<td>773</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_112008.1</td>
</tr>
<tr>
<td>Klebsiella pneumoniae subsp. ozzaeae strain ATCC 11296 16S ribosomal RNA gene, partial sequence</td>
<td>767</td>
<td>767</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_119276.1</td>
</tr>
<tr>
<td>Klebsiella quasipneumoniae subsp. simulipneumoniae strain 07A044 16S ribosomal RNA, partial sequence</td>
<td>763</td>
<td>763</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_114663.1</td>
</tr>
<tr>
<td>Klebsiella pneumoniae strain DSM 30104 16S ribosomal RNA gene, partial sequence</td>
<td>763</td>
<td>763</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_114715.1</td>
</tr>
<tr>
<td>Klebsiella aerogenes strain KCTC 2190 16S ribosomal RNA, complete sequence</td>
<td>761</td>
<td>761</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_102493.2</td>
</tr>
<tr>
<td>Enterobacter cloacae strain ATCC 13047 16S ribosomal RNA gene, partial sequence</td>
<td>761</td>
<td>761</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_118568.1</td>
</tr>
<tr>
<td>Enterobacter kongliangense strain 10-17 16S ribosomal RNA gene, partial sequence</td>
<td>761</td>
<td>761</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_126208.1</td>
</tr>
<tr>
<td>Enterobacter aerogenes strain ATCC 13048 16S ribosomal RNA gene, partial sequence</td>
<td>761</td>
<td>761</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_118556.1</td>
</tr>
<tr>
<td>Klebsiella pneumoniae strain NBRC 14940 16S ribosomal RNA gene, partial sequence</td>
<td>761</td>
<td>761</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_113702.1</td>
</tr>
<tr>
<td>Enterobacter aerogenes strain NBRC 13534 16S ribosomal RNA gene, partial sequence</td>
<td>761</td>
<td>761</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_113614.1</td>
</tr>
<tr>
<td>Klebsiella pneumoniae strain JCM 1682 16S ribosomal RNA gene, partial sequence</td>
<td>761</td>
<td>761</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_113249.1</td>
</tr>
<tr>
<td>Klebsiella pneumoniae strain ATCC 13883 16S ribosomal RNA gene, partial sequence</td>
<td>761</td>
<td>761</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_114506.1</td>
</tr>
<tr>
<td>Enterobacter aerogenes strain NCTC10006 16S ribosomal RNA gene, partial sequence</td>
<td>761</td>
<td>761</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_114737.1</td>
</tr>
<tr>
<td>Enterobacter aerogenes strain JCM1235 16S ribosomal RNA gene, partial sequence</td>
<td>761</td>
<td>761</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_024643.1</td>
</tr>
<tr>
<td>Enterobacter asburiae strain JCM16051 16S ribosomal RNA gene, partial sequence</td>
<td>761</td>
<td>761</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_024640.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Description</th>
<th>Max score</th>
<th>Total score</th>
<th>Query cover</th>
<th>E value</th>
<th>Ident</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Klebsiella pneumoniae</em> strain ATCC 13883 16S ribosomal RNA gene, partial sequence</td>
<td>758</td>
<td>758</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_119278.1</td>
</tr>
<tr>
<td><em>Klebsiella pneumoniae</em> strain DSM 30104 16S ribosomal RNA gene, partial sequence</td>
<td>756</td>
<td>756</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_117686.1</td>
</tr>
<tr>
<td><em>Klebsiella pneumoniae</em> strain DSM 30104 16S ribosomal RNA gene, partial sequence</td>
<td>756</td>
<td>756</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_117686.1</td>
</tr>
<tr>
<td><em>Klebsiella pneumoniae</em> strain DSM 30104 16S ribosomal RNA gene, partial sequence</td>
<td>756</td>
<td>756</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_117686.1</td>
</tr>
<tr>
<td><em>Klebsiella pneumoniae</em> strain DSM 30104 16S ribosomal RNA gene, partial sequence</td>
<td>756</td>
<td>756</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_117686.1</td>
</tr>
<tr>
<td><em>Enterobacter bugandensis</em> strain 247/BMC 16S ribosomal RNA, partial sequence</td>
<td>756</td>
<td>756</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_140049.1</td>
</tr>
<tr>
<td><em>Klebsiella alba</em> strain CW-0 3 16S ribosomal RNA, partial sequence</td>
<td>756</td>
<td>756</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_132590.1</td>
</tr>
<tr>
<td><em>Klebsiella variicola</em> strain F2R9 16S ribosomal RNA gene, partial sequence</td>
<td>756</td>
<td>756</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_026635.1</td>
</tr>
<tr>
<td><em>Klebsiella pneumoniae</em> strain JCM1662 16S ribosomal RNA gene, partial sequence</td>
<td>756</td>
<td>756</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_112009.1</td>
</tr>
<tr>
<td><em>Enterobacter cancerogenus</em> strain LMG 2693 16S ribosomal RNA gene, partial sequence</td>
<td>754</td>
<td>754</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_116756.1</td>
</tr>
<tr>
<td><em>Enterobacter cancerogenus</em> strain LMG 2693 16S ribosomal RNA gene, partial sequence</td>
<td>754</td>
<td>754</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_044977.1</td>
</tr>
<tr>
<td><em>Raoultella terrigena</em> strain ATCC 33257 16S ribosomal RNA gene, partial sequence</td>
<td>750</td>
<td>750</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_114503.1</td>
</tr>
<tr>
<td><em>Raoultella terrigena</em> strain 84 16S ribosomal RNA gene, partial sequence</td>
<td>750</td>
<td>750</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_037085.1</td>
</tr>
<tr>
<td><em>Klebsiella pneumoniae</em> strain DSM 30104 16S ribosomal RNA, partial sequence</td>
<td>750</td>
<td>750</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_036794.1</td>
</tr>
<tr>
<td><em>Klebsiella aerogenes</em> strain LX3 16S ribosomal RNA gene, partial sequence</td>
<td>749</td>
<td>749</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_029783.1</td>
</tr>
<tr>
<td><em>Klebsiella aerogenes</em> strain NBRBC 102467 16S ribosomal RNA gene, partial sequence</td>
<td>747</td>
<td>747</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_114108.1</td>
</tr>
<tr>
<td><em>Enterobacter soli</em> strain LF7 16S ribosomal RNA gene, partial sequence</td>
<td>747</td>
<td>747</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_117547.1</td>
</tr>
<tr>
<td><em>Klebsiella grimontii</em> strain SB73 16S ribosomal RNA, partial sequence</td>
<td>745</td>
<td>745</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_159317.1</td>
</tr>
<tr>
<td><em>Enterobacter tabaci</em> strain YIM Hb-3 16S ribosomal RNA, partial sequence</td>
<td>745</td>
<td>745</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_140607.2</td>
</tr>
<tr>
<td><em>Klebsiella miyaginensis</em> strain W14 16S ribosomal RNA gene, partial sequence</td>
<td>745</td>
<td>745</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_118335.1</td>
</tr>
<tr>
<td><em>Escherichia vulneris</em> strain NBRBC 102420 16S ribosomal RNA gene, partial sequence</td>
<td>745</td>
<td>745</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_114080.1</td>
</tr>
<tr>
<td><em>Raoultella terrigena</em> strain NBRBC 14941 16S ribosomal RNA gene, partial sequence</td>
<td>745</td>
<td>745</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_113703.1</td>
</tr>
<tr>
<td><em>Escherichia vulneris</em> strain Billing ES3 16S ribosomal RNA, partial sequence</td>
<td>745</td>
<td>745</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_104932.1</td>
</tr>
<tr>
<td><em>Cronobacter cClockens</em> strain LMG 23730 16S ribosomal RNA gene, partial sequence</td>
<td>745</td>
<td>745</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_104924.1</td>
</tr>
<tr>
<td><em>Escherichia vulneris</em> strain ATCC 33821 16S ribosomal RNA gene, partial sequence</td>
<td>745</td>
<td>745</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_041927.1</td>
</tr>
</tbody>
</table>
### Klebsiella pneumoniae strain DSM 30104 16S ribosomal RNA gene, partial sequence

**Sequence ID:** NR_117683.1  
**Length:** 1530  
**Number of Matches:** 1  
**Range:** 1 to 772

<table>
<thead>
<tr>
<th>Score</th>
<th>Expect</th>
<th>Identities</th>
<th>Gaps</th>
<th>Strand</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>773 bits (418)</td>
<td>0.00%</td>
<td>424/427 (99%)</td>
<td>0/4 (0%)</td>
<td>Plus/Plus</td>
<td></td>
</tr>
</tbody>
</table>

**Features:**

<table>
<thead>
<tr>
<th>Query 1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sBjct 346</td>
<td>60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 61</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sBjct 406</td>
<td>120</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 121</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sBjct 666</td>
<td>180</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 181</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sBjct 526</td>
<td>240</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 241</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sBjct 586</td>
<td>585</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 361</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sBjct 666</td>
<td>64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 361</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sBjct 666</td>
<td>360</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 361</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sBjct 766</td>
<td>705</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 421</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sBjct 766</td>
<td>765</td>
</tr>
</tbody>
</table>

### Klebsiella pneumoniae subsp. rhinoscleromatis strain ATCC 13884 16S ribosomal RNA gene, partial sequence

**Sequence ID:** NR_114507.1  
**Length:** 1436  
**Number of Matches:** 1  
**Range:** 1 to 767

<table>
<thead>
<tr>
<th>Score</th>
<th>Expect</th>
<th>Identities</th>
<th>Gaps</th>
<th>Strand</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>773 bits (418)</td>
<td>0.00%</td>
<td>424/427 (99%)</td>
<td>0/4 (0%)</td>
<td>Plus/Plus</td>
<td></td>
</tr>
</tbody>
</table>

**Features:**

<table>
<thead>
<tr>
<th>Query 1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sBjct 341</td>
<td>60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 61</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sBjct 461</td>
<td>120</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 121</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sBjct 461</td>
<td>180</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 181</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sBjct 521</td>
<td>240</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 241</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sBjct 581</td>
<td>580</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 361</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sBjct 641</td>
<td>700</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 361</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sBjct 761</td>
<td>760</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 421</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sBjct 761</td>
<td>767</td>
</tr>
</tbody>
</table>

### Klebsiella pneumoniae subsp. rhinoscleromatis strain R-70 16S ribosomal RNA gene, partial sequence

https:// Blast.ncbi.nlm.nih.gov/Blast.cgi
### 3.8.C. Bilophila

**BLAST Results**

**Job title:** 359872 105_242800 (428 letters)

<table>
<thead>
<tr>
<th>Query ID</th>
<th>Database Name</th>
<th>Description</th>
<th>Molecule type</th>
<th>Query Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Icl_Query_57127</td>
<td>wgs (3 databases)</td>
<td>359872 105_242800</td>
<td>nucleic acid</td>
<td>428</td>
</tr>
</tbody>
</table>

**Graphic Summary**

#### Distribution of the top 6 Blast Hits on 3 subject sequences

<table>
<thead>
<tr>
<th></th>
<th>&lt;40</th>
<th>40-50</th>
<th>50-80</th>
<th>80-200</th>
<th>&gt;=200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query</td>
<td>80</td>
<td>160</td>
<td>240</td>
<td>320</td>
<td>400</td>
</tr>
</tbody>
</table>

## Descriptions

Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>Description</th>
<th>Max score</th>
<th>Total score</th>
<th>Query cover</th>
<th>E value</th>
<th>Ident</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blophila wadsworthii ATCC 49250 T3760DRAFT_scaffold09175.175_C, whole genome shotgun sequence</td>
<td>785</td>
<td>785</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>JNPJ01000178.1</td>
</tr>
<tr>
<td>Blophila wadsworthii 3_1_c acCl2-supercord2.1 C1, whole genome shotgun sequence</td>
<td>780</td>
<td>3143</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>ADC02000901.1</td>
</tr>
<tr>
<td>Blophila sp. 4_1_50 cent1.122, whole genome shotgun sequence</td>
<td>785</td>
<td>785</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>ADC081008112.1</td>
</tr>
</tbody>
</table>

## Alignments

### Blophila wadsworthii ATCC 49250 T3760DRAFT_scaffold09175.175_C, whole genome shotgun sequence

Sequence ID: JNPJ01000178.1  Length: 1813  Number of Matches: 1
Range 1: 458 to 885

<table>
<thead>
<tr>
<th>Score</th>
<th>Expect</th>
<th>Identities</th>
<th>Gaps</th>
<th>Strand</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>785 bits(425)</td>
<td>0.00</td>
<td>427/428(99%)</td>
<td>0/428(0%)</td>
<td>Plus/Plus</td>
<td></td>
</tr>
</tbody>
</table>

Features:

<table>
<thead>
<tr>
<th>Query 1</th>
<th>458</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject 1</td>
<td>458</td>
</tr>
<tr>
<td>Query 2</td>
<td>510</td>
</tr>
<tr>
<td>Subject 2</td>
<td>510</td>
</tr>
<tr>
<td>Query 3</td>
<td>518</td>
</tr>
<tr>
<td>Subject 3</td>
<td>518</td>
</tr>
<tr>
<td>Query 4</td>
<td>578</td>
</tr>
<tr>
<td>Subject 4</td>
<td>578</td>
</tr>
<tr>
<td>Query 5</td>
<td>638</td>
</tr>
<tr>
<td>Subject 5</td>
<td>638</td>
</tr>
<tr>
<td>Query 6</td>
<td>697</td>
</tr>
<tr>
<td>Subject 6</td>
<td>697</td>
</tr>
<tr>
<td>Query 7</td>
<td>241</td>
</tr>
<tr>
<td>Subject 7</td>
<td>241</td>
</tr>
<tr>
<td>Query 8</td>
<td>698</td>
</tr>
<tr>
<td>Subject 8</td>
<td>698</td>
</tr>
<tr>
<td>Query 9</td>
<td>361</td>
</tr>
<tr>
<td>Subject 9</td>
<td>361</td>
</tr>
<tr>
<td>Query 10</td>
<td>818</td>
</tr>
<tr>
<td>Subject 10</td>
<td>818</td>
</tr>
<tr>
<td>Query 11</td>
<td>421</td>
</tr>
<tr>
<td>Subject 11</td>
<td>421</td>
</tr>
<tr>
<td>Query 12</td>
<td>878</td>
</tr>
<tr>
<td>Subject 12</td>
<td>878</td>
</tr>
</tbody>
</table>

### Blophila wadsworthii 3_1_c acCl2-supercord2.1 C1, whole genome shotgun sequence

Sequence ID: ADC02000901.1  Length: 3168074  Number of Matches: 4
Range 1: 5303981 to 534408

<table>
<thead>
<tr>
<th>Score</th>
<th>Expect</th>
<th>Identities</th>
<th>Gaps</th>
<th>Strand</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>785 bits(425)</td>
<td>0.00</td>
<td>427/428(99%)</td>
<td>0/428(0%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
</tbody>
</table>

Features:

<table>
<thead>
<tr>
<th>Query 1</th>
<th>534408</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject 1</td>
<td>534408</td>
</tr>
<tr>
<td>Query 2</td>
<td>534348</td>
</tr>
<tr>
<td>Subject 2</td>
<td>534348</td>
</tr>
<tr>
<td>Query 3</td>
<td>534348</td>
</tr>
<tr>
<td>Subject 3</td>
<td>534348</td>
</tr>
<tr>
<td>Query 4</td>
<td>534288</td>
</tr>
<tr>
<td>Subject 4</td>
<td>534288</td>
</tr>
<tr>
<td>Query 5</td>
<td>534228</td>
</tr>
<tr>
<td>Subject 5</td>
<td>534228</td>
</tr>
</tbody>
</table>

3.8.D.i. *S. aureus* NCBI BLAST

**Job title:** Nucleotide Sequence (427 letters)

RID: 0A5WYST614 (Expires on 12-04 18:46 pm)
Query ID: iod|Query_20279
Description: None
Molecule type: nucleic acid
Query Length: 427
Database Name: rRNA_typestrains/prokaryotic_16S_riboso
Description: 16S ribosomal RNA (Bacteria and Archaea)
Program: BLASTn 2.8.1+

**Graphic Summary**

Distribution of the top 100 Blast Hits on 100 subject sequences

<table>
<thead>
<tr>
<th>Color key for alignment scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;40</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>80</td>
</tr>
<tr>
<td>160</td>
</tr>
<tr>
<td>240</td>
</tr>
<tr>
<td>320</td>
</tr>
<tr>
<td>400</td>
</tr>
</tbody>
</table>
## Descriptions

Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>Description</th>
<th>Max score</th>
<th>Total score</th>
<th>Query cover</th>
<th>E value</th>
<th>Ident</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>Staphylococcus aureus strain S33 R 16S ribosomal RNA, complete sequence</td>
<td>784</td>
<td>784</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_037097.2</td>
</tr>
<tr>
<td>Staphylococcus aureus strain ATCC 12600 16S ribosomal RNA, complete sequence</td>
<td>784</td>
<td>784</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_118997.2</td>
</tr>
<tr>
<td>Staphylococcus aureus strain NBRC 100910 16S ribosomal RNA gene, partial sequence</td>
<td>784</td>
<td>784</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_113386.1</td>
</tr>
<tr>
<td>Staphylococcus similis strain CCM 7213 16S ribosomal RNA gene, partial sequence</td>
<td>784</td>
<td>784</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_043146.1</td>
</tr>
<tr>
<td>Staphylococcus aureus subsp. anserine strain MYP-7 16S ribosomal RNA gene, partial sequence</td>
<td>784</td>
<td>784</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_036628.1</td>
</tr>
<tr>
<td>Staphylococcus aureus strain ATCC 12600 16S ribosomal RNA gene, partial sequence</td>
<td>784</td>
<td>784</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_115668.1</td>
</tr>
<tr>
<td>Staphylococcus petrasii subsp. praegensis strain CCM 8529 16S ribosomal RNA, partial sequence</td>
<td>773</td>
<td>773</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_136466.1</td>
</tr>
<tr>
<td>Staphylococcus petrasii strain CCM 8418 16S ribosomal RNA gene, partial sequence</td>
<td>773</td>
<td>773</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_118450.1</td>
</tr>
<tr>
<td>Staphylococcus jaellensis strain SEQ110 16S ribosomal RNA gene, partial sequence</td>
<td>773</td>
<td>773</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_113248.1</td>
</tr>
<tr>
<td>Staphylococcus petrasii subsp. croocolysicus strain MCC10046 16S ribosomal RNA, partial sequence</td>
<td>773</td>
<td>773</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_132590.1</td>
</tr>
<tr>
<td>Staphylococcus epidemidis strain NBRC 100911 16S ribosomal RNA gene, partial sequence</td>
<td>767</td>
<td>767</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_113387.1</td>
</tr>
<tr>
<td>Staphylococcus haemolyticus strain JCM 2416 16S ribosomal RNA gene, partial sequence</td>
<td>767</td>
<td>767</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_113345.1</td>
</tr>
<tr>
<td>Staphylococcus hominis subsp. novobiosepticus strain GTC 1226 16S ribosomal RNA gene, partial sequence</td>
<td>767</td>
<td>767</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_041323.1</td>
</tr>
<tr>
<td>Staphylococcus epidermidis strain DSM 20908 16S ribosomal RNA gene, partial sequence</td>
<td>767</td>
<td>767</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_119252.1</td>
</tr>
<tr>
<td>Staphylococcus capitis subsp. urealyticus strain RAW 8438 16S ribosomal RNA gene, partial sequence</td>
<td>767</td>
<td>767</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_027819.1</td>
</tr>
<tr>
<td>Staphylococcus epidemidid strain Fussel 16S ribosomal RNA gene, partial sequence</td>
<td>767</td>
<td>767</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_036504.1</td>
</tr>
<tr>
<td>Staphylococcus women strain AIS 25 16S ribosomal RNA gene, partial sequence</td>
<td>767</td>
<td>767</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_026922.1</td>
</tr>
<tr>
<td>Staphylococcus capitis strain ATCC 27840 16S ribosomal RNA gene, partial sequence</td>
<td>761</td>
<td>761</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_117006.1</td>
</tr>
<tr>
<td>Staphylococcus capitis strain JCM 2420 16S ribosomal RNA gene, partial sequence</td>
<td>761</td>
<td>761</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_113348.1</td>
</tr>
<tr>
<td>Staphylococcus saccharolyticus strain JCM 1769 16S ribosomal RNA gene, partial sequence</td>
<td>761</td>
<td>761</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_113465.1</td>
</tr>
<tr>
<td>Staphylococcus daviesi strain KS-SP_60 16S ribosomal RNA gene, partial sequence</td>
<td>761</td>
<td>761</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_116627.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Description</th>
<th>Max score</th>
<th>Total score</th>
<th>Query cover</th>
<th>E value</th>
<th>Ident</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>Staphylococcus lugdunensis strain ATCC 43809 16S ribosomal RNA gene, partial sequence</td>
<td>761</td>
<td>761</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_924668.1</td>
</tr>
<tr>
<td>Staphylococcus cohnii subsp. urealyticus strain CK37 16S ribosomal RNA gene, partial sequence</td>
<td>761</td>
<td>761</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_937046.1</td>
</tr>
<tr>
<td>Staphylococcus caprae strain ATCC 35539 16S ribosomal RNA gene, partial sequence</td>
<td>761</td>
<td>761</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_924665.1</td>
</tr>
<tr>
<td>Staphylococcus escherolyticus strain S 1 16S ribosomal RNA gene, partial sequence</td>
<td>761</td>
<td>761</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_929158.1</td>
</tr>
<tr>
<td>Staphylococcus capitis strain FK 499 16S ribosomal RNA gene, partial sequence</td>
<td>761</td>
<td>761</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_936775.1</td>
</tr>
<tr>
<td>Staphylococcus hominis strain DM 122 16S ribosomal RNA, partial sequence</td>
<td>761</td>
<td>761</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_936956.1</td>
</tr>
<tr>
<td>Staphylococcus agnetis strain 164 16S ribosomal RNA gene, partial sequence</td>
<td>756</td>
<td>756</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_117263.1</td>
</tr>
<tr>
<td>Staphylococcus pasteuri strain ATCC 51129 16S ribosomal RNA gene, partial sequence</td>
<td>756</td>
<td>756</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_114438.1</td>
</tr>
<tr>
<td>Staphylococcus pasteuri strain ATCC 51129 16S ribosomal RNA gene, partial sequence</td>
<td>756</td>
<td>756</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_924669.1</td>
</tr>
<tr>
<td>Staphylococcus hyicus strain D. Sompolinsky no. 1 16S ribosomal RNA gene, partial sequence</td>
<td>756</td>
<td>756</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_936905.1</td>
</tr>
<tr>
<td>Staphylococcus gallinarum strain VIII 1 16S ribosomal RNA gene, partial sequence</td>
<td>756</td>
<td>756</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_936903.1</td>
</tr>
<tr>
<td>Staphylococcus cohnii strain GH 137 165 ribosomal RNA gene, partial sequence</td>
<td>756</td>
<td>756</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_936902.1</td>
</tr>
<tr>
<td>Staphylococcus chromogenes strain C8CC 1482 16S ribosomal RNA gene, partial sequence</td>
<td>756</td>
<td>756</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_936901.1</td>
</tr>
<tr>
<td>Staphylococcus aureus strain 151 16S ribosomal RNA gene, partial sequence</td>
<td>756</td>
<td>756</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_936997.1</td>
</tr>
<tr>
<td>Staphylococcus haemolyticus strain SM 131 16S ribosomal RNA, partial sequence</td>
<td>756</td>
<td>756</td>
<td>100%</td>
<td>0.0</td>
<td>99%</td>
<td>NR_936955.1</td>
</tr>
<tr>
<td>Staphylococcus nepalensis strain CW1 16S ribosomal RNA gene, partial sequence</td>
<td>752</td>
<td>752</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_928996.1</td>
</tr>
<tr>
<td>Staphylococcus intermedius strain JCM 2422 16S ribosomal RNA gene, partial sequence</td>
<td>750</td>
<td>750</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_113351.1</td>
</tr>
<tr>
<td>Staphylococcus xylosus strain JCM 2418 16S ribosomal RNA gene, partial sequence</td>
<td>750</td>
<td>750</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_113350.1</td>
</tr>
<tr>
<td>Staphylococcus muscace strain MB4 16S ribosomal RNA gene, partial sequence</td>
<td>750</td>
<td>750</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_194752.1</td>
</tr>
<tr>
<td>Staphylococcus microfl strain CCM 4903 16S ribosomal RNA gene, partial sequence</td>
<td>750</td>
<td>750</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_116494.1</td>
</tr>
<tr>
<td>Staphylococcus rostri strain ARI 262 16S ribosomal RNA gene, partial sequence</td>
<td>750</td>
<td>750</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_116912.1</td>
</tr>
<tr>
<td>Staphylococcus pseudointermedius strain CM 80 16S ribosomal RNA, partial sequence</td>
<td>750</td>
<td>750</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_842294.1</td>
</tr>
<tr>
<td>Staphylococcus sucinulus subsp. casei strain SBB72 16S ribosomal RNA gene, partial sequence</td>
<td>750</td>
<td>750</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_937053.1</td>
</tr>
<tr>
<td>Staphylococcus delphini strain ATCC 45171 16S ribosomal RNA gene, partial sequence</td>
<td>750</td>
<td>750</td>
<td>100%</td>
<td>0.0</td>
<td>98%</td>
<td>NR_924666.1</td>
</tr>
</tbody>
</table>
### 3.8.D.ii. *S. aureus* SILVA ACT

<table>
<thead>
<tr>
<th>Results of SILVA ACT Analysis: <a href="https://www.arb-silva.de/aligner/">https://www.arb-silva.de/aligner/</a></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>job_id</strong></td>
</tr>
<tr>
<td><strong>sequence_number</strong></td>
</tr>
<tr>
<td><strong>sequence_identifier</strong></td>
</tr>
<tr>
<td><strong>full_name</strong></td>
</tr>
<tr>
<td><strong>sequence_score</strong></td>
</tr>
<tr>
<td><strong>bp_score</strong></td>
</tr>
<tr>
<td><strong>cutoff_head</strong></td>
</tr>
<tr>
<td><strong>cutoff_tail</strong></td>
</tr>
<tr>
<td><strong>identity</strong></td>
</tr>
<tr>
<td><strong>quality</strong></td>
</tr>
<tr>
<td><strong>startpos</strong></td>
</tr>
<tr>
<td><strong>stoppos</strong></td>
</tr>
<tr>
<td><strong>ecolipos</strong></td>
</tr>
<tr>
<td><strong>bps</strong></td>
</tr>
<tr>
<td><strong>gene_bps</strong></td>
</tr>
<tr>
<td><strong>turn</strong></td>
</tr>
<tr>
<td><strong>lca_tax_embl</strong></td>
</tr>
<tr>
<td><strong>lca_tax_greengenes</strong></td>
</tr>
<tr>
<td><strong>lca_tax_ltp</strong></td>
</tr>
<tr>
<td><strong>lca_tax_rdp</strong></td>
</tr>
<tr>
<td><strong>lca_tax_slv</strong></td>
</tr>
</tbody>
</table>
Appendix 3.9. MinION Metagenomic Sequencing Library Tape Station Report

Filename: 2019-06-21 - 12.28.09.GDNA

Sample Info

<table>
<thead>
<tr>
<th>Well</th>
<th>DN</th>
<th>Conc. [pg/µl]</th>
<th>Sample Description</th>
<th>Alert</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>-</td>
<td>39.2</td>
<td>Ladder</td>
<td>▲</td>
<td>Ladder peak detected (two peaks detected)</td>
</tr>
<tr>
<td>B1</td>
<td>-</td>
<td>75.8</td>
<td>2</td>
<td>▲</td>
<td>Ladder peak detected (two peaks detected)</td>
</tr>
<tr>
<td>C1</td>
<td>-</td>
<td>39.5</td>
<td>A</td>
<td>▲</td>
<td>Ladder peak detected (two peaks detected)</td>
</tr>
<tr>
<td>D1</td>
<td>-</td>
<td>64.2</td>
<td>C</td>
<td>▲</td>
<td>Ladder peak detected (two peaks detected)</td>
</tr>
</tbody>
</table>
**AI: Ladder**

![Image of DNA ladder electrophoresis](image)

**Sample Table**

<table>
<thead>
<tr>
<th>Well</th>
<th>DIN</th>
<th>Conc. [ng/μl]</th>
<th>Sample Description</th>
<th>Notes</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>-</td>
<td>39.2</td>
<td>Ladder</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>▲</td>
<td></td>
</tr>
</tbody>
</table>

**Peak Table**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>6.30</td>
<td>6.30</td>
<td>-</td>
<td>23</td>
<td>192</td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>3.16</td>
<td>-</td>
<td>8.36</td>
<td>200</td>
<td>343</td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>3.24</td>
<td>-</td>
<td>8.59</td>
<td>343</td>
<td>524</td>
<td></td>
<td></td>
</tr>
<tr>
<td>600</td>
<td>3.41</td>
<td>-</td>
<td>9.05</td>
<td>524</td>
<td>700</td>
<td></td>
<td></td>
</tr>
<tr>
<td>800</td>
<td>3.14</td>
<td>-</td>
<td>8.34</td>
<td>700</td>
<td>1080</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>3.23</td>
<td>-</td>
<td>8.63</td>
<td>1080</td>
<td>1770</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1500</td>
<td>3.68</td>
<td>-</td>
<td>9.75</td>
<td>1770</td>
<td>1810</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>3.16</td>
<td>-</td>
<td>8.39</td>
<td>1810</td>
<td>2315</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2500</td>
<td>3.16</td>
<td>-</td>
<td>7.82</td>
<td>2315</td>
<td>2813</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3000</td>
<td>3.09</td>
<td>-</td>
<td>8.18</td>
<td>2813</td>
<td>3676</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4000</td>
<td>2.85</td>
<td>-</td>
<td>7.54</td>
<td>3676</td>
<td>6000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7000</td>
<td>2.64</td>
<td>-</td>
<td>7.06</td>
<td>6000</td>
<td>11923</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15000</td>
<td>3.14</td>
<td>-</td>
<td>8.33</td>
<td>11923</td>
<td>20530</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
BI: 2

Sample Table

<table>
<thead>
<tr>
<th>Well</th>
<th>BIN</th>
<th>Conc. [ng/ul]</th>
<th>Sample Description</th>
<th>Alert</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI</td>
<td>-</td>
<td>75.8</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Peak Table

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>8.50</td>
<td>8.50</td>
<td>-</td>
<td>30</td>
<td>106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2495</td>
<td>46.5</td>
<td>-</td>
<td>100.00</td>
<td>1017</td>
<td>4868</td>
<td>Loose Marker</td>
<td></td>
</tr>
</tbody>
</table>

Region Table

<table>
<thead>
<tr>
<th>From [bp]</th>
<th>To [bp]</th>
<th>Average Size [bp]</th>
<th>Conc. [ng/ul]</th>
<th>Region Stability [median]</th>
<th>% of Total</th>
<th>Region Comment</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>-00000</td>
<td>5180</td>
<td>74.5</td>
<td>99.3</td>
<td>98.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>1500</td>
<td>983</td>
<td>2.64</td>
<td>7.16</td>
<td>3.40</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CI: A

Sample Table

<table>
<thead>
<tr>
<th>Well</th>
<th>DN</th>
<th>Conc. [ng/µl]</th>
<th>Sample Description</th>
<th>Short</th>
<th>Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>-</td>
<td>39.5</td>
<td>A</td>
<td>▲</td>
<td>None w/ ladder peak detected flow forward direction</td>
</tr>
</tbody>
</table>

Peak Table

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>6.50</td>
<td>6.50</td>
<td>-</td>
<td>57</td>
<td>183</td>
<td></td>
<td>Low Marker</td>
</tr>
<tr>
<td>3659</td>
<td>35.8</td>
<td>-</td>
<td>100.00</td>
<td>1761</td>
<td>11412</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Region Table

<table>
<thead>
<tr>
<th>From [bp]</th>
<th>To [bp]</th>
<th>Average Size [bp]</th>
<th>Conc. [ng/µl]</th>
<th>Region Maturity [mmHg]</th>
<th>% of Total</th>
<th>Region Comment</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>60000</td>
<td>4557</td>
<td>39.3</td>
<td>53.6</td>
<td>99.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>1500</td>
<td>804</td>
<td>1.39</td>
<td>0.89</td>
<td>5.51</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Sample Table

<table>
<thead>
<tr>
<th>Well</th>
<th>DN</th>
<th>Conc. [ng/ul]</th>
<th>Sample Description</th>
<th>Alert</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>-</td>
<td>64.2</td>
<td>C</td>
<td>▲</td>
<td>Load Well 102 despite peak detection in R00 spurious data.</td>
</tr>
</tbody>
</table>

Peak Table

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>8.50</td>
<td>8.50</td>
<td>-</td>
<td>74</td>
<td>175</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1362</td>
<td>55.8</td>
<td></td>
<td>100.00</td>
<td>1545</td>
<td>4332</td>
<td>Loss Marker</td>
<td></td>
</tr>
</tbody>
</table>

Region Table

<table>
<thead>
<tr>
<th>From [bp]</th>
<th>To [bp]</th>
<th>Average Size [bp]</th>
<th>Conc. [ng/ul]</th>
<th>Region Similarity [mmoll]</th>
<th>% of Total</th>
<th>Region Comment</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>500</td>
<td>4572</td>
<td>63.1</td>
<td>96.3</td>
<td>96.36</td>
<td></td>
<td>♂</td>
</tr>
<tr>
<td>200</td>
<td>1500</td>
<td>888</td>
<td>2.81</td>
<td>8.32</td>
<td>4.59</td>
<td></td>
<td>♀</td>
</tr>
</tbody>
</table>
Appendix 3.10. MinION Sequencing Report

A. The First 24 hrs Sequencing
B. The Second day (24hrs) Sequencing
Appendix 4. Publications


