Application of bacterial genomics and culture-free microbiota diagnostics to respiratory infections

A thesis submitted for the degree of Doctor of Philosophy

February 2022

Aimee Sara Bettridge

Summary

Polymicrobial infections drive the progression of chronic lung diseases and can dramatically affect clinical outcomes. Clinical laboratories routinely employ culture-based tools to diagnose lung infection. Nevertheless, culture cannot capture the full extent of bacterial diversity and is prone to misidentification. The suitability of 16S rRNA gene sequencing and ribosomal intergenic spacer analysis (RISA) was explored in context of respiratory infection diagnosis in the Public Health Wales (PHW) microbiology service.

Optimised RISA and PCR protocols were used to analyse mock communities simulating respiratory bacterial pathogen assemblages. Sequence-based community abundance and diversity was distinct between each mock community, although community composition substantially deviated from expected proportions. RISA proved useful in predicting species identification. A genomic DNA extraction protocol was then validated for microbiota analysis using silica-based gDNA extraction employing bead-beating.

These optimised protocols were applied to profile bronchiectasis, cystic fibrosis, and chronic obstructive pulmonary disease sputa, and the results compared to PHW routine culture-based analyses. Respiratory microbiota was highly individualised, and bacterial diversity did not statistically correlate with reduced lung function. Combining both culture-free and culture-dependent methodologies proved beneficial in the identification of potentially pathogenic microorganisms missed in culture and for resolving poor sequence identification in a minority of samples.

Whole genome sequencing (WGS) was deployed for the accurate characterisation of the emerging cystic fibrosis pathogens *Achromobacter* spp., and was compared to conventional MALDI-TOF. WGS was instrumental in determining sources of *Achromobacter insuavis* acquisition and the multidrug-resistant nature of these species, suggesting the presence of previously unreported metallo-beta-lactamases. Phenotypic traits suggest that *Achromobacter xylosoxidans*' competitive fitness was demonstrably similar to traditional CF Gram-negative species. These results also suggested these species interacted as an intransitive hierarchy.

Overall, these DNA-based methods show promise for routinisation in the PHW however WGS is still largely a research tool in context of lung infection.

Contents

List of figures

List of Tables

Abbreviations

AATD	Alpha-1 antitrypsin deficiency
ACE	Abundance based coverage estimator
ANOVA	Analysis of variance
ARU	Anaerobe reference unit
BAL	Bronchoalveolar lavage fluid
BCC	*Burkholderia cepacia* complex
BHI	Brain Heart Infusion
BLF	British lung foundation
bp	Base pair
BR	Bronchiectasis
CAP	Community-acquired pneumonia
CBA	Columbia blood agar
CDS	Coding sequences
CF	Cystic fibrosis
CFBCC	Cystic fibrosis *Burkholderia cepacia* complex
CFPa	Cystic fibrosis *Pseudomonas aeruginosa*
CFTR	Cystic fibrosis transmembrane conductance regulator
CL3	Containment level 3
CLED	Cystine–lactose–electrolyte-deficient agar
COPD	Chronic obstructive pulmonary disease
CU	Cardiff university
CV	Crystal violet
DIN	DNA integrity number
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleoside triphosphate
dsDNA	Double stranded DNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid

ENaC	Epithelial sodium channel
ENFGN	Emerging non-fermenting Gram-negative
FEV_1	Forced expiratory volume (in 1 second)
gDNA	Genomic DNA
GRH	Genomics research hub
GTC	Guanidinium thiocyanate
GTR	Generalised time-reversible
HCCA	Alpha-cyano-4-hydroxycinnamic acid
HMW	High Molecular Weight
HS	High sensitivity
InvSimpson	Inverse Simpson
IQR	Interquartile range
ITS	Intergenic transcribed spacer
LB	Luria–Bertani broth
LoD	Limit of detection
MALDI-TOF	Matrix-assisted laser desorption ionisation time-of-flight
MIC	Minimum inhibitory concentration
MLSA	Multilocus sequence analysis
MLST	Multilocus sequence typing
NAD	Nicotinamide adenine dinucleotide
NGS	Next generation sequencing
NMDS	Non-metric multidimensional scaling
NRF	Normal respiratory flora
OD	Optical density
OTU	Operational taxonomic unit
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PenGU	Pathogen genomics unit
PERMANOVA	Permutational multivariate analysis of variance
PHW	Public Health Wales
PMM	Predictive mean matching

PPM	Potentially pathogenic microorganism
QC	Quality control
qPCR	Quantitative polymerase chain reaction
RISA	Ribosomal intergenic spacer analysis
RMSE	Root mean square error
ROS	Reactive oxygen species
SACU	Specialist Antimicrobial Chemotherapy Unit
SNP	Single nucleotide polymorphism
SOP	Standard operating procedure
ST	Sequence type
TAE	Tris-acetate-EDTA buffer
T-RFLP	Terminal restriction fragment length polymorphism
TSA	Tryptic soy agar
TSB	Tryptic soy broth
UPGMA	Unweighted pair group method with mean arithmetic
V4	16S rRNA gene hypervariable region 4
WGS	Whole genome sequencing

Acknowledgements

I would like to express my gratitude to the following people:

- My funders KESS2 who have supported me throughout this project.
- Cardiff university supervisory team: Professor Eshwar Mahenthiralingam and Professor Thomas Connor for their guidance and support throughout this project. Being on the front line, I appreciate this was no small task.
- My Public Health Wales supervisor Dr Rishi Dhillon for making so much work possible and for brightening my day.
- Public Health Wales: Joanne Watkins, Dr Sally Corden, Bree Gatica-Wilcox, and healthcare scientists of the Pathogen Genomics laboratory team for your support, patience, encouragement, and training. I would like to thank Dr Lim Jones, Dr Mandy Wootton, and the Specialist Antimicrobial Chemotherapy Unit for providing me with opportunity and valuable knowledge. Special thanks to the Anaerobe Reference Unit and the Mycology Regional Reference Unit for getting my project started, for the fantastic training, and for making my experience enjoyable. A huge thanks is due for Dr Rhys Davies for saving the day with my data, Drew Mack for great advice and encouragement, and last but not least, containment level 3 laboratory staff for who enabled me to carry out my research.
- Professor Julian Marchesi for helping me think outside the box!
- Cardiff university Microbiomes, Microbes, and Informatics division: with special thanks to Dr Laura Rushton, Dr Gordon Webster, Dr Beky Weiser, Dr Laura Osborne, and Dr Jane Mikhail for sharing valuable knowledge and your continuous support, no matter how big or small the task. I am forever grateful for my peers and friends Kasia Parfitt, Yoana Petrova, and Teresa Paradell. Thanks so much for your continuous support when I needed more time. I won't forget this. Thanks for all the favours and fun times together. I'm proud of you all! Kasia, you have shared a mountain of ideas over this time which has contributed to polishing this thesis, especially the final results chapter.
- Angela Marchbank: I cannot thank you enough for all your support and training throughout these years. You have always made time for me and provided all essential knowledge and training to survive this course! The Genome Hub group is just phenomenal, and I'd like to express my gratitude towards your team for your time and teachings.
- Kai Murphy: It's been a pleasure working alongside you for all these years. You have helped me in more ways than you know. I would most likely be in a different place to where I am now if you didn't join microbiology in the first place! You have always brought order and calm on both the good days and the bad days and have been a great friend to me.
- Dr Sarah Christofides: Where do I start! You are one of the main reasons how I survived my degree. With your help, my writing is better, and a little less irrational! Importantly, you have kept me keenly interested in many scientific topics. Some of my most memorable times at university involves working with you and I am very lucky to have you once again as a supervisor! I owe this to you. Thanks for showing me the fun side of statistics and providing me with some very complex R code for linear mixed effects models featured in the final results chapter.

- Huge thanks to my friends: Connor for keeping me sane, Drew for your support and many work-related favours. Thank you, James, for improving my proofing and language skills, and for all the glorious gelato!
- To my mam and dad: I wouldn't be where I am now without your everlasting love and support. Now I'm older, I appreciate how big a sacrifice you have both made to support my education and in pursuit of my happiness and health. I owe this all to you. Thanks to my sister Sophie for giving me strength even in the face of the many challenges we have faced in recent years. you have been a great role model.
- Nan: You have been such an important figure in my life, and I am incredibly lucky to have been so close for all these years. I finally made it. Wish you could see this to the end.
- My extended family for being there always.
- Ben: You took on a lot pretty early on, so I appreciate the challenges this has brought you. Thanks so much for your acceptance, understanding and love. Thanks to your family for treating me as their own. You have single-handedly got me through long nights of work with an infinite supply of coffee, tea, sweets, and interesting facts.
- Last but not least: to my very best friends Percy and Bailey for their unconditional love and support. They always know and have always been there.

Chapter 1 General introduction

1.1 Social and economic burden of respiratory disease and infection

Respiratory tract infections remain the commonest cause for GP consultation and hospital admissions in developed nations (Stanton et al. 2010). Respiratory diseases severely impair patient quality of life, accounting for 6.5% of hospital admissions in addition to 24% of all mortalities in the UK (Naser et al. 2021). In England and Wales, chronic lower respiratory diseases, and lower respiratory infections constitute 26.6% and 14.9% of respiratory related hospital admissions (Naser et al. 2021). Despite this, only £8.9 million of research funding has been allocated to researching chronic respiratory disease, including cystic fibrosis (CF), chronic obstructive pulmonary disease (COPD) and non-CF bronchiectasis (BR) (Head et al. 2014; Burki 2017). This underfunding means there has been less progress in developing infrastructure and technologies needed to diagnose and treat these chronic respiratory diseases compared to other chronic illnesses such as tuberculosis, lung cancer, and heart disease (Holgate 2007; Head et al. 2014).

Today, CF, COPD, and BR lung disease pathology and epidemiology is well understood, and methods to better understand lung disease epidemiology, infection, and disease progression have been steadily evolving to improve patient outcome. Still, the limitations in sensitivity, throughput, and low turnaround time of routine clinical laboratory diagnostics emphasises the urgency of diagnosing lung infection with greater diagnostic efficiency (Qian et al. 2020). As the main focus of this thesis is to improve current bacterial lung infection diagnosis, this chapter reviews the role of bacterial infection and polymicrobial community diversity in BR, CF, and COPD disease, in addition to exploring the performance and application of current versus high-resolution diagnostic tools, and outstanding methodological challenges affecting sputum microbiota profiling. Following the recent deployment of whole genome sequencing (WGS) as a clinical research tool, *Achromobacter xylosoxidans* is also introduced as an emerging potentially pathogenic microorganism (PPM) in context of respiratory infection.

1.2 Respiratory disease pathogenesis and patient demographics

1.2.1 Cystic fibrosis

CF is the most common lethal autosomal recessive genetic disorder among the Caucasian population, affecting approximately 1 in 3,000 live births. Its incidence is also reported in around 1 in 4,000 to 1 in 10,000 live births among Latin Americans and every 1 in 15,000 to 1 in 20,000 African American live births (Ratjen et al. 2015). CF is a multisystem disease

experienced from infancy where disease of the lung is the main source of morbidity and mortality among patients (Figure 1.1). At present, the median age of survival is approaching 50 years (Bierlaagh et al. 2021). CF was not recognised until 1938 as it shares the same clinical manifestations as coeliac disease, such as intestinal malabsorption syndrome and pancreatic insufficiency (Ramos et al. 2016).

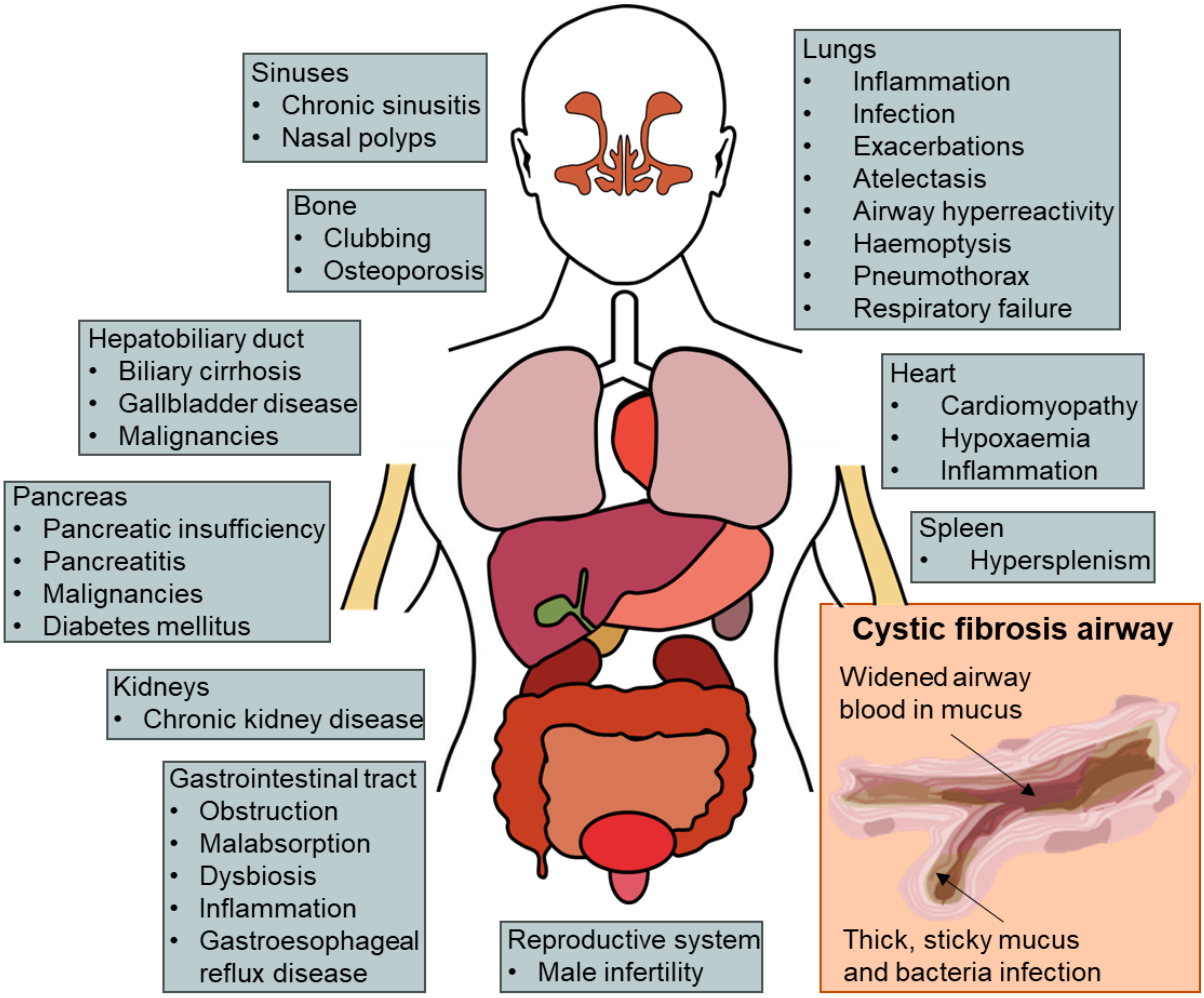


Figure 1.1 Cystic fibrosis clinical manifestations. Information about CF pathology was obtained from (Quon et al. 2011; Labombarda et al. 2016; Ooi and Durie 2016; Anabtawi et al. 2019).

In 1953, sweat electrolyte deficiency in CF was discovered, and defects in the chloride ion transport were identified accompanied by increased sodium reabsorption (Ramos et al. 2016). The cause was elucidated with the identification of the cystic fibrosis transmembrane conductance regulator (CFTR) gene spanning 250 kb of chromosome 7 in 1989 (Pérez-Frías et al. 2019; Rang and Wilson 2019). A functional CFTR protein along with the epithelial sodium channel (ENaC) is essential for airway surface hydration and mucociliary clearance (Figure 1.2). To date, more than 2100 CFTR mutations have been linked to CF. Of these, fewer than

150 CF mutations cause disease, and these disease-causing mutations are split into 6 categories (Ratjen et al. 2015). The *ΔF508* mutation (class II) causing deletion of phenylalanine at residue 508 accounts for 66% of mutations worldwide (Ratjen et al. 2015).

Loss of CFTR expression on the cell surface or loss of function are generally associated with a severe CF phenotype while mutations with residual CFTR function are often associated with preserved pancreatic function (Ratjen et al. 2015). CF genotype does not always predict disease severity (O'Neal and Knowles 2018). This phenotypic variability could also be attributed to sex; females have a reduced median survival of 3 years and have been noted to become infected sooner by major CF pathogens, especially *Pseudomonas aeruginosa* (Weiler and Drumm 2013; Harness-Brumley et al. 2014). This disparity is postulated to be influenced by factors including sex hormones, a higher prevalence of CF related diabetes, and the fact that females are often diagnosed with CF later than males (Weiler and Drumm 2013; Kayani et al. 2018; Somayaji and Chalmers 2022).

Functional CFTR allows chloride ions to move across the apical membrane of epithelial cells while reducing the transport of sodium ions into the cell through the ENaC (Figure 1.2). This exchange results in a net increase of water flux into the periciliary fluid layer which maintains the airway surface liquid (ASL) volume (Ratjen et al. 2015). pH and bicarbonate also influence normal mucus production and are proposed to play a role in bacterial clearance within this microenvironment (Pezzulo et al. 2012). Mutations in the CFTR gene adversely affect epithelial ion and water transport, leading to increased influx of chloride ions and increased sodium and water absorption via ENaC. This leads to ASL dehydration, resulting in thicker mucus secretions and impaired mucociliary clearance (Cohen-Cymberknoh et al. 2013). These dry and thick secretions obstruct the airways, promote an inflammatory environment, and prevent bacterial clearance from the lung, leading to infection and increased lung inflammation (Cohen-Cymberknoh et al. 2013). Bacterial infection subsequently leads to neutrophil-dominated inflammation that avoids apoptosis and produces proteases and radical oxygen species, thus continuing a self-perpetuating cycle of lung inflammation and infection (Conese et al. 2017). Chronic pulmonary infection inevitably leads to lung function decline, becoming more severe following exacerbations. Respiratory failure and death ensue as disease severity worsens (Filkins and O'Toole 2015).

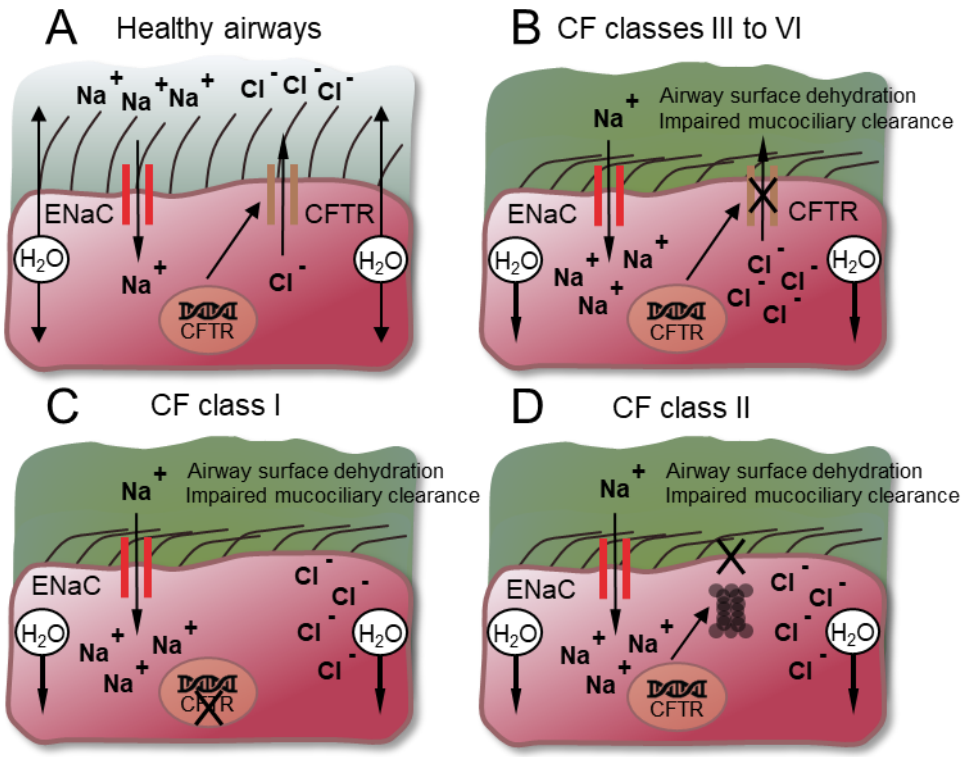


Figure 1.2 Role of CFTR in healthy lungs and CF lungs. A) CFTR is expressed at the apical surface of epithelial cells with the ENaC. Functional CFTR is crucial for chloride, bicarbonate, and fluid secretion. ENaC absorbs sodium and fluid across the airway epithelium. B) Missense and splicing mutations produce CFTR channels which reach the apical surface of epithelia but are not fully functional. This covers classes III (defective channel regulation), class IV (diminished protein maturation and ion conductance), class V (reduced functional CFTR production), and class VI (reduced time the CFTR is at the cell apical surface). C) CFTR nonsense or splicing mutations completely prevents CFTR production. D) Several missense mutations lead to impaired folding of CFTR, and it is degraded by the proteosome. Information was derived from (Gentzsch and Mall 2018).

1.2.2 Non-cystic fibrosis bronchiectasis and chronic obstructive pulmonary disease

COPD is an umbrella term encompassing chronic bronchitis and emphysema (Vindhyvasni and Gupta 2020). These conditions are characterised by progressive airflow limitation and decline in respiratory function. COPD is associated with structural lung changes due to chronic inflammation from prolonged exposure to harmful gases or particles, most commonly cigarette smoke (Agarwal et al. 2021). Approximately 5% of COPD sufferers are genetically predisposed due to alpha-1 antitrypsin deficiency (Brode et al. 2012). An estimated 328 million people have COPD worldwide and it is already the third global leading cause of death according to the Global Burden of Disease (Quaderi and Hurst 2018). Smokers and adults >40 years of age are primarily affected, and prevalence increases with age. Global prevalence is likely underestimated due to the underdiagnosis of this disease (Agarwal et al. 2021). COPD is more prevalent among females despite smoking less than males (Somayaji and Chalmers

2022). Females also develop airway obstruction after smoking fewer numbers of cigarettes and are associated with worsened clinical outcome; severe COPD stemming from smoking carries an increased risk of death in females compared to males (Harness-Brumley et al. 2014). Pollutants also disproportionately affect females (Somayaji and Chalmers 2022). Reasons for the greater susceptibility of females developing COPD is largely unknown, however Barnes (2016) suggested a link between female sex hormones and altered inflammatory response to cigarette smoke in female mouse models.

Environmental triggers, particulates, and microorganisms trigger an inflammatory cascade in the COPD lung. This inflammation is heterogenous and key inflammatory cell types include macrophages, neutrophils, and T cells (King 2015). Macrophages and epithelial cells release growth factors that activate fibroblasts and recruited immune cells secrete cytotoxic granular contents, reactive oxygen species (ROS), and proteinases into these tissues (Dey et al. 2022). These cascades of events are associated with the development of mucus secretion, emphysema, and progression of COPD (Dey et al. 2022). Chronic disease is often marked by periods of increased symptoms that are known as acute exacerbations, accelerating disease progression (Beasley et al. 2012). Acute COPD exacerbations are defined as major causes of hospital admission and mortality. As exacerbations accelerates, lung function declines and patient prognosis worsens (Mayhew et al. 2018). Pulmonary inflammation and reduced mucus clearing also leads to cycles of infection where the presence of bacteria in the lower airways has been correlated with exacerbation frequency (Rogers et al. 2009).

BR has been described as one of the most neglected respiratory diseases in medicine where few studies have explored its pathogenesis (McCallum and Binks 2017). BR is differentiated from COPD by permanent thick-walled and abnormally dilated bronchi that are often chronically infected (Cox et al. 2017). BR is both caused by and associated with several systemic disorders, and in numerous cases, the underlying aetiology is unknown (Lazarus et al. 2008). In known cases, viral and bacterial infection, immunodeficiencies such as HIV and IgA deficiency, allergic bronchopulmonary aspergillosis, AATD, and aspiration are known BR risk factors (De Soyza and Aliberti 2017; Carreto et al. 2020; Chen et al. 2020; Mitchelmore et al. 2020). BR-related incidence, prevalence, and mortality have increased to ~212,000 people living with BR in the UK and 60% of diagnoses are given to patients >70 years of age (Snell et al. 2019). BR affects a smaller portion of the US population, affecting ~340,000 to 522,000 people (Weycker et al. 2017). BR is another female-dominated disease where it has been theorised females are more prone to small airway inflammation in response to toxic particles (Somayaji and Chalmers 2022). On the other hand, males are presumed to experience increased co-morbidities because smoking and COPD is more common among this demographic (Finch et al. 2018).

BR pathogenesis is understood as the following: the walls of the small airway become infiltrated by inflammatory cells causing obstruction and anatomical changes to the bronchi that lead to progressive worsening of pulmonary physiology and increased exacerbation frequency (King et al. 2010; Chandrasekaran et al. 2018). BR exacerbation and disease progression is also driven by overgrowth of one or few microbial species (Cox et al. 2017). The host immune response to infection is primarily neutrophilic and driven by neutrophil derived proteases which result in dilated bronchi through large airway damage, amplifying this recurrent cycle of infection and inflammation (King et al. 2010; Chandrasekaran et al. 2018).

1.3 The respiratory microbiome

Culture-independent tools have revolutionised characterisation of the lung microbiome, defined as a polymicrobial community of commensal, mutualistic, and pathogenic microorganisms that are present in the lung (O'Dwyer et al. 2016). The lungs were historically considered sterile, and only in recent decades have respiratory microbiota been considered important in health and disease (Dickson et al. 2013). Additionally, gut microbiota has been shown to affect lung immunity through the gut-lung axis by the exchange of endotoxins, metabolites, cytokines, and hormones through the bloodstream (Enaud et al. 2020; Zhang et al. 2020b).

The lungs harbour significantly lower bacterial biomass than the gut. There are approximately 10-100 bacteria cells per 1,000 human cells in the lung (Sze et al. 2012). Lung community composition depends on microbial colonisation from the oropharynx and upper respiratory tract, host mucociliary clearance and cough, host immune system, microbe-microbe interactions, and physical or chemical gradients favouring or demoting growth (Enaud et al. 2020; Zhang et al. 2020b) The maintenance of these small numbers of bacteria may be a critical determinant of good health where a constant balance is maintained between microbial immigration and elimination (Mathieu et al. 2018).

The oral cavity plays an important role in seeding the lower respiratory tract of healthy individuals with diverse microbes, typically dominated by *Proteobacteria, Firmicutes,* and *Bacteroidetes* (Mathieu et al. 2018). Within these phyla, *Prevotella, Haemophilus, Streptococcus, Neisseria, Veillonella, Fusobacterium,* and *Rothia* are often highly abundant (Filkins and O'Toole 2015; Venkataraman et al. 2015). During lung disease, the balance between bacterial immigration and elimination is interrupted, resulting in microbial composition changes and with the bacteria exhibiting competitive advantages becoming predominant (Mathieu et al. 2018). The key to understanding disease pathogenesis is to understand these complex interactions between host, pathogen, and resident microbiota and to define causal

relationships between lung microbiota composition, diversity, and disease progression (O'Dwyer et al. 2016). The remainder of this section will explore the lung microbiota considered to contribute towards BR, CF, and COPD disease, focusing on traditional pathogens, i.e., those historically isolated through culture, emerging respiratory pathogens, and the oropharyngeal microbiota often referred to as normal respiratory flora (NRF) in conventional bacteriology.

1.3.1 Cystic fibrosis microbiota

Historically, both *Haemophilus influenzae* and *Staphylococcus aureus* were considered primary pathogens infecting CF infants, which then progressed to recurrent infections dominated by the *Burkholderia cepacia* complex (BCC) and *P. aeruginosa* in adulthood (Garcia-Nuñez et al. 2020). These pathogens still play a significant role in clinical outcome where *P. aeruginosa* is the predominant cause of morbidity and mortality in CF (Hatziagorou et al. 2020). Moreover, *P. aeruginosa,* BCC, and *Achromobacter xylosoxidans* dominate microbiota composition during end-stage CF disease. (Zemanick and Hoffman 2016). Today, these classical CF pathogens are increasingly being supplemented with new and emerging pathogens that are rarely reported to cause disease in humans (Parkins and Floto 2015). These emerging pathogens include *Acinetobacter* spp*., Achromobacter* spp*., Inquilinus* spp., *Ralstonia* spp., *Cupriavidus* spp., and *Pandoraea* spp., and are referred to as emerging non-fermenting Gram-negative (ENFGN) species (LiPuma 2010; Parkins and Floto 2015). These organisms are often misidentified as *P. aeruginosa* or BCC and are notably multidrug-resistant (Parkins and Floto 2015). Despite the poorly-understood clinical relevance of several ENFGN species, these species have been implicated in immunocompromised infections and CF infections (Waters and Lipuma 2020). Meanwhile, prevalence of the *Streptococcus milleri* group and nontuberculous *Mycobacteria* (NTM) is steadily increasing worldwide (Ratnatunga et al. 2020).

Hatziagorou et al. (2020) have recently reported *S. aureus, P. aeruginosa,* and *Stenotrophomonas maltophilia* as having the highest prevalence among CF patients. Overall prevalence and incidence of *P. aeruginosa* has reduced significantly over recent years while the observed decrease in BCC prevalence and incidence was not significant. An opposite trend was reported with *S. aureus,* NTM, and *S. maltophilia* (Raidt et al. 2015). These changes partially reflect improved infection control following patient segregation in hospitals to prevent patient-patient transmission. Due to extensive antimicrobial therapy required in CF, a substantial increase in multi-drug resistant CF isolates have been identified, however this has not resulted in the detection of novel pathogens in CF recently (Hatziagorou et al. 2020).

16S rRNA gene sequence studies have led to a detailed account of CF core microbiota not identified through routine culture. This includes *Actinomyces, Fusobacterium, Prevotella, Veillonella, Gemella, Granulicatella, Neisseria*, *Rothia,* and *Streptococcus* which are traditionally classified as NRF (Parkins and Floto 2015; Mahboubi et al. 2016; Caverly and LiPuma 2018). These "commensal" NRF species can act as synergins, supporting the growth and proliferation of pathogens, or act as facultative pathogens themselves depending on inter and intra-bacterial species signalling, non-bacterial components of sputa, host immune response, environmental triggers, and antimicrobial therapy (Parkins and Floto 2015).

An increasing number of reports have investigated the role of mucin-degrading anaerobic lung microbiota (for example *Prevotella*, *Veillonella*, *Streptococcus*, and *Fusobacterium*) as synergins modulating and increasing the pathogenicity of *P. aeruginosa* in the CF lung (Duan et al. 2003; Flynn et al. 2016). Mucin degradation has been hypothesised to provide nutrients for primary pathogens that are unable to efficiently obtain a carbon source in the lung (Duan et al. 2003; Flynn et al. 2016). Another example includes *Streptococcus salivaris* directly antagonising *P. aeruginosa* IL-8 secretion from human bronchial epithelial cells (Rogers et al. 2010). The *S. milleri* group have also been associated with increased mucoid *P. aeruginosa* virulence in *Galleria mellonella* infection model (Waite et al. 2017). This evidence highlights the necessity to understand these complex interactions between respiratory microbiota where definitions of synergins, and commensals should be reconsidered.

1.3.2 Chronic obstructive pulmonary disease microbiota

Early COPD microbiological studies based on culture and serological tests reported approximately one third of pulmonary exacerbations were related to viral infection whereas disease aetiology was unknown in all other cases (Sethi 2010). At the time, the causal relationship between bacteria and exacerbations was not considered since COPD traditional pathogens *Streptococcus pneumoniae* and *H. influenzae* were present in similar numbers during phases of both acute exacerbations and stable disease (Sethi 2010). It is now accepted that the majority of exacerbations are infectious, and bacteria are responsible for 30-50% of these cases (Sharan 2015).

Sputum cultures often reveal *P. aeruginosa, Escherichia coli, Klebsiella pneumoniae, S. aureus, Acinetobacter baumannii, Enterobacteriaceae, Moraxella catarrhalis, Haemophilus parainfluenzae,* and *H. influenzae* to be present during acute exacerbations (Matkovic and Miravitlles 2013; El-Korashy and El-Sherif 2014; Wang et al. 2016). *P. aeruginosa* is the bacterium most commonly isolated during acute exacerbations (Wang et al. 2017). Studies have correlated microbial infection to loss of lung function, exacerbation, and disease progression (Sethi 2004; Pragman et al. 2019; Bouquet et al. 2020) although the mechanisms

by which this occurs are largely unknown (Sze et al. 2015). The large-scale prospective studies AERIS (Mayhew et al. 2018) and COPDMAP (Wang et al. 2018) are the biggest contributors to understanding lung microbiota composition during clinical stability and exacerbations. These studies among others have observed an increase in *Proteobacteria* abundance and a reduction in *Bacteroidetes* abundance (mainly *Prevotella*) during exacerbations and severe COPD disease (%FEV_1 predicted 40-50%) compared to clinical stability where *Moraxella* and *Haemophilus* are dominant components (Garcia-Nuñez et al. 2014; Dickson et al. 2016; Pragman et al. 2019; Haldar et al. 2020; Ramsheh et al. 2021). Gram-negative *Enterobacteriaceae* and nonfermenting bacilli are also characteristic of severe COPD disease (Ruby 2020). Similarly, sputum cultures often reveal *S. pneumoniae, H. influenzae, K. pneumoniae, M. catarrhalis, S. aureus* and *E. coli* being responsible for community acquired pneumonia (CAP), even though *H. influenzae* is also a core COPD component (Wang et al. 2017). These CAP infections are important as the principal reason for acute COPD exacerbations hospitalisations and are associated with poor patient outcomes (Cavallazzi and Ramirez 2020).

Importantly, bacteria overrepresented in microbiome analysis that may be causal to acute exacerbations are not necessarily recovered in sputum culture. One important example describes patients showing chronic colonisation by *P. aeruginosa* who actually experience exacerbations that are due to other potentially pathogenic microorganisms (Monsó 2017). These PPMs may represent traditional infecting pathogens, facultatively pathogenic oropharyngeal microbiota, or synergistic oral-associated microbiota. 16S rRNA gene sequence analysis has expanded the definition of core COPD microbiota to include *Corynebacterium, Prevotella, Staphylococcus, Streptococcus, Veillonella, Haemophilus, Neisseria, Fusobacterium, Rothia,* and *Serratia* (Hilty et al. 2010; Erb-Downward et al. 2011; Cabrera-Rubio et al. 2012; Garcia-Nuñez et al. 2014; Sze et al. 2015). From these analyses, non-traditional pathogens have been identified. Examples include *Fusobacterium nucleatum,* which has been reported to coaggregate with *P. aeruginosa* to facilitate invasion and modulate the inflammatory toxicity of lung epithelial cells (Li et al. 2021). The oral pathogen *Porphyromonas gingivalis* was observed to increase *S. pneumoniae* adhesion through the upregulation of platelet-activating factor receptor (PAFR) from alveolar epithelial cells (Imai et al. 2021). Additionally, *Prevotella intermedia* has been isolated from the sputum of patients with acute exacerbation of chronic bronchitis and similarly increased the expression of PAFR which led to severe pneumococcal pneumonia (Nagaoka et al. 2014; Kelly et al. 2021). The bottom line is that distinguishing between carriage of microbes not contributing toward disease and pathogens is often unclear (Filkins and O'Toole 2015).

1.3.3 Bronchiectasis microbiota

Traditional sputum culture has implicated *H. influenzae, S. pneumoniae, S. aureus* and *P. aeruginosa* as the most commonly-isolated BR pathogens (Bopaka et al. 2015; Amati et al. 2019). Like the COPD lung, *P. aeruginosa* and *H. influenzae* dominated microbiomes have been linked to severe disease and frequent exacerbations (Richardson et al. 2019; Mac Aogáin et al. 2021). Nontuberculous *Mycobacteria, M. catarrhalis, Klebsiella* spp., *S. maltophilia, A. xylosoxidans, Serratia marcescens, Alcaligenes* spp., *A. baumannii* and *Escherichia* spp. have also been occasionally isolated from sputa through both culture and culture-independent tools and have been linked to lung inflammation and poorer clinical outcomes (Verduin et al. 2002; Shimizu et al. 2015; O'Donnell 2017; Sulaiman et al. 2018; Amati et al. 2019; Lin et al. 2019; Stepman et al. 2020; Dicker et al. 2021). BR is also a heterogenous disease displaying periods of clinical stability and exacerbation.

Despite evidence pointing towards clinical and biological roles in BR exacerbations, the role of microbiota in exacerbations remain poorly understood (Amati et al. 2019). This onset could be triggered by changes in bacterial community composition, non-traditional pathogen infection, emergence of new strains, or spread of infection to a new niche in the lung (Tunney et al. 2013). 16S rRNA gene sequencing have identified *Fusobacterium, Actinomyces, Haemophilus, Neisseria, Pseudomonas, Prevotella, Rothia, Streptococcus, Serratia, Staphylococcus,* and *Veillonella* as core BR microbiota, and occasionally NTM (Rogers et al. 2013; Tunney et al. 2013; Lee et al. 2018). At present, there is an unknown link between NTM lower airway infection and oropharyngeal microbiota exhibiting increased inflammatory biomarkers (Sulaiman et al. 2018). *Prevotella* and *Veillonella* have also been postulated to be risk factors for BR exacerbations but have not been significantly associated with BR severity (Lee et al. 2018).

1.3.4 *Achromobacter* species as emerging respiratory pathogens

Achromobacter spp. are Gram-negative, lactose non-fermenting rods that are widely distributed in the environment (Jeukens et al. 2017). To date, 22 species have been validly published (Parte et al. 2020). *A. xylosoxidans* was the first species of this genus to be described in 1971 (Yabuuchi and Oyama 1971) and is most renowned for various healthcare infections and infection in immunocompromised patients (Edwards et al. 2017; Marion-Sanchez et al. 2019). This species is also most recognised for being an emerging pathogen in CF (Mahenthiralingam 2014) and has been associated with COPD and BR infection (Renom et al. 2010; Tunney et al. 2013; Liu et al. 2017). *Achromobacter* spp. are described as uncommon respiratory pathogens, however their clinical relevance has been disputed; several reports have drawn associations between *A. xylosoxidans* colonisation and lung function

decline (Pereira et al. 2011; Talbot and Flight 2016) while others found no association (Cuthbertson et al. 2020).

Worryingly, *Achromobacter* prevalence has been increasing and has been reported up to 30% in centres, which is partly attributed to high-resolution molecular tools (Somayaji et al. 2017). WGS has also established that *A. xylosoxidans* is highly multidrug resistant and is becoming resistant to last-line antimicrobials; however, its pathogenic potential remains unclear (Traglia et al. 2012; Nielsen et al. 2019). The pathogenic potential and AMR of other *Achromobacter* species are also yet to be determined.

1.4 Role of microbial diversity and microbial interactions in lung function decline

1.4.1 Microbial diversity and lung function decline

Clinical correlations have been drawn between loss of respiratory microbial diversity and poor lung function (Flight et al. 2015; Jorth et al. 2019; Richardson et al. 2019). Despite this, potential causative factors driving this association, such as presence or absence of key microbes, microbe-microbe interactions, or microbe-host interactions have not yet been confirmed (Filkins and O'Toole 2015). Furthermore, no causal link between altered microbial composition, loss of microbial diversity, and immunological disease has been drawn either from the more extensively-studied gut microbiome (Round and Palm 2018), emphasising the challenges of validating these associations.

Reduced microbial diversity and poorer lung function have often been reported in chronic lung disease where a single taxon or small number of taxa in the lung becomes dominant, for example, *P. aeruginosa*, BCC, and *S. aureus* (Dickson et al. 2013; Flight et al. 2015; Cuthbertson et al. 2020). This concept is also extended to lung microbiota traditionally classified as normal respiratory flora. For example, *Streptococcus* spp., *Granulicatella* spp., *Prevotella* spp., and *Veillonella* spp., have also been reported be less abundant in more severe CF disease (Flight et al. 2015). *P. aeruginosa* and BCC species are known to directly contribute to declining microbial diversity due to a high competitive fitness in the CF lung, which can lead to exclusion of competitors and also limit the acquisition of other pathogens (Welp and Bomberger 2020). Competition has been largely documented between traditional respiratory pathogens and rarely in emerging pathogens, e.g., *S. maltophilia* and *A. xylosoxidans* (Collins 2019; Menetrey et al. 2020). Competitive fitness is considered important in establishing airway colonisation, warranting further research into these species and additional PPMs (Menetrey et al. 2020).

1.4.2 Microbial interactions potentially influence lung function decline

It is important to note interspecies interactions can lead to synergistic or antagonistic effects that are highly specific between species and are considered to influence the community as a whole (Vandeplassche et al. 2019). Interpreting the significance of these interactions remains a challenge as species may be co-localised in the lung or derive from the oropharynx (Menetrey et al 2020). It is worth investigating lung microbiota functional response diversity during exacerbations and periods of stability. Functional response diversity refers to sensitivity and reaction to ecosystem change among species that contribute to the same ecosystem function (Leslie and McCabe 2013). Further research into genotypic (virulence and AMR) and phenotypic (biofilm production, motility, growth requirements, and metabolic capabilities including enzyme production) functional response diversity may play a key role in understanding respiratory microbiota composition and diversity changes that are associated with pulmonary exacerbations (Escalas et al. 2019).

In the wake of WGS, culture based phenotypic assays are still informative in determining bacterial niche adaptation, AMR, and virulence traits. Key examples include flagellar motility that is essential for bacterial adhesion and biofilm formation (Menetrey et al. 2020). Swimming motility has been shown to stimulate inflammation through the toll-like receptor 5 and is involved in epithelial invasion by BCC (Zlosnik et al. 2014). Swarming motility is another trait of interest, being a quorum sensing regulated phenotype that is associated with increased AMR, virulence, and host immune evasion (Kearns 2010). Bacterial biofilm formation shares a reciprocal relationship with motility and is significant in establishing chronic infections, contributing towards AMR and resistance to antiseptics (Jakobsen et al. 2013; Filipic 2017). The most renowned example in the context of CF infection is *P. aeruginosa* persisting in the lung for decades due to alginate and Psl protecting biofilms from antibiotic penetration (Kovach et al. 2017). Proteases are also considered to contribute to the pathology of CF by impairing mucociliary clearance, immune function, and perpetuating neutrophilic inflammation (Voynow et al. 2008). As a final example, bacterial toxins such as the Panton-Valentine leucocidin that is produced by *S. aureus* can lead to necrotising pneumonia (Lucas et al. 2020). While the pathogenic potential of several traditional respiratory pathogens is well known, these traits are largely understudied in emerging respiratory pathogens including *Achromobacter* spp., necessitating further research.

1.4.3 Does antimicrobial therapy contribute to poorer lung function?

Chronic respiratory infection requires aggressive antimicrobial treatment and is considered to contribute to lung function decline. Despite this long-term exposure to potent antimicrobials, CF pathogens fail to clear, and treatment failure is a common occurrence in COPD infection (Filkins and O'Toole 2015). Antimicrobial efficacy is influenced by antimicrobial resistance (AMR) as well as low oxygen environments in the lung, host epithelial cells, immune cells, pH, biofilms, and secreted metabolites to name a few (Pragman et al. 2016; Vandeplassche et al. 2019). Inhaled antimicrobials such as aztreonam and tobramycin are traditionally used as a 28-day intermittent treatment strategy (Taccetti et al. 2021). Oral antimicrobials are administered based on sputum culture positivity while intravenous antibiotics are recommended for chronic infections such as *P. aeruginosa* (Smith et al. 2014). This long-term administration of these antimicrobials has led to significant changes in lung microbial composition and diversity but there is evidence that these changes are transient (Smith et al. 2014; Bevivino et al. 2019; Taccetti et al. 2021). It is unclear whether potent antimicrobials select for the most resistant microbes associated with chronic lung disease which then contributes to lung function decline (Zemanick and Hoffman 2016). The significance of long-term antimicrobial therapy on bacterial community composition and diversity is uncertain since baseline community composition has been reported to recover in 30 days (Filkins and O'Toole 2015). Changes in sputum microbiota is not always necessarily an effect of antimicrobial therapy, thus careful consideration is required when assigning community changes as a causal relation to treatment success or lung function decline (Bevivino et al. 2019).

1.5 Routine respiratory pathogen identification and high-resolution lung microbiota profiling

1.5.1 Culture-based microbial identification and Gram stain

In vitro culture of respiratory specimens is routinely used to guide antimicrobial therapy. (Mahboubi et al. 2016). Culture is commonplace in clinical laboratories, focusing on the selective recovery of traditional respiratory pathogens that are assumed to contribute to disease, for example, *P. aeruginosa* and *Burkholderia cepacia* complex in CF infections in addition to *H. influenzae, K. pneumoniae,* and *S. pneumoniae* in BR and COPD infections (**see section 1.2**). Respiratory infection is initially diagnosed alongside signs and symptoms of cough, fever, sputum production and chest radiography (Giuliano et al. 2019). Once microorganisms are isolated, colonies are first identified by morphology, pigment production, haemolysis on blood agar, oxidase production, and catalase production (Burns and Rolain 2014). Selective culture is frequently relied on to isolate fastidious microorganisms and those

which are easily overgrown by fast growing species like *P. aeruginosa* (Table 1.1). Selective culture is also highly recommended to isolate slow growing bacteria of high clinical relevance and high transmissibility, for example, the BCC and nontuberculous *Mycobacteria* (Burns and Rolain 2014; Table 1.1). Selective culture can also offset costs and time required for subculture and confirmatory tests which encompasses antigen testing and serological testing (Laupland and Valiquette 2013).

Table 1.1 Commonly used selective media used for respiratory pathogens[a]

Organism/s	Agar
Pseudomonas aeruginosa	Cetrimide agar
Burkholderia cepacia complex	Oxidative-fermentative polymyxin B-bacitracin-lactose agar
	Pseudomonas cepacia agar
	MAST *B. cepacia* selective agar
	Remel *Burkholderia cepacia* selective agar
H. influenzae/H. parainfluenzae	Blood agar supplemented with haemin and bacitracin
	Chocolate agar supplemented with bacitracin or cefsulodin
S. aureus	Mannitol salt agar
	Chromogenic agar
S. milleri group	McKay agar
Enterobacteriaceae Lactose fermenting vs. non-fermenting bacteria	Cystine–lactose–electrolyte-deficient agar

[a]Derived from Burns and Rolain (2014).

Gram staining is also routinely applied to respiratory specimens to quickly obtain presumptive information about Gram-positive, Gram-negative, and non-bacterial infections to guide initial antimicrobial therapy (Nagendra et al. 2001). While Gram staining has been shown to be highly specific to community-acquired *S. pneumoniae* and *H. influenzae,* (Ogawa et al. 2020), a poor correlation more often exists between culture data and quantitative Gram stain (Nair et al. 2002). Sputum Gram stains are prone to variability of smear preparation where Nagendra et al. (2001) found 56% of sputum samples had at least one inconsistency and are prone to contamination from the oropharynx (Thomson 2016).

Though culture is rapid and cost-effective, it is not suitable to differentiate closely-related clinically relevant species due to phenotypic heterogeneity, excludes slow growing and strict anaerobic species and often fails to consider non-conventional pathogenic species (Bittar and Rolain 2010; Capizzani et al. 2018). Commonly isolated respiratory pathogens including *S. aureus* and *P. aeruginosa* are renowned for adaptive phenotypic changes in the CF lung including small colony variants (SCVs), hypermutable and mucoid strain variants (Caballero et al. 2014). SCVs are particularly slow growing and are associated with increased

antimicrobial resistance (Al Ahmar et al. 2020). These species often need additional subculture across various media and temperatures, and further biochemical and chromogenic testing that may take up to 72-h after isolation to confirm identification (Burns and Rolain 2014).

Matrix assisted laser desorption time of flight (MALDI-TOF) is a more recent diagnostic method adopted by clinical laboratories to resolve culture data. For example, a high level of colistin resistant non-fermenting Gram-negative species including *Achromobacter spp., S. maltophilia* and *Inquilinus limosus* are misidentified by culture and are often routinely identified through this method (Burns and Rolain 2014). MALDI-TOF works by ionising and separating particles according to their mass-to-charge ratio. This results in a unique peptidic spectrum which is matched to the spectra database of reference organisms and indicates the probability of correctly identifying the genus and species being tested (Giuliano et al. 2019). Although MALDI-TOF has been previously reported to accurately identify CF Gram-negative species compared to 16S rRNA gene sequence data (Alby et al. 2013; AbdulWahab et al. 2015), MALDI-TOF usually fails to identify several clinically important species belonging to the NTM, *Achromobacter, Acinetobacter, Ralstonia,* and *Burkholderia* (Rychert 2019).

1.5.2 Polymerase chain reaction amplification and next generation sequencing

Because culture-based methods often fail to resolve bacterial species, higher resolution DNA-based techniques including polymerase chain reaction (PCR) and molecular typing are used instead. PCR is an *in vitro* method used to exponentially amplify DNA through repeated cycles of DNA denaturation, primer annealing, and polymerase extension (Chuang et al. 2013). PCR is a powerful and rapid diagnostic tool as it detects target gene/s that are either broadly shared between bacterial species or are specific to few or individual species. Clinical laboratories often employ multiplexing to target multiple sequences/species in a single PCR and use species-specific PCR probes in their standard operating procedures (SOPs). Numerous species-specific PCR probes have been developed for the detection of respiratory pathogens including *recA* PCR amplification of *Burkholderia* species (Payne et al. 2005), *nrdA* gene amplification for *Achromobacter* spp., (Coward et al. 2016) and *rcsA* targeting *K. pneumoniae* (Dong et al. 2015) to name a few. Clinical laboratories may also opt for PCR kits designed to detect common respiratory pathogens for high throughput analysis, for example, the BioMérieux BioFire® Respiratory Panel and the PathoFinder RespiFinder® SMART 22 FAST kit, however these kits have limited use with bacteriology as they primarily detect respiratory viruses (Rytter et al. 2020).

To obtain a broad overview of microbial community composition and diversity,16S rRNA gene sequencing is traditionally employed. 16S rRNA gene sequencing has been used extensively

in the taxonomic assignment of bacterial and archaeal communities from terrestrial ecosystems, aquatic ecosystems, and the human body for its ubiquity and sensitivity in the detection of low abundance species (Chen et al. 2017; Edgar 2017; Ross et al. 2018). Next generation sequencing (NGS) platforms such as the Illumina MiSeq or Ion Torrent Personal Genome Machine (PGM) are the most popular benchtop sequencers for 16S rRNA gene sequencing (Pollock et al. 2018). 16S rRNA gene NGS typically involves amplifying short hypervariable regions of this gene using barcoded primers (Salipante et al. 2014). Although there is no consensus on which hypervariable gene region is most suitable for lung microbiota profiling, V4 amplicon sequencing is very popular in lung microbiome sequencing as it generally provides greater resolution over other hypervariable regions (Ghyselinck et al. 2013; Tremblay et al. 2015; Fouhy et al. 2016; Yang et al. 2016). Sequencing of V1-V3 regions have also been recommended (Wang et al. 2020) while V1-V2 sequencing has been discouraged due to poor resolution. (Doud et al. 2010). Regardless of choice, hypervariable gene sequencing can only identify bacteria to the genus level or family level. High resolution molecular typing and/or whole genome sequencing (WGS) often supplements 16S rRNA gene-based NGS for species-level inference.

WGS is typically selected for rapid detection of AMR genes, strain/clone level resolution, detection of virulence genes, and for outbreak investigation in hopes of leading to the earlier detection of highly transmissible pathogens (Burns and Rolain 2014). In context of lung pathogens, WGS has characterised *P. aeruginosa* pathoadaptive genes which were mostly related to AMR, gene regulation, and cell envelope components (Smith et al. 2006). WGS is also useful in a hospital outbreak scenario where (Lewis et al. 2010) determined the transmission of multidrug-resistant *A. baumannii* from military casualties to civilians. WGS is also a powerful tool to survey respiratory microbiota (metagenomics) without the bias of PCR amplification that inevitably skews relative abundances (Hauser et al. 2014). Because WGS is limited by resource allocation, high costs, and expertise, this technology is primarily a research tool; however clinical laboratories have begun to adopt this technology to target PPMs where clinical relevance is not well understood.

1.5.3 Molecular typing in the identification of respiratory pathogens

Molecular typing encompasses a large category of methods that creates a unique “fingerprint” for individual microorganisms based on extracted or amplified gDNA. Pulse field gel electrophoresis (PFGE) is considered the “gold standard” for bacteria typing according to hospital epidemiologists (Burns and Rolain 2014). This method compares genome-wide fingerprints using rare-cutter restriction enzymes, (Ballarini et al. 2012) and has been successful in the high-resolution study of the CF patient-patient transmission of *P. aeruginosa,*

S. aureus, and *H. influenzae* (Ridderberg et al. 2016). PFGE does however lack comparability between laboratories, thus alternative typing methods are commonly employed in research, depending on the goal of each study, throughput required etc. (Table 1.2; Burns and Rolain 2014). Examples include random amplified polymorphic DNA (RAPD) which has been useful in DNA typing of *P. aeruginosa* and BCC in CF (Mahenthiralingam et al. 1996a; Mahenthiralingam et al. 1996b). 16S rRNA gene terminal restriction fragment length polymorphism (T-RLFP) and ribosomal intergenic spacer analysis (RISA) offer distinct advantages in indicating community diversity and dominance, while also being capable of predicting bacterial species through unique molecular fingerprints either by *in silico* PCR amplification or through PCR of reference samples (Rogers et al. 2009; Flight et al. 2015; **see section 3.1**). RISA has been proven valuable in the identification of ENFGN missed in culture in addition to correlating loss of bacterial diversity to reduced lung function (Flight et al 2015). Multilocus sequence typing (MLST) is a high-resolution typing method often used for strain-level resolution from sequencing several housekeeping genes and grouping species into sequence types. Species-specific multilocus sequence analysis schemes exist for numerous respiratory pathogens including *Achromobacter* spp., (Spilker et al. 2012), BCC (Baldwin et al. 2005), and *K. pneumoniae* to name a few (Diancourt et al. 2005).

Table 1.2 Molecular typing application, advantages, and disadvantages

Typing method	Target	Process	Pros	Cons	Reference
Pulse field gel electrophoresis (PFGE)	Whole genome	Agarose plugs containing bacteria are lysed and genomic DNA is digested with a rare-cutter restriction enzyme. The digested DNA (~10-800 kb) is separated by alternating the electric field between electrodes. DNA reorients and migrates at different speeds through gel pores, creating unique DNA fingerprints.	Cost effective analysis Easy to interpret Highly discriminatory Widely available	Cannot differentiate all unrelated isolates Expensive equipment Labour intensive Low reproducibility	(Ballarini et al. 2012) (Sharma-Kuinkel et al. 2016)
Random Amplified Polymorphic DNA (RAPD) typing	Whole genome	Random amplified polymorphic DNA uses arbitrary 10 bp primers that may or may not amplify segments of DNA, depending on positions that are complementary to primer sequence. This results in different patterns of amplified DNA segments on a gel.	Inexpensive Highly discriminatory Rapid	Requires optimisation for each species Very sensitive to changes in PCR cycling conditions	(Mahenthiralingam et al. 1996) (Babu et al. 2021)
Ribosomal intergenic spacer analysis (RISA)	Intergenic spacer region between the 16S and 23S rRNA subunits	RISA is a rapid single step PCR that shows variation in the ITS region. ITS products can then be viewed using various electrophoretic methods.	Bacterial species prediction tool Inexpensive High throughput Rapid Reproducible Microbial diversity can be estimated through the number and concentration of ITS amplicons within a sample	Can only identify dominant and highly abundant taxa Not a definitive measure of diversity Variation in amplicon size from same species	(Flight et al. 2015)
Terminal restriction fragment-length polymorphism (T-RFLP)	Whole genome	One of the primer pairs is labelled with a fluorescent dye and amplifies a gene of interest through PCR. This amplicon is digested with restriction enzymes and the terminal restriction fragments are separated with a DNA analyser	Bacterial species prediction tool Inexpensive High throughput Microbial diversity can be estimated by analysing the number and peak heights of terminal restriction fragments Reproducible	Best suited for low to medium community diversity	(Ercolini and Cocolin 2014)
Multilocus sequence typing	Housekeeping genes	MLST includes a PCR of several housekeeping genes that can be made specific to bacterial species. Amplified housekeeping genes are then sequenced, and different sequences are provided with its own allelic profile and sequence type. MLST can identify SNPs in addition to genomic rearrangements. rMLST can also be performed where 53 ribosomal genes are analysed	High throughput Highly reproducible Highly discriminatory	Sequencing costly Time consuming	(Ballarini et al. 2012)

1.6 Outstanding methodological challenges

1.6.1 Sputum is a heterogenous and hard-to-lyse environment

Sputum is a complex mixture of mucus, inflammatory cells, epithelial cells, and cellular degradation products that is often contaminated with a substantial proportion of human gDNA, saliva, and squamous epithelial cells (Kelly et al. 2001). Sputum is a heterogenous environment, often with areas of purulence and mucus plugs that trap microorganisms resulting in uneven distribution (Burns and Rolain 2014). Sputum requires specialised extraction protocols to maximise pathogen recovery using mucolytic agents such as dithiothreitol (DTT) which is commonly used to digest sputum in order to carry out culture, Gram staining, microscopy, and gDNA extractions.

Inefficient cell lysis primarily occurs due to the physically robust peptidoglycan Gram-positive cell wall, yeasts, endospores, microbial growth status, sample matrix and host-derived extracellular components hampering the extraction process (Silhavy et al. 2010; Lopes et al. 2018; Nagler et al. 2018; Li et al. 2020). It is important to consider which type of gDNA extraction process is needed for a more even representation of microbiota abundance depending on sample material and the aim of extraction (Table 1.3). Common methods include liquid homogenisation that broadly lyse bacteria, plant, and mammalian cells; the French press is a preferred choice for lysing liquid bacterial culture by forcing liquid through a small valve at high pressure and is highly efficient (Burden 2008). Heating, freeze-thaw cycles, and cavitation (sonication) are also traditionally used to lyse bacterial cells where cavitation can also be used to disrupt endospores and fine tissue (Islam et al. 2017). As a milder lysis methodology, osmotic shock is considered more suitable for mammalian cell lysis due to the fragile structure of cell membrane (Islam et al. 2017). Mechanical lysis is by far the most popular choice to lyse cells and is typically used to disrupt microbial cells and soft, solid tissues (Islam et al. 2017). Bead-beating is generally preferred for its high lysis efficiency and can be integrated into any type of gDNA extraction protocol (Table 1.3; Table 1.4). Additionally, diverse lysis matrices are offered to lyse microorganisms and tissues allowing for flexibility in optimisation (Table 1.4).

Table 1.3 Overview of advantages and drawbacks of cell lysis methods[a]

Lysis method	Advantages	Disadvantages
Physical/mechanical		
Bead-beating	Easy optimisation parameters	Difficult gDNA purification due to complete cell disintegration
	High lysis efficiency	Heat generation leading to sample/DNA degradation
	Used for various materials and microorganisms	Higher costs
Cavitation	Can be conducted on a large scale	Difficult gDNA purification due to complete cell disintegration
	Energy efficient machinery generating less heat	Expensive
	Higher cell density can be simultaneously lysed	Heat generation leading to sample/DNA degradation
Freeze-thaw	Efficiently extract highly expressed recombinant proteins	Time consuming
		Unable to extract temperature-sensitive components
Heat shock	Recover vast quantities of intracellular components	Expensive to conduct on a large scale
	Simplicity	Heat generation leading to sample/DNA degradation
	Suitable in use with microfluidic devices	
	Used to recover large amounts of bacteria plasmid DNA	
Liquid homogenisation	Handling small volumes	Can be slow
	High lysis efficiency	Difficult gDNA purification due to complete cell disintegration
	Higher cell density can be simultaneously lysed	Heat generation leading to sample/DNA degradation
	Reproducibility	Higher costs
	Scalability	Reproducibility of sample handling
	Used for various materials and microorganisms	Scalability
Osmotic shock	Recovery of recombinant proteins	Inclusion of periplasmic proteins of unwanted organisms
	Recovery of sensitive intracellular products	Not suitable for all cell types
	Result in high lipid yield	Not efficient
Chemical lysis		
Alkaline chemicals/chaotropes	More reproducible than physical/mechanical lysis	Expensive for large scale operations
	Suitable for all cell types	Potential human health and safety risk
	Suitable for the extraction of sensitive proteins and enzymes	Very time consuming
	Used widely for bacteria gDNA and plasmid DNA extraction	
Detergents	More reproducible than physical/mechanical lysis	Expensive for large scale operations
	Rapid protein denaturation	Lower lysis efficiency
	Strength of lysing agents varies greatly, good for sensitive protein isolation to lysing tough cells	Not suitable for recovering sensitive enzymes
	Used for various materials and microorganisms	Potential human health and safety risk
		Removal of detergent following lysis is difficult
Biological lysis		
Enzymes	High specificity to target organisms	Expensive even for small batches
	More reproducible than physical/mechanical lysis	Incomplete lysis
	Suitable for extracting proteins	May be subject to long incubation times
	Used to isolate high molecular weight gDNA	Must be used with detergents to lyse bacteria

[a]Information derived from Islam et al. (2017).

Table 1.4 Application of common bead types for cell lysis[a]

Bead type	Characteristics	Application
Ceramic/porcelain	Moderate density and strength, durable, hard, does not bind nucleic acids	Bacteria, soft animal tissue, plant tissue, insects
Garnet	Dense and aggressive, highly durable, hard, does not bind nucleic acids, modifiable	Almost any sample, often fibrous animal tissue
Glass	Low density, low durability, does not bind nucleic acids, modifiable, low cost	Lyse tough cell walls of bacteria, microalgae, and fungi. Usually added with other beads to reduce shearing intensity
Silica	Least dense and softest material, low durability, small	Bacteria and yeast mainly
Silicon carbide	Low density, high strength, hard, highly durable	Lyse tough, hard, or brittle cell membranes, isolate intact organelles, super-molecular structures, RNA, and protein
Stainless steel/tungsten carbide	Most dense and hardest of bead types, highly durable, may react with DNA extraction chemicals	Lyse plant tissues, seeds, fibrous animal tissue, elastic animal tissue, resilient animal tissue, fungal fruiting bodies and RNA extraction from tough samples
Zirconium oxide	High density, highly durable, hard, modifiable	Bacteria, yeast, microalgae, soft animal tissue, leaf tissue, fibrous animal tissue, pollen, dense exterior matrices
Zirconium silicate	Moderate density, durable, hard, modifiable	Yeast, microalgae, soft animal tissue, leaf tissue, fibrous animal tissue

[a]Information derived from SPEX SamplePrep (2021).

1.6.2 PCR optimisation and standardisation of a next generation sequencing approach

It is important to understand what factors bias sequence-based analysis and how to best limit these consequences. In context of PCR, this encompasses sample processing, gDNA extraction method (see above), PCR reagents, primer design, and target hypervariable regions (**see section 1.5.2** Fouhy et al. 2016). 16S rRNA gene PCR relies on multiplexing to maximise the number of taxa detected in a bacterial community. The optimisation of multiplex PCR can pose several challenges: poor sensitivity and specificity, strong preferential amplification of certain taxa, and primer-dimer formation due to primer self-complementarity and spurious amplification as the number of degenerate primers within a reaction increases (Elnifro et al. 2000). These biases can also impede the measure of community structure (Silverman et al. 2021). Suboptimal PCR cycling conditions can also lead to erroneous conclusions regarding community composition and diversity. Low biomass samples like sputum are at a greater risk of insufficient amplification using low PCR cycle numbers while the recovery of contaminating sequences is substantial when the numbers of cycles increase (Kim et al. 2017). Automation of gDNA extraction and PCR has been recommended to limit

contamination in low biomass samples (Flight et al. 2015), though this step is not often feasible due to its incompatibility with handling many commercially available reagents and kits. Choosing reference strains relevant to the field of study is also recommended to gauge the specificity and sensitivity of PCR in a particular environment. Sequencing of mock community DNA is also common practice to estimate sequencing error and observe deviations in community composition and diversity that can be used to improve PCR (Pollock et al. 2018).

1.7 Project aims

The overall aim of this thesis was to develop a 16S rRNA gene sequencing protocol and a RISA PCR with the goal to translate this research into routine Public Health Wales diagnostics of BR, CF, and COPD lung infection. The application of high-resolution whole genome sequencing to target poorly-characterised *Achromobacter* species in CF was explored along with phenotypic testing to begin elucidating its clinical relevance. This work set out to answer key questions in respiratory microbiota research such as:

1) What type of PCR, sequencing, and gDNA extraction methods is most suitable for routine analysis of respiratory microbiota in the Public Health Wales?

2) Can 16S rRNA hypervariable gene sequencing identify a causal link between altered airway microbiota diversity and disease progression?

3) Can WGS better inform clinicians to treat and survey *Achromobacter* spp. infection?

4) To what extent can polymicrobial interactions be documented and what are the resulting implications on disease progression and antimicrobial treatment?

Specific project aims are outlined as follows:

Chapter 3 set out to optimise and validate a 16S rRNA gene sequencing protocol and a RISA PCR for the routine diagnosis of respiratory bacterial lung infection in Public Health Wales laboratories. Optimised protocols were then tested on 4 mock communities simulating CF, and COPD microbiota to streamline bioinformatic analysis, differentiate these disease-specific communities on an OTU basis and by diversity, address the impact of PCR amplification bias on sequence-based diversity and abundance analysis, and evaluate the usefulness of RISA as a prediction tool for dominant and abundant bacterial species.

Chapter 4 explored 2 types of automated and manual silica-based gDNA extraction methods and 2 types of lysis methods (bead-beating and enzymatic lysis) that most evenly represented

major respiratory pathogens including the Gram-positive *S. aureus* and Gram-negative *B. cepacia.* The superior extraction method was afterwards tested in sputum to examine how PCR detection limits differed and as a preliminary analysis to sputum microbiota profiling.

Chapter 5 harnessed optimised gDNA extraction, PCR, and sequencing methods from **chapters 3 and 4** to profile the respiratory microbiota of patients who suffer from BR, CF, and COPD. The performance of these DNA-based methods was compared against routine culture-based methods. Key themes regarding bacterial diversity influencing lung function decline and specific taxa contributing to poorer lung function were also addressed.

Chapter 6 explored the application of high-resolution whole genome sequencing to accurately characterise emerging *Achromobacter* spp., in CF. These results were compared against current routine diagnostics for this genus, MALDI-TOF. As the prevalence of *Achromobacter* species is increasing worldwide and its clinical relevance is uncertain, WGS uncovered its mode of transmission and AMR combined with antimicrobial susceptibility data. Its pathogenic potential and competitive fitness were also explored through phenotypic testing and was compared to traditional CF pathogens.

Chapter 2 General methodology

2.1 Microorganisms, culture, and MALDI-TOF

2.1.1 Microorganism collection

A panel of 20 CF and COPD bacterial species was created to include major respiratory pathogens, commensals, emerging non-fermenting Gram-negative (ENFGN) species and clinically important pathogens that are infrequent in CF/COPD (Table 2.1). This panel was selected to represent a diverse and representative group of species, covering both aerobic and anaerobic bacteria to enable the validation of a 16S rRNA gene sequencing protocol and RISA PCR (Kozich et al. 2013; Weiser et al. 2021). All aerobic species were collected from the Specialist Antimicrobial Chemotherapy Unit (SACU) and all strict anaerobic species were collected from the Anaerobe reference unit (ARU) within Public Health Wales (PHW) laboratories, University Hospital of Wales. For whole genome sequencing (WGS) studies, 28 CF-derived *Achromobacter* spp. isolates were collected from SACU (Table 2.2; **see chapter 6**). In conjunction with the phenotypic identification of these CF *Achromobacter* isolates, a smaller panel of CF pathogens was used for coculture studies which was collected from the Cardiff University (CU) E. Mahenthiralingam culture collection (Table 2.3; **see chapter 6**).

Table 2.1 Respiratory bacterial species used for PCR and genomic DNA extraction validation

Species	Origin	Featured	Role	Reference
Achromobacter xylosoxidans 26400	Clinical	Chapters 3 & 6	ENFGN in CF Opportunistic pathogen[a]	(Mahenthiralingam 2014); (Rocchetti et al. 2018)
Acinetobacter baumannii NCTC 13301	Reference	Chapter 3	ENFGN in CF Opportunistic pathogen[a] Linked to COPD exacerbation	(Falagas et al. 2005); (Grochowalska et al. 2017)
Actinomyces odontolyticus 45714	Clinical	Chapter 3	Anaerobic commensal CF microbiota member May induce lung inflammation[b] Rare COPD lung infection	(Sherrard et al. 2016); (Higashi et al. 2017)
Anaerococcus prevotii 45400	Clinical	Chapter 3	Anaerobic commensal Common respiratory flora Opportunistic pathogen[a, b]	(Murphy and Frick 2013)
Burkholderia cepacia NCTC 10661	Reference	Chapters 3 & 4	Major CF pathogen Faster lung function decline	(Fauroux et al. 2004); (Elborn 2008)
Cupriavidus pauculus 26728	Clinical	Chapter 3	ENFGN in CF rare opportunistic pathogen[a]	(LiPuma 2010)
Cutibacterium acnes 54672	Clinical	Chapter 3	Anaerobic commensal Opportunistic pathogen[a, b] Rarely linked to COPD infection	(Cobo et al. 2018)
Haemophilus influenzae ATCC 49247	Reference	Chapter 3	A commensal Major CF and COPD pathogen Reduces lung function Acute COPD exacerbation	(Sethi 2004); (Erwin and Smith 2007) (Elborn 2008)
Moraxella catarrhalis 29213	Clinical	Chapter 3	A commensal Major COPD pathogen Acute COPD exacerbation	(Verduin et al. 2002); (Sethi 2004)
Mycobacterium abscessus 1719	Clinical	Chapter 3	Emerging respiratory pathogen Faster lung function decline in CF	(Degiacomi et al. 2019); (Flight et al. 2020)
Pandoraea sputorum 19000	Clinical	Chapter 3	ENFGN in CF linked to severe lung disease[a]	(Mahenthiralingam 2014)
Prevotella melaninogenica 44495	Clinical	Chapter 3	Anaerobic commensal CF microbiota member Implicated in COPD infection May induce lung inflammation[b]	(Mahenthiralingam 2014); (Sherrard et al. 2016)
Pseudomonas aeruginosa ATCC 27853	Reference	Chapter 3	Major CF pathogen Faster lung function decline Severe pneumonia in COPD	(Elborn 2008); (Sze et al. 2015)
Ralstonia mannitolilytica 11333	Clinical	Chapter 3	ENFGN in CF Opportunistic pathogen[a] Rarely linked to COPD infection	(LiPuma 2010); (Zong and Peng 2011); (Sze et al. 2015)
Staphylococcus aureus 29213	Reference	Chapters 3 & 4	A commensal Major CF pathogen CF and COPD lung inflammation	(Sethi 2004); (Elborn 2008); (Hurley 2018)
Stenotrophomonas maltophilia 13636	Reference	Chapter 3	ENFGN in CF Opportunistic pathogen[a] Rarely linked to COPD infection	(Mahenthiralingam 2014); (Rocchetti et al. 2018); (Oladunjoye et al. 2020)
Streptococcus anginosus 25876	Clinical	Chapter 3	Anaerobic commensal CF microbiota member Acute CF exacerbation Acute COPD exacerbation	(Mahenthiralingam 2014); (Navratilova et al. 2016)
Streptococcus intermedius 25875	Clinical	Chapter 3	Anaerobic commensal CF microbiota member Acute CF exacerbation Acute COPD exacerbation	(Mahenthiralingam 2014); (Navratilova et al. 2016)
Streptococcus pneumoniae ATCC 49619	Reference	Chapter 3	Anaerobic commensal CF microbiota and linked to COPD Severe CF lung function decline Major COPD pathogen Acute COPD exacerbation	(Sethi 2004); (Paganin et al. 2015)
Veillonella parvula 22720	Clinical	Chapter 3	Anaerobic commensal CF microbiota member May induce lung inflammation[b]	(Sherrard et al. 2016); (Poppleton et al. 2017)

[a]Cause opportunistic infections but its role in lung inflammation and lung function decline is uncertain or controversial.

[b]Linked to lung inflammation however it is uncertain whether these anaerobic species are/become main causative agents of disease.

Table 2.2 Clinical CF *Achromobacter* strains obtained from SACU

Strain	Featured	Role	Reference
Achromobacter xylosoxidans 7901	Chapter 6 genomic	ENFGN in CF Opportunistic pathogen	(Mahenthiralingam 2014); (Rocchetti et al. 2018)
Achromobacter xylosoxidans 11529	Chapter 6 genomic	ENFGN in CF Opportunistic pathogen	(Mahenthiralingam 2014); (Rocchetti et al. 2018)
Achromobacter xylosoxidans 13945	Chapter 6 genomic	ENFGN in CF Opportunistic pathogen	(Mahenthiralingam 2014); (Rocchetti et al. 2018)
Achromobacter xylosoxidans 16967	Chapter 6 genomic	ENFGN in CF Opportunistic pathogen	(Mahenthiralingam 2014); (Rocchetti et al. 2018)
Achromobacter xylosoxidans 18454	Chapter 6 genomic & phenotypic	ENFGN in CF Opportunistic pathogen	(Mahenthiralingam 2014); (Rocchetti et al. 2018)
Achromobacter xylosoxidans 19179	Chapter 6 genomic	ENFGN in CF Opportunistic pathogen	(Mahenthiralingam 2014); (Rocchetti et al. 2018)
Achromobacter xylosoxidans 19519	Chapter 6 genomic	ENFGN in CF Opportunistic pathogen	(Mahenthiralingam 2014); (Rocchetti et al. 2018)
Achromobacter xylosoxidans 20408	Chapter 6 genomic	ENFGN in CF Opportunistic pathogen	(Mahenthiralingam 2014); (Rocchetti et al. 2018)
Achromobacter xylosoxidans 20777	Chapter 6 genomic & phenotypic	ENFGN in CF Opportunistic pathogen	(Mahenthiralingam 2014); (Rocchetti et al. 2018)
Achromobacter xylosoxidans 22087	Chapter 6 genomic	ENFGN in CF Opportunistic pathogen	(Mahenthiralingam 2014); (Rocchetti et al. 2018)
Achromobacter xylosoxidans 22114	Chapter 6 genomic	ENFGN in CF Opportunistic pathogen	(Mahenthiralingam 2014); (Rocchetti et al. 2018)
Achromobacter xylosoxidans 22569	Chapter 6 genomic	ENFGN in CF Opportunistic pathogen	(Mahenthiralingam 2014); (Rocchetti et al. 2018)
Achromobacter xylosoxidans 23826	Chapter 6 genomic	ENFGN in CF Opportunistic pathogen	(Mahenthiralingam 2014); (Rocchetti et al. 2018)
Achromobacter xylosoxidans 24739	Chapter 6 genomic	ENFGN in CF Opportunistic pathogen	(Mahenthiralingam 2014); (Rocchetti et al. 2018)
Achromobacter xylosoxidans 25103	Chapter 6 genomic	ENFGN in CF Opportunistic pathogen	(Mahenthiralingam 2014); (Rocchetti et al. 2018)
Achromobacter xylosoxidans 26400	Chapters 3 & 6 genomic	ENFGN in CF Opportunistic pathogen	(Mahenthiralingam 2014); (Rocchetti et al. 2018)
Achromobacter ruhlandii 23681	Chapter 6 genomic & phenotypic	ENFGN in CF Clinical importance unknown	(Edwards et al. 2017)
Achromobacter ruhlandii 23696	Chapter 6 genomic & phenotypic	ENFGN in CF Clinical importance unknown	(Edwards et al. 2017)
Achromobacter insuavis 18179	Chapter 6 genomic	ENFGN in CF Clinical importance unknown	(Edwards et al. 2017)
Achromobacter insuavis 19065	Chapter 6 genomic & phenotypic	ENFGN in CF Clinical importance unknown	(Edwards et al. 2017)
Achromobacter insuavis 19114	Chapter 6 genomic	ENFGN in CF Clinical importance unknown	(Edwards et al. 2017)
Achromobacter insuavis 19115	Chapter 6 genomic	ENFGN in CF Clinical importance unknown	(Edwards et al. 2017)
Achromobacter insuavis 19116	Chapter 6 genomic	ENFGN in CF Clinical importance unknown	(Edwards et al. 2017)
Achromobacter insuavis 19117	Chapter 6 genomic	ENFGN in CF Clinical importance unknown	(Edwards et al. 2017)
Achromobacter insuavis 19118	Chapter 6 genomic	ENFGN in CF Clinical importance unknown	(Edwards et al. 2017)
Achromobacter insuavis 20885	Chapter 6 genomic	ENFGN in CF Clinical importance unknown	(Edwards et al. 2017)
Achromobacter insuavis 20886	Chapter 6 genomic	ENFGN in CF Clinical importance unknown	(Edwards et al. 2017)
Achromobacter insuavis 23032	Chapter 6 genomic & phenotypic	ENFGN in CF Clinical importance unknown	(Edwards et al. 2017)

Table 2.3 CF pathogen panel used in the coculture of clinical CF *Achromobacter* isolates

Strain	Origin	Featured	Role	Reference
Achromobacter xylosoxidans LMG 1863	Clinical	Chapter 6	ENFGN in CF Opportunistic pathogen[a]	(Mahenthiralingam 2014); (Rocchetti et al. 2018)
Burkholderia multivorans ATCC 17616	Soil	Chapter 6	Major CF pathogen Faster lung function decline Cepacia syndrome	(Blackburn et al. 2004)
Candida albicans SC5314	Lab strain	Chapter 6	A common commensal Regarded as not clinically relevant Higher coexistence with CF bacterial pathogens	(Haiko et al. 2019)
Pseudomonas aeruginosa PA14	Clinical	Chapter 6	Major CF pathogen Faster lung function decline Highly motile Strong biofilm formation	(Elborn 2008); (Sze et al. 2015)
Pseudomonas aeruginosa LES-B58	Clinical	Chapter 6	Major CF pathogen Faster lung function decline Highly transmissible and virulent Enhanced AMR	(Salunkhe et al. 2005); (Elborn 2008); (Sze et al. 2015)
Staphylococcus aureus NCTC 12981	Lab strain	Chapter 6	A common commensal A major CF pathogen Linked to lung inflammation	(Sethi 2004); (Elborn 2008); (Hurley 2018)

[a]Cause opportunistic infections but its role in lung inflammation and lung function decline is controversial.

2.1.2 Growth media and stains

All agar and broth ingredients were purchased from Sigma-Aldrich Company Ltd. (Dorset, UK) supplying Oxoid culture media (Oxoid Ltd., Basingstoke, UK). Unless stated otherwise, growth media was prepared following manufacturer instruction. Growth media were prepared in Milli-Q ultrapure reverse-osmosis water (Merk Chemicals Ltd., Dorset, UK) and was sterilised by autoclaving (120°C, 138 kPa for 20 minutes). Plates were then allowed to cool at 50°C before pouring and setting for 30 minutes under laminar flow. Tryptic soy broth (TSB) and tryptic soy agar (TSA) was used for the routine culture of microorganisms excluding strict anaerobes, *Streptococcus* spp. and *H. influenzae* (Table 2.1-Table 2.4). Columbia blood agar (CBA) base supplemented with 5% horse blood was used for the transportation and culture of clinical CF *Achromobacter* species, containment level 3 (CL3) specimens and other aerobic/capnophilic bacteria from SACU (Table 2.1; Table 2.2; Table 2.5). CBA and TSB were used for the routine culture of *Streptococcus* species (Table 2.4; Table 2.5)*. Haemophilus influenzae* was cultured using chocolate agar supplemented with 175 g/l bacitracin or TSB supplemented with 0.02% (v/v) nicotinamide adenine dinucleotide (NAD) and 1% (v/v) haemin (Table 2.4; Table 2.5). Chocolate agar was also used for the culture of CL3 specimens. ARU anaerobic bacteria were supplied in brain heart infusion (BHI) broth with 10% serum (Table 2.6). Cystine Lactose Electrolyte Deficient (CLED) agar, cetrimide agar, mannitol salt agar, and *Burkholderia cepacia* selective agar was also used to culture CL3 specimens (Table 2.7-Table 2.10). Luria-Bertani (LB) broth tablets and 0.3% (w/v) and 0.5% (w/v) purified agar was used to create swimming and swarming motility agar (Table 2.11). 1% (w/v) skimmed-milk agar was used to

study bacterial protease activity (Table 2.12). To quantify the optical density (OD) of bacterial biofilms, a 0.1% (w/v) of crystal violet (CV) solution was prepared in Milli-Q water and was filter-sterilised using Sartorius™ Minisart™ 0.2 µm pore syringe filters (Sartorius Stedim Biotech, Surrey, UK).

Table 2.4 Tryptone soy agar/broth composition

	Components			
Weight g/l	Pancreatic digest casein	Tryptic digest soybean meal	NaCl	Agar[a]
	15	5	5	15

[a]Only applicable to TSA.

Table 2.5 Columbia blood agar base with 5% (v/v) horse blood composition[a]

	Components				
Weight g/l	Starch	Special peptone	NaCl	Agar	Horse blood[b,c]
	1	23	5	10	50 ml

[a]Both blood agar and chocolate agar were provided in pre-poured petri dishes.
[b]5% horse blood was added aseptically to agar after cooling to 50°C.
[c]Horse red blood cells lysed in chocolate agar.

Table 2.6 Brain heart infusion broth with 10% (v/v) horse serum composition

	Components						
Weight g/l	Brain infusion solids	Beef heart infusion solids	Proteose peptone	Glucose	NaCl	Na_2HPO_4	Horse serum
	12.5	5	10	2	5	2.5	100 ml

Table 2.7 Cystine lactose electrolyte deficient agar composition

	Components						
Weight g/l	Peptone	Lab-Lemco powder	Tryptone	Lactose	L-cystine	Bromothymol blue	Agar
	4	3	4	10	0.128	0.02	15

Table 2.8 *P. aeruginosa* cetrimide selective agar composition

	Components				
Weight g/l	Gelatine peptone	$MgCl_2$	K_2SO_4	Cetrimide	Agar
	20	1.4	10	0.3	13.6

Table 2.9 Mannitol salts agar composition

	Components					
Weight g/l	Lab-Lemco powder	Peptone	Mannitol	NaCl	Phenol red	Agar
	1	10	10	75	0.025	15

Table 2.10 *Burkholderia cepacia* selective agar composition

	Components											
Weight g/l	Peptone	Yeast extract	$C_3H_3NaO_3$	KH_2PO_4	Na_2HPO_4	Bile salts	$(NH_4)_2SO_4$	$MgSO_4$	Ammonium ferrous sulphate	Phenol red	Crystal violet	Agar
	5	4	7	4.4	1.4	1.5	1	0.2	0.01	0.02	0.001	12

Table 2.11 0.3% (w/v) and 0.5% (w/v) Luria-Bertani agar composition[a]

	Components			
Weight g/l	Pancreatic digest of casein	Yeast extract (low Na)	NaCl	Purified agar[b]
Standard LB agar[c]	10	5	5	15
0.3% (w/v) Swimming motility	10	5	5	3
0.5% (w/v) Swarming motility	10	5	5	5

[a]LB Lennox tablets used for lower salt content over Miller version.
[b]For high clarity and consistency in nutrient composition.
[c]Reference for regular composition of LB agar.

Table 2.12 1% (v/v) skimmed milk agar composition

	Components				
Weight g/l	Pancreatic digest casein	Tryptic digest soybean meal	NaCl	Agar	Skimmed milk[a]
	15	5	5	15	100 ml

[a]Prepared by autoclaving a 10X skimmed milk powder stock in Milli-Q water. Any resulting scum was removed, and the stock was added to TSA that was cooled to 50°C.

2.1.3 Culture maintenance and storage

Bacterial/*C. albicans* isolates were routinely incubated at 37°C, either statically or shaking at 150 rpm (**see section 2.1.2** for a detailed account of agars and broth used). *Streptococcus* spp., *H influenzae* and *Moraxella catarrhalis* required a 5% CO_2 incubator or equivalently a capnophilic jar. Up to 1 week old cultures were used to inoculate liquid media. Generally, 5 ml 18 h old broth cultures were used to set up experiments. For longer storage of bacterial working stocks (up to 1 month), cultures were incubated and stored at room temperature. On a monthly basis, microbial cultures were freshly incubated from frozen stocks. Frozen stocks were created from 24-h microbial broth cultures and 8% (v/v) dimethyl sulfoxide (DMSO; Sigma) and were stored at -80°C.

2.1.4 MALDI-TOF analysis

MALDI-TOF identification was achieved with the Bruker Biotyper MALDI-TOF mass spectrometer in SACU. A sterile wooden toothpick was used to directly transfer a single bacterial colony from 24-hour old pure cultures onto the Bruker target plate, creating a fine homogenous distribution of sample. 1 µl alpha-cyano-4-hydroxycinnamic acid (HCCA) matrix

was loaded onto the dried target plate prior to loading. All isolates were tested in duplicate and any result yielding poor identification was repeated.

2.2 Mock communities

2.2.1 Illumina MiSeq control

Mock microbial standard B (SI Table A1; BEI resources, Virginia, USA) was used to assess sequence error rates for 16S rRNA gene sequence data of CF, COPD, and BR sputa microbiota (**see chapter 5**) and was supplied through CU Genomics research hub (GRH).

2.2.2 CF/COPD mock community construction and rationale

All mock communities were constructed using extracted genomic DNA (gDNA) from 20 bacterial species (Table 2.1). Prior to construction, gDNA was quantified using a Qubit™ 2.0 fluorometer high sensitivity (HS) dsDNA kit (Invitrogen) and was diluted to 1 ng/µl. 3 mock communities were created to include a staggered community composition (Figure 2.1). Of these, 2 CF-like mock communities were designed to simulate one low-diversity community dominated by *P. aeruginosa,* and another more diverse community dominated by *B. cepacia.* A COPD-like mock community was also generated simulating stable COPD (theoretically most diverse of the 3 communities) that was dominated by *S. pneumoniae.* These staggered communities were created to assess 16S-based PCR bias at the genus level, how OTUs are clustered, and to ensure the detection of all high and low abundance species. Similarly, the RISA diversity profiles of these mock communities were compared to assess its application in the routine typing of respiratory samples. Refer to **sections 3.2.2, 3.2.4 and 3.2.5** for additional information regarding the PCR amplification, sequencing, and analysis of mock communities. For each community, the most abundant genera included major CF/COPD bacteria that are typically frequent and dominant. Medium and low abundance bacteria were represented by clinically relevant pathogens that are reportedly less prevalent in CF/COPD but capable of dominating clinical samples. Various other infrequent pathogens were included in the design of each mock community mainly for its clinical relevance or emergence in CF/COPD. Estimates of % relative abundance of species were largely derived from the results of existing studies while low abundant genera were included to ensure its detection. See Table 2.13-2.15 for a detailed account of mock community composition, rationale, and supporting literature. A fourth mock community was designed to include all 20 respiratory bacteria (Table 2.1) in equimolar amounts purely to assess 16S rRNA gene-based PCR amplification bias and OTU clustering.

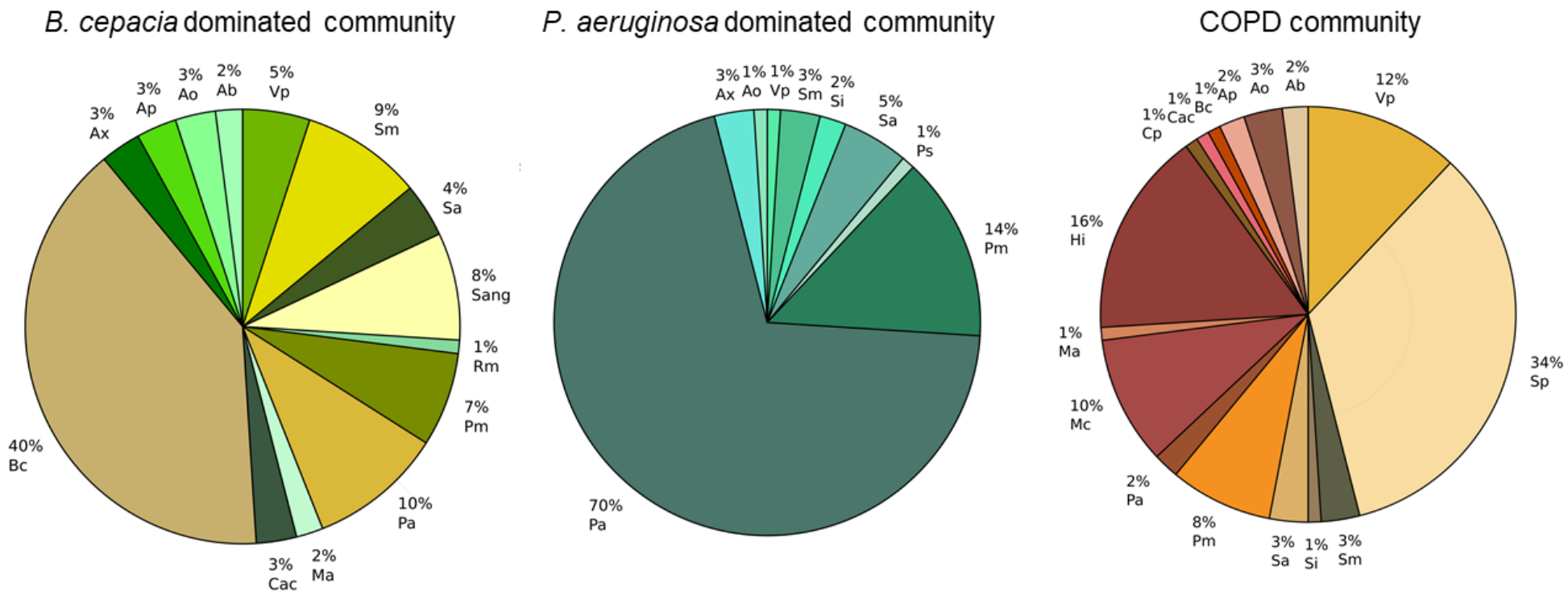


Figure 2.1 Theoretical mock microbial composition. The 3 staggered communities are illustrated where each abbreviation corresponds to the following: Ab: *A. baumannii,* Ao*: A. odontolyticus,* Ap*: A. prevotii,* Ax*: A. xylosoxidans,* Bc*: B. cepacia,* Cac*: C. acnes,* Cp*: C. pauculus,* Hi*: H, influenzae,* Ma*: M. abscessus,* Mc*: M. catarrhalis,* Pa*: P. aeruginosa,* Pm*: P melaninogenica,* Ps*: P. sputorum,* Rm*: R. mannitolilytica,* Sang*: S. anginosus,* Sa*: S. aureus,* Si*: S. intermedius,* Sm= *S. maltophilia,* Sp = *S. pneumoniae,* Vp= *V. parvula.*

Table 2.13 Rationale for CF *Burkholderia* dominated mock community construction

Genus	% abundance	Rationale[a]	References
Achromobacter xylosoxidans	3	An infrequent pathogen. May dominate in CF sputa[b]	(Mahenthiralingam 2014)
Acinetobacter baumannii	2	An infrequent pathogen that is emerging	(Sibley et al. 2010);
Actinomyces odontolyticus	3	Frequently isolated in high numbers, less abundant than *Prevotella* or *Veillonella*	(Delhaes et al. 2012);
Anaerococcus prevotii	3	Frequently isolated in high numbers, less abundant than *Prevotella* or *Veillonella*	(Verregghen et al. 2012);
Burkholderia cepacia	40	A dominant CF member and a problematic CF pathogen	(Flight et al. 2015);
Cutibacterium acnes	3	Frequently isolated in high numbers, less abundant than *Prevotella* or *Veillonella*	(O'Neill et al. 2015);
Mycobacterium abscessus	2	A rare pathogen isolated in <10% of the CF population	(Rocha et al. 2018);
Prevotella melaninogenica	7	Predominant CF anaerobe	(Ryan and Byrd 2018)
Pseudomonas aeruginosa	10	A dominant CF member and a major CF pathogen	
Ralstonia mannitolilytica	1	An infrequent pathogen. May dominate in CF sputa[b]	
Staphylococcus aureus	4	Major CF pathogen. Its abundance is reportedly variable	
Stenotrophomonas maltophilia	9	An infrequent pathogen. May dominate in CF sputa[b]	
Streptococcus anginosus	8	Its abundance is reportedly variable in CF	
Veillonella parvula	5	Dominant CF anaerobe	

[a]Information regarding bacterial species clinical relevance can be viewed in Table 2.1.
[b]Its frequency and abundance are supported by RISA data only. Both 16S rRNA gene sequence data and RISA supports the information for all other samples.

Table 2.14 Rationale for CF *Pseudomonas* dominated mock community construction

Genus	% abundance	Rationale[a]	References
Achromobacter xylosoxidans	3	An infrequent pathogen. May dominate in CF sputa[b]	(Mahenthiralingam 2014) (Sibley et al. 2010); (Delhaes et al. 2012); (Verregghen et al. 2012); (Flight et al. 2015); (O'Neill et al. 2015); (Martina et al. 2017)
Actinomyces odontolyticus	1	Frequently isolated in high numbers, less abundant than *Prevotella* or *Veillonella*	
Pandoraea sputorum	1	An infrequent CF pathogen that is emerging	
Prevotella melaninogenica	14	Predominant CF anaerobe	
Pseudomonas aeruginosa	70	A dominant CF member and the major CF pathogen	
Staphylococcus aureus	5	Major CF pathogen. Its abundance is reportedly variable	
Stenotrophomonas maltophilia	3	An infrequent pathogen. May dominate in CF sputa[b]	
Streptococcus intermedius	2	Its abundance is reportedly variable	
Veillonella parvula	1	Dominant CF anaerobe	

[a]Information regarding bacterial species clinical relevance can be viewed in Table 2.1.
[b]Its frequency and abundance are supported by RISA data only. Both 16S rRNA gene sequence data and RISA support the information for all other samples.

Table 2.15 Rationale behind the simulated COPD community abundance scores

Genus	% abundance	Rationale[a]	References
Acinetobacter baumannii	2	An infrequent pathogen that is emerging	(Lieberman 2003); (Erb-Downward et al. 2010); (Hilty et al. 2010); (Matthaiou et al. 2011); (Beasley et al. 2012); (Murphy and Frick 2013); (Li et al. 2017); (Higashi et al. 2017); (Cobo et al. 2018); (Oladunjoye et al. 2020)
Actinomyces odontolyticus	3	Infrequently reported in COPD	
Anaerococcus prevotii	2	Not reported as dominant in COPD	
Burkholderia cepacia	1	Extrapulmonary opportunistic infection in COPD patients. Rarely reported	
Cupriavidus pauculus	1	Unreported[b]	
Cutibacterium acnes	1	Infrequently reported in COPD	
Haemophilus influenzae	16	Part of the "core" COPD microbiota. Its abundance is reportedly variable	
Moraxella catarrhalis	10	Part of the "core" COPD microbiota. Its abundance is reportedly variable	
Mycobacterium abscessus	1	Rarely reported in COPD. A chronic coloniser	
Prevotella melaninogenica	8	Top 3 most abundant species in COPD	
Pseudomonas aeruginosa	2	Uncommon in stable COPD. May become dominant	
Staphylococcus aureus	3	Less frequently reported in COPD	
Stenotrophomonas maltophilia	3	Rarely reported in COPD	
Streptococcus intermedius[c]	1	Newly established presence in COPD. Prevalence is unclear	
Streptococcus pneumoniae[c]	34	Top 3 most abundant species in COPD	
Veillonella parvula	12	Top 3 most abundant species in COPD	

[a]Information regarding bacterial species clinical relevance can be viewed in Table 2.1.
[b]Included to ensure its detection at a low abundance and assess whether this isolate can be differentiated from other *Burkholderiales*.
[c]To assess OTU clustering of related species using short read sequencing.

2.3 gDNA extraction methods

2.3.1 Microbial colony gDNA extraction using Chelex®100 resin

Either 10 µl TBS bacterial culture or minimal microbial biomass from single colonies was transferred using a sterile pipette tip into 0.2 ml PCR tubes Starlab Ltd., Milton Keynes, UK) containing 50 µl 5% (w/v) Chelex® resin solution (Yang et al. 2008) that was autoclaved in a Milli-Q water solution prior to use. Samples were heated in a Veriti™ 96-well thermal cycler at 98°C for 5 minutes and then incubated in a refrigerated PCR tube rack that was placed in a -20°C freezer for 5 minutes. This process of heating and cooling was repeated once more. Samples were then centrifuged at 3000 rpm for 5 minutes to sediment the Chelex® resin and debris. 5 µl of the resulting supernatant was used as for RISA PCR amplification. Crude gDNA extracts were stored no longer than a month before PCR was performed.

2.3.2 Automated DNA extraction; Maxwell® 16 system

The Maxwell® 16 instrument (Promega, Wisconsin, USA) was used to extract bacterial gDNA from a panel of 20 respiratory bacterial species and CF *Achromobacter* species for WGS (Table 2.1; Table 2.2). Liquid cultures were either incubated overnight (24-h) in TSB or for 72-h in BHI media (strict anaerobes) at 37°C (Table 2.4; Table 2.6; **see chapter 3**) CBA plates were also prepared overnight at 37°C for the gDNA extraction of *Achromobacter* species (Table 2.5; **see chapter 6**). Liquid cultures were pelleted at 4000 rpm for 10 minutes and resuspended in 3 ml 1X sterile phosphate buffered saline (PBS). From plates, bacterial biomass was aseptically scraped and transferred into 3 ml 1X sterile PBS. Following resuspension, samples were homogenised and pelleted at 4000 rpm for 10 minutes. The supernatant was discarded, and pelleted bacteria were transferred in 300 µl 4M guanidine isothiocyanate (Invitrogen) or 500 µl guanidine isothiocyanate plus 500 µl 1X sterile PBS if a particular sample contained a significant amount of biomass. Samples were vortexed and allowed to incubate for 5 minutes at room temperature. Gram-positive bacterial samples were also lysed by bead-beating at 4.5 mph for 2 x 30 seconds with a 5-minute cooling interval (MP BIO FastPrep-24™ instrument; MP Biomedicals UK Ltd., Kent, UK) using MP BIO lysing matrix E (soil) lysis tubes with a 1.4 mm ceramic bead, 0.1 mm silica bead and 2 mm glass bead. The lysate was then centrifuged at 10000 rpm for 10 minutes and 400 µl of the resulting supernatant was loaded in the Promega Maxwell 16 platform. gDNA extraction was achieved using the Maxwell® 16 Tissue DNA purification kit following manufacturer's protocol. Extracted gDNA was quantified with the Invitrogen Qubit™ 2.0 fluorometer using the broad-range dsDNA kit.

2.3.3 Automated gDNA extraction; BioMérieux EMAG

All work involving EMAG gDNA extractions was undertaken in the PHW Pathogen Genomics Unit (PenGU). The predominant methods for extracting bacterial gDNA from the EMAG (BioMérieux Ltd., Basingstoke, UK) incorporated mechanical lysis for 2 x 90 seconds at 4000 rpm using 1 g of 1.3 mm ceramic beads (Qiagen, Maryland, USA) or chemical lysis using 25 µl MetaPolyzyme (Table 2.16; Sigma). Both methods required 250 µl of either pure TSB bacterial culture (incubated for 18 h at 37°C; **see chapter 4**) or 0.1% (w/v) Mucolyse™ pre-processed sputum (Pro-lab diagnostics, Ontario, Canada; **see chapter 4**). TSB samples were pelleted at 4000 rpm for 10 minutes before resuspending in 1X sterile PBS. All samples were inactivated in 0.9 ml 5M guanidine isothiocyanate (BioMérieux) prior to mechanical/chemical lysis and were incubated at room temperature for 10 minutes.

For bead-based lysis, the resulting lysate often rose above tube cap "O" rings, posing as a significant hazard when handling CL3 derived samples. As a result, samples were required to

be split into 2 lysis tubes before bead-beating, following PHW sample handling protocol. Samples were lysed at 4000 rpm (5 minutes of 2 x 30 seconds with 5-minute intervals of cooling in between) using the Precellys 24 tissue homogeniser (Bertin instruments, Montigny-le-Bretonneux, France; **see chapter 4** for lysis optimisation). Afterwards, tubes were allowed to stand for 10 minutes to settle the lysate which was then centrifuged at 13000 G for 1 minute. 400 µl supernatant was transferred into 2 ml 5M guanidinium isothiocyanate, following PHW protocol. To maximise gDNA recovery, ceramic beads were washed with 100 µl lysis buffer and centrifuged once again at 13000 G for 1 minute. Approximately 200 µl of additional sample was recovered and transferred into 2 ml lysis buffer. Using a sterile plastic Pasteur pipette, the total volume of lysis buffer and sample (600 µl) was transferred into a NucliSens® easyMAG® cartridge (BioMérieux) and automated gDNA extraction was performed as per on screen instructions (validated by PHW Mycology Regional Reference Unit) resulting in a 60 µl elution volume.

The pre-processing of bacterial samples differed when using MetaPolyzyme, where pelleted pure TSB cultures were washed twice in 1 ml of 1X sterile PBS. 250 µl of resuspended sample was transferred to an Eppendorf tube containing 25 µl MetaPolyzyme with 0.1% (v/v) of sodium azide. Samples were incubated for 3 hours at 37°C and then transferred to 5M lysis buffer as described above. 400 µl of the lysate was added to an NucliSens® easyMAG® cartridge for automated gDNA extraction. For this chemical-based lysis method, 250 µl thinned sputum was directly added to MetaPolyzyme and sodium azide which was transported in heat sealed bags. Further information regarding MetaPolyzyme sputum processing and sample information can be viewed in **chapter 4**.

Table 2.16 MetaPolyzyme composition and uses[a]

Enzyme	Purpose
Achromopeptidase	Strong lysis of Gram-positive aerobic bacteria, especially those with A1α and A3α cell wall chemotype
Chitinase	Degrades chitin enzymatically to N-acetylD-glucosamine. Effective against fungi
Lysostaphin	Contains hexosaminidase, amidase, and endopeptidase activities. Cleaves polyglycine crosslinks in the cell wall of *Staphylococcus* spp.
Lysozyme	Degrades the peptidoglycan cell wall of Gram-positive bacteria through the hydrolysis of β(1→4) linkages between N-acetylmuraminic acid and N-acetyl-D-glucosamine peptidoglycan residues and N-acetyl-D-glucosamine residues in chitodextrin
Lyticase	Degrades linear glucose polymers with β(1→3) linkages, effective against yeasts and sphaeroplast formation
Mutanolysin	A muramidase, particularly cleaves the β(1→4) bond in MurNAc-GlcNAc Gram-positive cell wall

[a]Information obtained from the product technical bulletin from Sigma.

2.3.4 Column based gDNA extraction; Zymo kit

Similar to EMAG extractions (**section 2.3.3**), ZymoBIOMICS™ DNA miniprep kit (Cambridge Bioscience, Cambridge, UK) gDNA extractions required 250 µl of either pure TSB bacterial culture (incubated for 18 h at 37°C; **see chapter 4**) or 0.1% (w/v) Mucolyse™ pre-processed sputum (Pro-lab diagnostics, Ontario, Canada; **see chapter 4 and chapter 5**). TSB samples were pelleted at 4000 rpm for 10 minutes before resuspending in 1X sterile PBS. Resuspended TSB cultures/Mucolyse™ treated sputum was transferred into 2 ml lysis beads containing 0.1 mm and 0.5 mm ZR BashingBead lysis matrix. 750 µl of an ethylenediaminetetraacetic acid (EDTA) disodium salt-based lysis solution was transferred into lysis tubes and was incubated at room temperature for 10 minutes to safely inactivate microorganisms. Following incubation, bead-based lysis was performed in the Precellys 24-homogeniser set to 4000 rpm for 5 minutes (2 x 30 seconds with 5-minute intervals of cooling in between; **see chapter 4** for lysis optimisation). Following lysis, samples were centrifuged at 13000 G for 1 minute and allowed to stand for 5 minutes to settle the lysate. gDNA extraction procedure largely followed manufacturer instruction where 400 µl of the resulting supernatant was transferred to a II-F spin column and after filtering, 1200 µl DNA binding buffer was added. The resulting solution was homogenised thoroughly by pipetting. gDNA elution was achieved by incubating a IICR column with nuclease-free water for 2 minutes and filtering, resulting in a 60 µl elution volume. Following gDNA elution, gDNA was purified using a ZymoBIOMICS™ HRC prep solution.

2.4 Routine PCR amplification

Routine PCR included the amplification of 5 µl gDNA in a 25 µl volume containing 1 µl primer (Table 2.17), 6.5 µl nuclease-free water (Severn Biotech Ltd., Kidderminster, UK), and 12.5 µl AccuStart II PCR ToughMix. This PCR master mix contained a 2X optimised solution of $MgCl_2$, deoxynucleoside triphosphate (dNTPs), reaction buffer and hot-start DNA polymerase (VWR). At all times, PCR amplification was performed in a Veriti™ 96-well thermal cycler (Applied Biosystems, California, USA). Here, the optimised PCR conditions for a RISA PCR and a 16S rRNA gene PCR that were routinely used are illustrated below (Table 2.18; Table 2.19; **see sections 3.2.2** and **3.2.3** for further information about optimisation). A *recA* gene PCR was also employed using AccuStart reagents to confirm the recovery of *B. cepacia* from gDNA extraction (Payne et al. 2005; Table 2.20; **see chapter 4**). The V4 region of the 16S rRNA gene was amplified with barcoded primers 16Sf and 16Sr for Illumina library preparation (Kozich et al. 2013; Table 2.17). The 16S-23S intergenic transcribed spacer (ITS) region was

targeted with primers 1406F and 23SR (Borneman and Triplett 1999; Table 2.17). The *recA* gene was amplified using primers BUR1 and BUR2 (Payne et al. 2005; Table 2.17).

Table 2.17 Primer sequences used for validation and routine PCR

Primer	Sequence 5'-3'
16Sf	GTGCCAGCMGCCGCGGTAA
16Sr	GGACTACHVGGGTWTCTAAT
1406F	TGYACACACCGCCCGT
23SR	GGGTTBCCCCATTCRG
BUR1	GATCGARAAGCAGTTCGGCAA
BUR2	TTGTCCTTGCCCTGRCCGAT

All resulting amplicons were first visualised by 1% agarose gel electrophoresis, run at either 90V for 30 minutes for RISA or 120V for 30 minutes for 16S rRNA gene and *recA* gene PCR products. For greater resolution,16S rRNA gene amplicons were also visualised using Agilent 2200 TapeStation D1000 or high sensitivity (HS) D1000 ScreenTapes, and D5000 ScreenTapes (Agilent Technologies, California, USA). The Qiagen QIAxcel was also used to visualise mixed RISA amplicons (Qiagen; **see 2.5.3**).

Table 2.18 Routine RISA PCR protocol using AccuStart II PCR ToughMix kit reagents

	Temperature	Time
Initial denaturation	94°C	2 minutes
Denaturation	94°C	30 seconds
Annealing	55°C	30 seconds
Extension	72°C	1 minute
Final extension	72°C	5 minutes
Hold	4°C	Indefinitely
Final primer concentration	200 nM of each forward and reverse primer	
DNA template	5 µl	
PCR cycles	35	

Table 2.19 Routine 16S rRNA PCR protocol using AccuStart II PCR ToughMix kit reagents

	Temperature	Time
Initial denaturation	94°C	2 minutes
Denaturation	94°C	30 seconds
Annealing	55°C	15 seconds
Extension	68°C	30 seconds
Final extension	68°C	5 minutes
Hold	4°C	Indefinitely
Final primer concentration	100 nM of each forward and reverse primer	
DNA template	5 µl	
PCR cycles	35	

Table 2.20 *recA* gene PCR modified protocol using AccuStart II PCR ToughMix kit reagents[a]

	Temperature	Time
Initial denaturation	94°C	2 minutes
Denaturation	94°C	30 seconds
Annealing	60°C	30 seconds
Extension	72°C	45 seconds
Final extension	72°C	5 minutes
Hold	4°C	Indefinitely
Final primer concentration	100 nM of each forward and reverse primer	
DNA template	5 µl	
PCR cycles	35	

[a]A modified protocol as a presence/absence test for *Burkholderia* species recovered from sputum. This assay did not require optimisation as it was not targeted for routine PCR detection of species and worked effectively using this protocol.

2.5 DNA visualisation methods

2.5.1 Gel electrophoresis

All PCR products were visualised using a 1% (w/v) agarose gel: 1 g of agarose powder (Severn Biotech Ltd.) in 100 ml 1X Tris-acetate-EDTA buffer (TAE) (Severn Biotech Ltd.) was heated in a microwave for 2 minutes. The dissolved solution was mixed and allowed to cool before adding 0.005% (v/v) SYBR™ Safe DNA Gel Stain (Invitrogen, California, USA). The resulting gel set at room temperature for 30 minutes before loading 5 µl PCR product with 1 µl of 10X bromophenol blue loading dye (Severn Biotech Ltd.). For reference, 5 µl of a 200 bp-10 kb ladder was included (Meridian Bioscience, Ohio, USA). To visualise 16S rRNA gene amplicons, gel electrophoresis was performed at 100V for 30 minutes while 80V was applied to visualise RISA PCR products for greater separation of multiple DNA fragments. Gels were visualised using the Bio-Rad VersaDoc™ MP 4000 molecular imaging system (Bio-Rad). To estimate DNA size and concentration more clearly and accurately, the Agilent 2200 TapeStation or Qiagen QIAxcel Advanced System was employed following successful visualisation of PCR amplicons (**see sections 2.5.2 and 2.5.3**).

2.5.2 TapeStation system capillary electrophoresis

The Agilent TapeStation was used to obtain more accurate estimations of DNA size, concentration, molarity, and to visualise possible DNA degradation or RNA carryover following gDNA extraction. The TapeStation 2200 was the principal instrument used. In use with these systems, a gDNA ScreenTape was used to quantify and assess the quality of gDNA to be used either for WGS or validating bacterial gDNA extraction methods (**see chapter 4 and chapter 5**). HS D1000 ScreenTapes (range of 15 bp-1 kb fragments between 10-1000 pg/µl)

were used for 16S rRNA gene PCR limit of detection studies (LoD), 16S rRNA gene library quality control (QC) and calculating average 16S rRNA gene library size and molarity. Standard D1000 ScreenTapes were used to analyse higher concentration DNA amplicons (range of 0.1-50 ng/µl) such as RISA products of ≤2 species in a sample (**see chapter 4**). D5000 ScreenTapes were also used to visualise RISA products (100 bp-5 kb) and for generating single species RISA references (Table 2.1; **see chapter 3**). Samples were processed following manufacturer instructions.

2.5.3 Qiagen QIAxcel Advanced system capillary electrophoresis

The QIAxcel Advanced System was used to analyse RISA profiles of mock microbial DNA, sputum microbiota, and the RISA profiles of CF *Achromobacter* species in coculture with several CF pathogens (Table 2.1; Table 2.2; Table 2.3; **see chapters 3, 5 and 6**). RISA profiles were analysed using the following: a high-resolution DNA cartridge, a 15 bp-3 kb QX DNA alignment marker and a 100 bp-2.5 kb QX DNA size marker. PCR amplification resulted in 10-100 ng/µl DNA amplicons which led to the selection of a 30 ng/µl DNA size marker along with a predefined "0M500" method for sample processing: a 5 kV sample injection voltage was used for 10 seconds, and a 5 kV separation voltage was applied for 500 seconds). Though not as sensitive as the TapeStation HS D1000 ScreenTapes, this method allowed for sufficient detection of low concentration amplicons (~0.1 ng/µl) and resolution between 500 bp-1 kb while retaining profiles of larger DNA fragments >1 kb.

2.6 Illumina 16S rRNA gene library preparation and sequencing

2.6.1 Dual-indexed library preparation

The Illumina MiSeq platform was used to sequence 16S rRNA barcoded libraries (Kozich et al. 2013; SI Table A2; **see chapters 3 and 5**). The V4 gene region was amplified using primers 16Sf and 16Sr following the procedure in **section 2.4**. The resulting amplicons were viewed on a 1% agarose gel and a D1000 ScreenTape using the Agilent 2200 Tapestation.

2.6.2 Library purification

Beckman Coulter™ Agencourt AMPure XP paramagnetic beads (Fisher) were used to purify 16S rRNA gene PCR amplicons using a ratio of 0.8X bead solution to sample. Library purification followed manufacturers' protocols. The quality and yield of purified amplicons were checked with a Qubit™ 2.0 fluorometer HS dsDNA kit and 2200 TapeStation HS D1000 ScreenTapes.

2.6.3 Library pooling, denaturation and loading

Following purification, libraries were individually diluted to approximately 4 nM each following Illumina's formula:

(concentration in ng/µl) / (660 g/mol x average library size in bp) x 10^6

DNA molarity for each sample was calculated using Qubit DNA concentration measurements whereas average library size was determined by averaging the entire area of an electropherogram containing peaks in the TapeStation 2200 software. Following dilution, libraries were pooled in equimolar amounts, and scaled to/near the average library molarity to allow for correct scaling of sample of ~5 µl of each library and nuclease-free water. Negative controls were "diluted" to 4 nM by using an averaged library concentration score, essentially treating these controls as a real sample. This was to avoid overrepresentation of sequences that would be rare or absent in real samples. Likewise, the molarity of negative controls was taken to be the average molarity of all libraries when pooling. Library denaturation and loading was undertaken by CU GRH staff where libraries were denatured using 0.2 N NaOH. Denatured libraries were diluted and loaded following Illumina MiSeq instructions. Libraries were diluted to 12.5 pM which resulted in good clustering density using a V2 nano cartridge (500 cycles for paired-end sequencing of a ~250 bp product). 15% PhiX control was added to denatured libraries due to low diversity of 16S rRNA gene sequences. Finally, this V4 16S rRNA gene sequencing protocol required the addition of 3 µl of 100 nM V4 sequencing primers to the cartridge (SI Table A2).

2.6.4 16S rRNA gene sequence data analysis

All 16S rRNA gene sequence data was analysed through the Mothur pipeline (version 1.44; Schloss 2010). Poor quality and erroneous sequences were removed using a threshold of a maximum of 8 homopolymers, 0 ambiguous sequences and sequences above 275 bp or below 250 bp. Erroneous sequences from sequencing were removed using a threshold of 2 mismatches from each sequence. VSEARCH (Rognes et al. 2016) was employed to remove chimeric sequences. Non-bacterial sequences were also identified and removed from the dataset.

The SILVA database (Pruesse et al. 2007) was used to classify OTU taxonomies to the genus level using the Wang approach (Wang et al. 2007). In addition to SILVA, both RDP V16 (Cole et al. 2009) and Greengenes v13.8.99 (DeSantis et al. 2006) were used to classify OTU taxonomies using the Wang approach in order to compare the performance of each reference database (**see chapter 3**). Sequences were grouped into OTUs using dist.seqs and

cluster.split Mothur commands based on a 0.03 distance label, equivalent to 97% sequence similarity. Percent relative abundance was calculated for all phyla and/or genera identified in each sample using Mothur consensus taxonomy and raw relative abundance data in R (R-Core-Team 2020).

To perform diversity analysis and Metastats (differential OTU comparison) function, all sequenced reads were subsampled to the group containing the lowest number of sequences. Rarefaction curves were calculated as the number of OTUs observed as a function of the number of reads. Bacterial alpha diversity was calculated by comparing observed OTU numbers. The Chao1 richness estimator (Chao1) and abundance-based coverage estimators (ACE) were also used to calculate richness while inverse Simpson (InvSimpson) and Shannon indices were used to measure diversity. Good's coverage estimates were included to note the percentage of total bacterial OTUs represented in a sample.

Dist.seqs, a command used to calculate uncorrected pairwise distances between aligned DNA sequences were used along with Clearcut (Evans et al. 2006) for a distance-based phylogenetic tree construction. This output was then used to calculate beta diversity using both weighted and unweighted UniFrac commands (Lozupone et al. 2007). Non-metric Multi-dimensional Scaling (NMDS) plots were generated from the square distance matrices from UniFrac distance matrices in R. The corr.axes command calculated the correlation coefficient (spearman method) for each column in the shared file and NMDS axes created in Mothur. This output could then be used to create a biplot to plot the coordinates of OTUs which significantly shifted community membership and structure.

2.6.5 Population biology statistics

All statistical tests were performed in (R Core Team 2020), apart from a population level analysis completed in Mothur using Metastats command (White et al. 2009) which performs non-parametric t-tests to assess which OTUs are differentially abundant in communities. Alpha diversity was compared either using a one-way ANOVA (**see chapter 5**), or a linear mixed effects regression model (**see chapter 3**). The R package *vegan* (Oksanen et al. 2019) was used to perform permutational multivariate analysis of variance (PERMANOVA) on the distance matrices (Anderson 2001). PERMANOVA was performed to assess whether the centroids of each category were significantly different, indicating the diversity of communities between categories is more diverse than within categories. Multivariate homogeneity of groups dispersions, a multivariate analogue of Levene's test for homogeneity of variances was run using the betadisper function of the *vegan* package.

2.7 Achromobacter whole genome Illumina sequencing

2.7.1 Illumina Nextera XT library preparation

The Nextera XT DNA library preparation kit was used to generate whole-genome libraries for Illumina MiSeq sequencing of CF *Achromobacter* isolates in the PHW Pathogen Genomics Unit (Table 2.2; **see chapter 6**). Library preparation followed the Illumina Nextera XT kit protocol except for the following: DNA libraries were diluted to 0.3 ng/µl using the Eppendorf epMotion 5075 (Eppendorf Ltd., Stevenage, UK) and diluted DNA was incubated with tagment mix for 3 minutes. Modifications to the kit protocol resulted in libraries yielding 800 bp fragments on average which was confirmed using D1000 and D5000 ScreenTapes on the 2200 TapeStation.

2.7.2 Whole genome library purification

Agencourt AMPure XP paramagnetic beads were also used for the purification of whole-genome libraries where purification followed Nextera XT library cleanup protocol, requiring a 0.5X ratio of beads to sample. The quality and yield of purified amplicons were checked with a Qubit™ 3.0 fluorometer HS dsDNA kit and 2200 TapeStation D1000 and D5000 ScreenTapes. Libraries generating fragments at ≥1000 bp were discarded prior to pooling and loading.

2.7.3 Library pooling, denaturation and loading

Library pooling, loading, and sequencing was undertaken by CU GRH staff following the Illumina Nextera XT protocol for library normalisation and pooling, then following the Illumina MiSeq protocol for denaturation (NaOH) and loading compatible with a V3 cartridge. With V3 reagents, coverage for libraries was estimated to be up to 60X. For library quantification, gDNA was quantified using the Invitrogen Qubit™ 3.0 fluorometer HS dsDNA kit.

2.8 CF Achromobacter phenotypic testing

2.8.1 Bacterial growth curve estimation

Bacterial optical density (OD) was measured using the Bioscreen C instrument (Labsystems, Finland). Bacterial isolates were incubated for 18 h at 37°C in TSB. Bacterial OD was measured using a spectrophotometer set to a 600 nm wavelength and the resulting isolates were diluted to approximately 10^6 CFU/ml. Afterwards, 200 µl diluted culture was transferred into the honeycomb wells of a Bioscreen microplate. Details of samples, temperature parameters and time parameters can be seen in **chapter 6**. For each experiment, OD

measurements were taken at 15-minute time intervals using a 450-580 nm wide band filter. Note the direct growth of *C. albicans* could not be determined using the same spectral bandwidth and was omitted from experiments. Every 10 minutes, the microplate was agitated to mix samples. Growth data was generated twice with 3-4 technical replicates per experiment (**see chapter 6**).

2.8.2 Growth curve data generation and statistical analysis

Growth data was exported into Microsoft Excel from the Bioscreen C EZExperiment software. Background correction was first achieved by subtracting the mean broth-only (negative control) OD values from all samples. Raw, mean OD values for each isolate was used to firstly generate scatterplots in R using *ggplot2* (Wickham 2016). Growthcurver (Sprouffske and Wagner 2016) was used to estimate growth parameters including carrying capacity (*k*), doubling time (*g*), growth rate (μ) and maximum OD reached. Growth parameters were estimated using logistic regression. Lag time (λ) was determined in Grofit (Kahm et al. 2010). Additionally, predicted OD values for each sample was included in each plot to visualise goodness of fit. Goodness-of-fit parameters, R^2 and root mean square error (RMSE) were calculated for each sample. Growth rates were statistically analysed with a kruskal-Wallice test which was analysed *post-hoc* using the Dunn test and an aligned Rank Transform ANOVA with Benjamini-Hochberg correction for multiple comparisons.

2.8.3 Swimming and swarming motility assays

Bacterial motility was assessed through its swimming and swarming capacity (**see chapter 6**). The methodology to these motility assays were derived from (Weiser 2015), who adapted the methodology from (Rashid and Kornberg 2000). 0.3% (w/v) swimming plates and 0.5% (w/v) swarming plates (Table 2.11) were uniformly poured (20 ml) and dried under laminar flow for 20 minutes to preserve comparable moisture content. Motility was first qualitatively assessed using 18 h undiluted TSB cultures of CF pathogens (Table 2.3) and a selection of CF *Achromobacter* isolates representing different diversity clades (data not shown). Later, the same CF pathogen panel and 6 *Achromobacter* isolates (Table 2.2) was standardised to 10^7 CFU/ml (using a spectrophotometer set to 600 nm) to statistically compare the motility capacity of the respective isolates in pure culture and in coculture. Swimming motility was assessed by inoculating liquid cultures with a sterile toothpick through the agar surface. In contrast, swarming plates were inoculated on the surface to promote bacterial growth over the agar. All plates were incubated for 48-h at 37°C and were stacked in triplicate in bags to prevent dehydration and maintain similar moisture content per plate. For the same rationale, plates were placed in the exact same area of the incubator. At 24-hour intervals, motility was measured across 4 radii using a ruler. Isolates were classified as non-motile (radius ≤5 mm),

motile (radius >5 mm and ≤30 mm) and highly motile (radius >30 mm; definitions modified from Weiser 2015). 2 biological replicates and 3 technical replicates were generated per experiment and per strain/interaction.

2.8.4 Biofilm formation on an abiotic surface experiments

Approximately 10^5 CFU/ml 18 h TSB bacterial cultures (incubated at 37°C) were used to quantify bacterial biomass produced by biofilm formation modified from (O'Toole and Kolter 1998) (**see chapter 6**). In pure culture assays, 100 µl liquid culture was transferred to a well of a clear Nunclon delta 96 well plate (Thermo Fisher Scientific, Massachusetts, USA) to yield 6 technical replicates per isolate and 3 biological replicates overall. For coculture studies, 6 *Achromobacter* isolates representing 3 species with varying biofilm formation capacity were selected to assess biofilm formation in coculture with a small CF pathogen panel (Table 2.2; Table 2.3). As previously mentioned, bacteria used in coculture studies were diluted to ~10^5 CFU/ml and mixed thoroughly as a 1:1 ratio prior to inoculating wells with 100 µl of this inoculum. For every plate, a control consisting of TSB absent of bacterial growth was incorporated to subtract the background OD and to check for contamination. Plates were incubated statically at 24-h and 48-h, and were bagged to prevent further evaporation. Plates were placed in the exact same incubator area and wells encompassing the plate edge were always kept empty to maintain comparatively similar moisture content.

Following incubation, the resulting bacterial growth in each well was carefully decanted and wells were washed twice with 200 µl sterile distilled water. To dispose of any residual broth, plates were submerged in Tupperware filled with sterile distilled water for approximately 10 seconds, mildly agitating plates. Plates were then blotted dry before adding 200 µl 0.1% crystal violet (CV). Samples were then left to incubate with CV for 30 minutes at room temperature. Following this step, excess CV was decanted from the wells and wells were washed twice in 200 µl sterile distilled water. Plates were blotted dry and allowed to air dry at room temperature overnight. To quantify CV OD, the dried stain was reconstituted in 200 µl 70% ethanol, allowing the stain to fully dissolve at room temperature for 15 minutes. A plate shaker was used to fully homogenise the solution for 1 minute at 500 rpm. CV OD was measured using a Tecan plate reader (Tecan Group Ltd, Männedorf, Switzerland). This method measured and averaged the OD of each well at 570 nm from 4 readings per well. OD obtained from broth controls were subtracted from each sample to obtain true readings.

2.8.5 Protease assay

Protease production was examined in a panel of 6 CF pathogens and 6 CF *Achromobacter* isolates (Table 2.2; Table 2.3; **see chapter 6**). Protease production was compared in single culture and coculture, taking note of isolates producing protease and how this production might differ in coculture (**chapter 6**). Protease assays were created following the protocol of (Ash and Ramteke 2018) where 18 h TSB cultures (incubated at 37°C) of each isolate were diluted to approximately 10^7 CFU/ml using a spectrophotometer reading at 600 nm. For coculture experiments, a 1:1 ratio of bacteria in coculture was created and homogenised prior to use. 10 µl bacteria inocula were spotted on 1% skimmed milk agar (Table 2.12) and plates were allowed to dry for 30 minutes before incubating at 37°C for 24-h and 48h. The resulting zones, if any, were measured across 4 radii using a ruler.

2.8.6 RISA profiling of the active growing edge of motility plates

RISA PCR was employed to assess the presence of 2 motile pathogens or whether a non-motile pathogen was also present in the active edge of bacterial growth of swimming and swarming motility plates (Collins 2019; Table 2.2; Table 2.3; **see section 2.8.3; see chapter 6**). Typically, 48-h growth was used with the exception of some interaction swimming plates with *P. aeruginosa* PA14: due to its rapid growth, 24-h growth was used instead of 48-h. 3 of the 6 total replicates per interaction were selected at random for this assay. For each sample, a sterile 200 µl pipette tip was inserted through surface bacterial growth and the surrounding agar underneath. This portion of bacterial culture was transferred to 50 µl Chelex® resin by gently scraping material into a sterile 0.2 ml PCR tube using a sterile pipette tip. DNA was extracted and RISA PCR amplified as described in **sections 2.3.1 and 2.4**. The resulting amplicons were visualised using a HS DNA cartridge using the Qiagen QIAxcel (**see 2.5.3**) where both DNA ITS size/s and concentration was noted.

2.8.7 RISA profiling of CF bacteria ratios in a biofilm

The exact methods outlined in **section 2.8.4** were used to create 96-well samples of CF bacteria in coculture (modified from Collins 2019; Table 2.2; Table 2.3; **see chapter 6**). Here, all growth at 24-h was taken for analysis whereas 48-h growth was taken only if a significant difference in biofilm OD measurements between the two timepoints occurred. Following incubation at the respective time and temperature (**see section 2.8.4**), wells were mixed gently by pipetting, and 10 µl of inoculum was transferred to a sterile 0.2 ml PCR tube containing 50 µl Chelex® resin. DNA extraction by chelation was performed (**see section 2.3.1**) and the resulting DNA was RISA PCR amplified (**see section 2.4**). 3 biological and 3 technical replicates were generated and of these, 3 replicates were chosen at random where resulting amplicons were visualised using the Qiagen QIAxcel (**see 2.5.3**).

Chapter 3 Development of a 16S rRNA gene sequencing protocol and ribosomal intergenic spacer PCR, and its validation in artificial bacterial communities

3.1 Introduction

Polymicrobial infections drive the progression of chronic lung diseases and can profoundly affect clinical outcomes. The composition of polymicrobial communities is unique to each patient, complicating individual response to treatment (Filkins and O'Toole 2015). To combat complex respiratory infections, it is vital to accurately define these polymicrobial communities.

In the context of PHW CF and COPD bacteriology, culture-based methods remain in the forefront of clinical diagnostics. Culture-based identification employs both selective and nonselective media to recover major and emerging respiratory pathogens that are associated with the evolution of pulmonary disease (Lipuma 2010; Zemanick and Hoffman 2016). Despite its uses, microbial culture is time-intensive, favours the growth of fast-growing organisms, neglects anaerobic and/or unculturable species and is uninformative about changes in microbial diversity associated with lung function decline (Flight et al. 2015; Watterson et al. 2020). Similarly, conventional biochemical tests fail to identify clinically important emerging non-fermenting Gram-negative (ENFGN) bacteria, *Burkholderia cepacia* complex (BCC) species and *P. aeruginosa* strains, among others (Caballero et al. 2014; Tabatabaei et al. 2019). Due to this lack of resolution, biochemical tests have been largely excluded from PHW CF and COPD bacteriology. Instead, matrix-assisted laser desorption ionisation time-of-flight (MALDI-TOF) is routinely employed to formally identify cultured specimens. This method rapidly produces species-specific peptidic spectra, however, it frequently fails to differentiate *Achromobacter, Acinetobacter, Ralstonia,* and *Burkholderia* species (Burns and Rolain 2014).

To achieve a more accurate and detailed insight, PCR-based NGS methods are recommended alongside bacteria culture. 16S rRNA gene sequencing is most frequently used to capture a broad range of bacteria genera while species-specific probes can be developed for species level resolution. Such examples comprise of a 16S rRNA gene and *recA* multiplexed qPCR of BCC species, and *ecfX* and *gyrB* qPCR targeting *P. aeruginosa* (Bittar and Rolain 2010). Owing to its simplicity and higher discriminatory power, 16S rRNA gene sequencing has largely superseded outdated and less accurate culture and biochemical methods in research. 16S rRNA gene NGS typically involves amplifying a short hypervariable region(s) of the gene using unique barcoded primers to enable highly multiplexed sequencing runs (Salipante et al. 2014). Historically, 16S rRNA gene-based metaprofiling relied on Roche 454 massively parallel sequencing, the first commercially available instrument that could investigate up to 700 bp of

a partial 16S rRNA gene fragment (Liu et al. 2012). Since its discontinuation in 2013, both Illumina and Ion Torrent have provided benchtop sequencers for 16S rRNA gene-based metabarcoding. Benchtop sequencers have been revolutionary to facilitate diversity studies within laboratories and perform longitudinal studies (Salipante et al. 2014). Platforms such as the Illumina MiSeq or Ion Torrent Personal Genome Machine (PGM) are most economically viable for 16S rRNA gene sequencing. The MiSeq is generally preferred as it generates a higher throughput and the lowest error rate for short read sequencers ~0.01%, meaning samples can be sequenced at a higher depth or lower cost per sample (Caporaso et al. 2011; Pollock et al. 2018). At present, 16S rRNA gene Illumina NGS is a standard research tool in PHW CF and COPD bacteriology, requiring a standardised workflow for routine use.

To validate a 16S rRNA gene-based NGS workflow, it is crucial to understand and identify what factors introduce bias to sequence data analysis. Examples include sample type, sample storage, DNA extraction method, primer selection, PCR reagents, choice of sequencing platform, sequence data quality, bioinformatic pipeline and 16S rRNA gene reference database (Fouhy et al. 2016). A standardised approach for sequencing is required to create reproducible data across sequencing runs and studies. Standardisation is also critical for clinical diagnostics as it promotes routinisation and increases process efficiency using in-house resources, and in turn, free up additional resources (Wears 2015; Wiegmann 2017).

A major drawback to hypervariable gene NGS includes limited 16S rRNA gene sequence resolution to the genus or family level. As a solution to this limitation, molecular typing methods are often employed alongside 16S rRNA gene-based metaprofiling for species-level resolution in research (**see section 1.3**). Molecular typing can also aid in the visualisation of abundant and dominant bacterial species. Ribosomal intergenic spacer analysis (RISA) is one example that investigates bacterial diversity in a single step PCR that amplifies the 16S and 23S rRNA intergenic transcribed spacer (ITS) region, defined as an area of non-coding RNA which is transcribed together with the 16S rRNA and 23S rRNA genes (Osorio et al. 2005; Flight et al. 2015). Variation in ITS size and number of fragments can then be viewed using various electrophoretic methods. These genomic regions exhibit high variability between species, and can be used to predict most abundant species. Flight et al. (2015) was the first to demonstrate the application of RISA in CF, who modified this type of protocol from (Borneman and Triplett 1997). Since then, minor adaptations have been made to this protocol, facilitating diversity analysis and respiratory microbiota profiling while identifying major and emerging CF pathogens that are missed through culture (Flight et al 2015; Weiser et al. 2021).

3.1.1 Aims and objectives

The main aim of this chapter was to optimise a 16S rRNA gene PCR and a RISA PCR using PHW resources to broadly detect respiratory pathogens. These DNA-based methods then underwent validation to assess their suitability as routine diagnostic tools.

The objectives of this chapter were:

1) Create an *in silico* and experimental reference database of RISA profiles for all bacteria species studied.
2) Optimise a 16S rRNA gene PCR and a RISA PCR protocol to broadly detect respiratory pathogens.
3) Analyse 16S rRNA gene-based sequence data in mock communities simulating CF and COPD microbiota with the aim of differentiating these communities at the genus and OTU level.
4) Evaluate the performance of RISA in its ability to differentiate diversity profiles and to provide species-level identification of mock community bacteria.

3.1.2 Hypotheses

Hypothesis for this chapter were:

1) Both 16S rRNA gene-based sequencing and RISA PCR are able to differentiate CF and COPD-specific communities by community composition and membership.
2) RISA PCR can generate clear profiles for a pure bacterial culture and single species, but is limited in its ability to classify constituents of a complex respiratory community.

3.2 Materials and methods

3.2.1 Microorganism culture and DNA extraction

A panel of 20 CF and COPD bacteria species was chosen to optimise a 16S rRNA gene PCR and a RISA PCR protocol (Table 2.1). All SACU aerobic and capnophilic species were incubated in 5 ml TSB for 24-h (37°C) prior to DNA extraction. The Anaerobic Reference Unit (ARU) strict anaerobic isolates were supplied in 10 ml BHI broth supplemented with 10% serum (37°C for 72-h incubation) for gDNA extraction. Samples were homogenised before centrifugation at 4000 rpm for 10 minutes and resuspension in 1X sterile phosphate buffered saline (PBS). Isolates were lysed with 300 µl guanidinium isothiocyanate where an additional bead-beating step was required for Gram-positive species (**see section 2.3.2**). Samples were then homogenised and stood at room temperature for 5 minutes to lyse. 400 µL of resulting lysate was loaded into a Maxwell® 16 platform and gDNA was extracted using the Maxwell® 16 Tissue DNA purification kit (**section 2.3.2**). Extracted gDNA was quantified using a Qubit broad-range dsDNA kit.

3.2.2 Validation of a RISA PCR protocol in mock communities

A RISA protocol was selected (Weiser et al. 2021) to visualise ITS profiles among the panel of respiratory bacteria species using primers 1406F (5'-TGYACACACCGCCCGT-3') and 23SR (5'GGGTTBCCCCATTCRG-3'). This protocol is an adaptation of Flight et al. (2015) who modified a 16S and 23S rRNA ITS PCR from Borneman and Triplett (1997) originally for CF lung infection microbiota analysis. Preliminary modifications to PCR cycling conditions were made to fit the recommended AccuStart PCR conditions (Table 3.1). From this preliminary work, 5 µl of genomic DNA was amplified in duplicate in a 25 µl mixture containing 200 nM of both forward and reverse primer for 35 cycles (Table 2.19). The resulting amplicons were first visualised on a 1% agarose gel and then on a 2200 Agilent TapeStation D5000 ScreenTape for greater resolution between multiple amplicons.

Following the generation of single-species ITS references, mock community DNA simulating CF-specific microbiota, COPD-specific microbiota, and an evenly distributed mock community was amplified in triplicate over 3 consecutive PCR assays and then pooled. Mock community DNA proportions and rationale for construction are illustrated in detail in **section 2.2.2,** Figure 2.1, and Table 2.13-Table 2.15. The resulting amplicons were visualised with the Qiagen QIAxcel high-resolution DNA cartridge using a 15 bp-3 kb QX DNA alignment marker and a 100 bp-2.5 kb QX DNA size marker to achieve greater resolution between DNA amplicons spanning ~500 bp-2.5 kb (**section 2.5.3**). Both experimental data and an *in-silico* PCR tool (Bikandi et al. 2004) were used to predict and identify mock bacteria community members.

Table 3.1 Outline of AccuStart recommended PCR cycling conditions compared to reagents and protocol used by Weiser et al. 2021

	AccuStart II PCR ToughMix		Qiagen *Taq* PCR Core Kit	
	Temperature	Time	Temperature	Time
Initial denaturation	94°C	1-3 minutes	95°C	5 minutes
Denaturation	94°C	10-30 seconds	95°C	1 minute
Annealing	55-65°C	15-30 seconds	54°C	30 seconds
Extension	68-72°C	1 minute per kb of product length	72°C	1 minute
Final extension	68-72°C	Optional	72°C	5 minutes
Hold	4°C	Indefinitely	10°C	Indefinitely
Reaction volume	25 µl		25 µl	
DNA template	5-10 µl		2 µl	
Primer concentration	100-500 nM for each forward and reverse primer		200 nM of each forward and reverse primer	
PCR cycles (n)	20-40 cycles		35 cycles	

3.2.3 Optimisation of a 16S rRNA gene PCR protocol

A 16S rRNA gene PCR protocol was optimised to ensure efficient and correct amplification of hypervariable region 4 (V4) region using primers 16Sf (5'GTGCCAGCMGCCGCGGTAA-3') and 16Sr (5'-GGACTACHVGGGTWTCTAAT-3') (Kozich et al. 2013). All 20 bacteria species including mock community DNA were chosen to optimise this protocol (Figure 2.1; Table 2.1). For primer concentration optimisation and for mock community library preparation, 5 µl gDNA was amplified with indexed primers in a 25 µl reaction for 30 cycles (**see section 2.2.2**)**.** Optimisation was performed in a stepwise fashion with the aim to eliminate nonspecific amplification and maintain high DNA amplification efficiency. Primer self-complementarity was checked through PRIMER-BLAST. All samples were amplified twice in duplicate per optimisation stage and isolates that exhibited spurious amplification were investigated further.

PCR optimisation began by modifying both denaturation and extension conditions to meet AccuStart recommended PCR conditions (Table 3.2). Optimal annealing conditions and PCR cycle number were already predetermined and therefore unchanged from the original protocol (Table 3.2). PCR extension modifications were required to improve PCR amplification efficiency and accuracy (**see section 3.3.1.2**). Afterwards, the effect of PCR primer concentration on PCR amplification efficiency and sequence library adapter carryover was compared by amplifying 2 ng/µl and 0.2 ng/µl of *S. aureus* gDNA with a combined 400 nM or a 200 nM forward and reverse primer concentration. PCR products were viewed by 1% agarose gel electrophoresis or a 2200 TapeStation HS D1000 ScreenTape.

Table 3.2 Outline of AccuStart recommended PCR cycling conditions compared to reagents and protocol outlined in Kozich et al. (2013)

	AccuStart II PCR ToughMix		AccuPrime™ Pfx SuperMix	
	Temperature	Time	Temperature	Time
Initial denaturation	94°C	1-3 minutes	95°C	2 minutes
Denaturation	94°C	10-30 seconds	95°C	20 seconds
Annealing	55-65°C	15-30 seconds	55°C	15 seconds
Extension	68-72°C	1 minute per kb of product length	72°C	5 minutes
Final extension	68-72°C	Optional	72°C	10 minutes
Hold	4°C	Indefinitely	4°C	Indefinitely
Reaction volume	25 µl		20 µl	
DNA template	5-10 µl		1 µl	
Primer concentration	100-500 nM of each forward and reverse primer		500 nM for each forward and reverse primer	
PCR cycles (n)	20-40 cycles		30 cycles	

3.2.4 RISA and 16S rRNA gene sequence analysis of mock communities

Following 16S rRNA gene PCR optimisation (**see sections 3.2.2 and 3.2.3**), Illumina library preparation was performed in triplicate over 3 consecutive sequencing batches using barcoded primers (Kozich et al. 2013; **see section 2.6;** SI Table A2). These samples were also RISA PCR amplified using an optimised protocol. Amplified sequencing libraries were purified using AMPure XP SPRI beads. DNA was quantified with a Qubit HS dsDNA kit and quality-checked with a 2200 TapeStation HS D1000 ScreenTape. Libraries were pooled to 4 nM and were denatured using 0.2 N NaOH prior to loading (**section 2.6**).

Sequenced reads were quality-checked with FastQC (Andrews 2010) and trimmed by Trim Galore! (Krueger 2021). Trimmed reads were processed in Mothur to filter poor quality or erroneous sequences and to assign taxonomy following the methods described in **sections 2.6.4 and 2.6.5**. Sequence error rates were estimated in Mothur from mock community sequence data which was compared to an *in silico* mock community containing V4 16S rRNA gene sequences of the 20 bacteria species (Table 2.1). OTUs were classified using the Wang approach (Wang et al. 2007) through SILVA v138 (Pruesse et al. 2007), RDP V16 (Cole et al. 2009), and Greengenes v13.8.99 (DeSantis et al. 2006) 16S rRNA gene reference databases. A phylogeny based on V4 16S rRNA gene sequences was constructed to evaluate SILVA, RDP, and Greengenes OTU clustering and species-level identification of *S. anginosus*, *S. intermedius*, and *S. pneumoniae* along with reference *Streptococcus* spp. 16S rRNA gene sequences that were downloaded from NCBI. These sequences were concatenated and aligned using the ClustalW algorithm in MEGA7 (Kumar et al. 2016). This alignment was then used to construct a maximum likelihood 16S rRNA gene phylogeny in IQ-TREE using the

generalised time-reversible (GTR) model under the gamma rate heterogeneity (gamma). The reference database which most accurately reflected community membership and composition was ultimately chosen for downstream analysis and for taxonomical assignation of patient sputum microbiota (**see section 2.6.4; see chapter 5**). Rare OTUs were not filtered to assess OTU inflation and reagent contamination. Resulting sequenced reads were subsampled to 20151 reads, equating to the group containing the lowest number of sequences. Rarefaction, Good's coverage, ACE, Chao1, inverse Simpson (InvSimpson), and Shannon alpha diversity and both unweighted and weighted UniFrac beta diversity was measured following the methods outlined in **section 2.6.4** and **section 2.6.5**. Both InvSimpson and Shannon were compared because InvSimpson index is more suited to measuring dominance while Shannon diversity is better suited for species richness. ACE index was chosen for comparison as it is better suited when observed richness is small whereas Choa1 is useful towards low-abundance classes and high observed richness. Alpha diversity values were also mathematically predicted and compared against observed values to compare differences in experimental data to answer whether 16S rRNA gene sequencing can differentiate distinct diversity profiles within each community type.

RISA PCR was employed alongside 16S rRNA gene sequencing to compare basic mock community diversity profiles, noting highly abundant and dominant species. Amplicons were visualised using a QIAxcel high-resolution DNA cartridge using a 15 bp-3 kb QX DNA alignment marker and a 100 bp-2.5 kb QX DNA size marker (**see section 2.5.3**). Bacterial species were inferred through comparing QIAxcel data against *in silico* ITS data and TapeStation reference data (Table 3.3). Dominant species were interpreted by having at least one highly concentrated RISA amplicon present, while a more diverse community was expected to include both numerous high and lower concentrated RISA amplicons.

3.2.5 Figure generation and statistical analysis

Figures were generated in R (R core team 2021) using the tidyverse (Wickham et al. 2019) package which included dplyr (Wickham et al. 2018) and ggplot2 (Wickham 2016). Hierarchical clustering of RISA profiles was achieved by calculating unweighted pair group method with arithmetic mean (UPGMA) and the Sørensen–Dice coefficient using GelJ software (Heras et al. 2015). A dendrogram was created and edited in FigTree and Inkscape along with average molecular weight dissimilarity scores.

Pairwise differences in mean OTU relative abundance scores between each mock community were compared using the mother Metastats function, discarding OTUs at <0.01% relative abundance (**see section 2.6.5**). This function was tested to see whether 16S rRNA gene sequencing can differentiate communities on an OTU basis in the event of substantial changes

to mock community composition arising from PCR amplification bias. Subsampled data was used for diversity analysis. Alpha diversity was statistically compared in the nlme package (Pinheiro et al. 2020) using a linear mixed effects model with sequencing batch as a random term. Beta diversity was statistically compared between each mock community type and sequencing batch using the vegan package to perform PERMANOVA (Anderson 2001; Oksanen et al. 2019). Multivariate homogeneity of group dispersions was calculated using the betadisper function.

3.3 Results

3.3.1 RISA PCR optimisation and species reference generation

A RISA PCR was employed to generate single species ITS references for 20 respiratory bacterial species (Table 2.1). These references were later used to predict these 20 species in mock communities to assess its suitability as a tool to profile lung infection microbiota that is missed through culture (Flight et al. 2015; **see chapters 4 and 5**). Initial experiments optimised the published RISA protocols for use within a PHW laboratory diagnostic context as follows. The protocol of Weiser et al. (2021) was modified by changing PCR denaturation to 94°C for 30 seconds, including a 2-minute initial denaturation step (Table 2.19; Figure 3.1; Table 3.3). This enabled clear and distinct amplification of ITS profiles, where it was possible to generate species-specific RISA references, fulfilling objectives 1 and 2 (Figure 3.1). Most amplicons fell between 400 to 1500 bp (Table 2.19; Figure 3.1; Table 3.3). 8 bacterial species generated distinct RISA profiles based on ITS amplicon molecular weight and number, for example, *S. aureus* and *A. prevotii* (Figure 3.1). The remaining 12 species presented with 1 clear concentrated amplicon between 577 bp and 935 bp where additional weaker amplicons ≥1500 bp appeared in 8 of these samples (Table 3.3).

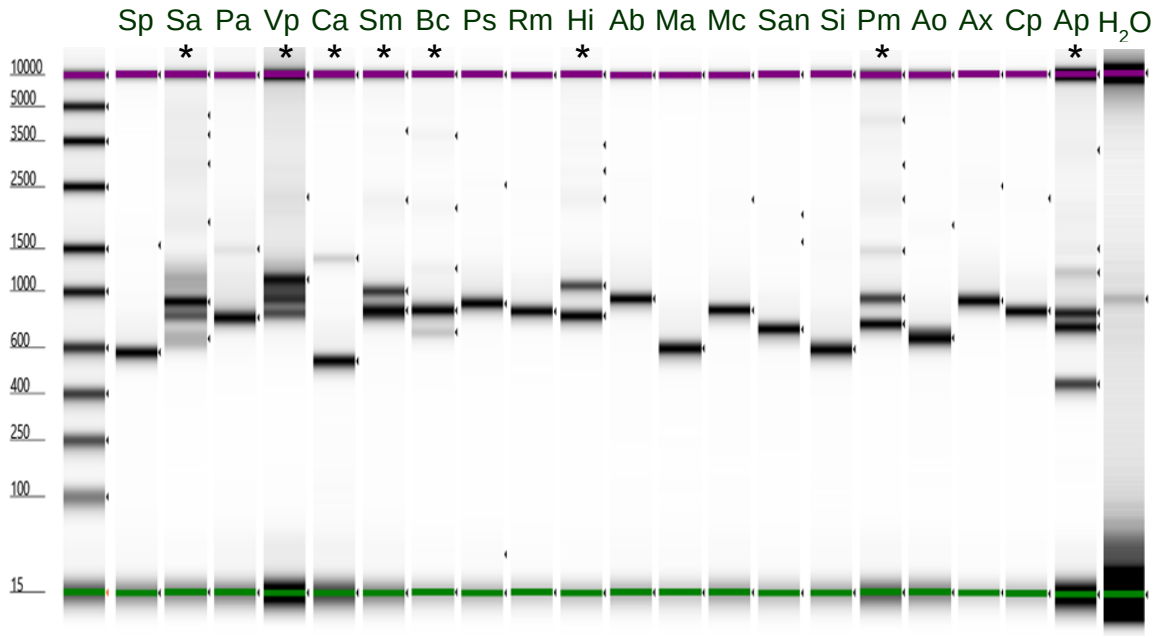


Figure 3.1 Reference single-species RISA profiles. Profiles were scaled individually to the most concentrated peak in each sample and were aligned to the Tapestation's lower marker (green lanes) and upper marker (purple lanes) to calculate sample DNA molecular weight in bp. Sample ID is abbreviated above this figure. Samples which contained multiple concentrated PCR products were considered distinct and were annotated with an asterisk. These samples included *S. aureus* (Sa), *V. parvula* (Vp), *C. acnes* (Ca), *S. maltophilia* (Sm), *B. cepacia* (Bc), *H. influenzae* (Hi), *P. melaninogenica* (Pm), and *A. prevotii* (Ap). The 12 remaining species presented with 1 clear concentrated amplicon ranging 577 bp and 935 bp. Due to substantial variation in Tapestation measurements between experiments, these samples could not be confidently described as distinct once repeated. These samples included *S. pneumoniae* (Sp), *P. aeruginosa* (Pa), *P. sputorum* (Ps), *R. mannitolilytica* (Rm), *A. baumannii* (Ab), *M. abscessus* (Ma), *M. catarrhalis* (Mc), *S. anginosus* (San), *S. intermedius* (Si), *A. odontolyticus* (Ao), *A. xylosoxidans* (Ax), and *C. pauculus* (Cp). A PCR nuclease-free water control is shown last.

Table 3.3 Single-species RISA profiles and *in silico* predictions

Species	ITS predicted size (bp)	ITS observed size (bp)
Achromobacter xylosoxidans	887-891, 1021[a]	925, 2512
Acinetobacter baumannii	889[b]	935
Actinomyces odontolyticus	609-612[b]	657, 1823
Anaerococcus prevotii	415, 609, 628-637[b]	441, 733, 834
Burkholderia cepacia	681, 830, 871[b]	688, 836, 1238, 2090, 3679
Cupriavidus pauculus	794, 922[b]	844, 2268
Cutibacterium acnes	496-487[b]	534, 1373
Haemophilus influenzae	995-1015, 757-759[a]	801, 1056, 2260, 2809, 3401
Moraxella catarrhalis	777[b]	842, 2253
Mycobacterium abscessus	581[b]	597
Pandoraea sputorum	848, 1911[b]	894, 2534
Pseudomonas aeruginosa	753[a]	789, 1500
Prevotella melaninogenica	651-652, 844[b]	747, 947, 1471, 2254, 2936, 4351
Ralstonia mannitolilytica	805[a]	834
Staphylococcus aureus	648, 743-757, 830[b]	649, 805, 904, 1128, 1862, 2951, 3720, 4539
Stenotrophomonas maltophilia	777-782, 826-829[a]	841, 1003, 2240
Streptococcus anginosus	669[b]	709, 1588, 1989
Streptococcus intermedius	672-673[b]	593
Streptococcus pneumoniae	529[b]	577, 1545
Veillonella parvula	883, 926, 1075[b]	823, 943, 1118, 2299

[a]Data obtained from Weiser et al. (2021).
[b]*in silico* estimations were obtained from http://insilico.ehu.eus/ (Bikandi et al. 2004).

3.3.2 16S rRNA gene PCR optimisation

A V4 16S rRNA gene sequencing protocol was optimised from Kozich et al. (2013) with the goal to integrate a higher-resolution method for microbial diversity analysis in context of PHW lung infection diagnostics who currently rely on culture-based tools. An optimised sequencing protocol could then be compared to current routine analysis of respiratory specimens, in addition to evaluating the usefulness of RISA PCR within a diagnostic laboratory (**see chapter 5**). 20 respiratory bacterial species (Table 2.1) and mock community DNA (Figure 2.1) were targeted to optimise PCR conditions, taking note of non-specific amplification. Optimisation was performed in a stepwise fashion, with the aim of detecting all respiratory reference species while simultaneously diminishing the amount of PCR artifacts that would impact the integrity of 16S rRNA gene-based microbiota analysis. As for RISA PCR, a denaturation step of 94°C for 30 seconds (and 2 minutes for initial denaturation) was found optimal and was also applied to this protocol. Previously optimised PCR annealing conditions and cycle number (Kozich et al. 2013) were used to optimise extension conditions and primer concentration (Table 2.19).

PCR extension at 72°C for 30 seconds excluding a final extension step (combined 400 nM forward and reverse primer concentration) resulted in strong PCR amplification of a ~250 bp 16S rRNA gene amplicon for all species except for *C. acnes,* which yielded ≤20 ng/µl of amplified DNA (SI Figure B2). Non-specific amplification of a ~800 bp product was visible in *A. xylosoxidans* and *A. baumannii* samples, albeit faintly. ~500 bp amplicons were also visible from amplified mock community DNA samples (SI Figure B2).

The reduction of PCR extension temperature to 68°C reduced but did not eliminate ~800 bp nonspecific amplicons from *A. xylosoxidans* and *A. baumannii* samples, or ~500 bp nonspecific amplicons from amplified mock community DNA samples. A 5-minute final extension at 68°C was also included to ensure complete replication although this still did not eliminate non-specific amplification (SI Figure B2). *C. acnes* PCR amplification efficiency did not improve, suggesting reduced primer complementarity to the *C. acnes* 16S rRNA gene compared to other species. The number of PCR cycles was not increased, to avoid increasing spurious amplification. It was concluded that further modifications to PCR extension parameters would not improve this outcome since A) a 30 second extension is theoretically optimal for a ≥400 bp dual-indexed Illumina library for sequencing, and B) extension temperatures beyond AccuStart recommended PCR guidelines would likely reduce amplification efficiency and/or increase undesirable PCR artifacts (Table 3.2). Ultimately, an extension of 68°C for 30 seconds including a 5-minute final extension was used in the finalised protocol, meeting objective 2.

Substantial PCR primer carryover was evident in post-PCR samples (SI Figure B2). Reducing primer concentration to 200 nM demonstrated no substantial reduction in amplified *S. aureus* DNA. Some primer carryover was still detected in these post-PCR samples, indicating there was sufficient primer for complete amplification (SI Figure B2).

3.3.3 16S rRNA gene sequence statistics and sample contamination

Before sequence filtering, approximately 7.8% of chimeric reads were discarded in addition to 9.7% of reads that were erroneous or poor quality. Following filtering, 660353 reads were retained from 3 consecutive sequenced batches. Overall sequence error rates were 1.1%, 1.2%, and 2.6% for each sequenced batch. For diversity analysis, and to meet objective 3, each sample was subsampled to 20151 reads, corresponding to the sample containing the least number of reads. Because sequence error was higher than expected, OTU counts were inflated, resulting in 130, 125, and 162 OTUs detected between each sequenced run respectively.

Reagent-derived sequences appeared consistently low for every sequenced batch where 12 unclassified genera were in 0.1% abundance and was thus not considered to substantially impact 16S rRNA gene sequence data analysis. *Haemophilus* (n=2), *Atopobium* (n=3),

Gemella (n=1), *Campylobacter* (n=1), and *Rothia* (n=1) also represented 0.005% community composition. *Prevotella* genogroup 6 (n=3) and *Moraxella* (n=1) constituted 0.02% and 0.08% community composition while *Acinetobacter* (n=3) could have affected 16S rRNA gene sequence data as it notably abundant in *P. aeruginosa*-dominated mock communities (5.1% community composition).

3.3.4 16S rRNA gene sequence mock microbial composition

3.3.4.1 CF *Burkholderia*-dominated mock community composition

Mock microbial community sequencing is essential to address how PCR bias could impact diversity analysis. Here, sequenced mock communities simulating CF and COPD were compared to investigate community diversity and composition changes, and whether V4 16S rRNA gene sequencing could differentiate these communities on an OTU basis. Gram-positive sequenced reads were generally recovered in higher amounts than Gram-negative sequenced reads in this mock community, excluding *Prevotella* (now 22%-22.1%) and *Acinetobacter* (now 7.2%; Figure 3.2; SI Figure B3; SI Figure B4). The % relative abundance of the dominant *Burkholderia* substantially reduced compared to its predicted figure (now 8%). Instead, *Veillonella* became dominant (now 26%-26.1%), followed by *Prevotella*, and *Streptococcus* (now 13.9%-14%). *Pseudomonas* % abundance also reduced unexpectedly from its predicted figure and was not detected to the genus level by Greengenes (Figure 3.2; SI Figure B3; SI Figure B4). Rarer taxa, for example, *Achromobacter*, *Cutibacterium*, *Ralstonia*, and *Mycobacterium* typically became underrepresented.

3.3.4.2 CF *Pseudomonas*-dominated mock community composition

Pseudomonas exhibited the most dramatic reduction in predicted sequenced reads compared to any mock community (now 21.8%-21.9%; Figure 3.2; SI Figure B3; SI Figure B4). Greengenes also failed to identify *Pseudomonas* to the genus level (SI Figure B4). *Prevotella* became dominant in this mock community (now 48.3%-48.4%), followed by *Staphylococcus* (now 11.1%), and *Veillonella*. Comparable to the CF *Burkholderia*-dominated community, the detection of Gram-positive sequenced reads increased in comparison to Gram-negative sequenced reads (Figure 3.2; SI Figure B3; SI Figure B4). The detection of rarer taxa decreased to different extents when compared to predicted abundance data.

3.3.4.3 COPD mock community composition

The COPD mock community bore the closest resemblance to its intended composition (Figure 3.2; SI Figure B3; SI Figure B4). *Haemophilus* was dominant (now 39%-39.1%) followed by *Streptococcus* (now 19.1%-19.2%). *Veillonella* (now 13.1%-13.2%) and *Prevotella* (now

11.9%-12%) was also highly abundant. The number of sequenced reads in all remaining species excluding *Acinetobacter* reduced compared to predicted proportions (Figure 3.2; SI Figure B3; SI Figure B4). The composition of rarer taxa declined similarly to both CF mock communities. Greengenes also failed to identify *Cupriavidus* and *Pseudomonas* to the genus level. Sequenced reads belonging to *Mycobacterium, Burkholderia* and *Cutibacterium* became highly underrepresented where *Cutibacterium* failed detection in 2 replicates.

3.3.4.4 Equimolar mock community composition

Most notable increases in % relative abundance scores compared to predicted figures occurred in *Veillonella* (now 16.9%-17.2%), *Haemophilus* (now 13.5%), *Prevotella* (6.23%-6.39% increase), *Acinetobacter* (now 11%-11.2%), and *Moraxella* (now 8.7%-8.9%; Figure 3.2; SI Figure B3; SI Figure B4). Comprising of 3 species, *Streptococcus* spp. remained dominant (now 17.8%). A higher number of sequenced reads belonging to *Anaerococcus* and *Staphylococcus* were detected compared to previously predicted figures whereas % relative abundance scores of all remaining species were lower than what was previously predicted (Figure 3.2; SI Figure B3; SI Figure B4). *Ralstonia*, *Mycobacterium*, and *Cutibacterium* were especially underrepresented.

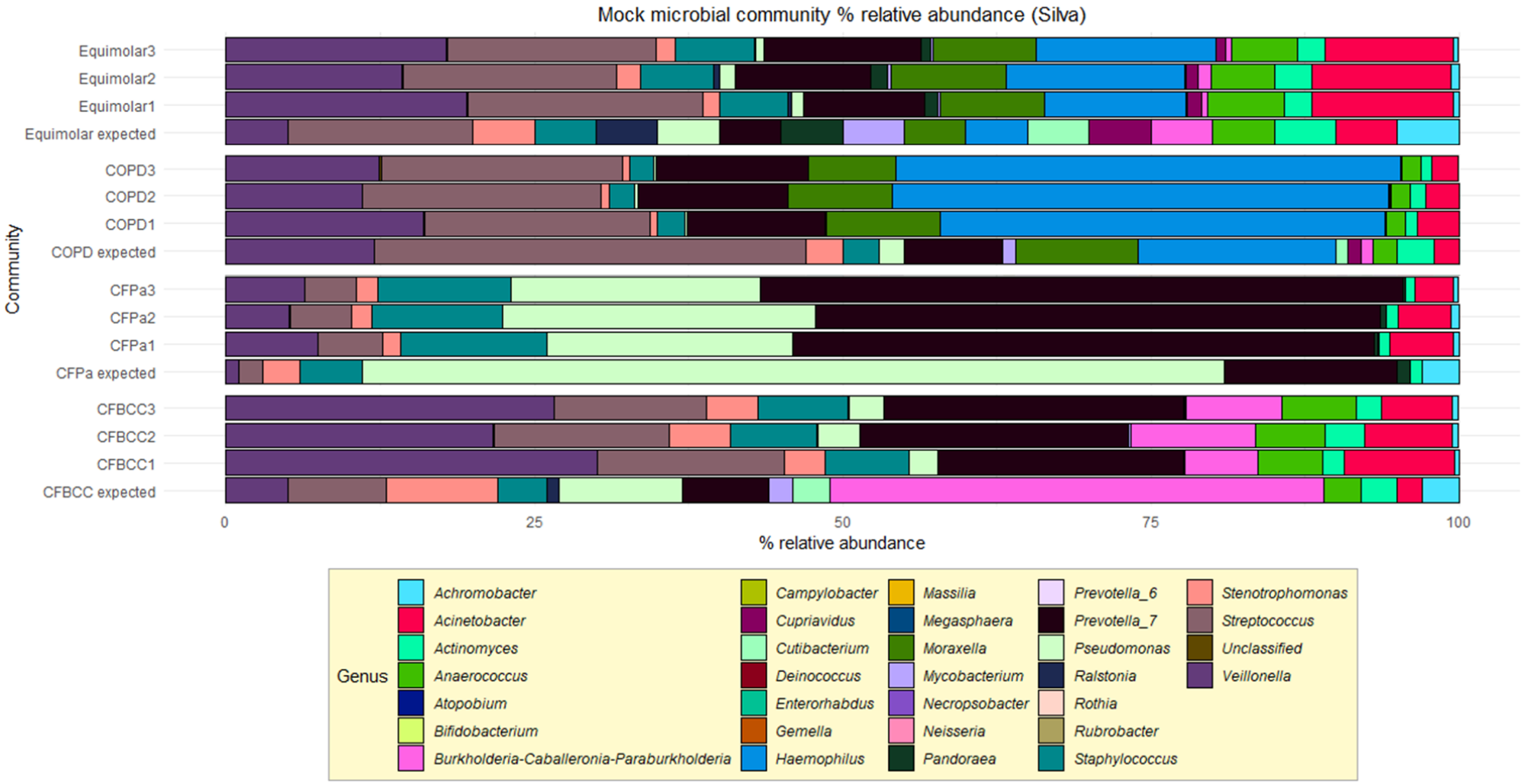


Figure 3.2 SILVA mock community genus composition. Due to the low number of genera detected, relative abundance was not capped prior to generating bar plots. Each sequenced run is shown to visualise batch variation in microbial composition in addition to expected community composition. All unclassified sequences were grouped together.Abbreviations include: Equimolar; equimolar community, COPD; COPD community, CFPa; CF *Pseudomonas*-dominated community, CFBCC; CF *Burkholderia cepacia* complex-dominated community. Bars were edited to make comparisons easier between groups.

3.3.5 OTU based analysis of community-defining microbiota

From the 4 sequenced mock communities, the 20 most proportionally abundant OTUs were extracted. In theory, these 20 OTUs should correspond to the 20 reference bacterial species used for mock community construction, offering insight into mock community composition changes (Figure 2.1; Table 2.1). Figure 3.3 illustrates the 20 most abundant OTUs by proportion across the entire sequenced dataset. In theory, these 20 OTUs should correspond to all 20 bacteria species used for mock community construction. Both SILVA and RDP correctly identified all bacteria genera and performed similarly in reporting OTU proportions whereas Greengenes lacked the same resolution. The 3 databases reported *Prevotella* (~21% of sequenced reads), *Veillonella* (~15% of sequenced reads), and *Haemophilus* (~14% of sequenced reads) as the most dominant OTUs in line with predicted community composition (Figure 3.3). SILVA was the only database to identify different *Prevotella* genotypes. In contrast, a significant portion of reads that was expected to correspond to the dominant *P. aeruginosa* and *B. cepacia* were lost (Figure 2.1; Figure 3.3). The reported proportion of remaining genera were similar between reference databases (Figure 3.3). The proportion of Gram-negative sequenced reads in all cases was lower than the predicted values; this was variable for Gram-positive genera (Figure 3.3).

Both SILVA and RDP reported a high proportion of *Streptococcus* OTU4 (8.5%-8.6% of sequenced reads) that clustered with *S. intermedius* reference strains. A significantly smaller portion of reads belonging to OTU20 *Streptococcus* (0.07%-0.09%) clustered with *S. anginosus* references by phylogeny (SI Figure B5). In contrast, Greengenes split *Streptococcus* sequences more evenly into OTU8 (5.37% of sequenced reads) and OTU11 (3.18% of sequenced reads). OTU8 clustered with *S. anginosus* and OTU11 clustered with *S. intermedius* references through phylogenetic construction (SI Figure B5). Additionally, OTU7 *Streptococcus* clearly clustered with *S. pneumoniae* and its proportion was highly similar between reference databases.

Due to poor PCR amplification, *C. acnes* was highly underrepresented (~0.01% of sequenced reads). Crucially, Greengenes failed to classify *C. pauculus* and *P. aeruginosa,* indicating poor alignment quality to the V4 region using this database (Figure 3.3). Unclassified reads were reported in each database and were higher using Greengenes where 5.79% and 0.35% of reads should have been assigned to *Pseudomonas* and *Cupriavidus,* respectively. The number of unclassified sequences reported by SILVA and RDP remained low, at 0.03% and 0.06% correspondingly. SILVA was ultimately determined to be most suitable for analysis of respiratory microbiota and was solely used for alpha and beta diversity analysis (see below).

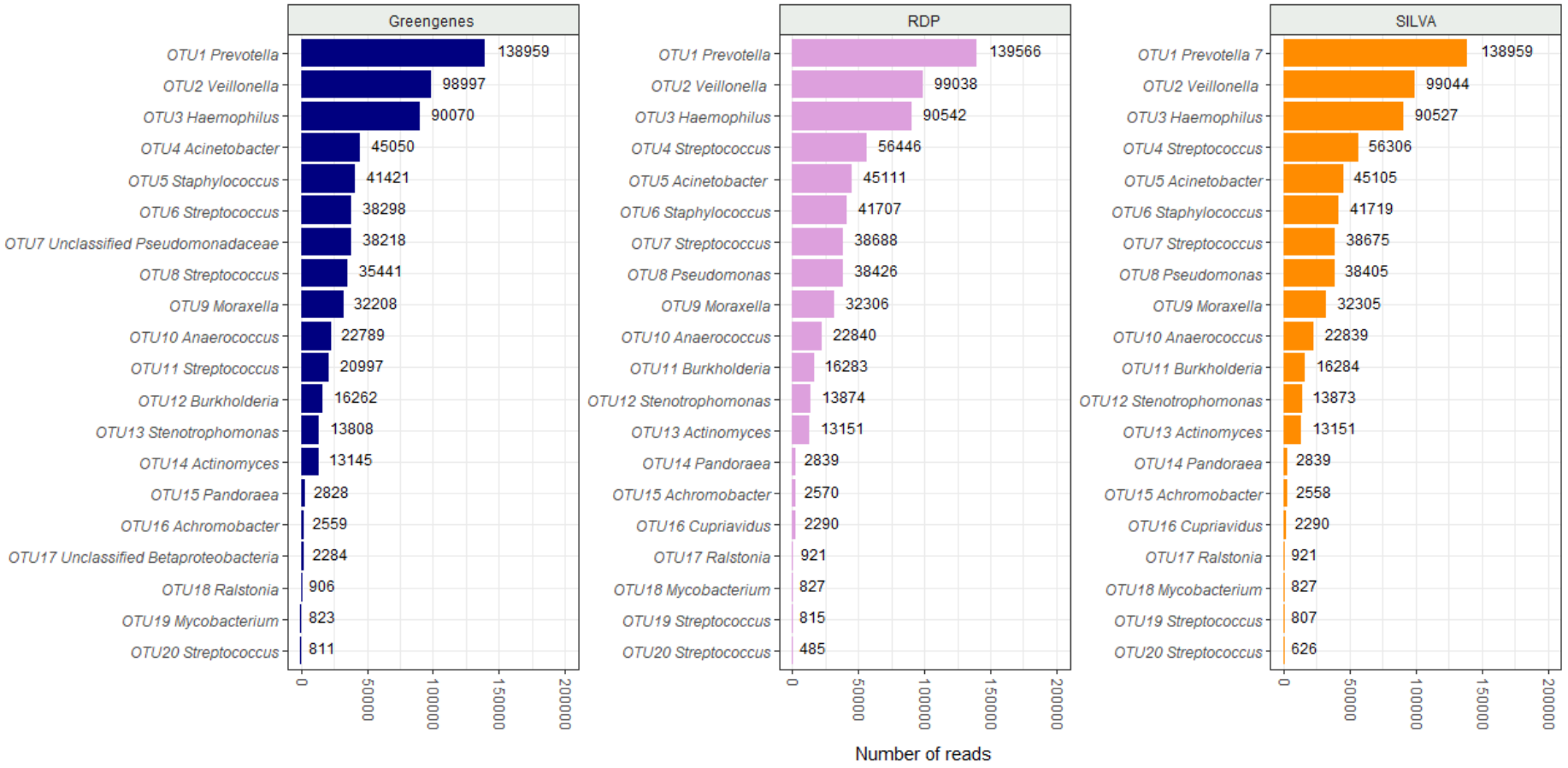


Figure 3.3 A Greengenes, RDP, and SILVA comparison of 20 most proportionally abundant OTUs. The proportion of reads belonging to each OTU was calculated by taking a percentage from the total number of reads (660353) and is displayed in the bar plots.

3.3.6 Mock community differentially abundant OTUs

Metastats was used to compare major distinguishing OTUs from each disease-specific mock community, not only to determine OTUs that significantly shifted mock community composition, but to assess whether disease-specific communities were distinguishable at an OTU basis. Due to deviations in sequenced mock community composition from their expected composition, Metastats confirmed whether dominant and abundant community members were still differentially abundant between staggered communities (part of objective 3). OTU12 *Stenotrophomonas,* OTU4 *Streptococcus,* and OTU11 *Burkholderia* were key OTUs defining the CF BCC dominated mock community which were significantly more abundant in this community compared to the others. OTU6 *Streptococcus,* OTU8 *Pseudomonas* were also significantly more abundant in this mock community compared to the COPD mock community. OTU1 *Prevotella,* OTU6 *Staphylococcus,* and OTU8 *Pseudomonas* were key CF *P. aeruginosa*-dominated community OTUs which were significantly more abundant in this community compared to the other two. OTU4 *Streptococcus*, and OTU12 *Stenotrophomonas* were also more abundant in this mock community compared to the COPD mock community OTU7 *Streptococcus,* OTU3 *Haemophilus* and OTU9 *Moraxella* were key OTUs defining the COPD mock community which were significantly more abundant in this community compared to the others. See Table 3.4 for all pairwise comparison p values.

Table 3.4 Metastats differentially abundant OTUs between mock communities[a,b]

OTU	Genus	Mean % abundance ± SE	Mean % abundance ± SE	P value
OTU1	*Prevotella* genogroup 7	21.9±1.2 (CFBCC)	48±1.9 (CFPa)	P=0.006
OTU2	*Veillonella*	26±2.4 (CFBCC)	6.3±0.7 (CFPa)	P=0.03
OTU4	*Streptococcus*	13.7±0.8 (CFBCC)	4.8±0.3 (CFPa)	P=0.02
OTU6	*Staphylococcus*	7±0.1 (CFBCC)	11.1±0.4 (CFPa)	P=0.03
OTU8	*Pseudomonas*	2.9±0.3 (CFBCC)	21.9±1.8 (CFPa)	P=0.01
OTU11	*Burkholderia*	7.9±0.04 (CFBCC)	0±0 (CFPa	P=0.03
OTU12	*Stenotrophomonas*	4.1±0.5 (CFBCC)	1.6±0.1 (CFPa)	P=0.005
OTU13	*Actinomyces*	2.3±0.4 (CFBCC)	0.9±0.1 (CFPa)	P=0.04
OTU1	*Prevotella* genogroup 7	21.9±1.3 (CFBCC)	11.8±0.4 (COPD)	P=0.04
OTU3	*Haemophilus*	0.003±0.002 (CFBCC)	39.1±1.5 (COPD)	P=0.01
OTU4	*Streptococcus*	13.7±0.8 (CFBCC)	0.7±0.1 (COPD)	P=0.03
OTU6	*Staphylococcus*	7±0.1 (CFBCC)	2.1±0.1 (COPD)	P=0.006
OTU7	*Streptococcus*	0.005±0.005 (CFBCC)	18.1±0.5 (COPD)	P<0.0001
OTU8	*Pseudomonas*	2.9±0.3 (CFBCC)	0.2±0.02 (COPD)	P=0.03
OTU9	*Moraxella*	0±0 (CFBCC)	8.2±0.6 (COPD)	P=0.03
OTU10	*Anaerococcus*	5.6±0.2 (CFBCC)	1.5±0.03 (COPD)	P=0.02
OTU11	*Burkholderia*	7.9±1.2 (CFBCC)	0.07±0.02 (COPD)	P=0.048
OTU012	*Stenotrophomonas*	4.1±0.5 (CFBCC)	0.6±0.05 (COPD)	P=0.045
OTU1	*Prevotella* genogroup 7	48±1.9 (CFPa)	11.8±0.4 (COPD)	P=0.02
OTU3	*Haemophilus*	39.1±0.07 (CFPa)	0.002±0 (COPD)	P=0.01
OTU4	*Streptococcus*	4.8±0.3 (CFPa)	0.7±0.1 (COPD)	P=0.04
OTU6	*Staphylococcus*	11.1±0.4 (CFPa)	2.1±0.1 (COPD)	P=0.02
OTU7	*Streptococcus*	0.008±0 (CFPa)	18.1±0.006 (COPD)	P=0.005
OTU8	*Pseudomonas*	21.9±1.8 (CFPa)	0.2±0.02 (COPD)	P=0.04
OTU9	*Moraxella*	0.03±0 (CFPa)	8.2±0.01 (COPD)	P=0.03
OTU12	*Stenotrophomonas*	1.6±0.1 (CFPa)	0.6±0.1 (COPD)	P=0.04

[a]Abbreviations for group comparisons include: CFBCC: CF *Burkholderia cepacia* complex-dominated mock community, CFPa: CF *P. aeruginosa*-dominated mock community, COPD: COPD mock community, Eq: equimolar mock community.

[b] Pairwise comparisons where OTU differential abundance was higher in equimolar mock communities were excluded from this analysis.

[c]OTU4 corresponding to *S. anginosus* group by phylogeny.

[d]OTU7 corresponding to *S. pneumoniae* by phylogeny.

3.3.7 Mock community 16S rRNA gene sequence diversity analysis

3.3.7.1 Mock community rarefaction and alpha diversity

Good's coverage (>0.99) suggested sample diversity was captured well although rarefaction curves of several replicates did not reach an asymptote, most likely as a result of varying sample sequence depth and batch variation (SI Figure B6). Community richness from highest to lowest was expected to be: equimolar mock community > COPD mock community > CF *Burkholderia*-dominated community > CF *Pseudomonas*-dominated community (SI Table B1). Overall alpha diversity measurements were statistically significant between each mock community (R^2 = 0.998, $F_{12,40}$=40.9, P<0.0001). The number of observed species (Sobs)

ranged from 48 to 142, and on average it was highest among the equimolar mock community (~111), followed by the COPD (~89) CF *Burkholderia*-dominated (~76), and CF *Pseudomonas*-dominated (~59) mock communities (Figure 3.4). Sobs reflected initial predictions of community richness and was significantly lower in the CF *Pseudomonas*-dominated mock community compared to the COPD mock community ($P<0.0001$). Sobs was additionally significantly lower in the CF *Pseudomonas*-dominated community compared to the equimolar mock community ($P<0.0001$). ACE values followed a similar trend to Sobs whereas Chao1 values were highest in the equimolar mock community, followed closely by CF *Burkholderia*-dominated, COPD, and CF *Pseudomonas*-dominated communities (Figure 3.4). Neither ACE nor Chao1 reported significant differences in community richness after correcting for batch variation.

InvSimpson and Shannon community diversity from highest to lowest was expected to follow the order of: equimolar mock community > COPD mock community > CF *Burkholderia*-dominated community > CF *Pseudomonas*-dominated community (SI Table B1). Shannon estimates of diversity generally better reflected predicted community diversity compared to community composition (SI Table B1). Mean InvSimpson scores positively correlated with Shannon diversity (Figure 3.4). Both diversity indices reflected Chao1 scores but were less affected by batch variation of each sequenced run, i.e., the variation between predicted and experimental data was smaller (Figure 3.4; SI Table B1). Because of this, both Shannon and InvSimpson diversity were significantly different between every community type ($P<0.0001$).

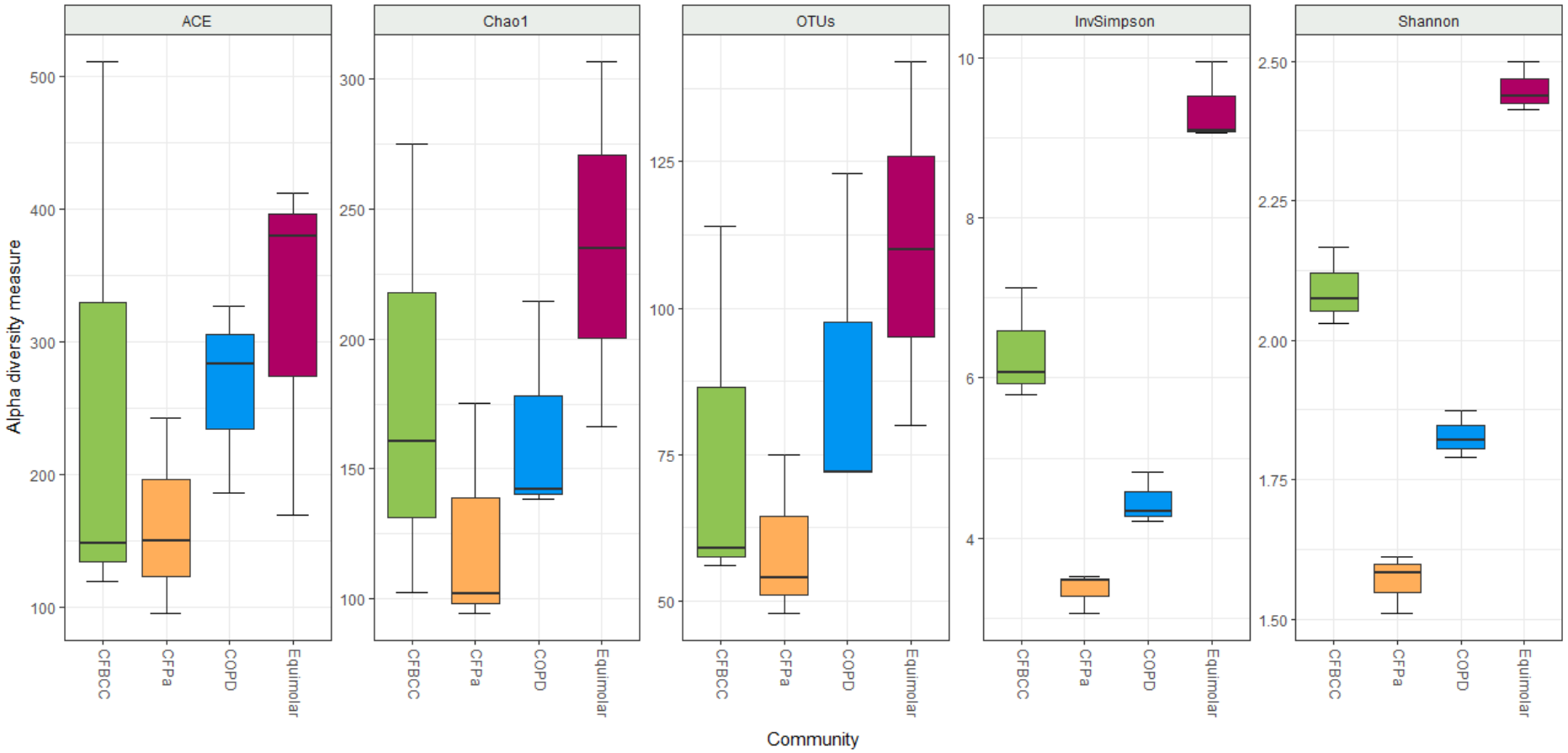


Figure 3.4 Mock community alpha diversity scores. Boxplots include average alpha diversity measurements for each sequenced triplicate which was grouped and coloured by community type. Abbreviations include: Equimolar; equimolar mock community, COPD; COPD mock community, CFPa; CF *Pseudomonas*-dominated mock community, CFBCC; CF *Burkholderia cepacia* complex-dominated mock community.

3.3.7.2 Beta diversity and community membership.

NMDS (lowest stress: 0.2; R^2: 0.5) using unweighted UniFrac showed clear and relatively even spatial separation of each mock community type with no overlap of centroids (the mean position of each community on a 2D scale) (Figure 3.5). This clear spatial separation suggested community membership was distinct between each mock community, and a significant PERMANOVA ($F_{3,8}$=1.5, R^2=0.4, P<0.001) confirmed this indication. Pairwise comparisons were however insignificant.

3.3.7.3 Beta diversity and community structure.

NMDS (lowest stress: 0.03; R^2: 0.995) using weighted UniFrac illustrated distinct separation in the structure of mock communities, with no overlap of centroids (Figure 3.5). Both equimolar and CF *Burkholderia* community spatial coordinates clustered closely together. In contrast, the spatial coordinates of both CF *Pseudomonas* communities and COPD communities were distantly dispersed (Figure 3.5). Unlike unweighted UniFrac NMDS, the spatial distance of each community from its centroid was short, suggesting community structure was more similar within each community than its membership (Figure 3.5). A significant PERMANOVA ($F_{3,8}$=88.2, R^2=0.97, P<0.001) confirmed significant differences in community structure. No significant batch effects were found by PERMANOVA or betadisper.

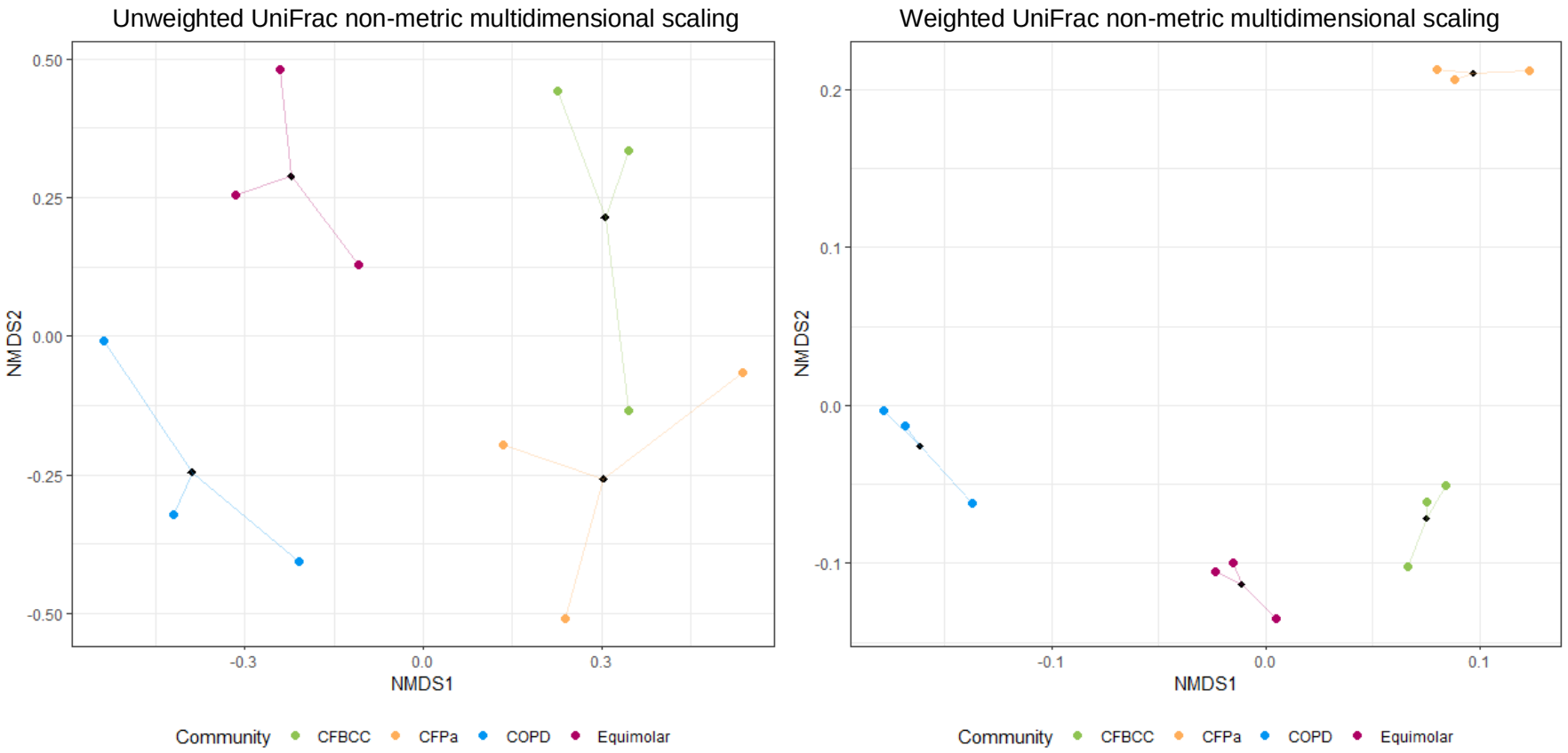


Figure 3.5 NMDS plots measuring unweighted and weighted UniFrac. NMDS axes were plotted for unweighted UniFrac (left) and weighted UniFrac (right) measurements. Each sequenced replicate has been coloured according to mock community type and centroids have been illustrated with a black dot. Abbreviations include: Equimolar; equimolar mock community, COPD; COPD mock community, CFPa; CF *Pseudomonas*-dominated mock community, CFBCC; CF *Burkholderia cepacia* complex-dominated mock community.

3.3.8. RISA PCR diversity analysis of mock communities

After using high resolution 16S rRNA gene sequencing to profile the diversity of the mock communities, RISA was employed as a basic microbiota profiling method to assess its bacterial species predictive capability in mock communities and to address PCR amplification bias. Figure 3.6 achieves objective 4 where RISA diversity profiles were distinct between each mock community and most replicates of each mock community clustered with their respective groups (Figure 3.6). RISA PCR generated reproducible community profiles based on comparable DNA molecular weight and concentration. A visual inspection of RISA profiles indicated the equimolar mock community was most diverse, followed by the CF *Burkholderia*-dominated mock community while both CF *Pseudomonas*-dominant and COPD mock communities appeared least diverse (Figure 3.6). This dendrogram did however highlight one CF BCC-dominated mock community sample (replicate 3) clustering with the equimolar mock communities. A more detailed account of predicted single-species presence is described in the following sections.

3.3.8.1 CF *Burkholderia*-dominated mock community RISA profiles

In contrast to the 16S rRNA gene sequence data where *Veillonella* was dominant, the dominant species was inferred to be *B. cepacia,* corresponding to a 63.7 ng/µl, ~883 bp amplicon (Table 3.3). A 52.2 ng/µl, ~774 bp amplicon was inferred to represent both *S. maltophilia* and *P. aeruginosa* due to its high concentration and resolution of the QIAxcel being unable to separate amplicons bearing similar molecular weights (Table 3.3). A 4.8 ng/µl, ~701 bp fragment suggested the presence of *S. anginosus* and a 4.6 ng/µl, ~729 bp fragment suggested a partial match to *P. melaninogenica* (Table 3.3). Both RISA and 16S rRNA gene sequence data indicated a high abundance of *P. melaninogenica* and *S. anginosus. V. parvula* could not be identified through RISA PCR although this species was abundant in sequence data.

3.3.8.2 CF *Pseudomonas*-dominated mock community RISA profiles.

P. aeruginosa was the dominant species of this community, represented by a 34.1 ng/µl, ~803 bp amplicon, once again contrasting 16S rRNA gene sequence data where *Prevotella* was dominant (Table 3.3). Similar to sequenced abundance data, *P. melaninogenica* was highly abundant, matching one 18.8 ng/µl, ~758 bp fragment and another 22.3 ng/µl, ~955 bp amplicon (Table 3.3). The smallest ITS profile was interpreted as *S. intermedius* (18.5 ng/µl, ~583 bp). A 6.4 ng/µl, ~676 bp fragment was postulated to belong to *S. aureus*, assuming its full profile was obscured by highly abundant Gram-negative species (Table 3.3). Similarly, the final 16.5 ng/µl, ~900 bp fragment could not be assigned to any species as it could have matched *S. maltophilia, A. xylosoxidans,* or *A. baumannii* which was assumed to have been

introduced through PCR reagents. It is equally possible that this ~900 bp amplicon obscured additional PCR amplicons sharing similar DNA molecular weights.

3.3.8.3 COPD mock community RISA profiles

The dominant species in this mock community could not be confidently assigned through RISA. *S. pneumoniae* was highly abundant, presenting as a 21.2 ng/µl, ~563 bp fragment (Table 3.3). The presence of a highly concentrated 60.5 ng/µl, ~819 bp amplicon, suggested a lack of QIAxcel resolution at separating amplicons that are similar in molecular weight and number which could include *Haemophilus, Streptococcus,* and *Veillonella* (Figure 3.2). *H. influenzae* or *V. parvula* could have also represented this ~819 bp fragment in addition to a 27.7 ng/µl, ~1080 bp amplicon, implicating its dominance (Table 3.3). These species were equally abundant in 16S rRNA gene sequence data (Figure 3.2). Similarly, *M. catarrhalis* and *P. melaninogenica* that were also abundant in 16S rRNA gene sequence data matched a 3.5 ng/µl, ~ 760 bp amplicon and partially matched a 2.7 ng/µl, ~665 bp amplicon respectively. Of note, the obscuring of multiple ITS profiles made this community appear less diverse than the CF *Pseudomonas*-dominated mock community (Figure 3.6).

3.3.8.4 Equimolar mock community RISA profiles

Either *V. parvula* and *H. influenzae* was proposed as the dominant species in this community sharing ITS amplicons that were present between 828-881 bp (56.4-68.5 ng/µl) and ~1151 bp (10.7 ng/µl) (Table 3.3). The smaller 6.4 ng/µl, ~587 bp amplicon and another 5.5 ng/µl, ~682 bp amplicon were assumed to have been partial matches to *Streptococcus* species which was predominant in 16S rRNA gene sequence data, *S. aureus* and *P. melaninogenica* (Table 3.3). A larger 12.1 ng/µl, ~772 bp amplicon could not be formally identified as it indicated the presence of *M. catarrhalis, P. aeruginosa,* and *P. melaninogenica,* assuming these species were abundant based on 16S rRNA gene sequence data (Table 3.3; Figure 3.2).

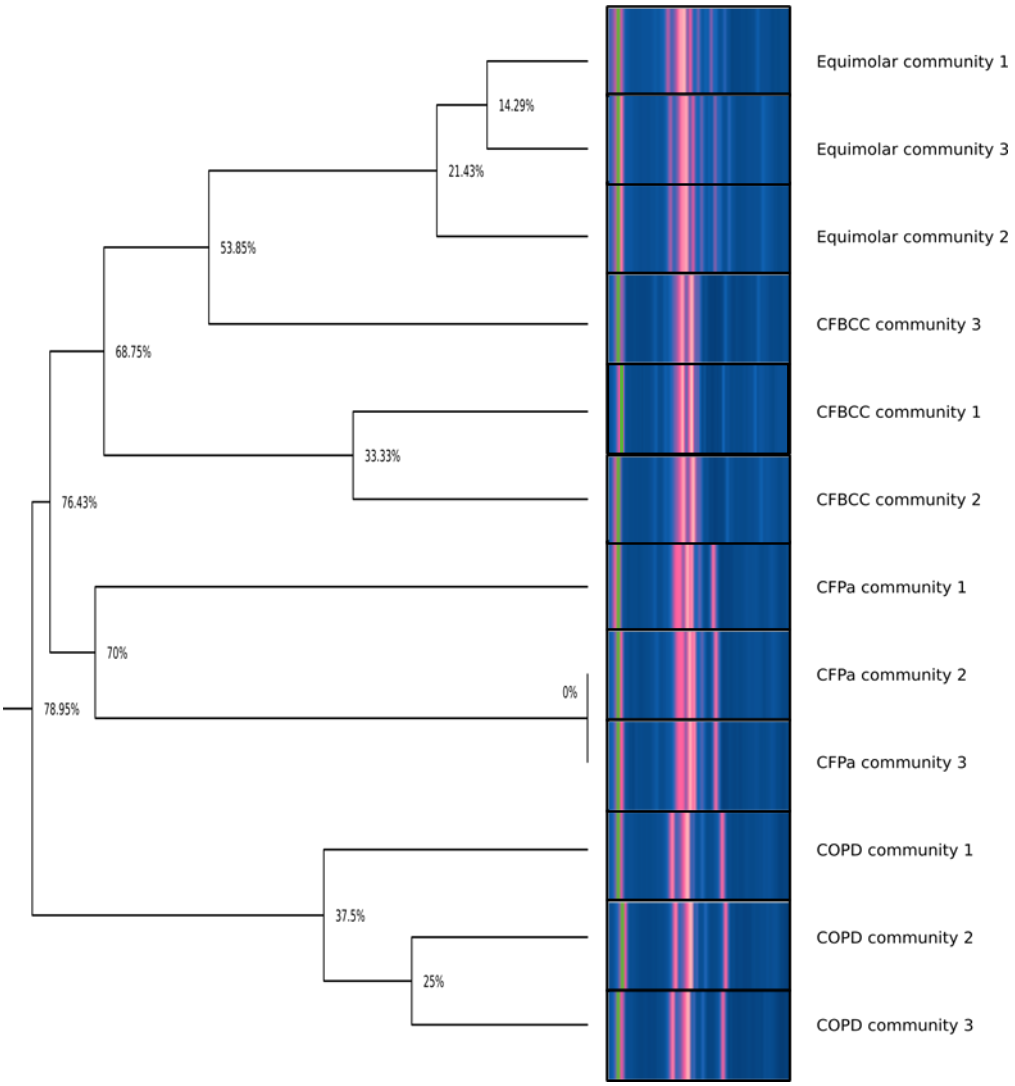


Figure 3.6 RISA mock community hierarchical clustering. A dendrogram using UPGMA clustering and the Sørensen-DICE algorithm to cluster RISA profiles by similarity of number and molecular weight of bands was generated. A tolerance threshold of 3% was used to separate DNA amplicons according to molecular weight. This dendrogram was manipulated in Figtree where node dissimilarity is showed. Raw QIAxcel gel images are shown on the right. Abbreviations include: CFBCC; CF *Burkholderia cepacia* complex-dominated mock community, CFPa; CF *Pseudomonas*-dominated mock community, COPD; COPD mock community, Equimolar; equimolar mock community. The QIAxcel upper marker is also visible in each sample and is highlighted in green.

3.4 Discussion

3.4.1 Overview

In this chapter, the performance of an optimised 16S rRNA gene sequencing protocol and RISA PCR was compared using mock bacterial communities prior to validating a gDNA extraction protocol based around these methodologies (**see chapter 4**). Bias introduced through choice of PCR primer, PCR amplification, and reference taxonomic database was also explored in detail. All remaining factors that bias sequence data including PCR reagents, choice of sequencer, and bioinformatic pipeline were either limited by PHW resource availability or chosen for its wide use in research, and will not be discussed further.

3.4.2 Evaluation of a 16S rRNA gene PCR and Illumina library preparation

The Illumina MiSeq was chosen to sequence the V4 portion of the 16S rRNA gene in artificial communities to evaluate its application in the routine sequencing of patient sputa (**see chapter 5)**. The V4 region was targeted for its ability to broadly detect respiratory microbiota associated with chronic pulmonary disease (Kitsios et al. 2018; Lucas et al. 2018; Mayhew et al. 2018). V4 amplicon sequencing is popular and has often been reported to provide greater taxonomic resolution than other hypervariable regions (Ghyselinck et al 2013; Tremblay et al. 2015; Fouhy et al. 2016; Yang et al. 2016; Sune et al. 2020).

The optimised 16S rRNA gene dual-indexing PCR protocol provided a simple and rapid method to prepare barcoded libraries in under 1.5 hours. This process also resulted in highly concentrated libraries that were consistently free from library preparation contaminants, facilitating optimal clustering efficiency and the production of high-quality sequence data. V4 primers 16Sf and 16Sr were specific to all target DNA, although the amplification of *C. acnes* was poor. Poor PCR amplification was attributed to a higher number of primer mismatches to this sample (Edgar et al. 2007). Primer mismatch bias is inescapable as no primer set is truly universal and may even fail to amplify relevant bacteria species (Frank et al. 2008; Klindworth et al. 2013; Salipante et al. 2014). Importantly, all clinically relevant respiratory bacteria tested could be amplified using these V4 primers.

Despite this success, non-specific amplification products persisted in *A. xylosoxidans, A. baumannii* and mock community DNA samples. Following the improvement of PCR cycling conditions, primer design was considered most likely responsible for these unwanted observations. Neither primer shared complementary sequences to unintended genomic targets in any sample, although primer 16Sf possesses a higher capacity to self-hybridise which could have potentially doubled the size of amplified PCR products. Additionally, the difference in forward and reverse primer melting temperature (T_m) was greater than 5°C, not

permitting flexibility in choosing annealing conditions without reducing amplification efficiency. This effect was not mitigated by using barcoded primers. The presence of minor non-specific PCR products did not pose concern following their removal through PCR purification and selective filtering of 16S rRNA gene sequences in Mothur.

3.4.3. The effect of 16S rRNA gene reference database choice on the taxonomic assignment of mock community bacteria

Mock communities are often used in the evaluation and validation of 16S rRNA gene-based metaprofiling studies. Made of predetermined ratios of DNA, mock communities enable the quantification of sequence error and estimation of bias introduced through sampling and library preparation (Pollock et al. 2018). A major process in 16S rRNA gene-based microbiome research is microbial taxonomic assignment through public 16S rRNA gene reference databases like SILVA (Pruesse et al. 2007), RDP (Cole et al. 2009), Greengenes (DeSantis et al. 2006), GenBank (Sayers et al. 2021), and EzTaxon (Kim and Chun 2014). Reference databases must collect relevant, up-to-date, and high-quality 16S rRNA gene sequences where raw data must align correctly to alignment coordinates. SILVA, RDP, and Greengenes are commonly used for these purposes and the performance of these databases vary depending on environmental niche, target region of the 16S rRNA gene and OTU clustering method (Ritari et al. 2015). Such differences encourage the comparison of multiple databases in order to choose one that most accurately represents each niche/microbiome.

SILVA was superior in the taxonomic assignment of respiratory microbiota followed closely by RDP. SILVA also resolved most unclassified sequences, standing at 0.03% of total reads. A low percentage of unclassified sequences are expected due to unclassified/unculturable bacteria present in a sample, chimeric sequences, and the deposition of poor-quality/erroneous sequences to 16S rRNA gene-based reference databases where estimates are thought to be between 0.2% to 2.5% (Kim and Chun 2014; Myer et al. 2020). In contrast, Greengenes failed to identify *C. pauculus* and *P. aeruginosa* to the genus level. Misidentification was partly attributed to an outdated database (last updated in 2013) and Greengenes aligning variable regions poorly which has been known to exaggerate OTU counts and diversity (Schloss et al. 2010). Contrariwise, Greengenes alone was more capable of clustering Streptococcal OTUs correctly as indicated through 16S rRNA gene-based phylogenetic construction. Both SILVA and RDP summed *S. anginosus* and *S. intermedius* into a single OTU, highlighting the inability of 16S rRNA gene sequencing in defining OTUs as species. Regardless of these reported differences, changes in mock community composition excluding unclassified genera were remarkably similar between each reference database.

3.4.4 Exploring 16S rRNA gene PCR and sequencing bias in mock communities

All species including low abundance taxa in each mock community were successfully identified, apart from *C. acnes* in two COPD mock community replicates. Each community could be differentiated according to major CF and COPD OTUs, supporting hypothesis 1 (Both 16S rRNA gene-based sequencing and RISA PCR are able to differentiate CF and COPD-specific communities by community composition and membership). PCR amplification bias caused significant deviations in dominant and abundant microbial composition. Gram-positive species were usually preferentially amplified except in COPD mock communities. This result signified the inability to predict PCR amplification bias using an even mock community distribution as this bias is dependent on the mix of species constituting a community. Generally, Gram-negative species were underrepresented and *B. cepacia* and *P. aeruginosa* lost the highest expected portion of sequenced reads. Gram-negative *P. melaninogenica, A. baumannii* and *H. influenzae* species were overrepresented.

These biased observations could be partly explained by primer degeneracy. This would have biased PCR amplification for all species due to different thermodynamics of primer variants annealing to template DNA with different strengths at 55°C. Moreover, primer 16Sr is a highly degenerate primer (Edgar et al. 2007 Tremblay et al. 2015) Template GC content and homopolymers have also been shown to influence polymerase efficiency and therefore amplification bias (Edgar et al. 2007). More significantly, PCR amplification bias is influenced by 16S rRNA gene copy number per cell, favouring the amplification of species containing more copies of the gene, for example, *H. influenzae, A. baumannii* and *Streptococcus* spp. 16S rRNA gene copy number did not however account for poorer amplification of ENFGN bacteria, *B. cepacia* or *P. aeruginosa.* Incorrect preparation of mock communities is unlikely as multiple aliquots were used for Illumina library preparation. PCR reagents could also have contributed to this skewed community composition; (Mallott et al. 2019) reported the preferential amplification of *Bacteroidetes* and underrepresentation of *Proteobacteria* using AccuStart reagents. Finally, differences in observed vs. expected microbial composition might have also been attributed to sequence error rates. One would expect MiSeq error rates ~0.01% in conjunction with filtering of poor quality and erroneous sequences (Pollock et al. 2018). Error rates fluctuate among hypervariable regions although V2 and V4 regions have been reported as most accurate (Vilo and Dong 2012). Nevertheless, error rates between 1.1%-2.6% were reported between each sequenced batch. This higher portion of erroneous sequences would have likely exaggerated OTU counts and misclassified more sequences and therefore cannot be ignored from analysis. Despite this, sequence data was highly replicable between batches.

3.4.5 The effect of V4 sequencing on bacteria diversity

Using several popular diversity indices, mock community alpha and beta diversity was measured to address whether diversity was qualitatively and quantitatively distinct between each community and whether these measures could be applied to 16S rRNA gene-based sequence analysis in patient sputa. The simplest way to describe and quantify microbial communities is the species richness concept (Escobar-Zepeda et al. 2015). Community richness was overinflated compared to previous predictions since richness is heavily influenced by OTU counts. This result is common where *de novo* OTU picking algorithms are applied and spurious OTUs generated through PCR are detected at low abundance (Myer et al. 2020). The introduction of spurious OTUs from higher sequencing error should also be considered. Chimeric sequences are another factor impacting diversity measure as closely-related 16S rRNA gene sequences form chimeras more readily and not all chimeric sequences pass filtering (Haas et al. 2011). Regardless of this inflated richness, observed richness largely reflected initial predictions, being significantly different between the most and least rich mock communities**.** Unlike richness, diversity measurements were largely unaffected by PCR bias, sequencing bias, and error, and was more accurate when comparing both InvSimpson and Shannon diversity. Likewise, diversity was statistically significant between all mock communities and largely followed previous predictions. This result demonstrates the ability of short-read NGS in differentiating respiratory pathogens according to disease profile, supporting hypothesis 1.

Beta diversity was measured using the qualitative unweighted UniFrac and quantitative weighted UniFrac algorithms. UniFrac was chosen for being robust as they account for degree of divergence between sequences, sample categories and across environmental gradients (Lozupone et al. 2007; Caporaso et al. 2011). Unweighted UniFrac detects the number of sequence changes from each lineage (community membership) and is better at assessing population founding effects and restricting conditions e.g., temperature. Weighted UniFrac also detects changes in which taxa are present (community structure). This ability is useful for understanding the factors that cause shifts in communities e.g., a nutrient limiting source (Lozupone et al 2007). UniFrac may also address the factors that change community structure and membership such as patient age, lung function and antibiotic use. 16S rRNA gene-based NGS was also successful in differentiating community membership and structure since the genetic diversity of each community was greater than its pooled diversity, supporting hypothesis 1. Both community membership and structure were found to be statistically different between community types, which was reproducible between each sequenced batch. This result demonstrated its suitability to address bacterial diversity in patient sputa in addition to addressing key questions regarding patient health, age, sex, and other relevant demographics.

3.4.6 Evaluation of RISA PCR in mock communities

To accompany routine 16S rRNA gene sequencing, the application of a RISA PCR for community diversity and to a lesser extent, species-level resolution was evaluated. RISA strongly amplified abundant species and proved to be a rapid and simple method of visualising species-specific ITS profiles; however, this tool did not reflect true community diversity. It is noteworthy to mention ITS profiles vary according to PCR protocol and electrophoretic method, therefore it is important to establish a set of reference RISA bands from both experimental and *in silico* predictions. A more relaxed criterion in interpreting RISA amplicons is necessary to infer species which amplify similarly sized RISA profiles

RISA was very effective at visualising dominant and abundant species in mock communities and appeared to have outperformed V4 primers in the identification of *B. cepacia* and *P. aeruginosa* and possibly *Proteobacteria.* This method cannot be used as a diagnostic tool alone as many abundant species identified through RISA could realistically be classed only to the genus level, in support of hypothesis 2 (RISA PCR can generate clear profiles for a pure bacterial culture and single species, but is limited in its ability to classify constituents of a complex respiratory community). This is due to partial matches to bacteria references and closely related species producing highly similar ITS profiles. Nevertheless, RISA could prove useful alongside 16S rRNA gene sequencing in inferring dominant and highly abundant bacterial species that produce simple RISA profiles, for example, *P. aeruginosa* and *S. maltophilia*. Furthermore, RISA can at least provide genus level discrimination of taxa where 16S rRNA gene sequencing might be limited to family level resolution. Additional caution is required when interpreting RISA PCR data for a number of reasons. Firstly, DNA molecular weights vary because capillary electrophoresis run samples sequentially. Secondly, hierarchical clustering of RISA profiles uses raw gel images to cluster profiles by similarity (amplicons manually assigned). Thirdly, the number of RISA bands does not necessarily correspond to high or low diversity. Finally, low abundance taxa or species generating similar ITS profiles cannot be confidently assigned using this method.

3.4.7 Future direction

In future, the creation of a 16S rRNA gene-based reference database specific to respiratory microbiota is recommended as the vast majority of bacteria 16S rRNA gene reference sequences originate either from the gut or the environment. This prioritisation would eventually lead to a more rapid and correct identification strategy. A gDNA extraction method suitable for high throughput and low turnaround time must be determined using both optimised DNA-based methods in pure bacteria cultures representing respiratory pathogens and is capable of

resisting DNA extraction and PCR inhibitory components of sputum; the development of such a method is covered in chapter 4.

While Illumina 16S rRNA gene-based metaprofiling is becoming commonplace in clinical microbiology laboratories, the adoption of long-read 16S rRNA gene sequencing is worth investigating to capture accurate community composition and diversity without amplification bias. Though promising, long-read sequencing technology will require improved throughput and significant reductions in costs and sequence error rates for this to be feasible.

3.4.8 Conclusions

16S rRNA gene NGS is valuable for routine laboratory diagnostics where there is a need for rapid turnaround time, maximised throughput, and minimised costs. The optimised 16S rRNA gene sequencing protocol was informative for the broad detection of respiratory microbiota, making clear distinctions in disease-specific community composition and diversity, accepting hypothesis 1. Limitations of this methodology in identifying species, and at times, genera were apparent. Incorrect clustering of OTUs from closely-related species, PCR bias, and sequencing artifacts led to inflated diversity and skewed community composition. It is vital to understand how these factors impact the validity of 16S rRNA gene sequence data and how standardisation of sequencing protocols largely mitigate but does not eliminate these factors. RISA is a promising tool to accompany 16S rRNA gene sequence analysis for its low cost, rapidity, its ability to provide basic species-level prediction and in certain circumstances, identify or compare deviations in community composition. Hypothesis 1 was disproven where for the most part, RISA could not differentiate all abundant mock community species however hypothesis 2 was accepted where high-quality single species references could be predicted. Its application in complex clinical samples should be investigated further where there was lower discriminatory power.

Chapter 4 Evaluation of automated and manual DNA extraction methods for the routine gDNA extraction of sputum microbiota

4.1 Introduction

Genomic DNA (gDNA) extraction plays an essential role in NGS for enabling the characterisation of all culturable and unculturable microorganisms within clinical samples (Yuan et al. 2012). This process has been simplified through the use of gDNA extraction kits which are designed to lyse microbial cells mainly through mechanical or physical lysis coupled with chemical lytic agents (Islam et al. 2017). Following lysis, gDNA is captured and purified to remove amplification inhibitors (Li et al. 2020). Bacterial gDNA extraction can be achieved through organic extraction (phenol-chloroform), inorganic extraction (salting-out, Chelex® 100 resin, CsCl density gradient method), and solid-phase adsorption, usually to a silica-coated membrane/paramagnetic bead (Ali et al. 2017). Most modern commercially available bacteria gDNA extraction kits employ solid-phase gDNA purification using silica matrices, glass, diatomaceous earth, or anion exchange (Tan and Yiap 2009). Solid-phase DNA extraction is a popular choice for being simpler to operate, more economical, safer, quicker, and higher throughout than traditional gDNA extraction methods (Faraji et al. 2019). In spite of this, few extraction kits are designed for complex microbial communities, thus understanding the limitations of these kits for clinical use is essential along with an understanding of how gDNA extraction bias impacts interpretability of downstream sequence-based analysis.

To gain an accurate representation of sputum bacterial community composition, gDNA extraction bias must be identified and eliminated where possible. Cell lysis is the principal factor causing gDNA extraction bias; it heavily impacts bacterial community composition as cell lysis efficiencies vary greatly depending on bacterial cell wall, growth status, sample type, and bacteria or host-derived extracellular components acting as inhibitory agents (Lopes et al. 2018; Nagler et al. 2018; Li et al. 2020). Oftentimes, unequal lysis efficiencies arise between Gram-positive and Gram-negative bacteria. This is due to the Gram-positive peptidoglycan cell wall which is more resistant to lysis compared to the thinner peptidoglycan layer and lipopolysaccharide outer membrane found in Gram-negative species (Silhavy et al. 2010). Similarly, *Mycobacteria* are heavily resistant to lysis due to their lipid-rich cell wall, containing long-chain mycolic acids and polysaccharides (Chiaradia et al. 2017; Kolia-Diafouka et al. 2018). Consequently, suboptimal lysis may lead to a heavy underrepresentation of these hard-to-lyse organisms, leading to erroneous representation of microbial communities which directly impacts patient treatment strategies.

To address cell lysis bias, generalised bacterial gDNA extraction methodologies typically incorporate bead-beating, chemical, and enzymatic lysis that target physically robust microorganisms at the cost of damage to Gram-negative gDNA (Gill et al. 2016). It is also important to identify and reduce other sources of variation that arise through gDNA extraction including reagent type, laboratory personnel, manual vs. automated gDNA extraction, and sample handling to name a few (Pérez-Brocal et al. 2020). Sputum handling is especially vulnerable to contamination from laboratory processing and the upper respiratory tract, depending on sampling method (Drengenes et al. 2019). Sources of contamination should be identified to rule out false positives that could lead to the incorrect interpretation of bacterial community composition and potential clinical outcomes henceforth. Both positive and negative controls are equally important. Extraction kit blanks and reagent blanks can help identify contaminating species where these sources of contamination cannot be eliminated (Salter et al. 2014; Pérez-Brocal et al. 2020). Identifying and quantifying contaminants is key to correct community analysis and to enable normalisation across personnel, extraction kit batches, experiments, and laboratories (Pérez-Brocal et al. 2020). In turn, normalisation reduces both endogenous and exogenous variation in the DNA extraction process.

Even so, analysis of sputum microbiota poses many challenges in terms of technical difficulties and reproducibility of gDNA extraction methodologies (Li et al. 2020). The majority of existing DNA extraction protocols are focused on 16S rRNA gene-based metaprofiling or metagenomic studies from high biomass faecal samples that are incompatible with low biomass samples such as blood, urine, skin, and sputum (Greathouse et al. 2019; Selway et al. 2020; Sui et al. 2020; Karstens et al. 2021). The human lung harbours an estimated $2.2x10^3$ bacterial genomes per cm (Pérez-Brocal et al. 2020). As a consequence, sputum bacterial load is too low to permit multiple gDNA extractions and therefore a single lysis step is required to disrupt physically robust bacteria without shearing gDNA from easily lysed cells (Li et al. 2020). In research, numerous DNA extraction methodologies have been applied to sputum, but this lack of standardisation prevents comparable and reproducible results between studies (Pérez-Brocal et al. 2020). Modifications to sputum DNA extraction protocol using specialised lytic or digestion steps are necessary due to the physical properties and purulence of sputum, limiting the application of widely used DNA extraction methods (Pérez-Brocal et al. 2020). To address these limiting factors, sample storage and handling should be optimised before standardising cell lysis, gDNA extraction, and gDNA isolation (Hallmaier-Wacker et al. 2018). Notwithstanding this principle, to adhere to a single standardised sputum gDNA extraction protocol for research and routine clinical diagnostics is practical due to ever increasing demands for high-throughput analysis, low turnaround time and automation.

4.1.1 Aims and objectives

The aim of this chapter was to design and validate a bacterial gDNA extraction method suitable for respiratory specimens within a public health diagnostic laboratory context. This protocol should also offer high throughput, low turnaround time, low cost, and high translatability for PHW routine gDNA extraction. Suitable technologies for this work included a PHW EMAG automated gDNA extraction platform that works by capturing gDNA to a silica membrane in the presence of chaotropic salts. This process allows for a simple hands-free, bind-wash-elute process that has been validated for the detection of respiratory pathogens. This platform was then compared against a ZymoBIOMICS™ DNA Miniprep kit which uses a manual spin column (silica-based) procedure designed for microbiome and metagenome analyses from various clinical samples.

The objectives of this chapter were:

1) Optimise a bead-beating protocol for automated and manual gDNA extraction.
2) Compare 16S rRNA gene PCR and RISA PCR limits of detection for a gDNA extraction kit using optimised bead-beating and enzymatic lysis protocols.
3) Preliminary RISA identification of sputum spike-in organisms using gDNA extraction methods that enabled lowest PCR limit of detection (LoD).
4) Assess 16S rRNA gene PCR and RISA PCR limit of detection in sputum using spike-in organisms.

4.1.2 Hypotheses

Hypotheses for this chapter were:

1) Bead-beating is required to efficiently lyse sputum and physically robust microorganisms.
2) A column-based gDNA extraction kit designed for microbiome analysis would better-represent Gram-positive and Gram-negative microbiota compared to the EMAG platform at the cost of lower quality extracts.

4.2 Materials and methods

4.2.1 Experimental overview and microorganism preparation

S. aureus and *B. cepacia* were selected to compare extraction efficiencies of EMAG and ZymoBIOMICS™ DNA Miniprep kits of the easy-to-lyse Gram-negative *B. cepacia* and the hard-to-lyse Gram-positive *S. aureus.* Cell lysis through bead-beating was optimised for both EMAG and ZymoBIOMICS™ platforms (**see sections 4.1.3 and 4.1.4**). Since no lysis bead has been previously validated for EMAG respiratory microbiota extractions, several materials were compared before optimising bead-beating (**see section 4.1.3**). Enzymatic lysis with MetaPolyzyme was also explored in EMAG gDNA extractions with the aim to improve Gram-positive and Gram-negative gDNA extraction efficiencies, gDNA yield and quality. These optimised lysis protocols were used to assess V4 16S rRNA gene PCR and RISA PCR limit of detection (LoD) in pure culture and in spiked sputum unless stated otherwise (see **sections 4.1.5 and 4.1.6**).

Pure bacterial TSB cultures were incubated for 18 h at 37°C (150 rpm shaking) before pelleting at 4000 rpm for 10 minutes to resuspend samples in 1X sterile PBS. A spectrophotometer measured bacteria OD_{600} before normalising samples to ~$2x10^8$ CFU/ml for further dilution.

4.2.2 Sputum digestion

Sputum was processed within 48 hours upon collection in a PHW containment level 3 (CL3) laboratory where 10 ml volumes were required to generate a dilution series. Large volumes were essential to provide a comparable microbial community profile to detect spike-in organisms from a sputum blank control. An equal volume of Mucolyse™ containing a 1X dithiothreitol (DTT) concentration was added to an equal volume of sputum following manufacturer instruction. Additional Mucolyse™ was slowly mixed in highly purulent samples until homogenous. Samples were homogenised for at least 30 seconds depending on purulence. Samples were then incubated at room temperature for 15 minutes prior to extraction.

4.2.3 EMAG lysing bead selection

For its intended application in microbiome analysis and comparability to the ZymoBIOMICS™ DNA Miniprep kit, ZR BashingBead™ tubes were selected to optimise EMAG bead-beating. ZR BashingBead™ tubes constituted of 0.5 mm and 0.1 mm beads of an undisclosed matrix to lyse bacteria, yeast, and mycelial fungus. Three additional types of lysing beads were simultaneously evaluated for the highest gDNA extraction efficiency between each type. These materials included 1.3 mm ceramic beads, 2 mm glass beads, and MP Biomedicals lysing matrix E beads containing 1.4 mm ceramic beads, 0.1 mm silica beads and a 2 mm

glass bead. Bead lysis efficiencies were estimated through quantifying V4 16S rRNA gene PCR amplicons that were generated through amplifying ~10^4-~10^5 CFU/ml of extracted *S. aureus* gDNA that was lysed through bead-beating (**see section 2.4**; Table 2.19). Amplified DNA was quantified instead of isolated gDNA as bacteria were too low in abundance to be detected. To achieve this, 250 µl of diluted *S. aureus* cultures were transferred into 5M 0.9 ml guanidinium thiocyanate (GTC) and samples were deactivated following methods outlined in **section 2.3.3**. Due to lysis buffer strength and to safely work in a CL3 laboratory, inactivated samples had to be split into 2 lysis tubes containing 1 g of aliquoted beads. Lysis was achieved using a general bead-beating setting of 4000 rpm for 2 x 30 seconds using a Precellys 24 tissue homogeniser. *S. aureus* gDNA was extracted following EMAG gDNA extraction procedure outlined in **section 2.3.3**. DNA quantity and quality was measured using a 2200 TapeStation HS D1000 ScreenTape. DNA quality was assessed through the TapeStation DNA integrity number (DIN) where scores range from 1 (lowest quality) to 10 (highest quality). For each lysis method, 2 replicates of extracted and amplified *S. aureus* DNA were generated.

4.2.4 EMAG and ZymoBIOMICS™ bead-based lysis optimisation

Of the 3 bead types tested in **section 4.1.3**, the one which resulted in greatest lysis efficiency (greatest gDNA yield and quality) was selected for comparison alongside ZR BashingBead™ lysing beads for EMAG bead-beating optimisation. All of these samples required being split into 2 lysis tubes prior to lysis to prevent sample leakage due to high GTC molarity. ZymoBIOMICS™ DNA Miniprep kit bead-beating was optimised using its own validated ZR BashingBead™ lysis tubes. Optimal bead-beating speed and duration were determined as the setting which resulted in highest gDNA yields, most even Gram-positive and Gram-negative gDNA extraction efficiencies, and/or highest gDNA quality. Bead-beating settings were typically assessed from lowest to highest (Table 4.1). 250 µl *S. aureus* and *B. cepacia* cultures were lysed (Table 4.1) and extracted following methodologies outlined in **sections 2.3.3 and 2.3.4**. gDNA quantity and quality (DIN) was measured using a Qubit™ 2.0 fluorometer HS dsDNA kit and a 2200 TapeStation gDNA ScreenTape. 2 biological and 2 technical replicates of *S. aureus* and *B. cepacia* extracts were generated for each setting. EMAG RNA co-extraction was quantified with a Qubit™ 2.0 fluorometer HS RNA kit and a 2200 TapeStation RNA ScreenTape in 2 randomly selected *S. aureus* extracts.

Table 4.1 Bead-beating settings used for EMAG and ZymoBIOMICS™ gDNA extractions

EMAG settings	ZymoBIOMICS™ kit settings
4000 rpm (2 x 30 s)	4000 rpm (2 x 30 s; 5 minutes total)[b]
4000 rpm (2 x 90 s)	
5000 rpm (2 x 30 s)	
5000 rpm (2 x 90 s)[a]	5000 rpm (2 x 90 s)
6000 rpm (2 x 30 s)[a]	6000 rpm (2 x 30 s)

[a]These settings were not explored for EMAG gDNA extractions using ZR BashingBead™ lysis tubes due to severe gDNA shearing.
[b]Optimal settings suggested by PHW, lower settings were not pursued because of this.

4.2.5 EMAG and ZymoBIOMICS™ DNA Miniprep kit pure culture PCR limit of detection and preliminary RISA of spiked sputum

S. aureus and *B. cepacia* pure cultures were first used to assess EMAG and ZymoBIOMICS™ 16S rRNA gene PCR and RISA PCR limit of detection (LoD) using optimal bead-beating settings. These results were then compared against EMAG PCR LoD following MetaPolyzyme processing guidelines, i.e., comparing mechanical lysis to enzymatic lysis (Table 2.16; **see section 2.3.3**). Cultures were prepared as outlined in **section 2.3.3** where 250µl of $2x10^8$-$2x10^2$ CFU/ml of culture was extracted and PCR amplified (**see section 2.3.3; section 2.4**; Table 2.18; Table 2.19). EMAG lysis tubes were also split to prevent sample spillage with GTC. Beads were washed with 100 µl lysis buffer and centrifuged at 13000 G for 1 minute to maximise gDNA recovery. This resulted in the recovery of an additional 200 µl lysate to transfer into the EMAG.

Following these experiments, 250 µl of $2x10^6$ CFU/ml *S. aureus* and *B. cepacia* was spiked into sputum and was extracted using both EMAG-enzymatic lysis and ZymoBIOMICS™ gDNA extraction protocols. This high cell density was first used to ensure target bacterial gDNA could be extracted from sputum and detected through RISA PCR amplification (Table 2.18). The reproducible amplification of core *S. aureus* and *B. cepacia* ITS amplicons were priority in these experiments. 2 types of pre-treatments were applied to EMAG extracts from separate samples involving pelleting or homogenisation of sputum for 30 seconds prior to inoculation of bacteria and the addition of 25 µl MetaPolyzyme and 0.1% (v/v) sodium azide. This pre-treatment differed to assess which strategy resulted in most reproducible RISA profiles. Pelleting was accomplished by sample centrifugation at 4000 rpm for 10 minutes, washing twice, and resuspending in 1 ml 1X sterile PBS before inoculation. A *recA* PCR was also performed on pelleted gDNA extracts to confirm the presence of *Burkholderia* where RISA profiles were unclear (Table 2.20). For each lysis method, dilution, PCR assay and bacteria species, 2 biological and 2 technical replicates were generated. Pure culture *S. aureus* and *B.*

cepacia extracts were generated as positive controls to match ITS profiles to each sample. Amplified DNA was quantified using a Qubit™ 2.0 fluorometer HS dsDNA kit and visualised using 1% agarose gel electrophoresis and a 2200 TapeStation a D1000 ScreenTape. Where required, putative species identification through RISA analysis was achieved by using existing *in silico* and experimental data (Table 3.3), and by creating a new PRIMER-BLAST *in silico* RISA database of bacteria species that could generate ITS profiles to spike-in organisms.

4.2.6 Full sputum spike-in 16S rRNA gene PCR and RISA PCR LoD

After comparing ZymoBIOMICS™ and EMAG preliminary spike-in extracts, the ZymoBIOMICS™ DNA Miniprep kit was selected to assess spike-in *S. aureus* and *B. cepacia* 16S rRNA gene PCR and RISA PCR LoD in sputum using $2x10^6$-$2x10^2$ CFU/ml of bacteria (Table 2.18; Table 2.19). This dilution series was chosen to cover above and below pure culture PCR detection limits as spike-in detection limits were postulated to have fallen between these values. For each species, dilution, and PCR assay, 2 biological and 2 technical replicates were generated. RISA PCR LoD was interpreted by examining the presence/absence of species-specific ITS profiles (most concentrated core ITS amplicons) while 16S rRNA gene PCR LoD was interpreted by comparing the presence or absence of *Burkholderia* and *Staphylococcus* 16S rRNA gene sequenced reads between spike-in samples and against pure culture positive controls, a community blank, a gDNA extraction kit blank, and reagent controls. Sequenced reads were classified through Mothur using the SILVA reference database (**see section 2.6**). RISA PCR amplicons were visualised using the Qiagen QIAxcel Advanced System high-resolution DNA cartridge, a 15 bp-3 kb QX DNA alignment marker and a 100 bp-2.5 kb QX DNA size marker. Here, this method replaced TapeStation analysis to achieve greater resolution in DNA fragments spanning ~500 bp-2.5 kb (**see section 2.5.3**). RISA profiles were interpreted using existing *in silico* and experimental data (Table 3.3) in addition to creating additional PRIMER-BLAST *in silico* RISA bacteria references that might also generate similar ITS profiles to spike-in organisms.

4.3 Results

4.3.1 Bead-beating optimisation

4.3.1.1 Ceramic, glass, and mixed bead lysis efficiencies

Initial evaluations optimised the number of bacteria cells and bead types required to produce sufficient DNA for the subsequent comparative analysis of EMAG DNA extraction methods. A minimum of ~$2x10^5$ CFU/ml pure culture extract was required for successful 16S rRNA gene PCR amplification. Glass bead-beating resulted in poorest *S. aureus* cell lysis where resulting amplified DNA was below the Qubit's quantitation range. MP Biomedicals mixed beads marginally improved cell lysis where ≤0.1 ng/µl DNA was amplified. Ceramic bead lysis yielded ≤0.2 ng/µl amplified DNA, and was the most effective bead-based lysis method of the three. As the most effective beads, ceramic beads were chosen to optimise the EMAG bead-beating protocol (objective 1).

4.3.1.2 EMAG Ceramic bead-beating optimisation

Ceramic bead lysis resulted in tenfold as much concentrated *B. cepacia* gDNA than *S. aureus* extracted gDNA. The recovery of *S. aureus* gDNA was in contrast consistently poor regardless of bead-beating speed and duration (Figure 4.1; SI Figure C1). While *S. aureus* ceramic bead lysis efficiency was poor, DIN scores generally reflected adequate gDNA extract quality for both *S. aureus* (5.6-6.8) and *B. cepacia* (7.0-7.9). DIN scores generally improved as bead-beating speed decreased however figures could not be generated for dilute *S. aureus* gDNA extracts ≤1.3 ng/µl (SI Figure C1). In contrast to the DIN scores, *S. aureus* gDNA molecular weight (ranging 11531-19282 bp) was typically higher than that of *B. cepacia* (ranging 9739-15541 bp; Figure 4.1; SI Figure C1). These discrepancies were influenced by RNA co-extraction of bacterial 16S and 23S rRNA subunits, thus lowering sample gDNA quality (SI Figure C2).

B. cepacia gDNA molecular weight was reduced through aggressive bead-beating at 5000 rpm (2 x 90 seconds) and 6000 rpm (2 x 30 seconds; SI Figure C1). While detrimental to gDNA integrity, this resulted in highest observed gDNA yields (1.6-4.1 ng/µl *S. aureus* gDNA and 32.7-35.8 ng/µl *B. cepacia* gDNA). Lysis at 5000 rpm (2 x 30 seconds) was considered optimal for *B. cepacia* for producing high gDNA yields while maintaining its integrity (SI Figure C1). Because of this short setting, *S. aureus* gDNA yields were among the lowest observed (1.1-1.3 ng/µl) although gDNA degradation was minimal (SI Figure C1). Higher *S. aureus* gDNA yields were achieved using a longer lysis setting (4000 rpm for 2 x 90 seconds or 5000 rpm for 2 x 90 seconds; Figure 4.1; SI Figure C1). To accommodate for improved *S. aureus* gDNA yield without compromising *B. cepacia* gDNA integrity and yield, a setting of 4000 rpm (2 x 90

seconds) was determined to be most appropriate for both species, fulfilling objective 1 (optimising an EMAG bead-beating protocol) (Figure 4.1).

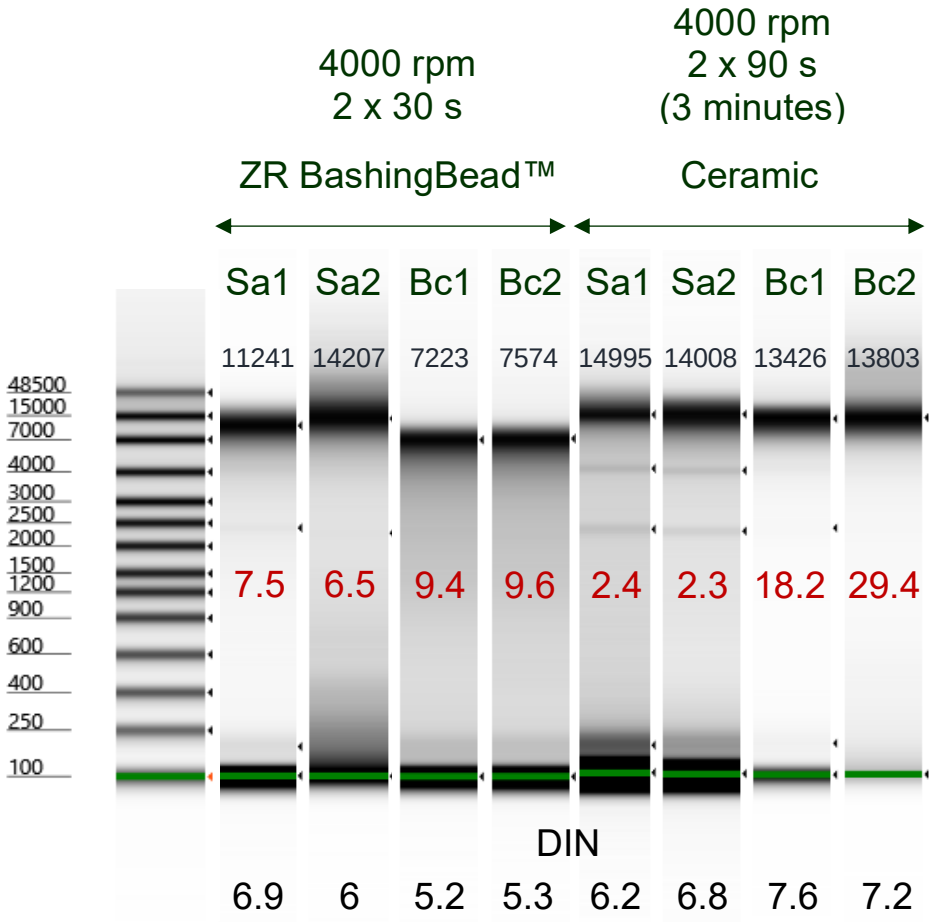


Figure 4.1 Best suited bead-beating settings for EMAG-ceramic and EMAG-ZR BashingBead™ gDNA extractions. *S. aureus* (Sa) and *B. cepacia* (Bc) replicates are shown on this composite image. Each sample has been scaled to the most concentrated electropherogram peak and contrast has been enhanced to observe gDNA degradation. The first lane shows the genomic DNA ScreenTape ladder, and the Tapestation lower marker is highlighted in green in each lane. Sample DNA molecular weight (in bp) has been included beneath each sample label (navy). gDNA concentration (in ng/µl) for each sample has been annotated in each lane (red). DNA integrity (DIN) has been included beneath each lane.

4.3.1.3 EMAG ZR BashingBead™ lysing bead bead-beating optimisation

Lysis efficiencies of ZR BashingBead™ lysis tubes were also compared to EMAG-ceramic bead-beating to determine which bead matrix was optimal to lyse bacteria with the EMAG and for subsequent PCR limit of detection assays. ZR BashingBead™ lysis coupled with EMAG chemistry resulted in significant gDNA degradation and loss of gDNA yield and quality for both *S. aureus* and *B. cepacia,* resulting in fewer bead-beating steps being performed. DNA

degradation was especially emphasised through reduced *B. cepacia* DIN scores (ranging 3.4-5.3) and sample gDNA molecular weights (ranging 5937-8120 bp). This also resulted in poor gDNA recovery overall (9.2-21.1 ng/µl; Figure 4.1; SI Figure C3). gDNA quality was severely impacted by extended bead-beating duration (4000 rpm for 2 x 90 seconds) whereas a reduction in bead-beating duration (2 x 30 seconds) resulted in highest gDNA integrity (DIN scores 5.2-5.3) at the cost of lowest gDNA yields (9.4-9.6 ng/µl; Figure 4.1; SI Figure C3).

Because ZR BashingBead™ lysing beads decreased *B. cepacia* gDNA yield and integrity, both Gram-positive and Gram-negative lysis efficiencies appeared more comparable following this methodology. Despite ZR BashingBead™ lysis recovering higher *S. aureus* gDNA yields (6.5-18.8 ng/µl), lower gDNA extract quality was observed compared to ceramic bead lysis (DIN 5.2-6.9 spanning 9221-14207 bp; Figure 4.1; SI Figure C3). Total degradation of *S. aureus* gDNA was observed at 5000 rpm, leaving only co-extracted RNA in these samples (SI Figure C3). Gentler bead-beating at 4000 rpm gave rise to nearly equal lysis efficiencies between *S. aureus* and *B. cepacia* where bead-beating for 2 x 30 seconds produced highest quality gDNA extracts (DIN 6-6.9) while extending bead-beating to 3 minutes produced greatest gDNA yields (14.7-18.8 ng/µl; Figure 4.1; SI Figure C3). A bead-beating setting of 4000 rpm for 2 x 30 seconds was determined to be most appropriate for the lysis of these species as a compromise for gDNA quality and yield.

4.3.1.4 ZymoBIOMICS™ DNA miniprep kit bead-beating optimisation

ZymoBIOMICS™ bead-beating was also optimised prior to comparing PCR LoD to EMAG extracts. ZymoBIOMICS™ bead-beating protocol greatly improved *S. aureus* and *B. cepacia* gDNA quality compared to EMAG-ZR BashingBead™ lysis, excluding the highest bead-beating setting which led to complete gDNA degradation (Figure 4.2). Because of this result, *B. cepacia* was not tested at this setting. This protocol resulted in a slight increase in gDNA shearing as demonstrated by reduced gDNA molecular weight of both *S. aureus* (range 7549-10786 bp) and *B. cepacia* (range 6970-11022 bp; Figure 4.2). DIN scores more closely reflected EMAG-ceramic gDNA extracts which were highly similar between *S. aureus* (range 6.4-6.8) and *B. cepacia* (range 6.3-6.7).

Unlike the EMAG, the ZymoBIOMICS™ kit produced concentrated *S. aureus* (range 20.4-36.7 ng/µl) and *B. cepacia* (range 21.3-31.2 ng/µl) gDNA extracts. Both Gram-positive and Gram-negative gDNA extraction efficiencies were most even when subjected to 4000 rpm bead-beating (2 x 30 seconds for 5 minutes) yielding 20.4 ng/µl *S. aureus* gDNA compared to 21.3-26.3 ng/µl *B. cepacia* gDNA (Figure 4.2). This setting also permitted higher quality *S. aureus* gDNA extracts (Figure 4.2). While gDNA extraction efficiencies at 5000 rpm (2 x 90 seconds) were less even, higher gDNA yields were obtained compared to 4000 rpm. Little difference

was observed in gDNA quality at this setting. A bead-beating setting of 4000 rpm for 5 minutes was ultimately determined optimal (meeting objective 1) for both *S. aureus* and *B. cepacia* for producing highly even Gram-positive and Gram-negative lysis efficiencies, high gDNA yields, and moderately good gDNA quality which surpassed the performance of EMAG gDNA extractions (objective 1).

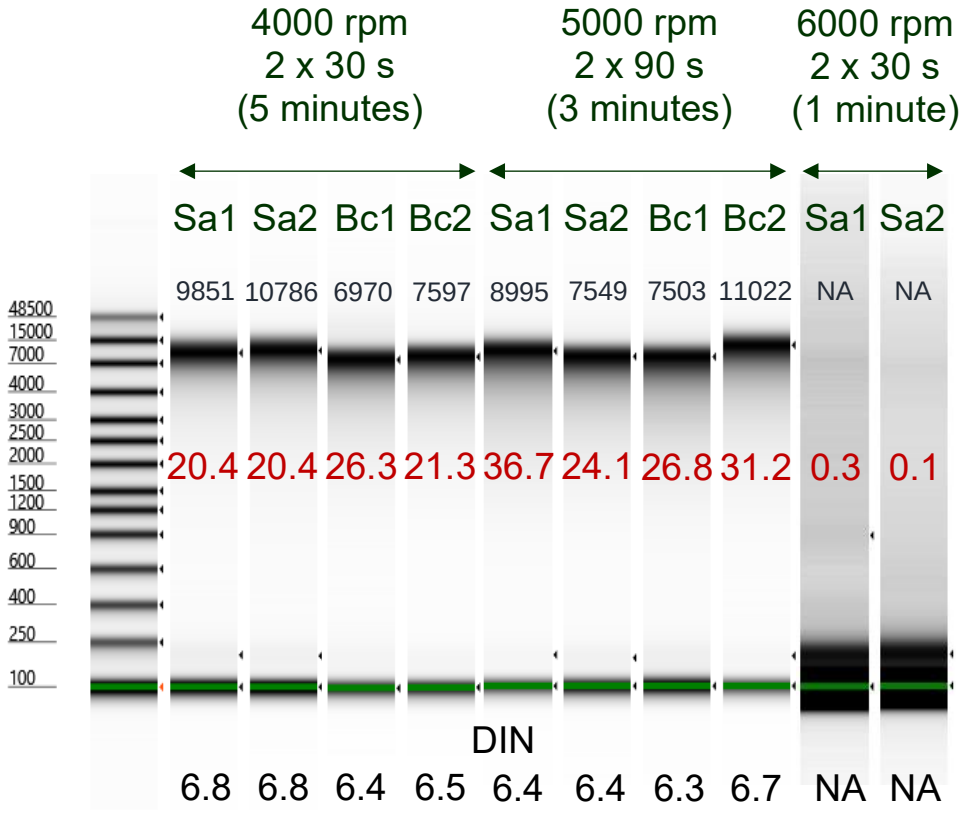


Figure 4.2 ZymoBIOMICS™ DNA miniprep kit bead-beating optimisation. This composite image has ordered *S. aureus* (Sa) and *B. cepacia* (Bc) replicates according to lysis setting. Samples have been scaled individually to the most concentrated electropherogram peak and contrast has been enhanced to highlight gDNA degradation. The first lane shows the genomic DNA ScreenTape ladder, and the Tapestation lower marker is highlighted in green in each lane. Sample DNA molecular weight (in bp) has been included beneath each sample label (navy). gDNA concentration (in ng/µl) for each sample has been annotated in each lane (red). DNA integrity (DIN) has been included beneath each lane. Note the lack of extracted gDNA using the highest bead-beating setting due to significant degradation.

4.3.2 Pure culture extract PCR limit of detection

The limit of RISA PCR and 16S rRNA gene PCR detection for all gDNA extraction kits and optimised treatments was explored by PCR, and by visualising resulting amplicons on agarose gels and ScreenTapes to answer objective 2. As previously shown (**see 4.3.1.2**), highly unequal Gram-positive and Gram-negative lysis occurred due to ceramic bead lysis, resulting

in a tenfold lower *B. cepacia* 16S rRNA gene PCR LoD (~2x10^6 CFU/ml) than *S. aureus* LoD (~2x10^6 CFU/ml; Figure 4.3). Uneven lysis efficiencies were also suggested by a markedly higher amount of *B. cepacia* gDNA being recovered compared to *S. aureus* (Table 4.2). The TapeStation indicated a lower absolute detection limit for both species compared to the agarose gel (Table 4.2; SI Figure C4). *B. cepacia* RISA PCR LoD also initially appeared tenfold lower than *S. aureus* using agarose electrophoresis (Table 4.2; Figure 4.3). The TapeStation similarly demonstrated all *S. aureus* core ITS amplicons were visible to ~2x10^5 CFU/ml while *B. cepacia* RISA profiles were visible to ~2x10^3 CFU/ml (SI Figure C6).

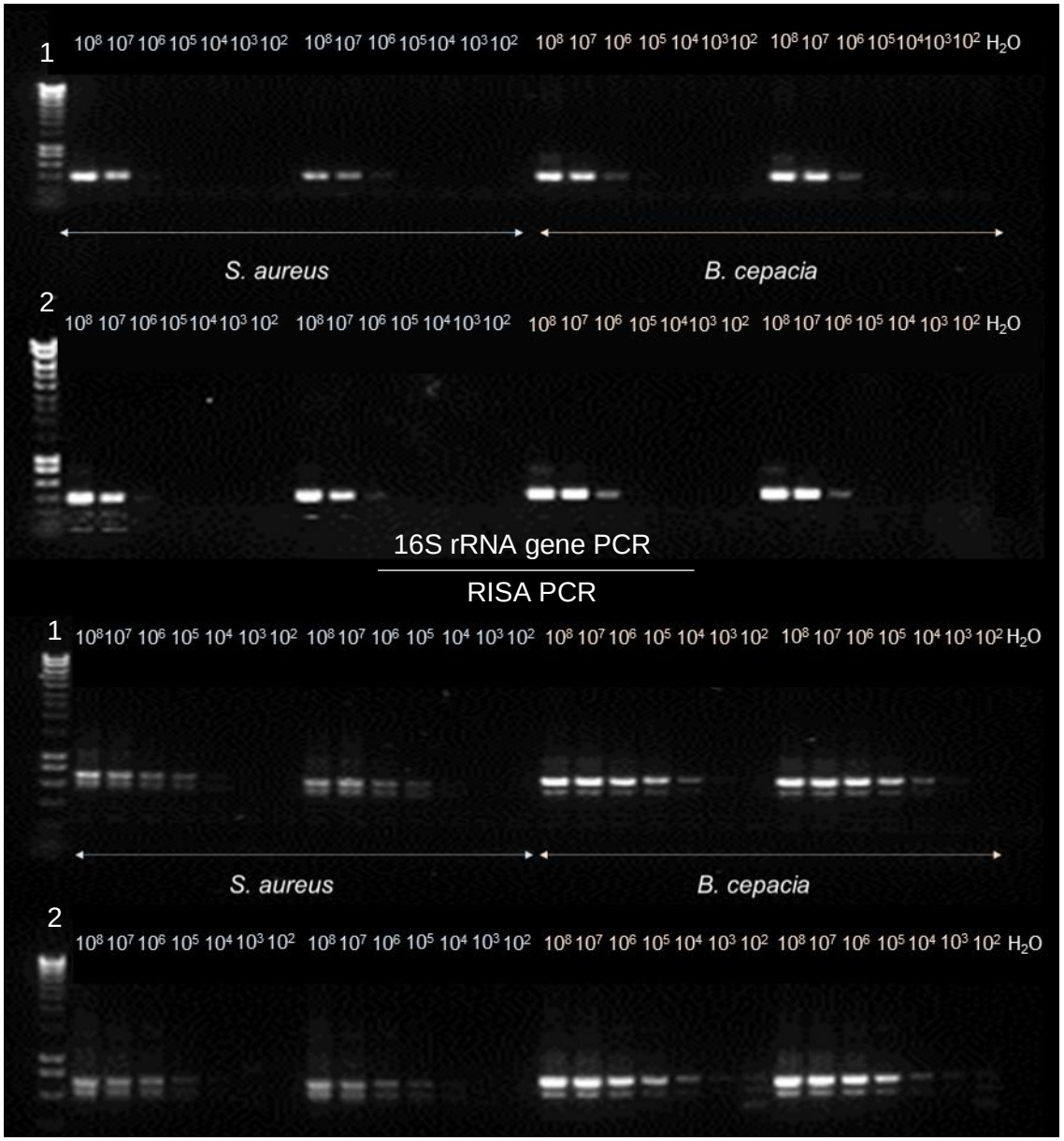


Figure 4.3 EMAG ceramic bead-beating *S. aureus* and *B. cepacia* 16S rRNA gene PCR and RISA PCR limit of detection. This composite image shows both biological replicates of *S. aureus* and *B. cepacia* samples in order from ~2x10^8-10^2 CFU/ml of extracted bacteria gDNA. A PCR nuclease-free water control has been included last and a 1 kb ladder has been included first.

Table 4.2 Summary of PCR detection limits[a,b,c]

	EMAG ceramic				EMAG MetaPolyzyme				ZymoBIOMICS™ kit			
	16S rRNA gene PCR[c]		RISA PCR[d]		16S rRNA gene PCR[c]		RISA PCR[d]		16S rRNA gene PCR[c]		RISA PCR[d]	
CFU count	Sa	Bc	Sa	Bc	Sa	Bc	Sa	Bc	Sa	Bc	Sa	Bc
10^8 CFU/ml	0.4	8.5	0.4	8.5	2.1	4	2.1	4	3	6.6	3	6.6
10^7 CFU/ml	0	0.6		0.6	0.2	0.2	0.2	0.2	0.2	0.6	0.2	0.6
10^6 CFU/ml										0.1		0.1
10^5 CFU/ml												
10^4 CFU/ml												
10^3 CFU/ml												
10^2 CFU/ml												

[a] Qubit gDNA concentration (ng/µl) are annotated in each cell where gDNA was quantifiable.
[b]Blank cells represent amplified DNA that could not be quantified by the Qubit.
[c]Dark shaded cells (green representing 16S rRNA gene PCR and red representing RISA PCR) shows both PCR detection limits when visualised on a 1% agarose gel, and lighter shaded cells represent absolute detection of amplified PCR products when visualised on a HS D1000 ScreenTape.

In comparison to EMAG-ceramic bead-beating, MetaPolyzyme achieved more even gDNA extraction efficiency between *S. aureus* and *B. cepacia* where both 16S rRNA gene PCR LoD and the absolute detection of RISA PCR products appeared equal (Figure 4.4). Enzymatic lysis increased *S. aureus* gDNA yield at the cost of lowering *B. cepacia* gDNA yield when bead-beating was applied (Figure 4.4; Table 4.2). Enzymatic treatment gave rise to more equal PCR LoD compared to ceramic bead-beating; both *B. cepacia* and *S. aureus* 16S rRNA gene PCR LoD was measured at ~$2x10^6$ CFU/ml from an agarose gel (Figure 4.4). The TapeStation suggested absolute detection for both species at ~$2x10^2$ CFU/ml (SI Figure C7). Both visualisation methods demonstrated the absolute detection of *B. cepacia* to ~$2x10^2$ CFU/ml while all core *S. aureus* ITS amplicons were present to ~$2x10^5$ CFU/ml (Table 4.2; Figure 4.4; SI Figure C8).

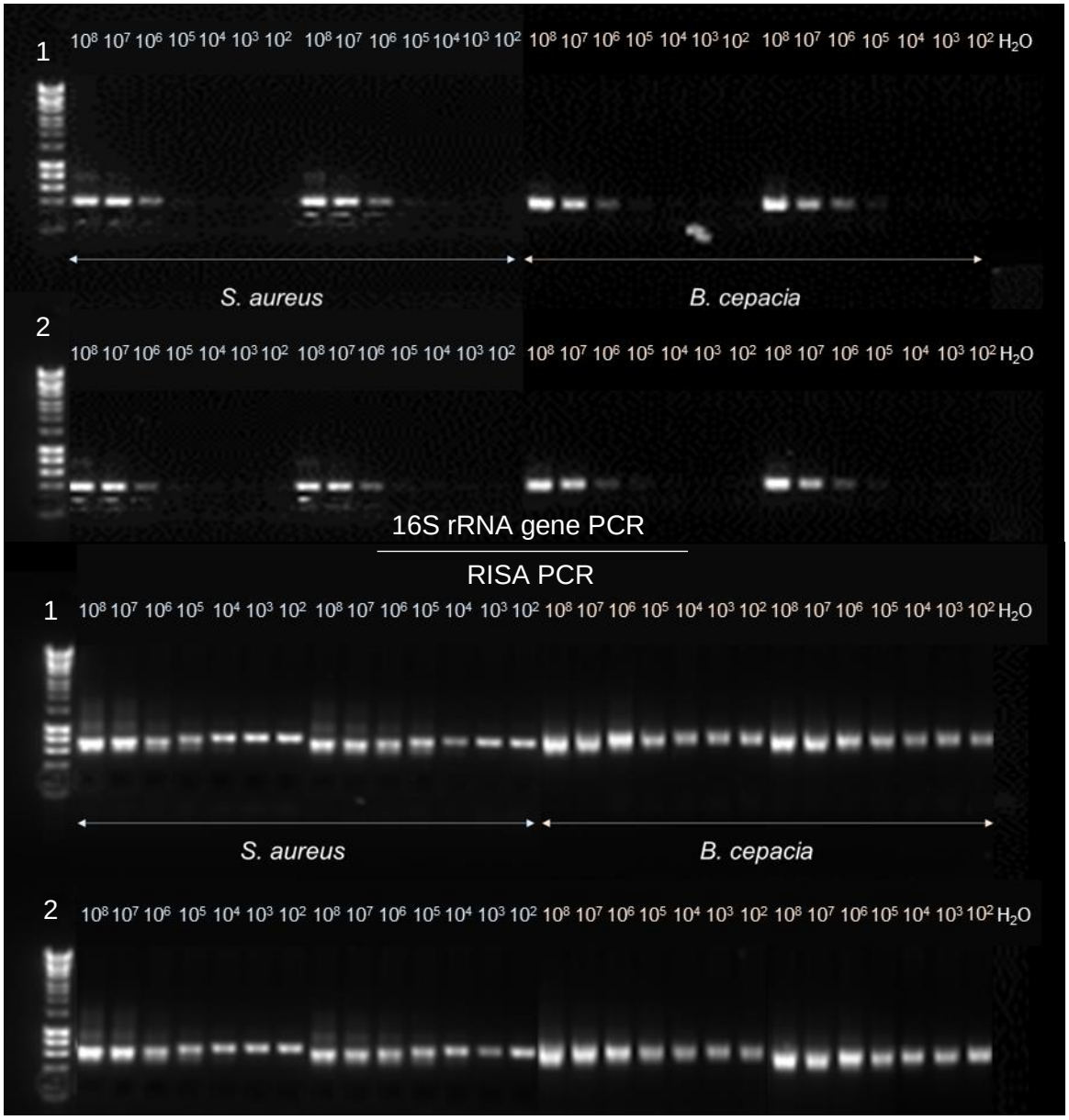


Figure 4.4 EMAG-MetaPolyzyme *S. aureus* and *B. cepacia* 16S rRNA gene PCR and RISA PCR limit of detection. This image has been edited to show a 1 kb reference ladder (left), all replicates of *S. aureus* and *B. cepacia* samples in order from ~$2x10^8$-10^2 CFU/ml of extracted bacteria gDNA and a PCR nuclease-free water control last.

ZymoBIOMICS™ extraction efficiencies mirrored EMAG-MetaPolyzyme detection limits where even gDNA extraction efficiencies were observed between *S. aureus* and *B. cepacia* which was comparable to EMAG enzymatic lysis (Figure 4.5; SI Figure C10). This methodology allowed a higher isolation of *S. aureus* gDNA, comparable to the recovery of *B. cepacia* gDNA where this methodology resulted in a slightly higher bias towards extracting Gram-negative gDNA in comparison to EMAG-MetaPolyzyme extracts (Table 4.2). Like EMAG-MetaPolyzyme gDNA extracts, both *S. aureus* and *B. cepacia* 16S rRNA gene PCR

LoD were visible to ~2x10^6 CFU/ml on a 1% agarose gel and were similarly detected to ~2x10^2 CFU/ml on a ScreenTape (Figure 4.5; SI Figure C10). All core *S. aureus* ITS amplicons were detected to ~2x10^3 CFU/ml (SI Figure C11). This was decided following the acknowledgement of a *S. aureus* replicate failing detection while being dilute. Both core *B. cepacia* ITS amplicons were detected to ~2x10^4 CFU/ml due to the stronger background signature of bacteria present in water dominating low dilution samples (SI Figure C12). Agarose gel similarly showed these species were detected to ~2x10^3 CFU/ml through RISA PCR (Figure 4.5).

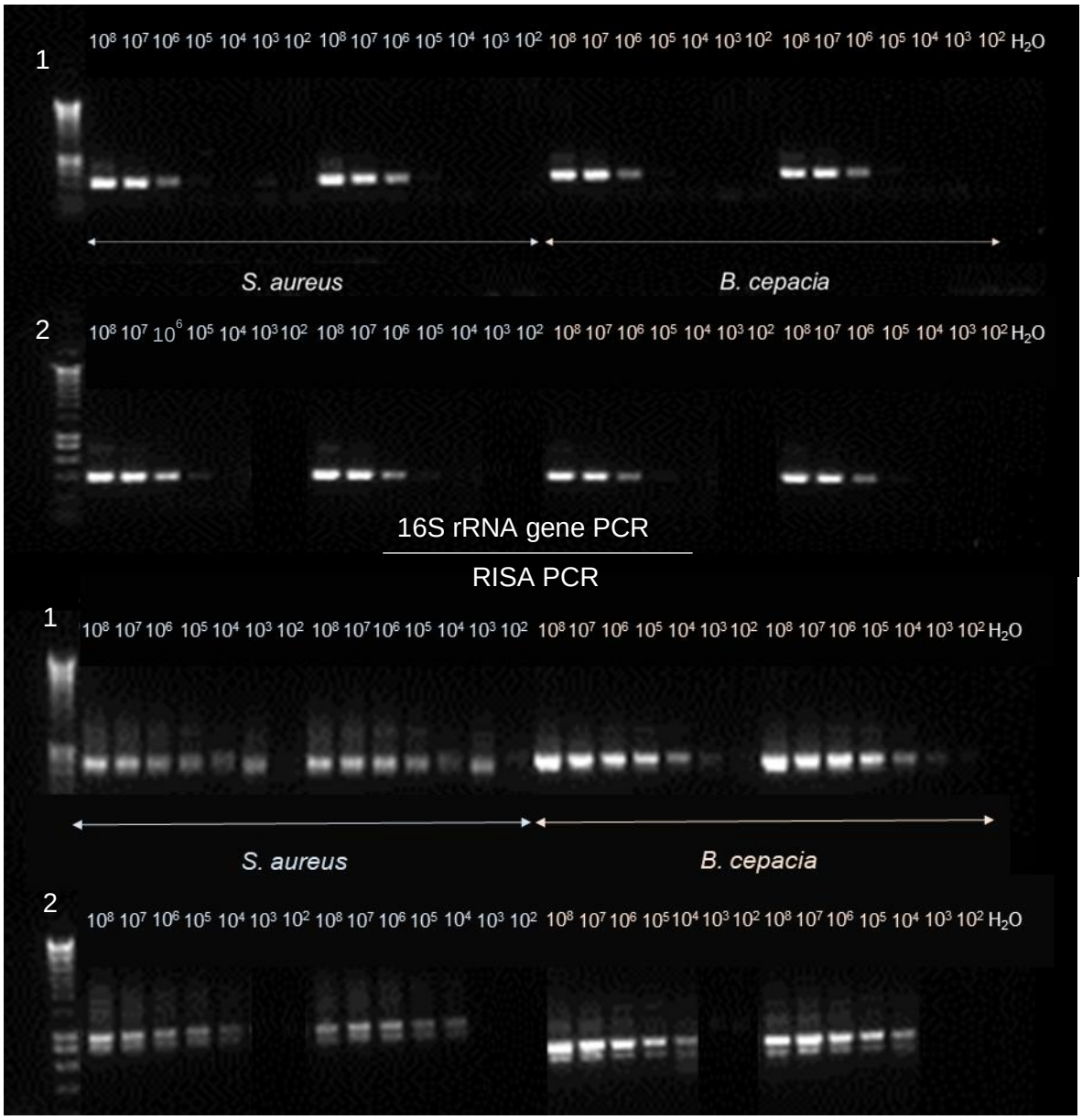


Figure 4.5 ZymoBIOMICS™ *S. aureus* and *B. cepacia* 16S rRNA gene PCR and RISA PCR limit of detection. V4 This gel image was edited to view each biological replicates of *S. aureus* and *B. cepacia.* 16S rRNA gene PCR LoD is shown in order from highest to lowest CFU count. A 1 kb ladder has been included next to ~2x10^8-10^2 CFU/ml of extracted bacteria gDNA. Nuclease-free water control has been included last.

4.3.3 EMAG and ZymoBIOMICS™ preliminary RISA analysis of spiked sputum

4.3.3.1 EMAG-MetaPolyzyme RISA PCR detection of *S. aureus* and *B. cepacia*

Before assessing spike-in *S. aureus* and *B. cepacia* limit of PCR detection in sputum, a high cell density (~$2x10^6$ CFU/ml) of spike-in organism was extracted using both gDNA extraction kits and optimal treatments, and was RISA PCR amplified to determine which extraction method enabled the lowest PCR limit of detection to answer objective 3. Following MetaPolyzyme processing guidelines, (not pelleting samples), both *S. aureus* and *B. cepacia* spike-in RISA profiles closely resembled the sputum blank (except for 2 *B. cepacia* replicates) (Figure 4.6). Partial *S. aureus* ITS profiles were inferred to have been present at ~877 bp (mean concentration 31.4 ng/µl), ~1131 bp (mean concentration 11.9 ng/µl), and occasionally at ~719 bp at a low concentration (Table 3.3; Figure 4.6). Importantly, those ~877 bp, ~1131 bp amplicons were concentrated and were not in the sputum blank. These samples were considered less likely to have been dominated by *H. influenzae* than *S. aureus* which could have accounted for the 2 largest amplicons as variation in DNA fragment sizes from existing *in silico* and experimental data were larger (Table 3.3). These samples also showed partial matches to the oral microorganisms *Capnocytophaga sputigena* or *V. parvula* (Figure 4.6; SI Table C1).

Two *B. cepacia* spike-in samples resembled its pure culture control where both of its core ITS products were visible at 662 bp and 830 bp. This 830 bp amplicon was particularly concentrated (~18 ng/µl), indicative of *B. cepacia* (Table 3.3; Figure 4.6). In the other 2 replicates, it was uncertain whether *B. cepacia* was isolated or PCR amplified due to the presence of only one of its corresponding core ITS profiles at ~882 bp (Table 3.3; Figure 4.6). This ~882 bp ITS amplicon could equally correspond to various Gram-negative bacteria such as *Pandoraea* spp., *Achromobacter* spp., *Acinetobacter* spp., and *Neisseria* spp. (Table 3.3; SI Table C1). Without sequence data, these species could not be identified.

Figure 4.6 Unpelleted EMAG-MetaPolyzyme extract RISA of spiked *S. aureus* and *B. cepacia*. RISA profiles of reference and spiked *S. aureus* (Sa), *B. cepacia* (Bc), and sputum blank controls (Sp1&2) have been included in this composite image. The first lane shows the ScreenTape ladder, and the Tapestation upper marker is highlighted in purple in each lane. A PCR nuclease-free water control is shown last.

Pelleted MetaPolyzyme samples led to more consistent *S. aureus* and *B. cepacia* RISA profiles however this did not necessarily translate to the correct isolation and amplification of these spike-in organisms (Figure 4.7). Pelleted *S. aureus* spike-in samples were observably similar to their unpelleted counterparts, however, not all 3 *S. aureus* core ITS amplicons (641 bp-1076 bp) were detected in each sample (Table 3.3; Figure 4.7). These concentrated ~649 bp (~81.1 ng/µl) and ~1047 bp (~34.2 ng/µl) amplicons may have prevented clear amplicon separation and measurements in the TapeStation. These core amplicons also may have constituted partial ITS amplicons of *H. influenzae* or *V. parvula* (Table 3.3).

Similarly to unpelleted *B. cepacia* spike-in samples, pelleted *B. cepacia* spike-in replicates did not resemble either control where its expected core ITS amplicons were measured at 659 bp and 829 bp (Figure 4.7). Upon closer inspection, highly concentrated ITS amplicons were detected at ~762 bp (mean 19.7 ng/µl) in spike-in samples, best-resembling *B. cepacia* RISA *in silico* estimates (Table 3.3). Weak amplification of a ~657 bp fragment in these samples was also indicative of *B. cepacia* and was absent from the sputum blank (Table 3.3; Figure 4.7). A *recA* gene PCR performed on these spike-in samples suggested the presence of *Burkholderia* where a ~870 bp fragment amplified. A ~1400 bp *recA* PCR amplicon was also present in these spike-in samples and sputum blanks but was ultimately disregarded for most likely being as a result of 2 *recA* fragments from different *Burkholderia* species or less specifically, from closely-related Gram-negative species. Bacteria species was not formally identified without accompanying sequence data.

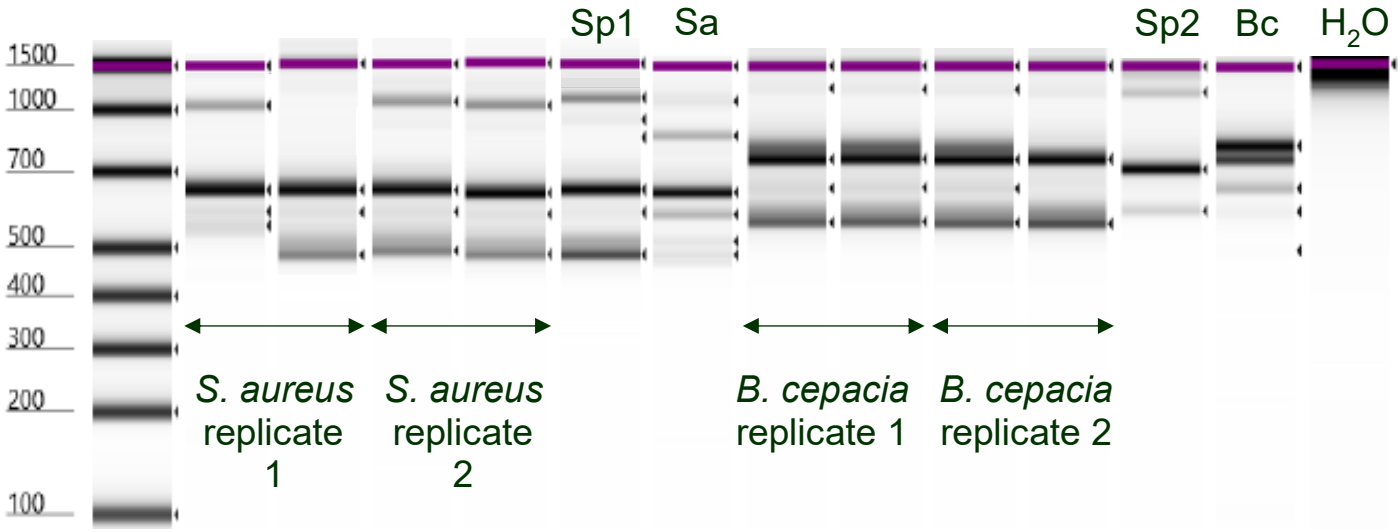


Figure 4.7 Pelleted EMAG-MetaPolyzyme extract RISA of spiked *S. aureus* and *B. cepacia*. Reference RISA profiles of *S. aureus* (Sa) and *B. cepacia* (Bc) and sputum blanks (Sp1&2) are shown next to corresponding spike-in samples in this composite image. A PCR nuclease-free water control has been included last. The first lane shows the ScreenTape ladder, and the Tapestation upper marker is highlighted in purple in each lane. Sample lanes 1, 2, and 5 were manually realigned and amplicons were remeasured using ImageJ software.

4.3.3.2 ZymoBIOMICS™ RISA PCR detection of *S. aureus* and *B. cepacia*

In the ZymoBIOMICS™ RISA spiked samples, numerous ITS profiles were amplified in all replicates, signifying either a more diverse sputum background, greater lysis and extraction efficiencies compared to EMAG methodologies, or the presence of few species that generated complex RISA profiles. Every spike-in sample and sputum control produced remarkably similar RISA profiles where the only notable difference was a higher concentration of a ~614 bp amplicon (~32.1 ng/µl compared to ~22.2 ng/µl background DNA; Figure 4.8). This ~614 bp amplicon however was too small to be considered as *S. aureus* (Table 3.3). *S. aureus* core amplicons were postulated to have been present at 689 bp and ~920 bp while its higher predicted ~1128 bp amplicon was totally absent. Like other extraction methods, it was uncertain whether this result occurred due to poor separation of amplicons or which species generated these amplicons (Figure 4.8). It is also possible that these ~614 bp and ~920 bp amplicons corresponded to common respiratory flora including *Rothia aeria* (~601 bp), *R. mucilaginosa* (~588 bp) and/or *Prevotella buccae* (~548 bp, ~913 bp), especially since no differences in amplicon concentration was observed between spike-in samples or the sputum blank (Figure 4.8; SI Table C1).

B. cepacia spike-in samples produced ~685 bp and ~917 bp amplicons (Figure 4.8). Amplicon molecular weight best resembled *B. cepacia in silico* predictions and it is possible that variation in TapeStation measurements could have resulted in this increased size and the lack of instrument resolution separating 800-900 bp fragments. On the other hand, no differences in amplicon size or concentration were observed in the sputum blank, suggesting these amplicons did not belong to spike-in *B. cepacia*. Instead, the dominant members of this mixed community may be represented by other *Burkholderia* species including *B. ambifaria* (~663 bp, 864-885 bp), or *B. cenocepacia* (~672-677 bp, 811-824 bp, 876 bp) for example (SI Table C1). Alternatively, emerging, and common respiratory species including *Acinetobacter* spp., *H. influenzae* or *Achromobacter* spp. could have accounted for these ~685 bp and ~917 bp fragments (Table 3.3; SI Table C1). Like EMAG spike-in samples, the lack of accompanying 16S rRNA gene sequence data could not confirm the identity of these dominant species.

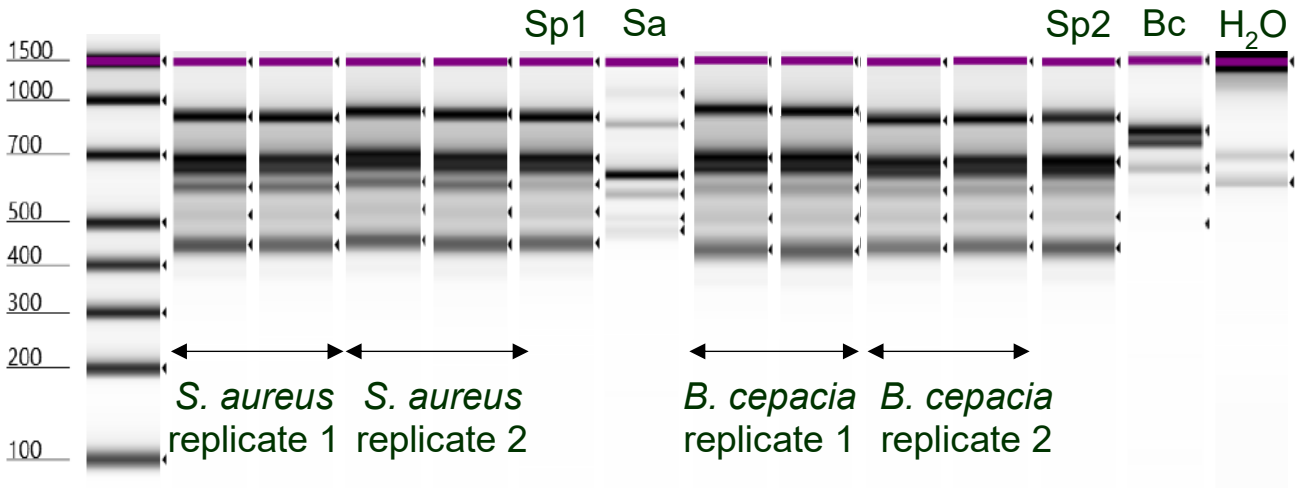


Figure 4.8 ZymoBIOMICS™ extract RISA of spiked *S. aureus* and *B. cepacia*. Reference RISA profiles of *S. aureus* (Sa) and *B. cepacia* (Bc) and sputum only controls (Sp1&2) have been edited next to corresponding spike-in samples A PCR nuclease-free water control has been included last. The first lane shows the ScreenTape ladder, and the Tapestation upper marker is highlighted in purple in each lane. Sample lanes 1 and 2 were realigned and measured separately in ImageJ.

4.3.4 ZymoBIOMICS™ spiked *S. aureus* and *B. cepacia* 16S rRNA gene PCR and RISA PCR limit of detection

4.3.4.1 *S. aureus* 16S rRNA gene PCR and RISA PCR detection limits

Due to slightly greater gDNA yields extracted through the ZymoBIOMICS™ kit and to low and more even PCR LoD between *S. aureus* and *B. cepacia* compared to other kits and treatments, the ZymoBIOMICS™ kit was selected to investigate spike-in RISA PCR and 16S rRNA gene PCR detection limits in sputum, meeting objective 4. Spike-in bacteria were first

identified among sequence and PCR data. Among the 10 identified *Staphylococcus* OTUs, *S. aureus* spike-in culture was identified as OTU6 (99.8% match to its pure culture control). This OTU was not abundant in the sputum blank, amounting to 0.02% of total reads (4/22496 reads). The composition of both spike-in samples and sputum blank were comparable, and dominant genera composed of common aerobic and anaerobic respiratory species including ~30.4% OTU2 *Veillonella*, ~19.2% OTU4 *Prevotella* 7, and ~13.6% OTU5 *Prevotella* 6). *S. aureus* OTU6 was not however abundant in any sample including non-spiked samples. Furthermore, OTU6 was detected in the sputum blank (4 reads), ZymoBIOMICS™ kit blank (3 reads), Mucolyse™ blank (7 reads), and PCR water blank (1 read; SI Table C2). Combined, OTU6 spike-in samples were postulated to have been detected to ~$2x10^6$ CFU/ml where there were between 177 and 198 sequenced reads (SI Table C2).

RISA portrayed diverse spike-in communities, indicating several abundant and dominant species (Figure 4.9). *S. aureus* RISA PCR profiles were predicted to be ~640-650 bp, ~740-750 bp, ~800-830 bp, ~900 bp, and ~1130 bp based on previous *in silico* and experimental data (Table 3.3). The QIAxcel measured its 3 corresponding amplicons at ~643 bp (~2.1 ng/µl), ~916 bp (~2.7 ng/µl), and ~1110 bp (~1.1 ng/µl), where 2 of which were among the most concentrated amplicons (Figure 4.9). These profiles were also comparable to pure *S. aureus* RISA profiles (664 bp at 1.4 ng/µl, 925 bp at 1.4 ng/µl, and 1038 bp at 3.5 ng/µl amplicons) and the sputum blanks (~649 bp at 2.5 ng/µl, ~948 bp at 1.9 ng/µl, and ~1138 bp at 0.8 ng/µl amplicons; Figure 4.9). Despite these similarities, there were no substantial differences in the concentration of these 3 core amplicons in any spike-in sample or sputum control, meaning RISA PCR LoD could not be determined. It is possible that spiked *S. aureus* RISA profiles did not appear in these samples as it was recovered in <1% relative abundance in 16S rRNA gene-based sequence data meaning RISA PCR could not visualise this low abundance organism.

Despite this this result, the ZymoBIOMICS™ kit was capable of extracting Gram-positive and Gram-negative organisms in sputum however RISA was mostly limited to genus-level resolution. Greater cohesion between 16S rRNA gene-based sequence data and RISA was observed compared to the sequencing of mock communities (**see chapter 3**). Here, the dominant *Veillonella* corresponded well to the most concentrated ~916 bp amplicon, and the ~643 bp or ~1110 bp amplicon, assuming *V. parvula* or *V. nakazawae* (652 bp, 723 bp, 925 bp, 950 bp predicted) was present (Table 3.3; SI Table C1). *P. buccae* could have also generated a ~570 bp fragment (548 bp predicted) and the 916 bp amplicon (913 bp predicted). RISA PCR suggested the dominance of *S. pneumoniae* (~534 bp, 2.4 ng/µl; Table 3.3). The remaining ~605 bp fragment was not considered to be part of *S. aureus*.

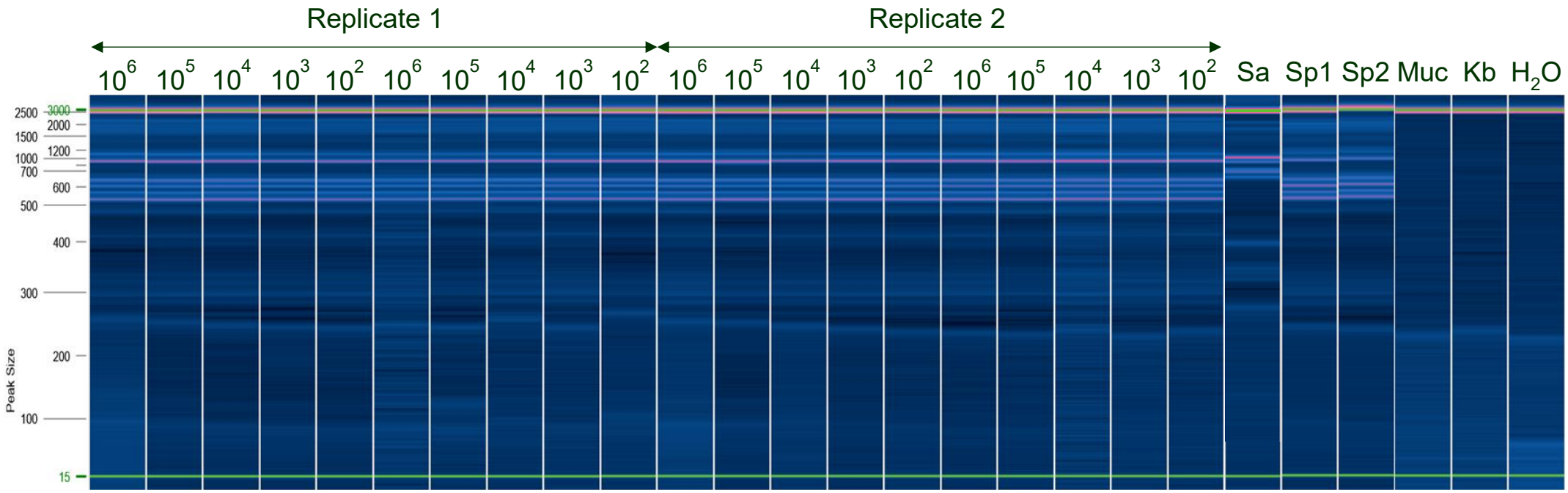


Figure 4.9 ZymoBIOMICS™ spiked *S. aureus* RISA PCR limit of detection. 2 biological replicates and 2 technical replicates of *S. aureus* samples ranging from 10^2 to 10^6 *S. aureus* CFU/ml of sputum are shown on this composite image. The leftmost lane shows the QIAxcel ladder, and the QIAxcel upper and lower markers are highlighted in green. Concentrated amplicons appear pink and less concentrated amplicons are a light blue. A positive control of pure *S. aureus* (Sa) has been included to compare its RISA profile to spike-in samples. Next, a sputum blank (Sp) control has been included for each biological replicate. Reagent controls for mucolyse (Muc), kit blank (Kb) PCR nuclease-free water has been included. *S. aureus* pure culture control was manually realigned and remeasured in ImageJ against the QIAxcel ladder.

4.3.4.2 *B. cepacia* 16S rRNA gene PCR and RISA PCR detection limits

Of the 2 identified *Burkholderia* OTUs, *B. cepacia* spike-in was identified as OTU8 (99.6% match to its pure culture control). OTU8 was also rare in the sputum blank, constituting 0.006% of total reads (1/16280 reads). This set of spike-in samples were abundant with common aerobic and anaerobic respiratory bacteria, which were dominated by 51.8% OTU1 *Haemophilus,* and abundant with 32% OTU3 *Streptococcus* and 3% OTU4 *Prevotella* 7, comparable to its sputum blank samples. Like *S. aureus* spike-in samples, *B. cepacia* OTU8 was not abundant in any sample due to high ratio of total bacteria load in sputum compared to *B. cepacia* CFUs. At ~$2x10^6$ CFU/ml, a significantly lower amount of *Burkholderia* sequenced reads were detected in comparison to spiked *S. aureus* at this dilution (SI Table C2). It was postulated that *B. cepacia* LoD in sputum was no lower than ~$2x10^6$ CFU/ml (~9 reads/0.05% relative abundance) due to its poor detection in 1 replicate (SI Table C2). In the other replicate, the proportion of OTU8 sequenced reads was equal to the number of OTU8 reads found in controls (1 read in ZymoBIOMICS™ kit blank and 1 read from the sputum blank).

B. cepacia spike-in RISA profiles partially matched its pure culture profile. This included a concentrated ~7.4 ng/µl ~830 bp amplicon similar to its 843 bp reference (Figure 4.10). Its second reference amplicon, measured at ~674 bp was completely absent from any spike-in sample and was instead replaced by a ~529 bp (~6.8 ng/µl) amplicon, indicative of *S. pneumoniae* RISA and 16S rRNA gene sequence data (Table 3.3; Figure 4.10; SI Table C1). A weakly amplified ~1172 bp fragment (1.5 ng/µl) was also present in each spike-in sample and the sputum blank (1174 bp, ~1.1 ng/µl), suggesting that the highly concentrated ~830 bp amplicon and the ~1172 bp amplicon most likely correlated to the dominant *H. influenzae* according to 16S rRNA gene-based sequence data (Table 3.3; SI Table C1). RISA PCR detection limit could also not be determined for this species due to incomplete ITS profile matching, no significant differences in OTU8 reads in any dilution, and the fact that OTU8 was rare in 16S rRNA gene sequence data suggesting RISA PCR would fail to amplify this low abundance species.

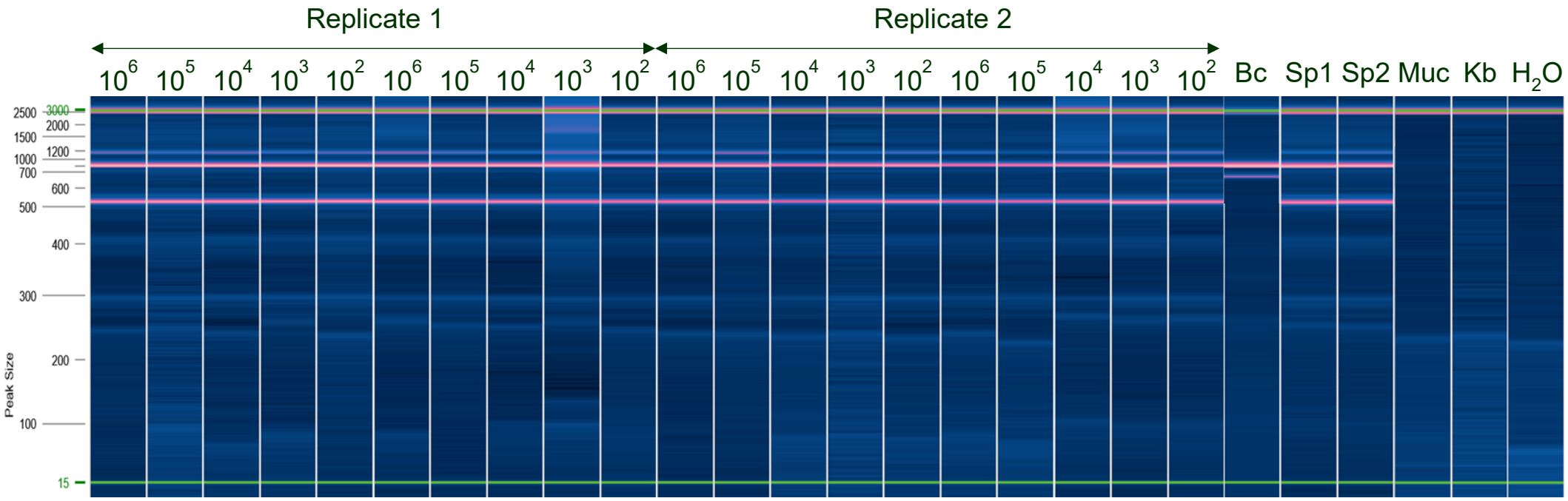


Figure 4.10 ZymoBIOMICS™ spiked *B. cepacia* RISA PCR limit of detection. 2 biological replicates and 2 technical replicates of *B. cepacia* samples ranging from 10^2 to 10^6 CFU/ml of sputum are shown in this composite image. The leftmost lane shows the QIAxcel ladder, and the QIAxcel upper and lower markers are highlighted in green. Concentrated amplicons appear pink and less concentrated DNA fragments are a light blue. A positive control of pure *B. cepacia* (Bc) culture has been included to compare its RISA profile to spike-in samples. A sputum background (Sp) control has been included for each biological replicate. Reagent controls for mucolyse (Muc), kit blank (Kb) PCR nuclease-free water has been included. *B. cepacia* pure culture control was manually realigned and remeasured in ImageJ against the QIAxcel ladder.

4.4 Discussion

4.4.1 Overview

Microbial cell lysis is a key factor in the DNA extraction process which introduces bias to PCR and metabarcoding data as a result of preferential or incomplete lysis. It is therefore vital to understand how lysis methods coupled with DNA extraction impacts gDNA extraction efficiency, yield, quality, and diversity analysis (Pollock et al. 2018). As an essential step in NGS workflows, DNA extraction methodologies must be translatable, reproducible, and comparable across laboratories and research, requiring full validation and standardisation for microbial community analysis (Tourlousse et al. 2021). As a prerequisite to achieve these objectives between CU and PHW laboratories, both EMAG and ZymoBIOMICS™ extraction kits were considered for validation as both kits are designed to efficiently extract high quality gDNA from low volume, low biomass samples like sputum. Additionally, both kits offered distinct advantages for routine gDNA extraction of clinical specimens (Table 1.3).

4.4.2 EMAG bead-beating optimisation

The EMAG features a wide range of functionalities for automated and high-throughput extraction of total nucleic acids from bacteria, viruses, and fungi (Garcia et al. 2017; Hindiyeh et al. 2019). For this reason, the EMAG is the PHW extraction platform of choice for routine PCR and sequencing workflows across disciplines. Additionally, the EMAG offers low turnaround time and greater flexibility in modifying cell lysis parameters compared to a spin column-based kit, relevant to sample type and target organism/s in accordance with PHW EMAG SOPs. Although these SOPs are routinely used to process sputum, the EMAG application is limited to few respiratory pathogens, mainly viruses and fungi. As a result, EMAG cell lysis methodology required optimisation to achieve high bacteria gDNA isolation and to reduce uneven lysis between a Gram-positive (*S. aureus*) and Gram-negative (*B. cepacia*) organism. Cell lysis optimisation commenced using bead-beating with the aim to achieve comparable lysis methodologies between EMAG and ZymoBIOMICS™ extraction kits. Bead-beating was also chosen for its convenience, high throughput, and has frequently been reported to increase gDNA yields, bacteria diversity and even lysis between Gram-positive bacteria, Gram-negative bacteria, and bacteria endospores (Islam et al. 2017; Pérez-Brocal et al. 2020; Zhang et al. 2020a; Tourlousse et al. 2021).

It was necessary to consider lysis bead material, density, shape, and hardness needed to achieve appropriate lysis of *S. aureus* and *B. cepacia* (Table 1.4). For bacteria cell lysis, the least aggressive materials are recommended for being small, spherical, low density, and less physically robust resulting in low shearing force with medium impaction (Table 1.4). In general,

the smaller the lysing matrix, the smaller the resulting lysed components (MP Biomedicals 2021). For this reason, 0.1-0.5 mm silica/zirconium beads are usually recommended for effective bacteria cell lysis (Vandeventer et al. 2011; Islam et al. 2017; Starke et al. 2019). Nevertheless, this "one size fits all" approach may be inadequate to disrupt polymicrobial communities of the lung due to the complexity of the sputum matrix and the presence of hard-to-lyse microorganisms (Pollock et al. 2018). To circumvent this obstacle, lysis beads designed for microbiome analyses incorporate a combination of these small 0.1-0.5 mm silica/zirconia beads as is the case for ZR BashingBead™ lysis tubes. Alternatively, ~0.1 mm beads are mixed with larger, ≥1 mm ceramic and/or glass beads. On this basis, 2 mm glass beads, 1.3 mm ceramic beads, and MP Biomedicals mixed beads (0.1 mm silica, 4 mm glass and 1.4 mm ceramic beads) were selected for the assessment of *S. aureus* lysis efficiencies before optimising bead-beating speed and duration.

Unsurprisingly, preliminary results suggested *S. aureus* could not be lysed efficiently using larger ceramic, glass, or mixed beads with EMAG reagents. Indeed, smaller beads with a higher surface area to volume ratio are preferable to lyse ~1 µm *S. aureus* cells (Monteiro et al. 2015; Passos et al. 2015; Proctor et al. 2019). Following this rationale, MP Biomedicals mixed beads should have most efficiently lysed *S. aureus*: however, this was not the case. The lack of highly concentrated PCR amplicons highlighted the possibility that this bead material was too gentle for EMAG extraction chemistry, or the ratio of mixed beads was suboptimal to lyse *S. aureus.* At most, only ~0.2 ng/µl of the partial 16S rRNA gene product was amplified using ceramic beads while amplified DNA was negligible following glass bead-beating. This poor lysis efficiency was further evidenced by the requirement of at least 10^5 CFUs/ml of extracted material for successful PCR amplification. By default, ceramic beads were chosen to optimise and compare EMAG bead-beating settings along with ZR BashingBead™ lysis tubes.

S. aureus and *B. cepacia* lysis efficiencies of both bead types were then compared to evaluate EMAG 16S rRNA gene PCR and RISA PCR limit of detection using the methodology that resulted in highest gDNA yields, most even *S. aureus* and *B. cepacia* gDNA extraction and least gDNA shearing. Optimal bead-beating parameters for both bead types required relatively gentle agitation (4000 rpm) where ceramic beads required a longer duration. Though described as optimal, these appointed bead-beating parameters for both ZR BashingBead™ lysis tubes and ceramic beads led to distinct problems during the DNA extraction process.

4.4.3 EMAG bead-beating pitfalls

Although mechanical lysis can disrupt all cell types and increase gDNA recovery from low biomass samples, suboptimal lysis conditions could either result in inefficient lysis biasing

Gram-negative overrepresentation or result in severely degraded gDNA biasing Gram-positive overrepresentation. DNA shearing is also highly problematic for reducing sequence quality and increasing PCR chimeric products leading to altered community composition (von Wintzingerode et al. 1997; Teng et al. 2018). On one side of this spectrum, ceramic bead lysis resulted in uneven and poor extraction of *S. aureus* gDNA regardless of bead-beating speed and duration. At best, ~4 ng/µl of *S. aureus* extracted gDNA was recovered through aggressive bead-beating whereas a tenfold amount of *B. cepacia* extracted gDNA was consistently isolated using this method. On the other end of this spectrum, ZR BashingBead™ matrix lysis efficiency was observably more even between *S. aureus* and *B. cepacia* indicating improved bacteria lysis efficiency with these beads. Still, the disparity in Gram-positive and Gram-negative gDNA yield was significant, suggesting incomplete *S. aureus* lysis and higher *S. aureus* gDNA degradation in addition to the severe degradation of *B. cepacia* gDNA as demonstrated through lower gDNA yields, significantly lower DIN scores and lower gDNA molecular weight per sample compared to ceramic bead-beating. DNA shearing was apparent in all extracts, becoming severe through more aggressive bead-beating. In contrast to ZR BashingBead™ lysis tube bead-beating, slight to moderate gDNA degradation in samples lysed using ceramic beads was indicated through DIN scores where paradoxically, gDNA shearing appeared worse in *S. aureus* extracts. After factoring in sample DNA average molecular weight and RNA co-extraction, ceramic bead-beating was not considered as a major source of gDNA degradation. As a product of total nucleic acid extraction, RNA carryover was not regarded as sample contamination, but it did emphasise poor mechanical lysis efficiency instead of indicating a suboptimal GTC concentration.

GTC is one of the strongest protein denaturants and is the most efficient chaotropic salt to bind nucleic acids to silica surfaces (Poeckh et al. 2008; Farrell Jr 2017; Shen 2019). When used at high concentrations, both inter-and intra-molecular covalent interactions in a sample are disrupted, leading to protein unfolding. This reaction leads to the solubilisation of proteins that are not otherwise soluble in water i.e., nucleases and peptidoglycan without denaturing DNA or RNA (Rabilloud et al. 2007; Hosomi et al. 2017). At 5M, GTC salts also permit effective nucleic acid binding to paramagnetic silica beads where RNA adsorbs more readily to silica compared to DNA under the same condition. It is perhaps due to different nucleic acid affinity to bind to silica that led to a higher amount of isolated RNA in *S. aureus* isolates where its lysis efficiency was low (Poeckh et al. 2008). It was not possible to reduce GTC molarity to increase gDNA binding affinity to silica for the following: a 4M lysis buffer is typically required to inactivate nucleases in a sample matrix to subsequently produce high-quality DNA/RNA extracts (Chomczynski and Sacchi 2006; Tan and Yiap 2009). Most importantly, reducing GTC

molarity in CL3 clinical specimens may not deactivate biological material and would therefore pose as a serious hazard to health during sample processing.

4.4.4 EMAG enzymatic lysis evaluation

Ceramic bead-beating was ultimately chosen to compare PCR LoD to EMAG gDNA extractions using enzymatic lysis. Bead-beating was demonstrably poor at extracting *S. aureus* gDNA and was therefore compared to EMAG-MetaPolyzyme PCR detection limits with the aim of improving uneven lysis efficiencies. A popular alternative to mechanical lysis, enzymes are employed to disrupt bacteria cells as chaotropes alone cannot solubilise the lipid bilayer of membranes (Rabilloud et al. 2007; Hosomi et al. 2017). In some instances, enzymatic lysis has reportedly achieved higher gDNA yields, HMW gDNA and lower LoD compared to bead-beating (Gill et al. 2016; Oriano et al. 2019; Li et al. 2020). A distinct advantage to enzymatic lysis is its specificity where mechanical lysis fails to disrupt hard-to-lyse microorganisms. For example, lysozymes are widely used to break the glycosidic bond within the peptidoglycan layer in Gram-positive bacteria cells (Islam et al. 2017). MetaPolyzyme was chosen as an enzymatic agent for its mixture of specific Gram-positive and yeast lytic enzymes covering a wide spectrum of respiratory microbiota. MetaPolyzyme also posed as an attractive option to continue expanding PHW diagnostic services towards metagenomic studies directly from sputum if proven to be an efficient lytic agent.

In contrast to EMAG mechanical lysis, the inclusion of MetaPolyzyme dramatically improved EMAG 16S rRNA gene PCR and RISA PCR LoD. Due to enzyme high lysis specificity but lower lysis efficiency, the recovery of *S. aureus* gDNA was noticeably higher than through bead-beating comparable to the ZymoBIOMICS™ kit. This greater Gram-positive lysis efficiency and lower Gram-negative lysis efficiency in the absence of bead-beating resulted in the most even gDNA extraction efficiencies between the 2 species from any gDNA extraction methodology. Although promising in pure culture, the addition of MetaPolyzyme to unpelleted sputum failed to generate reproducible RISA PCR profiles of spike-in samples, possibly suggesting poor lysis and PCR amplification of spike-in *B. cepacia*. Furthermore, while *S. aureus* spike-in samples indicated the presence of *S. aureus* RISA profiles, these profiles could have also represented other common Gram-negative respiratory bacteria. Inconsistent results may have also been attributed to DNA extraction/PCR inhibitors present in sputum and Mucolyse. Human clinical samples often contain proteases that can degrade polymerases and complex polysaccharides that hinder polymerases by mimicking the structure of nucleic acids (Schrader et al. 2012; Hall et al. 2013; Trombley Hall et al. 2013). Additionally, haemoglobin and mucolytic agents in sputum have been reported to inhibit PCR through lowering polymerase activity (Döşkaya et al. 2011; Sidstedt et al. 2018). These modes of inhibition are

largely due to poor gDNA purification during extraction giving rise to poor PCR amplification. Instead, the inclusion of 0.1% (w/v) DTT may have caused structural unfolding and denaturation of lysozyme in MetaPolyzyme required for Gram-positive cell disruption as PCR amplification still occurred (Wang et al. 2010).

This notion was further supported by generation of consistent and reproducible RISA profiles in pelleted spike-in samples. While spiked *B. cepacia* was detected by a RISA and *recA* PCR and spiked *S. aureus* RISA profiles appeared to match its reference, *S. aureus* spike-in samples could have equally been dominated by common respiratory species including *V. parvula* and *H. influenzae*. Although this method followed correct MetaPolyzyme processing guidelines, this method was not pursued further as sputum pelleting creates a heterogenous environment not representative of the original sample, was more costly to process samples (over £200 per 25 samples), and was highly time-consuming.

4.4.5 ZymoBIOMICS™ DNA Miniprep kit evaluation

ZymoBIOMICS™ DNA Miniprep kit was ultimately selected to extract gDNA from sputum for producing comparable yet slightly superior PCR LoD to EMAG-MetaPolyzyme extracts according to TapeStation data. Additionally, bead-beating offered higher and even *S. aureus* and *B. cepacia* gDNA yields from both *S. aureus* and *B. cepacia* without the need to pellet these samples like the EMAG, in accordance with hypothesis 1 (Bead-beating is required to efficiently lyse sputum and physically robust microorganisms). It is also worth mentioning ZymoBIOMICS™ extracts were visualised later than EMAG-MetaPolyzyme extracts, consequently leading to degradation of very dilute amplified DNA. Although ZymoBIOMICS™ extracted the highest *S. aureus* and *B. cepacia* gDNA yields, uneven lysis efficiencies were more pronounced in this kit compared to EMAG-MetaPolyzyme extracts due to higher lysis efficiencies of bead-beating at the cost of lower Gram-positive specificity and higher DNA shearing. Nevertheless, gDNA integrity and extraction yield was sufficient to generate short read Illumina libraries using optimal bead-beating settings in agreement with hypothesis 2 (A column-based gDNA extraction kit designed for microbiome analysis would better-represent Gram-positive and Gram-negative microbiota compared to the EMAG platform at the cost of lower quality extracts).

Preliminary RISA spike-in extracts indicated the presence of both Gram-positive and Gram-negative ITS profiles in each sample, appearing identical among each replicate except for a higher concentrated ~614 bp amplicon which was not thought to correspond to either spike-in organism. While *B. cepacia* spike-in RISA profiles could resemble *B. cepacia in silico* predictions, all spike-in samples were most probably dominated by both common and

emerging Gram-positive and Gram-negative species. Preliminary RISA analysis suggested greater bacteria diversity and greater Gram-positive and Gram-negative representation using this kit, even if spike-in organisms were not detected. A clear limitation to this experiment and other preliminary spike-in RISA analyses was presented with the lack of accompanying 16S rRNA gene sequence data to confirm the identity of these dominant microorganisms.

Nevertheless, on the basis of most diverse and most reproducible RISA profiles (through sample handling and gDNA extraction) the ZymoBIOMICS™ DNA Miniprep kit was selected for assessment of 16S rRNA gene PCR and RISA PCR LoD in sputum. While spike-in *S. aureus* and *B. cepacia* OTUs were identified among the sequenced reads, they amounted to <1% relative abundance in each spike-in sample. Instead, community composition of each spike-in sample was remarkably similar to the composition of sputum blanks which were dominated by common Gram-positive and Gram-negative genera including *Haemophilus. Streptococcus* and *Veillonella*. Notwithstanding these results, sample contamination was not thought to factor into the low recovery of *S. aureus* and *B. cepacia* sequenced reads due to the relative uniformity of spike-in and sputum blanks and an insignificant number of reads were introduced through negative controls. It was apparent that total bacteria load within sputum significantly overshadowed the proportion of spike-in organism present, leading to its underrepresentation. Though difficult to determine, spiked *S. aureus* 16S rRNA gene PCR LoD was estimated at ~$2x10^6$ CFU/ml and was expectedly higher than pure culture PCR LoD. Still, LoD estimates are susceptible to large random variation therefore determining a true result from these samples where spike-in samples and control samples differed by a few reads may be indeterminate using this methodology (Bernal 2014). Despite this flaw in methodology design, 16S rRNA gene sequence data did confirm that ZymoBIOMICS™ kit was capable of lysing Gram-positive and Gram-negative bacteria from sputum.

In contrast to spiked 16S rRNA PCR LoD, RISA PCR LoD could not be determined as spike-in bacteria were too low abundance to have been visualised through RISA PCR amplification. As the nature of this molecular test suggests, spike-in bacteria must have been in high abundance in the entirety of these sputum samples for successful amplification and visualisation. Alternatively, closely-sized ITS fragments may have not been correctly separated by the QIAxcel resulting in one highly concentrated amplicon. In this case, the former scenario was considered most probable. One notable outcome from the RISA analysis and 16S rRNA gene sequencing of ZymoBIOMICS™ extracts was the greater cohesion of 16S rRNA gene-based sequence data and RISA data in contrast to the sequencing and RISA analysis of mock community DNA (**see chapter 3**), indicating this kit may somewhat also alleviate PCR amplification bias between the 2 PCR methods.

4.4.6 Conclusions

This analysis validated that bead-beading of respiratory samples is vital to extract gDNA from a range of microorganisms and obtain even gDNA yields sufficient for sensitive downstream PCR. This result contrasted enzymatic lysis that appeared interfered by sputum and failed to generate reproducible RISA profiles in unpelleted samples. Because pelleting sputum would have altered microbial abundance, hypothesis 1 was accepted. The ZymoBIOMICS™ DNA Miniprep kit was the most appropriate kit utilising bead-beating, and more evenly represented Gram-positive and Gram-negative bacteria in pure culture in context of gDNA yield and PCR detection limit. ZymoBIOMICS™ bead-beating did however result in poorer quality gDNA extracts proving hypothesis 2. This kit also suggested a higher microbial diversity captured through RISA and demonstrated the potential for larger-scale standardisation of a cost gDNA extraction method that is capable of efficiently extracting respiratory microbiota. On the other hand, the ZymoBIOMICS™ kit poses challenges: contaminating sequences are more likely to be captured in low biomass samples when processes lack automation and would therefore also require additional laboratory personnel to achieve low turnaround times in a clinical setting. Finally, as gDNA extraction bias cannot be fully eliminated, it is essential to document how sample storage, sample preparation and gDNA extraction bias community composition of sputum microbiota using the ZymoBIOMICS™ kit and how this data might impact clinical decision making later on.

Chapter 5 Application of routine 16S rRNA gene sequencing and ribosomal intergenic spacer analysis for culture-free analysis of bronchiectasis, cystic fibrosis, and chronic obstructive pulmonary disease sputa

5.1 Introduction

Over the past decade, advancements in respiratory microbiota research have revealed its role in CF disease progression. The primary focus in this time has been detecting and combating traditional and emerging respiratory pathogens in hopes of improving disease trajectory and patient prognosis (Watson et al. 2019). Since then, several genotypic methods have been widely used to characterise traditionally recognised pathogens and non-conventional taxa. 16S rRNA gene-based metaprofiling is by far the most commonly-used respiratory microbiome tool which is typically used alongside culture and/or additional molecular methods including community fingerprinting, MLST, and species-specific PCR probes for species-level resolution (Bittar and Rolain 2010; Ditz et al. 2020).

Recent trends in CF lung infection has most notably shown that the annual percent change in *P. aeruginosa* prevalence is decreasing, contrasting to the rising prevalence of *S. aureus* and non-tuberculous *Mycobacteria* that are associated with significant morbidity (Rocchetti et al. 2018; Hatziagorou et al. 2020; Van den Bossche et al. 2021). The most prevalent *Burkholderia* species are largely recognised as *B. multivorans, B. cenocepacia,* and to a lesser extent, other *Burkholderia cepacia* complex (BCC) species and *B. gladioli* (Zlosnik et al. 2020). The emerging *S. maltophilia, A. xylosoxidans, Cupriavidus* spp., *Ralstonia* spp., and *Pandoraea* spp. are also increasingly reported to colonise or infect CF lung tissue. These Gram-negative non-fermenters are often misidentified by routine culture thus necessitating improved and continuous infection surveillance and control (Capizzani et al. 2018; Rocchetti et al. 2018; Díez-Aguilar et al. 2019).

While the COPD and non-CF bronchiectasis (BR) microbiome have been less extensively studied than CF, potentially pathogenic microorganisms (PPMs) have been recognised as causative agents of respiratory infection in both of these patient groups. These PPMs are most frequently identified as *P. aeruginosa, H. influenzae, H. parainfluenzae, H. haemolyticus, S. pneumoniae, M. catarrhalis, S. aureus,* and various members of *Enterobacteriaceae* (Matkovic and Miravitlles 2013; El-Korashy and El-Sherif 2014). *P. aeruginosa* is the most commonly isolated PPM from the COPD lung and is associated with acute exacerbations (Wang et al.

2017). Similarly, both *P. aeruginosa*- and *H. influenzae*-dominated BR microbiomes have been linked to severe disease and frequent pulmonary exacerbations (Richardson et al. 2019; Aogáin et al. 2021). While these PPMs are well documented, the emergence of *S. maltophilia, A. xylosoxidans* and *A. baumannii* in BR and COPD raises concern over potential underreporting of these nosocomial pathogens (Lin et al. 2019; Stepman et al. 2020; Dicker et al. 2021). Importantly, these organisms differ in terms of pathogenic potential and transmissibility therefore accurate species identification through culture-independent diagnostic tools are required for effective clinical management and disease control (Jorth et al. 2019).

Sequence-based analysis has also shed light on highly individual and diverse communities of non-conventional microorganisms that co-exist with these dominant major and emerging pathogens (Moffatt and Cookson 2017; Jorth et al. 2019). While strict anaerobic and other oral-associated bacteria may not be routinely cultured, the recent literature suggests these oral-associated and upper respiratory microbiota seed the lower airways (Mahboubi et al. 2016). When these bacteria are encountered by clinical bacteriology laboratories they are frequently recorded as normal respiratory flora (NRF) and are not classified further in comparison to the typical respiratory pathogens that may be present. In terms of microbial community analysis and ecology, NRF is not the best term for these respiratory microbiota species, but because of the linkage of this thesis to current bacteriology, it will be used to describe them collectively.

There is growing evidence in the potential role of NRF in modulating and dampening the pathogenicity of traditional respiratory pathogens, or independently contributing to disease (Duan et al. 2003; Flynn et al. 2016; Li et al. 2021). Community diversity may also be a key disease indicator where the relative abundance and diversity of NRF decrease as age and disease severity advances (Caverly et al. 2015; Einarsson et al. 2016; Jorth et al. 2019). Cross-sectional studies have also linked low community diversity and poor lung function across disease diagnoses (Erb-Downward et al. 2011; Rogers et al. 2013; Cuthbertson et al. 2020). Furthermore, pathogens dominating lung microbiota such as *P. aeruginosa,* BCC, *A. xylosoxidans, and H. influenzae* were found to be associated with severe lung disease (Flight et al. 2015; Richardson et al. 2019; Cuthbertson et al. 2020). Currently no causal relationship between dysbiosis of lung microbiota and disease progression has been established (Martin-Loeches et al. 2020). Understanding the changes of microbial diversity and composition that may occur at the onset of exacerbation has become an area of interest. Advancements in microbiome analysis may facilitate personalised treatment strategies regarding the type and

number of antibiotics to treat dominant microbe/s and exacerbation. In this scenario, the mode, specific type, and number of administered antibiotics may significantly shape community structure in the long term, thus driving or preventing dysbiosis (Caverly et al. 2015; Françoise and Héry-Arnaud 2020).

5.1.1 Aims and objectives

Culture-based methods remain in the forefront of Public Health Wales (PHW) respiratory bacteriology (**see section 3.1**), using selective growth media and subsequent microbial identification by MALDI-TOF. These conventional methods are not suitable to differentiate all the respiratory bacteria present due to inconsistent colony morphology, the exclusion of unculturable and strict anaerobic species, and bias towards fast-growing organisms (Capizzani et al. 2018). Consequently, a high level of non-fermenting Gram-negative and BCC misidentification is commonly reported, and non-conventional respiratory pathogens are often overlooked in this growth-based bacteriology (Wright et al. 2001; Bittar and Rolain 2010; Martinucci et al. 2016). Moving forward, the overarching goal of this chapter was to investigate a culture independent diagnostic strategy within a clinical diagnostic setting. This translated and evaluated the optimised gDNA extraction, PCR, and 16S rRNA gene-based NGS methodologies (**see chapters 3 and 4**) for routine laboratory examination of respiratory specimens.

The objectives of this chapter were:

1) Compare culture-independent bacterial community composition and diversity to bacteriological culture and MALDI-TOF identification results.
2) Define disease-specific microbiota profiles.
3) Correlate bacterial diversity to patient clinical parameters including disease diagnosis and lung function.
4) Identify pathogenic species that correlate with reduced lung function.

5.1.2 Hypotheses

Hypotheses for this chapter were:

1) Culture-independent methods identify clinically important microorganisms missed through culture.
2) Bacterial diversity is not an informative predictor of lung function decline due to high intra-individual and inter-individual variations in core and dominant respiratory microbiota.

5.2 Materials and methods

5.2.1 Sample collection, sputum processing, and culture-based pathogen identification

Sputum was collected from 13 BR, 12 CF, and 13 COPD patients which was processed within 48-h of sampling by PHW CL3 laboratory staff. A 1:1 ratio of 1X dithiothreitol (DTT) was added to an equal volume of sputum for liquefaction. For highly purulent samples, additional Mucolyse™ was slowly added until homogenous. Samples were homogenised for at least 30 seconds depending on sputum purulence. Samples were then incubated at room temperature for 15 minutes prior to plating out and bacterial gDNA extraction.

All culture-based methods and analyses were conducted by PHW staff where liquified sputum was firstly inoculated onto agar media (supplied through Oxoid Ltd.) This included plating onto CBA supplemented with 5% horse blood (general plating), chocolate agar supplemented with 175 g/l bacitracin (*H. influenzae*) and cystine lactose electrolyte deficient (CLED) agar (coliform bacteria) for all respiratory specimens. CF specimens also used cetrimide selective agar to isolate *P. aeruginosa*, mannitol salts agar to isolate *S. aureus*, and *Burkholderia cepacia* selective agar. All plates were incubated at 37°C aerobically except for CBA and chocolate agar which were incubated under 5% CO_2. Incubation times varied according to manufacturers' instructions. Afterwards, MALDI-TOF was conducted using the Bruker Biotyper MALDI-TOF mass spectrometer from the direct transfer of fresh bacterial colonies (~24-h) excluding NRF identified through culture (**see section 2.1.4**). Each isolate was tested in duplicate.

5.2.2 ZymoBIOMICS™ bacterial gDNA extraction

Sputum bacterial gDNA extraction was achieved using the optimised ZymoBIOMICS™ DNA Miniprep kit (Cambridge Bioscience) extraction protocol. 250 µl of DTT-treated sputum, Mucolyse™, and nuclease-free water (ZymoBIOMICS™ kit blank) controls were directly transferred to ZR BashingBead™ lysis tubes with 750 µl lysis buffer and were incubated at room temperature for 10 minutes to inactivate CL3 microorganisms. Lysis and gDNA extraction followed methodologies outlined in **section 2.3.4** where 400 µl of lysate was extracted resulting in a 60 µl purified eluate. gDNA was quantified using a Qubit™ 2.0 fluorometer broad-range dsDNA kit (Invitrogen) prior to Illumina library preparation and RISA PCR.

5.2.3 Illumina library preparation and 16S rRNA gene sequencing

Illumina library preparation of the V4 16S rRNA gene region was achieved using primers 16Sf (5'-GTGCCAGCMGCCGCGGTAA-3') and 16Sr (5'-GACTACHVGGGTWTCTAAT-3') in combination with the optimised 16S rRNA gene PCR protocol (**see section 2.4; section 2.6;** Table 2.19). Library preparation encompassed all 38 sputum samples, 2 Mucolyse™ controls, a ZymoBIOMICS™ extraction kit blank, a nuclease-free water (PCR) control, and a BEI resources microbial standard B (mock community) to estimate sequence error rates (**see section 2.2.1;** SI Table A1). gDNA was not diluted prior to PCR amplification. Resulting libraries were purified using Agencourt AMPure XP paramagnetic beads (**see section 2.6.2**) and library QC was assessed using a Qubit™ 2.0 fluorometer HS dsDNA kit and 2200 TapeStation HS D1000 ScreenTapes. Libraries were then diluted to 4 nM, pooled, and denatured with 0.2 N NaOH following methodologies outlined in **section 2.6.3**. Library loading was undertaken by CU GRH staff (**see section 2.6.3**).

5.2.4 16S rRNA gene-based sequence analysis

Sequenced reads were quality-checked by FastQC (Andrews 2010) and trimmed by Trim Galore! (Krueger 2021). Poor quality, erroneous, and chimeric sequences were removed through the Mothur pipeline following the methods described in **sections 2.6.4 and 2.6.5**. Sequence error was estimated in Mothur using a microbial standard B mock community (BEI resources; **see section 2.2.1;** SI Table A1). This estimation was achieved by comparing V4 16S rRNA gene sequences of an *in silico* mock community replicating microbial standard B ATCC strains to the sequenced mock community. ATCC strain nucleotide sequences were downloaded from NCBI.

SILVA v138 was used to classify OTU taxonomies to the genus level using the Wang approach (Wang et al. 2007) and percent relative abundance scores were calculated (**see section 2.6.4**). Median relative abundance scores of the 5 overall most proportionally abundant phyla were compared across disease diagnoses while community composition at the genus level was analysed individually using a 1% relative abundance cap to filter rare taxa. Rare taxa were later examined to check for gDNA extraction kit or reagent contamination and/or PCR amplification/sequencing issues.

For diversity analysis, reads were subsampled to 4955 reads equating to the sample containing the lowest number of reads. Rarefaction, Good's coverage, alpha diversity and beta diversity were measured following the methods outlined in **section 2.6.5**.

5.2.5 RISA PCR and amplicon visualisation

RISA PCR was conducted using optimal cycling conditions (**see section 2.4;** Table 2.18). All 38 sputum samples were amplified in addition to all extraction kit and reagent negative controls (**see section 5.1.3**). Amplicons were visualised by 1% agarose gel electrophoresis and with a Qiagen QIAxcel high-resolution DNA cartridge using a 15 bp-3 kb QX DNA alignment marker and a 100 bp-2.5 kb QX DNA size marker (**see section 2.5.3**). Bacteria species were inferred through matching ITS amplicons to the closest *in silico* RISA references based on abundant 16S rRNA gene-based OTU data and culture positivity. Predictions were accomplished through PRIMER-BLAST and an *in silico* PCR tool (Bikandi et al. 2004).

5.2.6 Statistical analysis and figure generation

Phylum composition by its % relative abundance across BR, CF, and COPD disease categories was statistically compared using a Kruskal-Wallis test. Pairwise differences in OTU relative abundance scores between each disease category were examined using the Mothur Metastats function (**see section 2.6.5**). This was used to define major distinguishing OTUs/species, if any, through comparing mean relative abundance scores. The mean was chosen in order to include the comparison of abundant OTUs which were present in <50% of samples, for which the median could not be calculated. Subsampled data was used for this statistical design where OTU relative abundance was capped at 0.01% to discard a majority of contaminating or spurious OTUs.

Patient parameters including age, sex, lung function, disease diagnoses, and sputum purulence were collected by PHW staff. Lung function, as measured by % predicted forced expiratory volume in 1 second (%FEV_1) was noted. Multiple imputation via predictive mean matching (PMM) was employed to estimate missing %FEV_1 data using the R MICE package (Van Buuren and Groothuis-Oudshoorn 2011). Severity of CF lung disease was recorded following the US CF Foundation Patient Registry guidelines (U.S. Cystic Fibrosis Foundation 2019). Severity of COPD and BR lung disease was recorded following the Global Initiative for Chronic Obstructive Pulmonary Disease schema (GOLD 2021), and the British Thoracic Society Guideline for Bronchiectasis in Adults (British Thoracic Society 2018) respectively.

Multiple linear regression was conducted to examine the overall effect of patient clinical data on both Shannon and inverse Simpson (InvSimpson) alpha diversity indices. Both indices were compared because InvSimpson index better reflects species dominance while Shannon diversity takes species richness into account. Differences in bacterial richness including number of observed OTUs (Sobs), ACE, and chao1 indices were also compared between all

disease categories with an ANOVA. ACE index was chosen for comparison as it typically performs best when observed richness is small whereas Choa1 is useful towards low-abundance classes and high observed richness. A 2 tailed t-test was used to compare both Shannon and InvSimpson diversity between communities that were dominated by lung microbiota not considered pathogenic and those dominated by traditional and emerging respiratory pathogens (Flight et al. 2015). Where available, lung function data from both upper and lower quartiles was extracted and statistically compared (ANOVA) against percent relative abundance of the dominant pathogen or lung microbiota from each sample (Flight et al. 2015). A Bonferroni correction was applied to allow for multiple pairwise comparisons. Beta diversity was compared against all patient clinical data with a PERMANOVA using the R vegan package (Anderson, 2001; Oksanen et al. 2019). Multivariate homogeneity of groups dispersions was then calculated for each PERMANOVA model using the betadisper function in vegan.

All figures were generated in R (R core team 2021) using the tidyverse (Wickham et al. 2019) package which included dplyr (Wickham et al. 2018) and ggplot2 (Wickham 2016). NMDS plots were generated from UniFrac distance matrices. A biplot was created from the base NMDS plots where significant ($P<0.05$) OTU correlation coefficients (calculated using Mothur corr.axes) were plotted along each axis to denote the spatial position of each OTU driving differences in community structure and membership (**see section 2.6.5**). Hierarchical clustering of RISA profiles was achieved by calculating unweighted pair group method with arithmetic mean (UPGMA) and the Sørensen–Dice coefficient using GelJ software (Heras 2015). Amplicon molecular weight dissimilarity scores annotating the resulting dendrogram was edited in Inkscape while the dendrogram dimensions were edited in FigTree.

5.3 Results

5.3.1 Patient cohort and clinical parameters

The patient cohort comprised 22 females and 16 males (n=38; mean age 56; range 20-88) who were divided according to disease diagnosis. From this total, 13 patients were diagnosed with BR (mean age 72; range 29-88), 12 with CF (mean age 31; range 20-42), and 13 with COPD (mean age 62; range 32-80). Disease severity ranged from mild (BR=3, CF=5, COPD=1), to moderate (BR=4, CF=2), severe (BR=2, CF=5, COPD=1) and very severe (COPD=3). At the time of sampling, 13 patients were prescribed concomitant antibiotics. These patients primarily had COPD and were most frequently administered doxycycline as a response to culture positivity. All available clinical parameters are in SI Table D1.

5.3.2 Culture-dependent identification of sputum microbiota

Eleven patients were noted with 1 culture-positive result (BR=4; COPD=7), another 11 with 2 culture-positive results (BR=5; CF=3; COPD=3), and 13 patients had at least 3 culture-positive results (BR=4; CF=8; COPD=1). No culture data was obtained from the remaining 3 patients. Bacteria were successfully isolated from 35 sputum samples where most frequently isolated species included NRF (n=25), *P. aeruginosa* (n=13), *H. influenzae* (n=7), *S. aureus* (n=4), and *S. maltophilia* (n=3). *Klebsiella pneumoniae* and *S. pneumoniae* were isolated twice while *Klebsiella oxytoca*, *Neisseria meningitidis*, *Burkholderia vietnamiensis*, *Escherichia coli*, *Streptococcus agalactiae*, *Corynebacterium striatum*, and *Serratia marcescens* were all isolated once. Yeast and mycelial fungus were additionally isolated from 13 and 5 sputum samples respectively.

The BR cohort contained the highest number of identified bacterial species (n=9) in addition to NRF, followed by the CF (n=6) and COPD cohorts (n=5). Within these groups, NRF was notably most frequently isolated among the BR cohort (n=10), *P. aeruginosa* was most frequently isolated among the CF cohort (n=9), and *H. influenzae* was most frequently isolated among the COPD cohort (n=4). All available culture data has been summarised in SI Table D1 along with accompanying MALDI-TOF identification results for 25 cultures which complemented these findings.

5.3.3 16S rRNA gene sequence QC, low-biomass contamination, and PCR amplification bias

Illumina sequencing produced high quality reads where the overall sequence error rate was 0.00012%. Following sequence filtering, 680876 reads were retained. From this total, approximately 8.5% of reads were flagged as chimeric and another 11.2% erroneous or poor-quality reads were removed. For diversity analysis, each sample was subsampled to 4955 reads, corresponding to the sample containing the least number of reads.

The impact of gDNA extraction kit and reagent contamination on this sequenced dataset was considered minimal. For instance, most prevalent and notable rare taxa most likely represented both pathogenic and non-pathogenic respiratory microbiota in low abundance. This included *Fusobacterium* (n=25), *Atopobium* (n=24), *Capnocytophaga* (n=21), *Pseudomonas* (n=19), *Actinomyces* (n=18), *Leptotrichia* (n=17), and *Staphylococcus* to name a few (n=17; SI Table D2). In these taxa, the number of corresponding reads in sputum samples were notably higher than those identified in PCR reagent controls, if at all (SI Table D3).

The impact of PCR amplification bias could also be investigated for rare taxa that were covered in chapter 3. Relevant taxa included *Actinomyces* (see above), *Stenotrophomonas* (n=3), *Pseudomonas* (see above), and *Burkholderia* (n=1) which previously amplified poorly. *Acinetobacter* (n=3), *Anaerococcus* (n=3), *Staphylococcus* (see above), *Prevotella* (n=7), *Veillonella* (n=7), and *Haemophilus* (n=14) were instead associated with increased PCR amplification (SI Table D2). Sequenced reads belonging to these genera were generally higher than those of both gDNA extraction kit and PCR reagents meaning these taxa were most likely a genuine component of respiratory microbiota (SI Table D2; SI Table D3). Furthermore, 8 samples were dominated by *Pseudomonas*, *Stenotrophomonas*, and *Burkholderia*, in contrast to previous findings where these species amplified poorly (see chapter 3; see section 5.3.4.2). Conversely, 4 culture-positive samples for *P. aeruginosa* and another culture-positive sample for *S. aureus* showed few sequences corresponding to these taxa.

5.3.4 Bronchiectasis, cystic fibrosis, and chronic obstructive pulmonary disease 16S rRNA gene sequence microbiota composition

5.3.4.1 Phylum-level community composition

Community composition was explored at the phylum level, genus level, and OTU level and was later compared to culture-based data (objective 1) and to define disease-specific microbiota (objective 2; **see section 5.3.4.3; section 5.3.2; and section 5.3.6**). Within the sample set, a total of 623 OTUs were classified at 97% sequence similarity representing 14 phyla (SI Table D4). The proportion of the 5 most dominant phyla across this dataset along with phylum median % relative abundance divided between each disease category is shown in Figure 5.1. *Firmicutes* and *Proteobacteria* constituted the majority of reads across this dataset, accounting for 55.1% and 15.8% of reads respectively. Additionally, *Bacteroidetes*, *Actinobacteria* and *Fusobacteria* appeared in >5% of total sequenced reads combined (Figure 5.1; SI Table D4). *Proteobacteria* was most abundant in the CF lung (41.7%), while *Firmicutes* were most abundant in the BR (48.6%) and COPD (76.5%) lung. Overall composition of *Bacteroidetes, Actinobacteria,* and *Fusobacteria* were also highest among the CF cohort (Figure 5.1). Considerable variation in these relative abundance scores resulted in no significant difference in phylum-level relative abundance scores between each disease cohort.

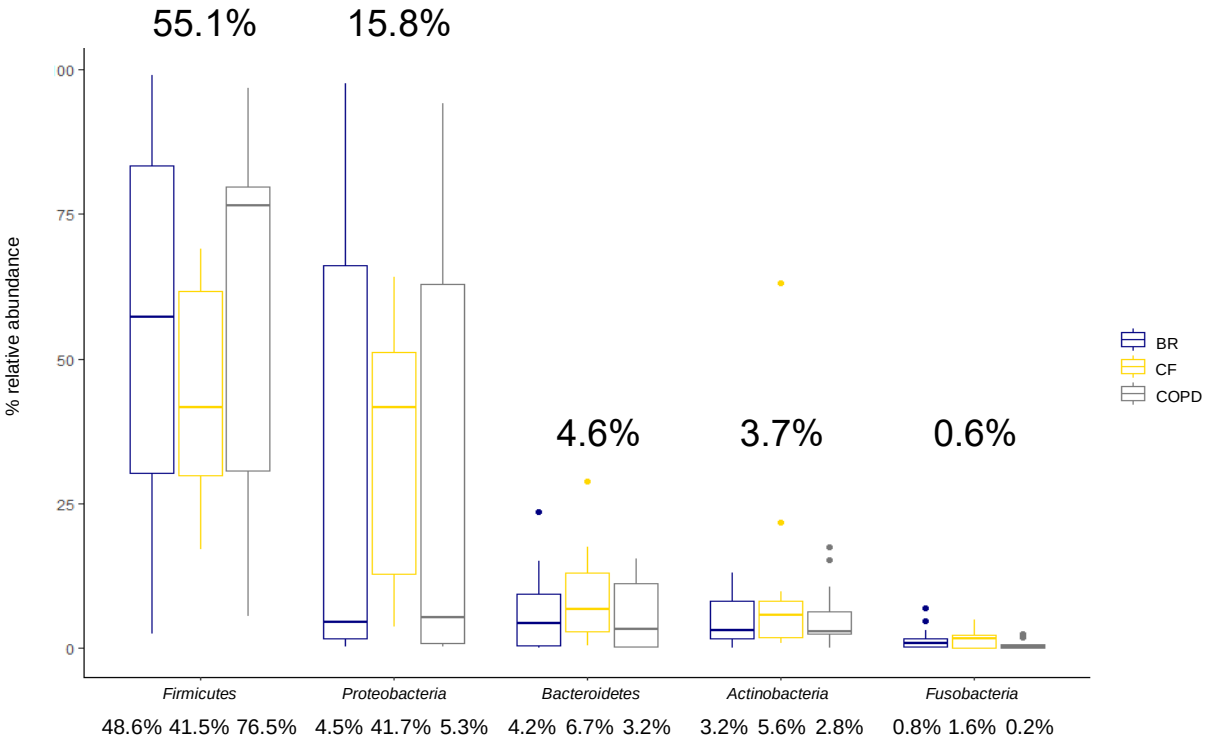


Figure 5.1 Phylum-level community composition. Data is ordered from highest to lowest proportionally abundant genera across this dataset. Median phylum percent relative abundance scores for each disease category are shown below the plot. Median percent relative abundance scores calculated across the entire dataset are shown above plots. Outliers have been illustrated by a dot.

5.3.4.2 Genus-level community composition

In sputum, 189 genera were identified along with 50 unclassified taxa. To examine disease-specific differences in bacterial community composition, genus % relative abundance was capped at 1% to filter rare OTUs. Following filtering, 19 OTUs representing 12 genera and 3 unclassified taxa remained. Predominant genera were defined as the most abundant genus in each sample. Co-dominant genera, if noted, were defined as highly abundant genera that constituted over half the % relative abundance score of the predominant genus (definition modified from (Lucas et al. 2018). Figure 5.2 illustrates abundant genera where *Streptococcus* was predominant in 17 samples (44.7%), accounting for the majority of BR (n=7; median relative abundance 33%) and COPD samples (n=8; 39.4%). *Streptococcus* was also co-dominant in 3 BR and 2 CF samples. In contrast, *Staphylococcus* (19.4%), closely followed by *Pseudomonas* (6.2%) were predominant in 3 CF samples each (25%). *Staphylococcus* was predominant only among the CF cohort and was co-dominant in 1 BR sample. *Pseudomonas* was also predominant in 1 BR sample, 1 COPD sample, and was co-dominant in a CF sample. *Haemophilus* was predominant in 2 BR and 2 COPD samples and was predominant in a single CF sample. *Stenotrophomonas* was predominant in 1 BR and 1 CF sample. *Gemella* was predominant in 1 BR sample, *Burkholderia* and *Rothia* were predominant in 1 CF sample each and *Granulicatella* was predominant in 1 COPD sample. Two unclassified predominant genera were also identified within the *Enterobacterales* and *Enterobacteriaceae*, respectively. *Neisseria* was co-dominant in 1 BR and 1 COPD sample, the latter corresponding to *N. meningitidis* (SI Table D1). *Prevotella* was also co-dominant in 1 BR and 1 CF sample.

Mean relative abundance scores were also calculated to estimate the average BR, CF, and COPD community. Mean calculations were necessary to represent abundant genera appearing in <50% of samples where medians could not be calculated. BR, CF, and COPD communities were all dominated by *Streptococcus* (BR n=13; mean relative abundance 35.9%; CF n=12; 22.1%; COPD n=13; 43.5%; Figure 5.2). These communities were also abundant with *Neisseria* (BR n=6; 1.8%; CF n=4; 2.6%; COPD n=2; 1.8%), *Rothia* (BR n=9; 3.1%; CF n=10; 9%; COPD n=8; 3.1%), *Veillonella* (BR n=11; 5.4%; CF n=10; 3.4%; COPD n=10; 5.4%), *Prevotella* (BR n=9; 4.9%; CF n=10; 7.2%; COPD n=8; 4.9%), *Granulicatella* (BR n=8; 4.9%; CF n=7; 2.4%; COPD n=7; 4.9%), *Actinomyces* (BR n=7; 1.6%; COPD n=5; 1.6%), and unclassified *Pasteurellaceae* (CF n=7; 1.9%; Figure 5.2). Other notable genera that were abundant yet infrequent in sputum samples included *Alloprevotella* (n=2), *Bergeyella* (n=1), *Capnocytophaga* (n=2), *Corynebacterium* (n=1), *Fusobacterium* (n=7), *Lachnoanaerobaculum* (n=1), *Lactobacillus* (n=5), *Lautropia* (n=1), *Leptotrichia* (n=6), *Porphyromonas* (n=2), *Sneathia* (n=1), *Treponema* (n=1), unclassified *Lactobacillaceae* (n=1),

and unclassified *Prevotellaceae* (n=3; Figure 5.2). Mean community composition of each disease category was highly similar to one another where notable differences included an increased abundance of *Pseudomonas* in the CF group (18.5%) and an increased abundance of *Haemophilus* within the COPD cohort (14.9%; Figure 5.2).

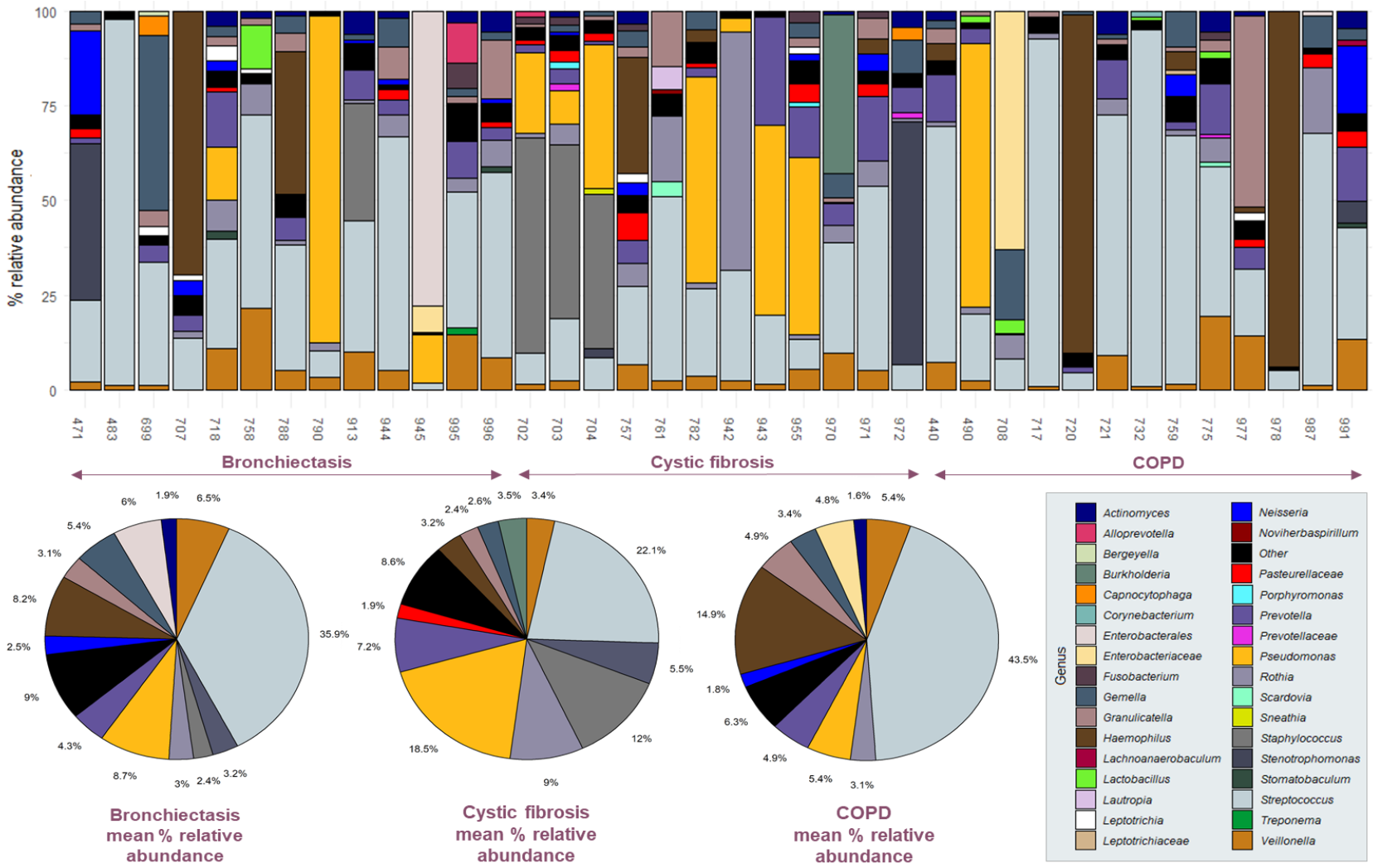


Figure 5.2 Bacterial genus level community composition. Stacked barplots illustrate bacterial genera that were capped at a 1% relative abundance. Each pie chart represents the mean % relative abundance scores of each genus divided by each disease category which were also capped at 1% relative abundance. Rare taxa were grouped into “other” category.

5.3.4.3 OTU-based analysis and OTU differential abundances

Of the 623 classified OTUs, OTU1 *Streptococcus* was most frequently detected across this dataset, accounting for 21.7% of total reads (Figure 5.3). This large portion of sequenced reads suggested OTU1 *Streptococcus* was a widespread and/or dominant component of lung microbiota. Moreover, the majority of the 20 most frequent and abundant OTUs were hypothesised to represent core respiratory microbiota characteristic of the lung microenvironment based on sequence identification and previous culture positivity and identification as NRF (SI Table D1). Accounting for a significant portion of reads, OTU2 *Haemophilus* (8.7% of reads), followed closely by OTU3 *Pseudomonas* (7.6% of reads) were hypothesised to correspond to the culture-positive *H. influenzae* and *P. aeruginosa* based on samples with a high abundance of/dominated by *Pseudomonas* and *Haemophilus* sequenced reads (Figure 5.2; Figure 5.3; SI Table D1).

Following this rationale, OTU7 *Staphylococcus* (4.6% of reads) presumably corresponded to *S. aureus* for being highly dominant in CF samples. Additionally, OTU13 *Stenotrophomonas* (2.3% of reads), and OTU19 *Burkholderia* (1% of reads) appeared less frequent in samples and were postulated to correspond *S. maltophilia* and *B. vietnamiensis* in culture respectively (Figure 5.2; Figure 5.3; SI Table D1). According to sequence data, OTU15 *Enterobacteriaceae* was largely dominant in 1 COPD sample and likely corresponded to *K. pneumoniae* in culture however OTU17 *Enterobacterales* which was dominant in 1 BR sample could not be assigned to either *K. oxytoca* or *S. marcescens* (Figure 5.2; Figure 5.3; SI Table D1). Figure 5.3 also highlights OTU14 *Pseudomonas* which was present in 2% of sequenced reads suggesting the presence of another *Pseudomonas* species or strain.

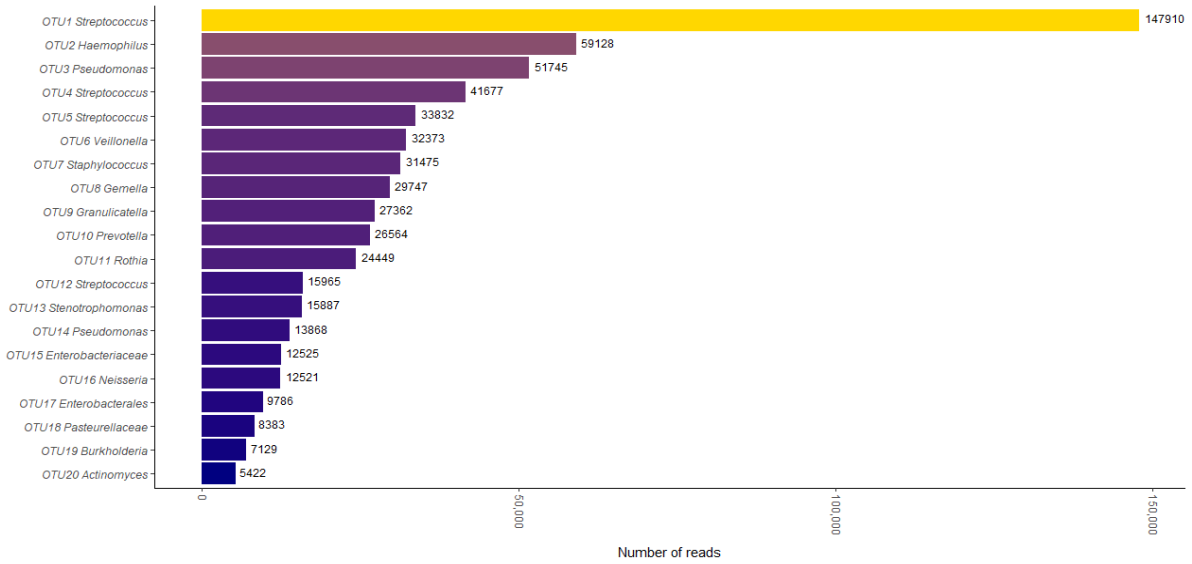


Figure 5.3 Twenty OTUs that were most proportionally abundant overall. The total number of reads each OTU represented has been annotated to the right. This plot was coloured as a gradient using the Viridis package to emphasise proportionally high number of reads.

When comparing both BR and CF cohorts, OTU4 *Streptococcus* was significantly more abundant among the BR cohort (Table 5.1). This OTU was indicative of core Streptococcal microbiota based on its high overall proportion of sequenced reads which also correlated to frequent NRF culture positivity (Figure 5.3; SI Table D1). Sequence data could however not conclusively determine whether OTU4 did or did not represent pathogenic *Streptococci* (Figure 5.3; SI Table D1). Between CF and COPD cohorts, OTU3 *Pseudomonas* (*P. aeruginosa* in culture) and OTU7 *Staphylococcus* (*S. aureus* in culture) were more abundant components of CF microbiota. OTU4 *Streptococcus* and OTU15 unclassified *Enterobacteriaceae* (*K. pneumoniae* in culture) were instead significantly more abundant among COPD patients (Table 5.1). OTU14 *Pseudomonas* was hypothesised as the dominant *Pseudomonas* OTU in the COPD cohort however species identification could not be pursued through culture due to the lack of reported growth (SI Table D1).

Table 5.1 Metastats differentially abundant OTUs between disease categories

OTU	Genus	CF OTU mean % relative abundance ± SE	BR/COPD OTU mean % relative abundance ± SE	P value
OTU4	***Streptococcus***	2.6±0.8	6.5±1.4 (BR)	P=0.002
OTU3	***Pseudomonas***	18.5±6.4	0.04±0.01 (COPD)	P=0.003
OTU4	***Streptococcus***	2.6±0.8	7.9±1.7 (COPD)	P=0.002
OTU7	***Staphylococcus***	12±6.3	0.02±0.01 (COPD)	P=0.02
OTU15	Unclassified ***Enterobacteriaceae***	0.002±0.002	4.9±4.8 (COPD)	P=0.02

5.3.5 Ribosomal intergenic spacer analysis of the composition and diversity of bronchiectasis, cystic fibrosis, and chronic obstructive pulmonary disease communities

RISA PCR bacteria species identities were interpreted through matching the most concentrated ITS amplicons to the closest *in silico* RISA profile reference based on abundant 16S rRNA gene-based OTU data and culture positivity. These results were also compared to culture and MALDI-TOF to meet objective 1. RISA PCR largely confirmed the culture-positive identities of conventional and emerging pathogenic bacteria in addition to enabling species-level inference of several predominant and highly abundant genera not identified through culture (for example *Streptococcus* spp., and *P. monteilii;* see below for more detail). RISA species assignation relied on matching *in silico* ITS size approximations to ITS amplicons visualised through the QIAxcel (see SI Table D5; SI Table D6). In concordance with 16S rRNA gene sequence abundance data, *Streptococcus* spp. was the most frequent and abundant or dominant taxon detected through RISA PCR (Figure 5.2). Of the most abundant Streptococcal OTUs, OTU1 *Streptococcus* was observed in 33 samples where the predominant RISA ITS amplicon spanned 524-545 bp (median concentration 1.5 ng/µl; IQR 2.1 ng/µl-5.2 ng/µl), and collectively corresponded to *S. gordonii*, *S. mitis, S. oralis,* and *S. pneumoniae*. OTU4 *Streptococcus* was also present in 15 samples which closely corresponded to the RISA PCR profiles of *S. anginosus, S. intermedius, S. mutans,* and occasionally *S. suis* (profiles spanning 649 bp-693 bp; median concentration 1 ng/µl; IQR 0.9 ng/µl -1.3 ng/µl). *S. gallolyticus* was postulated to be a less frequent microbiota component (corresponding to OTU5 *Streptococcus;* spanning 551 bp-575 bp; median concentration 3.1 ng/µl; IQR 1.5 ng/µl-5.8 ng/µl), while OTU12 *Streptococcus* was identified as *S. agalactiae* or *S. dysgalactiae* through both culture-dependent and culture-independent methods (SI Table D1; SI Table D5; SI Table D6).

RISA PCR profiles belonging to *Granulicatella, Gemella, Rothia, Neisseria, Prevotella,* and *Veillonella* were observed however this analysis provided too little species-level resolution for accuracy beyond most abundant and dominant versus diverse microbiota. This notion is supported by its higher overall frequency of sequenced reads and abundance in sputum over other abundant oral-associated members (Figure 5.2). OTU23 *Prevotella* was abundant and dominant in a single sample which most closely resembled complete *P. intermedia* and *P. melaninogenica* RISA PCR profiles (a 1.7 ng/µl, 663 bp and a 2.5 ng/µl, 819 bp fragment). More commonly, OTU10 *Prevotella* was abundant in 25 samples but sample ITS profiles partially matched to common oral/upper respiratory flora including *P. buccae, P. denticola,*

and *P. oris* to name a few (SI Table D5; SI Table D6). This result further complicated species identification of abundant *Veillonella* species (most commonly OTU6 and OTU27), *Rothia* (OTU11 and OTU26), or *Neisseria* (OTU16) which share similar ITS profiles (**see chapter 3;** SI Table D6). *G. adiacens* was observed as a 497 bp-517 bp amplicon (0.5 ng/µl-3.9 ng/µl) in 12 samples while *Gemella* was also identifiable through its smallest RISA amplicon (a 0.3 ng/µl-2.9 ng/µl, 477 bp-501 bp and a 0.5 ng/µl-4.5 ng/µl, 653 bp-693 bp amplicon). OTU8 *Gemella* was observed completely or partially in 17 samples which corresponded to *G. haemolysans, G. massiliensis, G. morbillorum,* and *G. sanguinis* (SI Table D5; SI Table D6).

Importantly, RISA PCR detected major and emerging respiratory pathogens that were either abundant or dominant such as *S. aureus* (corresponding to a 0.6 ng/µl-1.3 ng/µl, 642-690 bp fragment, a 1.6 ng/µl-13 ng/µl, 708-748 bp amplicon, and a 1.5 ng/µl-1.7 ng/µl, 917-937 bp amplicon), *P. aeruginosa* (a 0.6 ng/µl -15.8 ng/µl, 708-748 bp amplicon), *S. maltophilia* (a 7 ng/µl-22.6 ng/µl, 708-740 bp amplicon), and *B. cenocepacia, B. multivorans,* or *B. vietnamiensis* (a 3.4 ng/µl, 653 bp amplicon and a 8.2 ng/µl, 767 bp amplicon; SI Table D5; SI Table D6). The resulting RISA amplicons of all these species (excluding *P. aeruginosa*) generated only partial matches compared to their predicted *in silico* RISA profiles, meaning it was not possible to determine the composition of communities where more than one of these pathogens were present. In contrast, RISA enabled the confident identification of OTU14 *Pseudomonas* as *P. monteilii* (a 2.4 ng/µl, 623 bp amplicon) while *K. oxytoca, K. pneumoniae,* and *S. marcescens* could be identified with less certainty through partial RISA profile matching; SI Table D5; SI Table D6). Importantly, RISA could not resolve 1 sample that was dominated by unclassified *Enterobacteriaceae* sequenced reads. In this sample, culture identified *S. marcescens* and *K. oxytoca* which belong to the same family and were thus difficult to differentiate by amplicon-based methods (SI Table D1). Community diversity was indicated through the number and concentration of RISA profiles where highly dominant species presented with 1 or few concentrated amplicons. Diverse profiles were indicated by a higher number of RISA amplicons that were less concentrated than those of dominant species (Figure 5.4). Sample ITS profiles did not cluster according to disease diagnosis or according to disease severity (Figure 5.4; SI Table D1).

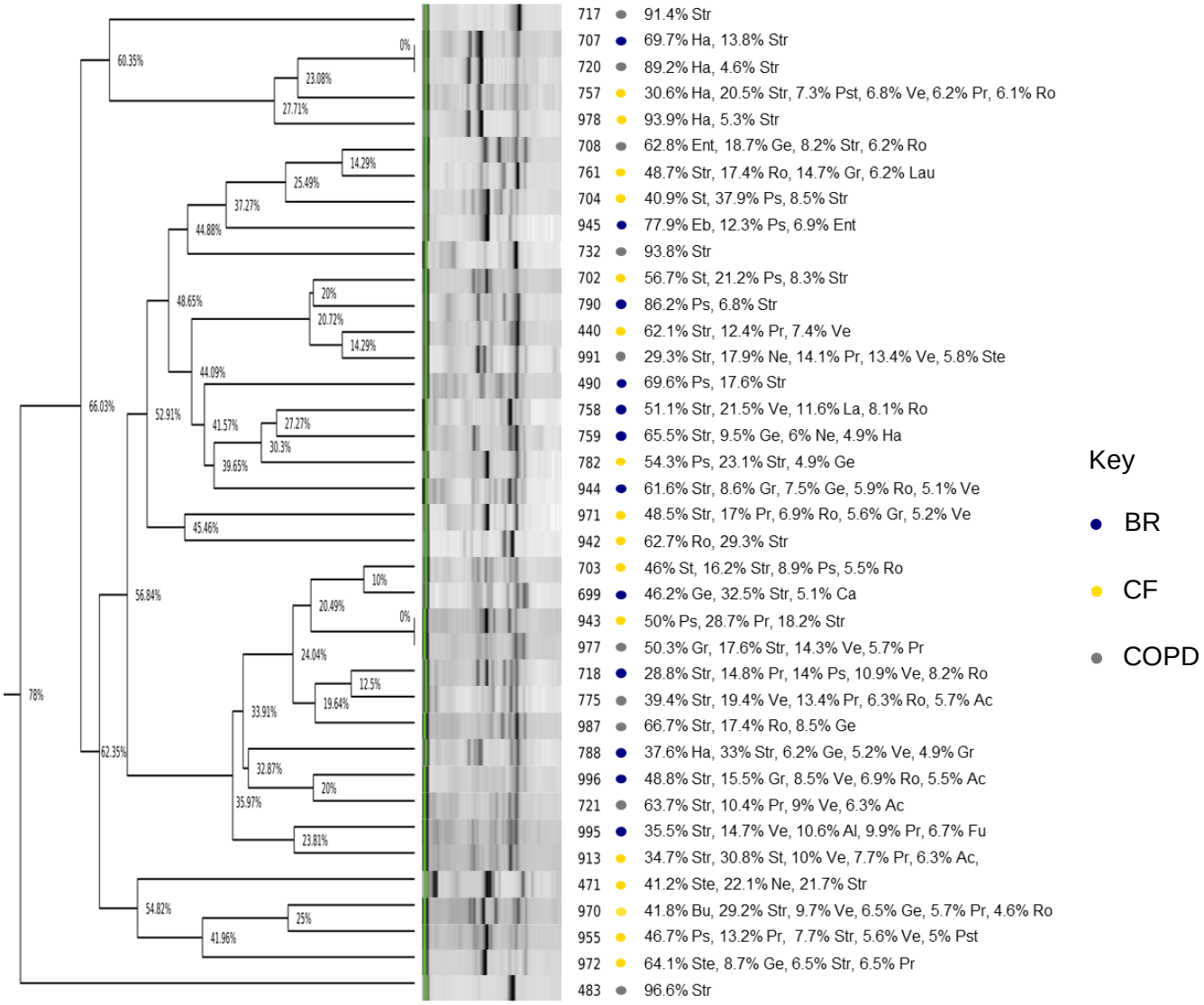


Figure 5.4 RISA hierarchical clustering of sputum microbiota. GelJ software was used to draw a dendrogram using UPGMA clustering and the Sørensen-DICE algorithm to cluster RISA profiles by similarity according to DNA molecular weight. A tolerance threshold of 2% was used to differentiate RISA fragments of a similar molecular weight. The dendrogram was manipulated in Figtree to include node dissimilarity. Raw QIAxcel gel images have been included to adjacent sample ID, diagnosis (coloured dots denoting BR in navy, CF in yellow, and COPD in grey), and 16S rRNA gene-based sequence data of highly abundant genera. The QIAxcel upper marker is also visible in each sample and is highlighted in green. Abbreviations include: Ac (*Actinomyces*), Al (*Alloprevotella*), Bu (*Burkholderia*), Ca (*Capnocytophaga*), Eb (*Enterobacterales*), Ent (*Enterobacteriaceae*), Fu (*Fusobacterium*), Ge (*Gemella*), Gr (*Granulicatella*), Ha (*Haemophilus*), La (*Lactobacillus*), Lau (*Lautropia*), Ne (*Neisseria*), Pr (*Prevotella*), Ps (*Pseudomonas*), Pst (*Pasteurellaceae*), Ro (*Rothia*), St (*Staphylococcus*), Ste (*Stenotrophomonas*), Str (*Streptococcus*), and Ve (*Veillonella*).

5.3.6 A comparison of 16S rRNA gene-based sequence data, culture, and RISA for species identification

Microbial culture complemented 16S rRNA gene-based sequence data of predominant and abundant genera from 29 samples, enabling species identification of a limited number of conventional and emerging respiratory pathogens (Figure 5.2; SI Table D1). Discrepancies in sequence data and culture arose from a minority of samples. The samples that were sequence-positive and culture-negative (false-negative cultures based on classified 16S rRNA gene sequence reads) variously failed to detect NRF (n=10), *Pseudomonas* (n=2), *Haemophilus* (n=2), *Staphylococcus* (n=1), and *Stenotrophomonas* (n=1; Figure 5.2; SI Table D1). Two samples were sequence-negative and culture-positive (false-negative sequences based on a positive culture result): 1 *C. striatum* and 1 *E. coli*. Greater cohesion was observed between RISA PCR and 16S rRNA gene sequence data, where all discernible RISA profiles were comparable to 16S rRNA gene-based predominant (n=32) and abundant (n=6) genera in spite of contradicting differences seen in chapter 3 (Figure 5.2; SI Table D6). In 5/6 of samples where RISA PCR data did not reflect most abundant species, *Streptococcus* spp., most likely representing common oral-associated species was predominant according to RISA PCR *in silico* estimations while *Pseudomonas* most likely became predominant in 1 sample (Figure 5.4; SI Table D6). Furthermore, RISA supported the sequence-negative *C. striatum* and *E. coli* result, implying these species were not present in sputum. The discriminatory power of RISA overall was akin to 16S rRNA gene sequencing where most abundant species could only be identified to the genus level.

5.3.7 Bacterial alpha diversity

Rarefaction curves and Good's coverage suggested sample diversity was captured well as most rarefaction curves reached their asymptote and coverage for each sample was >99% (SI Figure D1). The number of observed species (Sobs) ranged from 34 to 266, and on average it was highest in the COPD group (~111), followed closely by BR (~104), and finally CF (~94) groups (Figure 5.5). ACE values followed a similar trend to Sobs (Figure 5.5) Chao1 values were highest in the BR cohort, then COPD, and lastly CF (Figure 5.5). Nevertheless, no significant differences were found in bacterial richness between each disease category due to large variations in community-specific richness.

Shannon and InvSimpson indices estimate diversity through species richness and evenness. Similar to Chao1 estimates, InvSimpson diversity was highest among the BR cohort and was jointly followed by CF and COPD groups (Figure 5.5). Lower values indicate less diversity and

greater dominance by a few OTUs. According to Shannon estimates, the BR and CF cohorts were most diverse, followed by the COPD group (Figure 5.5). These differences were attributed to the Shannon index being more sensitive to rare taxa while InvSimpson is more sensitive to common/dominant taxa (Chiarucci et al. 2011). Comparable to richness, diversity scores varied widely within groups. Because of this, no significant differences in community diversity were determined between either disease category.

InvSimpson and Shannon diversity were also compared between samples that were dominated by either respiratory microbiota not considered as typical pathogens (n=18) or major and emerging respiratory pathogens (n=20). The former group encompassed *Streptococcus*, *Gemella, Rothia,* and *Granulicatella* while the latter group comprised of *H. influenzae, Pseudomonas* spp., *S. maltophilia, Klebsiella* spp., *S. aureus, B. vietnamiensis, S. agalactiae* and *S. pneumoniae.* The presence of dominant respiratory pathogens significantly reduced both InvSimpson (*t*=3.8; df=36; *P*=0.0006) and Shannon (t=3.2; df=36; *P*=0.003) diversity measures (Figure 5.6).

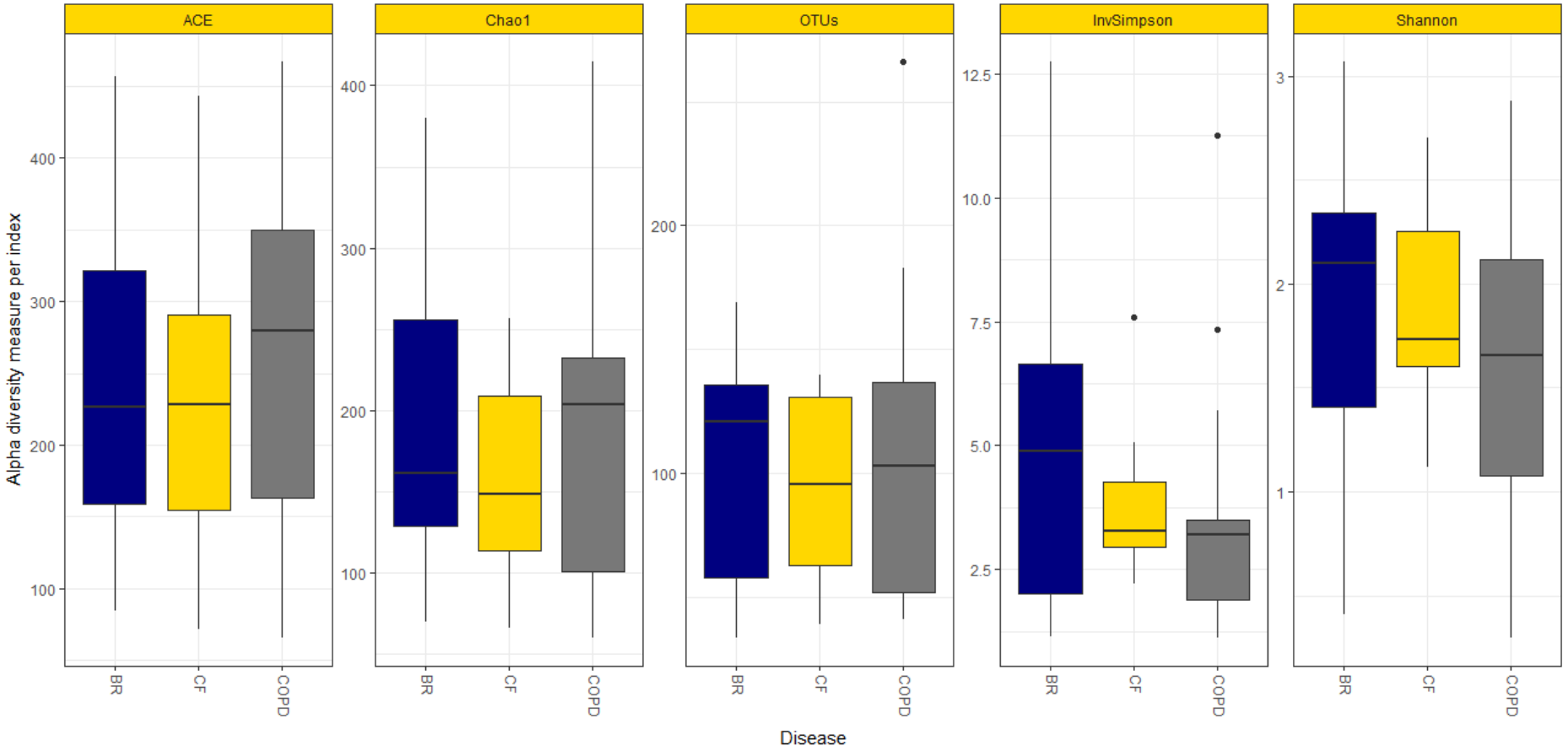


Figure 5.5 Alpha diversity scores grouped by disease category. Boxplots include species richness and diversity scores of samples that have been coloured according to disease diagnosis. Outliers have been denoted by a dot.

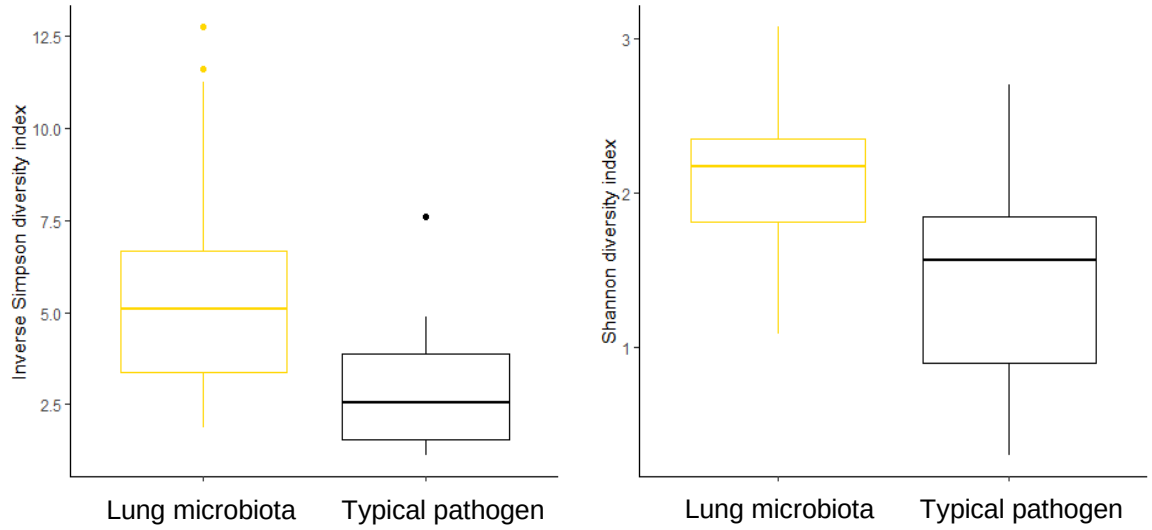


Figure 5.6 Impact of dominant respiratory pathogens on microbial diversity. Overall microbial diversity of respiratory microbiota excluding typical pathogens and bacterial pathogens was plotted against InvSimpson diversity index (left) and the Shannon diversity index (right). Median values are indicated by a horizontal line which do not overlap both boxplots. Outliers have been denoted with a dot.

5.3.7.1 Is alpha diversity a useful indicator of disease progression?

Objective 3 was achieved by correlating both Shannon and InvSimpson diversity indices to the age, sex, % FEV_1 predicted, disease diagnosis, and sputum purulence of 38 patients through multiple linear regression (SI Table D1). As lung function data could not be obtained for 12 patients, multiple imputation through predictive mean matching was employed to not further restrict the dataset of a small patient cohort. Correlation of lung function (% FEV_1 predicted) to bacterial diversity showed a positive trend between lung function and bacterial diversity however this result was not significant (Figure 5.7). A negative but insignificant correlation was observed between patient age and bacterial diversity (SI Figure D2). Higher sputum purulence was also associated with lower bacterial diversity, as was the male sex, however these effects were also statistically insignificant (SI Figure D2). There was no significant interaction between disease and lung function nor age and lung function.

All highly abundant and predominant genera were also compared against lung function to determine what pathogens, if any, correlated to reduced lung function (objective 4). Data was extracted from both upper quartile and lower quartile % predicted FEV_1 values, representing mild pulmonary disease and severe/very severe pulmonary disease (Flight et al. 2015). Genus % median relative abundance (only from the higher and lower quartile groups) was compared between each group. Increased *Streptococcus* (median relative abundance 16.2%), *Pseudomonas* (8.9%), *Granulicatella* (1%), *Rothia* (5.5%), *Veillonella* (3.4%), and

Staphylococcus (0.08%) were associated with higher lung function but were statistically insignificant. On the other hand, increased relative abundance of *Gemella* (3%), *Haemophilus* (0.2%), and *Prevotella* (5.7%) were associated with poorer lung function however these effects were also statistically insignificant (Figure 5.8).

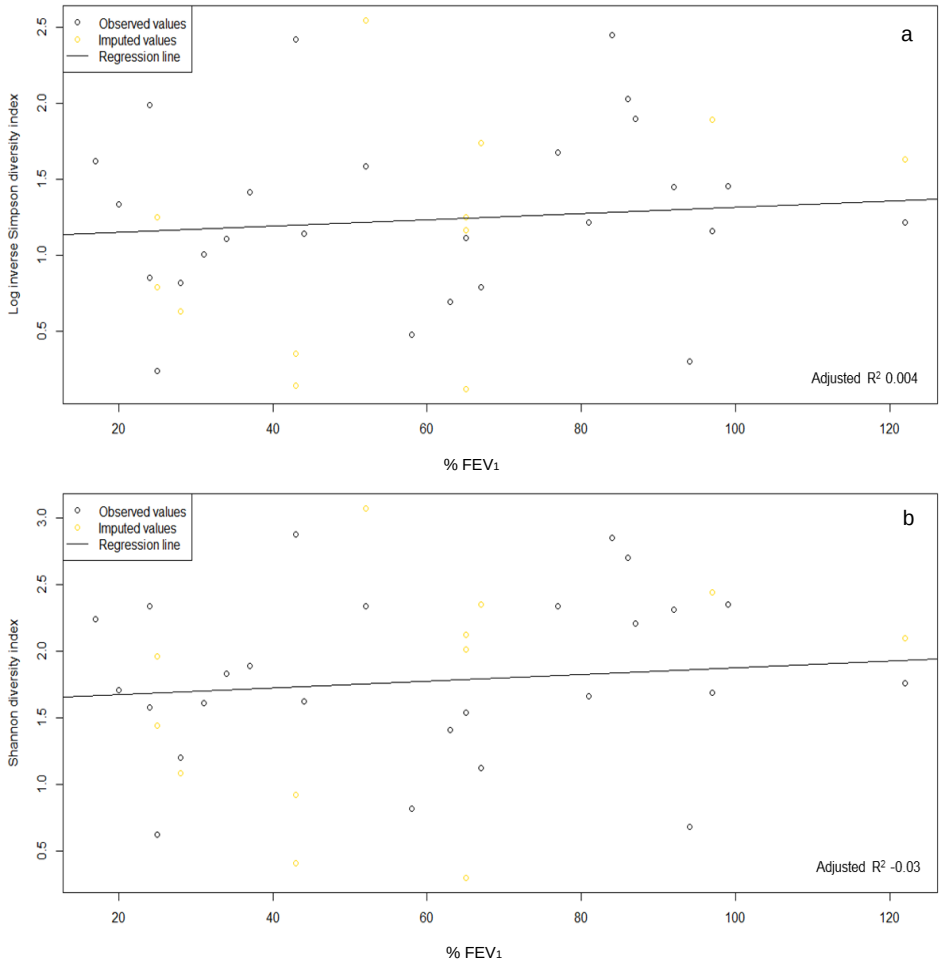


Figure 5.7 Is lower alpha diversity an indicator of poorer lung function? Both Inverse Simpson (a) and Shannon diversity (b) has been plotted against % FEV_1 predicted values for multiple linear regression. A regression line has been included along with the regression model R^2 value. Imputed values through predictive mean matching have been highlighted in gold.

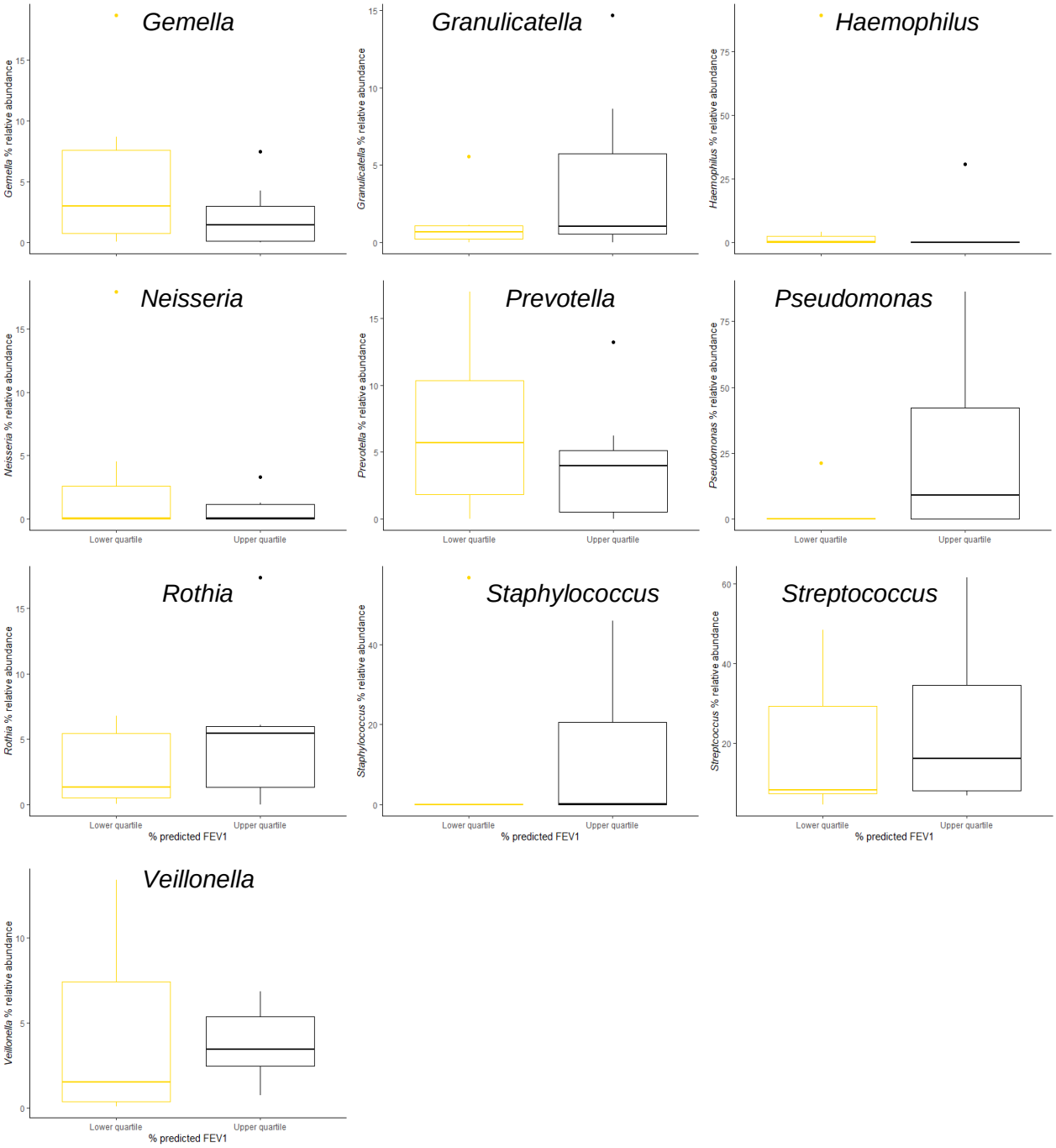


Figure 5.8 Are predominant and highly abundant genera associated with higher or lower lung function? All predominant and highly abundant respiratory microbiota including conventional pathogens were plotted against lowest and highest lung function quartiles and are labelled. Outliers have been denoted with a dot.

5.3.8 Bacterial beta diversity

5.3.8.1 Patient demographics, disease, and community membership (unweighted UniFrac)

Community membership was visualised with an unweighted UniFrac NMDS, annotated by OTUs that significantly altered NMDS spatial coordinates ($P<0.05$). NMDS (lowest stress: 0.3; R^2: 0.5) illustrated an overlap of the spatial coordinates across all 3 disease categories suggesting community membership was not significantly different among these groups (Figure 5.9). The large number of significant OTUs reflected high individuality of the respiratory microbiome, although up to 49 of these 110 significant OTUs were rare taxa (e.g. OTU22 *Fusobacterium,* OTU53 *Lactobacillus,* OTU45 *Oribacterium,* OTU85 *Mogibacterium* and OTU33) which suggests they are infrequent members of respiratory microbiota or contaminants (SI Table D3). A PERMANOVA suggested community membership was not significantly different between disease categories. The test of homogeneity of multivariate dispersions (betadisper) was also insignificant. Community membership and inter-group dispersion were also statistically insignificant between patient age, sex, lung function, except for sputum purulence ($P=0.001$) (Figure 5.10; SI Figure D3). This result was due to a large imbalance in group size causing significantly different levels of dispersion per group. There were no significant interactions between disease and lung function nor age and lung function.

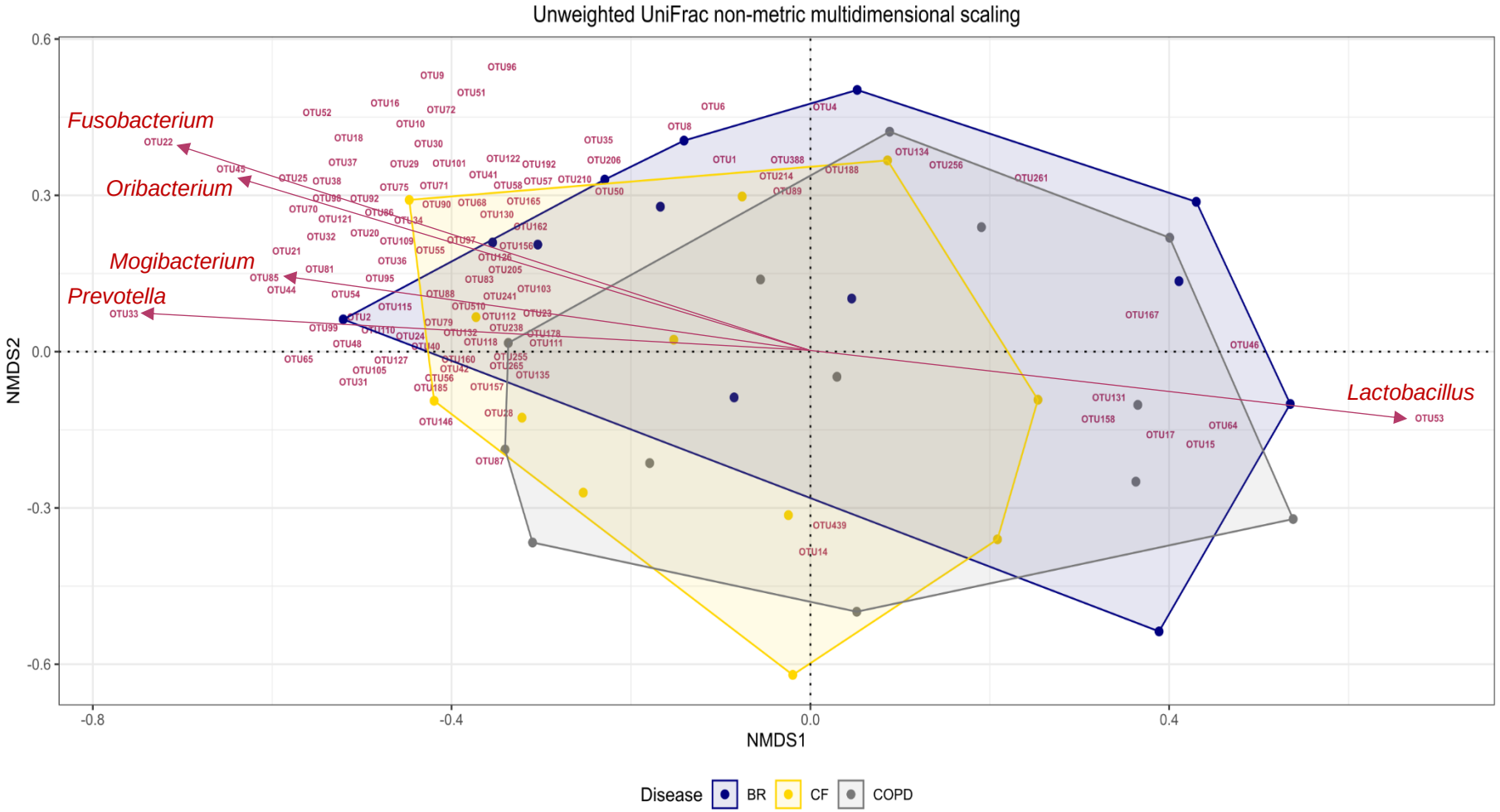


Figure 5.9 Unweighted UniFrac non-metric multidimensional scaling. NMDS axes were generated through Mothur where measurements from the first 2 axes were used to draw a 2D biplot. Each point has been coloured by disease category and connected via hulls. These hulls overlap, suggesting community membership was not significantly different per disease category. OTUs which significantly shifted spatial coordinates (P<0.05) have been included approximate to its true coordinates to avoid clustering of labels. Most significant of these OTUs (P<0.001) have been labelled and marked with an arrow.

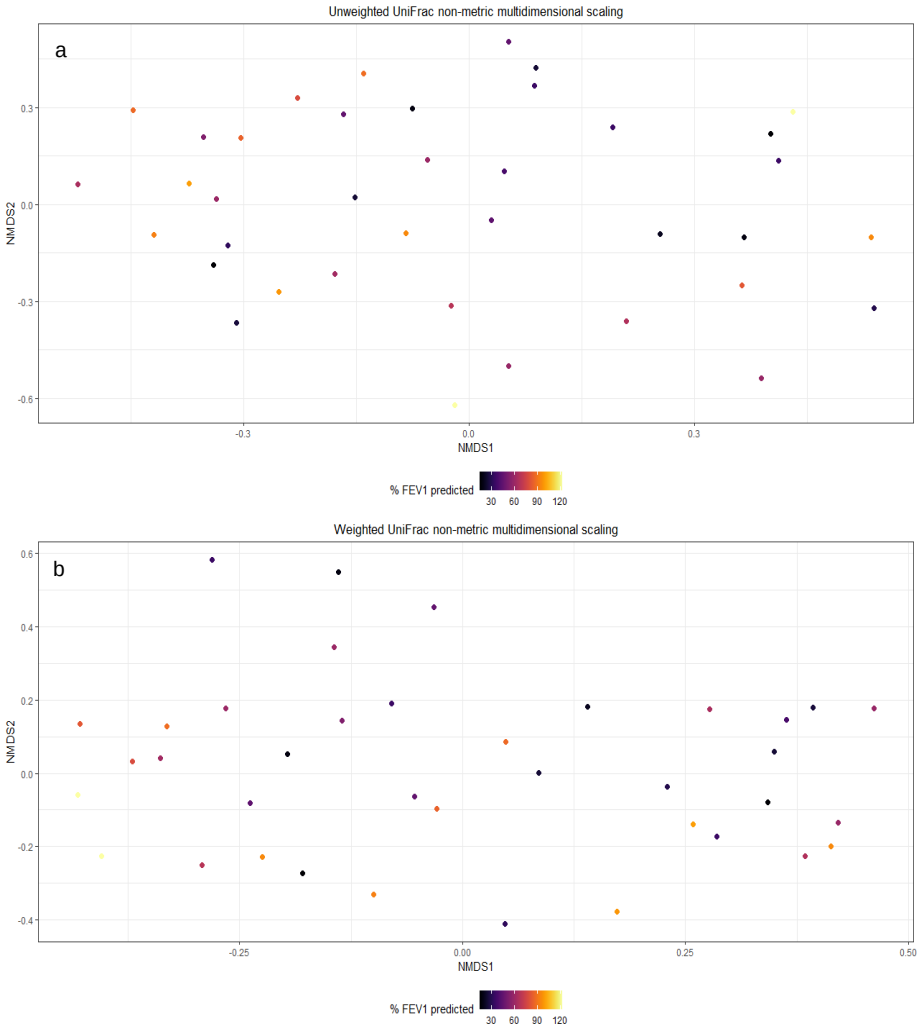


Figure 5.10 Does bacterial community membership and structure differ according to lung function? NMDS coordinates from unweighted UniFrac (a) and weighted UniFrac (b) were plotted and coloured in a gradient using R Viridis package where a lighter colour denoted higher lung function (Garnier et al. 2021).

5.3.8.2 Patient demographics, disease, and community structure (weighted UniFrac)

Community structure was visualised using a weighted UniFrac NMDS, annotated with OTUs which significantly altered NMDS spatial coordinates ($P<0.05$). NMDS (lowest stress: 0.2; R^2: 0.8) again illustrated a clear overlap in the community structure of each disease (Figure 5.11). In contrast to the unweighted UniFrac, this plot highlighted that the OTUs most responsible for shifting spatial coordinates were highly abundant or dominant in the 16S dataset: OTU5 *Streptococcus*, OTU11 *Rothia,* OTU7 *Staphylococcus*, and OTU3 *Pseudomonas* ($P<0.0001$). This shift in coordinates partially reflected differentially abundant OTUs mainly between CF and COPD communities. This plot potentially demonstrated the effect of rare OTUs shaping community structure that could have originated from sputum or gDNA extraction and sequencing reagents (SI Table D3). Neither PERMANOVA nor betadisper detected significant differences in community structure between disease categories, patient age, sex, or lung function. PERMANOVA determined sputum purulence as the only significant variable ($P=0.04$) while betadisper was also insignificant (Figure 5.10; SI Figure D3). This result was once again influenced by a large imbalance in group size. Interactions between disease and lung function or age and lung function were also insignificant.

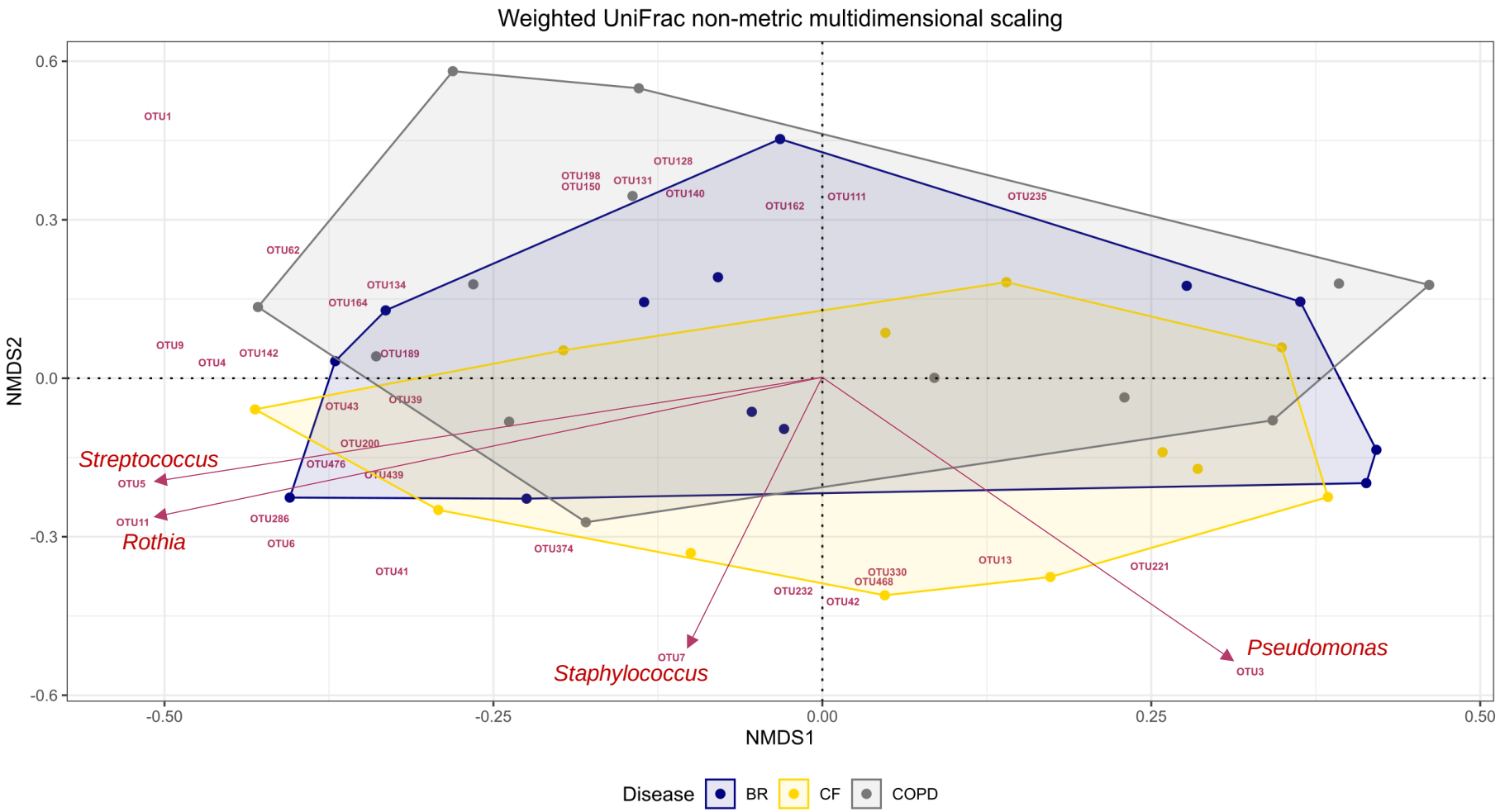


Figure 5.11 Weighted UniFrac non-metric multidimensional scaling. NMDS axes were generated through Mothur where measurements from the first 2 axes were used to draw a 2D biplot. Each point has been coloured by disease category and connected via hulls that overlap. OTUs which significantly shifted spatial coordinates (P<0.05) have been included approximate to its true coordinates to avoid clustering of labels. Most significant of these OTUs (P<0.001) have been labelled and marked with an arrow.

5.4 Discussion

5.4.1 Rationale for incorporating DNA-based methods in routine public health diagnostics

Clinical microbiological and public health laboratories are increasingly incorporating NGS platforms in routine practice to complement or even replace conventional diagnostic tools (Gargis et al. 2016). Deep sequencing has also become a popular tool to identify typical and unconventional pathogens that can help explain the onset of culture-negative pulmonary exacerbations and clinical courses (Caverly et al. 2015). To date, PHW bacteriology does not supplement routine culture with DNA-based methods which is needed for more accurate and rapid identification of respiratory microbiota contributing to disease progression. Worryingly, these culture-based methods are based on an outdated presumption that infection is caused by one or few conventional pathogens that are selectively isolated (Jorth et al. 2019).

This chapter proposes the integration of highly translatable and validated 16S rRNA gene-based NGS and RISA PCR protocols into PHW respiratory bacteriology to be used alongside routine culture. Here, both culture and molecular-based methods provided a comprehensive insight into respiratory microbiota composition and diversity, highlighting the potential role and limitations of NGS and RISA PCR for routine microbiota profiling. Differences in disease-specific community composition and diversity were investigated where NGS provided a broad overview of total genera present. In support of hypothesis 1 (culture-independent methods identify clinically important microorganisms missed through culture), bacterial species identification was enhanced by both culture and RISA PCR (**see section 5.4.2**). NGS also enabled a key question to be addressed concerning the effectiveness of bacterial diversity measure in predicting clinical course (hypothesis 2; bacterial diversity is not an informative predictor of lung function decline due to high intra-individual and inter-individual variations in core and dominant respiratory microbiota; **see section 5.4.3**).

5.4.2 16S rRNA gene-based metaprofiling reflected current epidemiological data and identified key respiratory genera missed in culture

As detected by NGS, core CF respiratory microbiome *Actinomyces, Fusobacterium, Prevotella, Veillonella, Gemella, Granulicatella, Neisseria*, *Rothia,* and *Streptococcus* were frequently abundant in sputum, comparable to the findings of Filkins and O'Toole (2015); Mahboubi et al. (2016); Caverly et al. (2018).These oral-associated species, especially *Prevotella* and *Veillonella* are among the most dominant anaerobic species of the respiratory tract and may constitute at least 50% of the bacterial community in the absence of

predominating emerging non-fermenting Gram-negative (ENFGN) and conventional pathogens such as *A. xylosoxidans,* BCC, *P. aeruginosa, S. aureus* and *S. maltophilia* (Goddard et al. 2012; Moran Losada et al. 2016; Muhlebach et al. 2018). Furthermore, this data reflected current CF reported studies where *Proteobacteria* and *Firmicutes* were dominant phyla. Analysed CF communities were characteristically dominated by *S. aureus, P. aeruginosa*, common oral-associated species, and less frequently, *H. influenzae, S. maltophilia, B. vietnamiensis*. This study highlighted the rise of Gram-negative non-fermenters and the reduction of *P. aeruginosa* among the adult population due to improved clinical care and infection control (Hatziagorou et al. 2020).

The microbial composition of BR and COPD communities in this work bore a strong semblance to abundant CF microbiota within this study and the wider literature. Consequently, there were no significant disease-specific differences in phylum-level or genus-level community composition. BR and COPD core microbiota comprised of *Fusobacterium, Haemophilus, Neisseria, Pseudomonas, Prevotella, Rothia, Streptococcus, Serratia, Staphylococcus,* and *Veillonella* in abundance, comparable to the results of Erb-Downward et al. (2011); Cabrera-Rubio et al. (2012); Rogers et al. (2013); Tunney et al. (2013); Garcia-Nuñez et al. (2015); Sze et al. (2015); Lee et al. (2018). COPD communities were largely dominated by traditional respiratory pathogens, common oral-associated microorganisms and a non-conventional pathogen, *P. monteilii.* Similarly, BR predominant species mostly comprised of *Streptococcus* spp. (n=7 including *S. agalactiae*), conventional and emerging pathogens including *K. oxytoca/S. marcescens* and *S. maltophilia*. These results also support recent epidemiological studies where *S. maltophilia* and other ENFGN species are being increasingly detected in BR respiratory secretions (Inchingolo et al. 2021). Unusually, *M. catarrhalis* was not abundant in COPD sputa but this may be due to a small patient cohort and high individual variability.

Besides key genera (i.e., core lung microbiota, recognised emerging pathogens, and major respiratory pathogens), several abundant unconventional pathogens or PPMs were detected in 16S rRNA gene sequence reads that were absent from culture. Notable examples included *Leptotrichia* (n=7), *Capnocytophaga* (n=2), and *Lautropia* (n=1). Although its clinical importance is unclear, the presence of these species may be significant. For instance, *Leptotrichia* spp. have been previously isolated from bronchiolar lavage (BAL) fluid of patients with respiratory distress and are associated with community-acquired pneumonia in addition to acute COPD exacerbations (Lo 2012; Su et al. 2015). Additionally, *L. mirabilis* has been implicated in the contribution of CF disease while *Capnocytophaga* spp., has been reported in rare cases of pneumonia (Ben Dekhil et al. 1997; Gosse et al. 2019). Whether these species are infrequent or underrepresented pathogens of the respiratory tract is unclear, however,

current epidemiological and clinical data has overlooked these species as novel emerging respiratory pathogens.

While it is important to rapidly identify and combat both recognised and potential pathogens, there is also growing evidence for mucin-degrading anaerobic microbiota (for example *Prevotella, Veillonella, Streptococcus,* and *Fusobacterium*) modulating and increasing the pathogenicity of *P. aeruginosa* in the CF lung (Duan et al. 2003; Flynn et al. 2016; Li et al. 2021). The pathogenic potential of oral-associated microbiota may be overlooked when patients are concurrently treated for *P. aeruginosa* and other multidrug resistant Gram-negative species. Such is the case for *Gemella* spp., that has been associated with severe BR disease, in addition to *Rothia* and *Neisseria* species that are known causative agents of pneumonia (Cho et al. 2013; Humbert and Christodoulides 2019; Sharma 2019). While amplicon sequencing might not be able to elucidate species pathogenic potential, this research highlights the necessity of using deep sequencing to observe respiratory microbiota composition changes that could necessitate higher-resolution DNA-based methods (such as metagenomics) to further investigate functional abilities.

Although RISA performed poorly at differentiating complex diversity profiles and less abundant taxa, this method provided greater resolution at identifying *Streptococcus* species in combination with 16S rRNA gene-based OTU abundance data rather than being used alone. In particular, the *S. milleri* group (*S. anginosus, S. constellatus* and *S. intermedius*) and *S. pneumoniae* may have been more widespread in respiratory samples than was initially suggested through culture and NGS. This complex of species is difficult to culture or differentiate from non-pathogenic Streptococci and are linked to pulmonary exacerbations (Grinwis et al. 2010). This merits further investigation, as ITS profiles of the *S. milleri* group could have been mistaken for other abundant species while ITS profiles of *S. pneumoniae* could not be differentiated from non-pathogenic streptococci. RISA also enabled the identification of *P. monteilii* that was absent from culture data. *P. monteilii* is rarely isolated from clinical specimens and is frequently misidentified in culture, therefore its pathogenic potential is uncertain. Nevertheless, its association in BR exacerbations and fatal meningoencephalitis raises concern over the accurate identification of this species (Aditi et al. 2017; Gupta et al. 2018). In this case, RISA could act as a valuable tool to detect dominant pseudomonal species that can be differentiated from basic ITS profiles.

5.4.3 Bacterial diversity was an uninformative predictor of lung function decline

Previous studies have shown marked differences in CF, COPD, and BR lung microbiota diversity where declining diversity has been associated with worsened lung function (Flight et al. 2015; Cox et al. 2017; Fenker et al. 2018; Tiew et al. 2021). Whether microbial diversity can predict disease progression or be controlled to prolong exacerbations is debatable, yet greater microbial diversity in adulthood has been associated with a less severe and slower progression of lung function decline. Despite this, the causal relationship between these variables is still unclear (Caverly et al. 2015).

Although the CF microbiome is better defined than the COPD and BR lung microbiome, recent studies using NGS have highlighted consistencies in core lung microbiota composition and bacteria diversity across disease diagnoses akin to this study (Faner et al. 2017). These more recent findings are significantly influenced by sampling location, sampling method (e.g., cough swabs, sputum, BAL fluid) and sequencing protocol which introduce inconsistencies among the existing literature (Prevaes et al. 2017; Ronchetti et al. 2018). NGS has also illustrated highly individual polymicrobial communities on an OTU basis. This high individuality was reflected within this study cohort where no disease-specific or demographic-specific distinction were found in bacterial richness, diversity, community membership, and community structure. These results did however highlight a significant loss of diversity when traditional and emerging pathogens dominated respiratory samples. Nevertheless, these predominant pathogenic species did not singly correlate to poorer lung function.

Similarly, a limited yet growing number of studies have reported a poor correlation between bacterial diversity and lung function decline (Stressmann et al. 2012; Paganin et al. 2015; Zemanick et al. 2017). Moreover, when compared between disease diagnoses this relationship is further complicated by disease pathology and how disease severity is defined between diagnoses. As a result, bacterial diversity as a standalone should not guide therapeutic strategies. Reasons for this poor correlation between diversity, patient clinical parameters and clinical course is attributed to reduced statistical power due to a small study cohort, highly individual microbiota as aforementioned, and may reflect microbiome transcriptomic and metabolomic activity changing independently to bacterial load, or non-bacterial microorganisms driving exacerbations (Cox et al. 2017). Sequence error and spurious OTUs could have also inflated diversity within these samples which could have led to erroneous conclusions although inflated OTUs should be consistent across groups (Caporaso et al. 2011). Similarly, the grouping of multiple bacterial species into single OTUs

is a common problem for 16S rRNA hypervariable gene regions leading to a misrepresentation of true bacterial diversity.

5.4.4 Evaluation of culture and amplicon-based analysis in public health microbiology

DNA-based methodologies are often portrayed as an appropriate substitute to culture however significant, and at times opposing, limitations exist between the two (Jorth et al. 2019). These conflicting limitations necessitate the integration of both culture-dependent and culture-independent diagnostic tools in PHW microbiology. A major drawback of sequencing is that 16S rRNA gene hypervariable regions often result in poor phylogenetic resolution at the genus level (Yang et al. 2016). Because of this poor resolution, clinical microbiology will often supplement NGS with extended culture (culturomics) to resolve sequence data (Françoise and Héry-Arnaud 2020). In the present study, culture proved valuable at resolving unclassified *Enterobacteriaceae* (*Klebsiella* spp., and *S. marcescens*) while providing species-level discrimination for several major and emerging pathogens. These results were supported by MALDI-TOF; however MALDI-TOF misidentification rates may also be significant and vary between Gram-negative species, *Streptococcus* spp., and *Staphylococcus* spp. (Florio et al. 2019). Another major drawback to culture is that it significantly underrepresents diverse polymicrobial communities within the lung and the full extent of microorganisms that may also contribute towards disease progression. This disparity was evident when comparing the 623 sequenced OTUs to the 13 bacterial species identified through culture. Sequence-based analysis is therefore required for a more robust identification of clinically important species.

The resolution of RISA PCR was limited to predominant species, *Streptococcus* spp., and *Pseudomonas* spp. Non-Streptococcal respiratory microbiota was generally difficult to identify, often resulting in genus-level identification. RISA was also mostly unable to differentiate ITS profiles in polymicrobial communities. This lack of resolution was due to the complexity and number of amplified ITS fragments, inability to differentiate species with highly similar percent relative abundance scores, incomplete profile matches to *in silico* references, obscuring of similar RISA profiles, and variations in QIAxcel measurements. This meant a more accurate estimation of most abundant species that amplified through RISA PCR was not possible. Regardless of this limitation, RISA resolved culture-negative and false-positive culture data on accordance with NGS data. This also suggested DNA amplification bias was not as significant as **chapter 3** sequencing of mock communities. In spite of this, further optimisations to RISA could improve the sensitivity and specificity of this assay to detect respiratory microorganisms.

5.4.5 Limitations of experimental design and experimental bias

Microbial DNA contamination arising from gDNA extraction kits, reagents, and sampling should be heavily scrutinised in low-biomass samples as it obscures clinically relevant taxa and alters diversity (Pollock et al. 2018). For example, the ZymoBIOMICS™ gDNA extraction kit has previously shown to overrepresent *Pseudomonas, Salmonella, Escherichia-Shigella* and underrepresent *Enterococcus, Staphylococcus, Listeria* and *Bacillus* that may have impacted this study (Claassen-Weitz et al. 2020). Several reagent-originating OTUs may have significantly shifted spatial coordinates of community membership and structure therefore these results must especially be taken seriously should between-group community membership and structure be significantly different.

Human gDNA is another source of contamination that should be considered for elimination as it impedes microbial DNA amplification. Although prior removal of human gDNA is generally recommended, it is also possible for bacterial DNA to overwhelm the ratio of human gDNA which could drastically reduce the sensitivity of PCR-based assays (Pérez-Brocal et al. 2020). Without a prior human gDNA depletion step, amplification of undiluted DNA samples was successful where non-bacterial sequences were eliminated downstream. On the other hand, how human gDNA removal influence microbial abundances should be investigated to improve the cohesion between culture and DNA-based methods if possible.

Another common drawback to 16S rRNA gene-based studies is attributed to sequencing relatively high proportions of dead bacteria which varies by site and time (Bellali et al. 2019; Bellali et al. 2021). Without an accurate estimation of live and dead bacteria, metabolically active species contributing to disease progression cannot be determined (Caverly et al. 2015). This factor may explain sequence positive and culture negative results which could be influenced by antipseudomonal and anti-*Haemophilus* treatments, however not all of these culture negative samples could be explained by concurrent antibiotic use. Similarly, sample storage could have led to these conflicting results. All samples were stored for at least 24-h which is known to distort community composition (Wielmann et al. 2017). On the other hand, these differences could also be attributed to sample contamination, poor or differential PCR amplification using degenerate primers, unequal cell lysis or Gram-negative gDNA degradation, and sampling heterogeneity of clinical samples. Though timing of sample collection and processing could not be controlled for, both culture and culture-free methods should be employed simultaneously to reduce this source of bias which could see certain taxa being systematically underreported or should be stored in sodium azide or cryoprotectants to halt bacterial growth prior to gDNA extraction.

5.4.6 Future direction

This study was restricted to a small retrospective cohort meaning the number of observed clinical parameters were too small to draw robust statistical conclusions. Lung function must be especially interpreted with caution due to multiple imputation of missing values. Additionally, temporal data for individual microbiota change could not be generated and antimicrobial treatment history was not controlled for. As such, larger-scale longitudinal studies that compare bacterial diversity to underlying disease, dominant pathogens and antibiotic treatment may lead to stronger correlations between respiratory microbiota and exacerbations/disease progression (Rogers et al. 2013; Cox et al. 2017). A larger cohort would also allow statistically robust correlations between patient demographics and clinical course. Repeated sampling could establish whether changes in respiratory microbiota composition and diversity is largely due to clinical changes (i.e., changes in disease trajectory), sampling variability or community variability not affecting patient outcome. However, this is dependent on costs and resource allocation (Acosta et al. 2017; Cox et al. 2017). Alongside sampling diseased individuals, a better understanding of the healthy lung microbiome is crucial for managing pulmonary exacerbations (Cox et al. 2017; Françoise and Héry-Arnaud 2020). Drivers of loss of bacterial diversity including interspecies competition, antibiotic treatment and host-microbiome responses must be investigated further to determine its use as an indicative tool of exacerbations (Faner et al. 2017). Since OTU analysis cannot define pathogenicity, species and strain level resolution is required to link pathogens to disease prognosis (through whole genome sequencing) while a better estimation of live bacteria is required to determine metabolically active species (propidium monoazide/fluorescence *in situ* hybridisation). RISA PCR used primers that were originally developed for soil bacterial communities (Borneman and Triplett 1997). This method could be refined by redesigning primers more specific to respiratory microbiota. These primers could alter species ITS amplicon size and complexity, giving rise to more interpretable community composition and diversity.

5.4.7 Conclusions

This research demonstrates the potential of combining culture-based and culture-independent methods for routine identification of respiratory microbiota. Both DNA-based methods provided a broad and detailed analysis of community diversity highlighting PPMs missed in culture, accepting hypothesis 1. RISA PCR complemented sequence data through providing greater bacterial species discrimination alongside culture and MALDI-TOF, yet was limited when comparing more complicated and diverse communities. Bacterial diversity was a poor predictor of lung function decline and is thus not likely to be useful to guide clinical therapeutic

strategies. Combined, these findings reflect a growing evidence base of highly individual respiratory microbiota where there is a poor correlation between bacterial diversity and clinical course, supporting hypothesis 2. Reasons for this poor correlation are complex and multifactorial, and non-bacterial components of sputa should also be considered for investigation to better understand these associations. In summary, DNA-based methods should not be used to guide clinical therapeutic strategies alone, but should supplement culture-dependent diagnostic methods.

Chapter 6 Comparative genomics, pathogenicity traits and antimicrobial resistance of *Achromobacter* spp.

6.1 Introduction

Achromobacter spp. are versatile aerobic, Gram-negative rods that are widely distributed in the environment where to date, 22 species have been validly published (Jeukens et al. 2017; Parte et al. 2020). *A. xylosoxidans* is a problematic species responsible for various healthcare associated infections including blood, urine, catheters, cerebrospinal fluid, sputum, and burn wounds (Kanellopoulou et al. 2004). This species is notably documented as an emerging CF pathogen (Mahenthiralingam 2014; Swenson and Sadikot 2015; Firmida et al. 2016), while also being associated with episodes of disease in patients suffering from COPD and BR (Renom et al., 2010, Tunney et al., 2013, Liu et al., 2017).

A. xylosoxidans colonisation in the lung has previously been shown to be able to transition from transient carriage to chronic infection in CF (Firmida et al., 2016). Dominant *A. xylosoxidans* clones previously reported in chronically colonised CF patients have also been associated with a severe loss of bacterial lung diversity and lung function from microbiota studies (Pereira et al., 2011, Talbot and Flight, 2016). *A. xylosoxidans* clinical relevance is however disputed due to varying clinical outcomes associated with *A. xylosoxidans* colonisation (Edwards et al. 2017; Somayaji et al. 2017; Cuthbertson et al. 2020). *Achromobacter* spp. transmission events are similarly disputed where a few case studies have suggested patient-patient transmission (Kanellopoulou et al. 2004; Van Daele et al. 2005; Raso et al. 2008) along with the argument that transmission events derive from a common environmental source (Pereira et al. 2011; Ridderberg et al. 2011). Differences in reported source of transmission are supported by more recent genomic and/or molecular typing data in addition to clinic visit-based epidemiology (Gabrielaite et al. 2021).

For decades, CF lung infection has been determined by traditional bacterial and fungal cultures, focusing on common, culturable microorganisms and major pathogens such as *P. aeruginosa* (Burns and Rolain, 2014). *Achromobacter* species have been difficult to identify as using conventional bacteriology. More recently, the wider use of high resolution identification techniques such as MALDI-TOF and multilocus sequence typing (MLST), and sequencing of 16S rRNA and *nrdA* genes has led to a greater understanding of *A. xylosoxidans* emergence in CF (LiPuma, 2010, Spilker et al. 2012; Mahenthiralingam, 2014; Coward et al. 2016; **section 1.5; chapter 5**). These epidemiological studies show the

prevalence of *A. xylosoxidans* CF lung infection has increased from between 1.1% (De Baets et al. 2007) to 30% (Somayaji et al. 2017) across multiple centres. Overall, *A. xylosoxidans* is of significant interest because of its increased prevalence, unclear clinical relevance, and also because it is innately highly multidrug-resistant and its response against last line drugs is variable (Traglia et al., 2012, Nielsen et al., 2019).

While numerous DNA-based tools can detect *Achromobacter* spp., species within the genus are often misidentified as *A. xylosoxidans* since conventional biochemical tests, 16S rRNA gene-based sequencing, and MALDI-TOF lack species-specific discriminatory power (Garrigos et al. 2021). Following recent advances in whole genome sequencing (WGS), PCR, and DNA typing, *A. aegrifaciens, A. animicus, A. anxifer, A. deleyi, A. denitrificans, A. dolens, A. insolitus, A. insuavis, A. marplatensis, A. mucicolens, A. piechaudii, A. pulmonis, A. ruhlandii,* and *A. spanius* have also been recovered from CF sputum while 9 of these species have rarely been detected in non-CF sputa (Vandamme et al. 2013; Edwards et al. 2017; Veschetti et al. 2021). After *A. xylosoxidans, A. ruhlandii* is now the second commonest *Achromobacter* species in the US (23.5% of reported cases) while *A. insuavis* (6.7%-12.4% of cases) and *A. dolens* (8% of cases) are more prevalent in Europe (Spilker et al. 2013; Amoureux 2016; Coward et al. 2016). Like *A. xylosoxidans,* these less understood species also exhibit multidrug resistance and are able to chronically colonise the CF lung (Dupont et al. 2015; Pereira et al. 2017; Isler et al. 2020).

Despite the increasing prevalence of *Achromobacter* spp. infection and concern over potential dissemination of multi-drug resistant clones, research establishing its phenotypic pathogenic potential is lacking (Marion-Sanchez et al. 2019; Isler et al. 2020). Furthermore, the phenotypic adaptation of non-*A. xylosoxidans* species to the lung environment remain less clear. Due to its rare occurrence and changing taxonomy, species-specific pathogenic features have not been fully defined and epidemiological data is scarce (Isler et al. 2020).

6.1.1 Aims and objectives

The goal in this chapter was to use WGS to identify genomic and phenotypic features associated with *Achromobacter* spp. pathogenicity, transmission, antimicrobial resistance (AMR) and host adaptation beyond the well-studied *A. xylosoxidans* species.
Few studies have conducted WGS or phenotypic assays for this purpose while this data does not yet fully exist for *A. insuavis* and *A. ruhlandii* (Ridderberg et al. 2011; Hu et al. 2015; Edwards et al. 2017; Gade et al. 2017; Fleurbaaij et al. 2018). To better-understand *Achromobacter* spp. infection in CF and its degree of multi-drug resistance, a retrospective

analysis of 28 suspected clinical *A. xylosoxidans* isolates previously identified by MALDI-TOF was conducted. This collection also represented infections encountered within Wales.
The objectives of this chapter were:

1) Evaluate the performance of MALDI-TOF against WGS for the correct identification of *Achromobacter* species.
2) Determine possible transmission events of *Achromobacter* strains among the Welsh CF population through single nucleotide polymorphism (SNP) analysis.
3) Determine AMR profiles by antimicrobial susceptibility testing and *in silico* AMR gene detection.
4) Research *Achromobacter* spp. phenotypic traits to elucidate its pathogenic potential and competitive fitness.

6.1.2 Hypotheses

Hypotheses for this chapter were:

1) WGS significantly outperforms routine MALDI-TOF in the accurate identification of *Achromobacter* species.
2) WGS provides the resolution required to guide appropriate antimicrobial therapies.
3) *Achromobacter xylosoxidans* competitive fitness (as inferred from phenotypic traits) is demonstrably similar to major CF Gram-negative species.

6.2 Materials and methods

6.2.1 Culture collection, maintenance and MALDI-TOF identification

Twenty-eight CF *Achromobacter* spp. isolates were obtained from 16 patients spanning 5 hospitals including one isolate from a clinical study (Table 2.2). Each isolate was grown in SACU on Columbia blood agar with 5% horse blood (Oxoid, UK) to be incubated aerobically for 24-h at 37°C for MALDI-TOF analysis, 48-h at 37°C for gDNA extractions, or for 18 h at 37°C for phenotypic analysis. MALDI-TOF was conducted on these cultures using the Bruker Biotyper MALDI-TOF mass spectrometer from the direct transfer of fresh bacteria colonies (24-hour old cultures). Each isolate was tested in duplicate (**see section 2.1.4**). For co-phenotypic analysis, 6 CF pathogens were selected from E. Mahenthiralingam CU culture collection: *A. xylosoxidans* LMG 1863, *B. multivorans* ATCC 17616, *C. albicans* SC5314, *P. aeruginosa* PA14, *P. aeruginosa* LES-B58, and *S. aureus* NCTC 12981. These strains were also incubated for 18 h on TSA (Table 2.3).

6.2.2 Genomic DNA extraction

Total bacteria biomass was scraped from agar plates and transferred into 1X sterile phosphate buffered saline (PBS). The bacteria suspension was homogenised and subjected to 4000 rpm centrifugation for 10 minutes. Each sample was resuspended in 500 µl guanidine isothiocyanate and 500 µl of 1X sterile PBS to fully emulsify samples following homogenisation. 400 µl of the resulting supernatant was loaded in the Promega Maxwell 16 platform and DNA was extracted following the DNA tissue protocol (**see section 2.3.2**). Extracted gDNA was quantified with the Invitrogen Qubit 2.0 fluorometer using the broad range dsDNA kit prior to library preparation.

6.2.3 Whole genome library preparation and sequencing

DNA normalisation, library preparation and purification were undertaken in the PHW PenGu department. The Nextera XT DNA library preparation kit was selected to generate Illumina whole genome libraries (**see section 2.7.1**). DNA was firstly normalised to 0.3 ng/µl by the Eppendorf epMotion 5075 and diluted DNA was incubated with a tagment mix for 3 minutes. Normalised DNA was quantified with the Invitrogen Qubit 3.0 fluorometer high sensitivity dsDNA kit. Pooling, loading, and sequencing of libraries were undertaken in the Genome Hub, Cardiff University (**see sections 2.7.2-2.7.3**). A final quality check was conducted to check library fragment size using the Agilent TapeStation D1000 and D5000 ScreenTape, yielding 800 bp fragments on average.

6.2.4 Sequence QC, *de novo* assembly and genome annotation

Each paired-end Illumina read was initially QC validated with FastQC (Andrews 2010). Adapter contamination was trimmed with Trim Galore! (Krueger 2021). Read quality was reassessed through Trim Galore! and subsequent validated reads were used in downstream processes. Sequence quality of assembled genomes was checked with Quast (Gurevich et al. 2013). Following removal of adapter contamination and low-quality reads, genomes were assembled into scaffolds *de novo* by Unicycler (Wick et al. 2017). Gene annotation was completed by Prokka (Seemann 2014) in order to calculate the *Achromobacter* pangenome using Roary (Page et al. 2015).

6.2.5 Whole genome *Achromobacter* identification and phylogenomics

Average nucleotide identity analysis (ANI) was employed to confirm species identity (Pritchard 2016) using the BLAST algorithm with the default percentage threshold for species boundary ID of 95%. For species assignation, SNP-sites (Page et al. 2016) generated clinical *Achromobacter* spp. SNP alignments with 16 *Achromobacter* GenBank whole genome references (accessed 04/2020) using all core genes identified by Roary. The SNP alignment was then used for the construction of a maximum likelihood (ML) phylogenetic tree using the generalised time-reversible (GTR) model under the gamma rate heterogeneity (gamma) in IQ-TREE with ascertainment bias correction (Nguyen et al. 2015). Additionally, a Roary SNP-sites ML tree was constructed in RAxML (Stamatakis 2014) using a GTR gamma model for all clinical *Achromobacter* isolates for a graphical representation of the *Achromobacter* pangenome. MLST was performed alongside the SNP alignment to identify known and novel sequence types (STs) following the MLST scheme of (Spilker et al. 2012). Shared STs between isolates provided preliminary indication of patient transmission. MLST (Seemann unpublished) was used to identify and extract the MLST sequences from clinical *Achromobacter* WGS data and existing *Achromobacter* MLST sequences deposited in PubMLST.

6.2.6 Gene characterisation, *in silico* antimicrobial resistance detection and clinical minimum inhibitory concentrations

ARIBA (Hunt et al. 2017) with the Comprehensive Antimicrobial Resistance database (CARD) (Jia et al. 2017) was used to provide a baseline for the prediction of AMR genes *in silico*. As *Achromobacter* are understudied, a manual analysis of genes annotated by PROKKA and clustered using Roary was necessary. This work enabled the identification of additional β-lactamases and carbapenemases. This data was then contrasted with antibiotic minimum

inhibitory concentration (MIC) data that was determined through agar plate dilution by the Specialist antimicrobial Chemotherapy Unit (SACU). MICs were interpreted using the CLSI "other non-*Enterobacteriaceae*" guidelines except for colistin breakpoints which were interpreted following *Pseudomonas* guidelines.

A protein alignment of the intrinsic *bla*OXA genes was created in MEGA7 (Kumar et al. 2016) using the ClustalW algorithm to compare the oxacillinase sequences of the CF *Achromobacter* isolates and all available GenBank *Achromobacter* oxacillinase variants (accessed 05/2020). A phylogenetic tree based on translated amino acid sequences was generated in RAxML using the protein GTR Gamma model; the resulting phylogeny was used to search for novel oxacillinase variants and correlate a cause for variable MICs.

6.2.7 *Achromobacter* transmission within the CF population

Due to the higher recovery of *A. insuavis* CF isolates sharing single STs compared to *A. xylosoxidans* and *A. ruhlandii*, its mode of transmission was explored by using Snippy (Seemann 2018b) to calculate SNP distances against an in-group *A. insuavis* reference belonging to each clade. These references were selected based on the lowest number of contigs and highest n50 values in each clade. Using the output data, SNP distances between and within groups were compared to determine whether SNP differences were significantly large to suggest environmental transmission or small enough to suggest recrudescence.

6.2.8 CF *Achromobacter* phenotypic testing of growth rates

Growth curves for 6 CF *Achromobacter* isolates were generated using the Bioscreen C instrument (Labsystems) following the methodology in **section 2.8.1** (Table 2.2). Following inoculation with ~10^6 CFU/ml of these *Achromobacter* isolates, 200 µl cultures were grown at 20°C for 5 days, 37°C for 2 days, and 41°C for 3 days. The growth of these 6 *Achromobacter* isolates were also tested in 1:1 ratio coculture with 6 CF pathogens at 37°C (Table 2.3). Two biological replicates of growth data including 4 technical replicates of single culture data was generated while 3 technical replicates were achieved for coculture assays.

6.2.9 *Achromobacter* motility

Bacteria swimming and swarming motility capacity was quantitively assessed on 6 *Achromobacter* isolates and 6 reference CF pathogens in pure culture and 1:1 ratio coculture (modified from Weiser 2015; **see section 2.8.3;** Table 2.2; Table 2.3). Isolates were first standardised to ~10^7 CFU/ml prior to inoculating on respective plates (**see section 2.8.3**). All plates were incubated for 48-h at 37°C when at 24-hour intervals, motility was measured across 4 radii using a ruler. Isolates were classified as non-motile (radius ≤5 mm), motile

(radius >5 mm and ≤30 mm) and highly motile (radius >30 mm; definitions modified from Weiser 2015). Two biological replicates and 3 technical replicates were generated per experiment and per strain/interaction.

In cases where motile mixed pathogens exhibited moderate to high motility, a RISA PCR (**sections 2.3.1 and 2.4**) was conducted on 48-hour old plates to profile which of the interacting species were present at the conclusion of motility interactions. In addition, a 24-hour timepoint was measured on *P. aeruginosa* PA14 mixed cultures to account for its rapid growth and exopolysaccharide production (**see section 2.8.6**). Samples were taken from the active edge of bacteria growth by scraping a small portion of bacteria biomass from a sterile 200 µl pipette tip and afterwards, gDNA was extracted using the Chelex® method followed by RISA PCR amplification (Collins 2019; **sections 2.3.1 and 2.4**). Three replicates were chosen at random for each plate/interaction. These resulting amplicons were visualised using a HS DNA cartridge using the Qiagen QIAxcel (**see 2.5.3**) where both DNA intergenic transcriber spacer (ITS) size/s and concentration were noted. Pure culture controls were also RISA PCR amplified and ran through the QIAxcel to predict mixed sample ITS profiles.

6.2.10 *Achromobacter* biofilm formation

A 96-well polystyrene plate/crystal violet stain methodology was used for biofilm quantitation (modified from O'Toole and Kolter 1998). 100 µl of ~10^5 CFU/ml bacteria cultures were used to quantify biofilm formation of 6 *Achromobacter* isolates and 6 CF microorganisms in pure culture and in 1:1 ratio mixed cultures (Table 2.2; Table 2.3; **see section 2.8.4;** Samples were homogenised by vortexing before transferring 100 µl of inoculum into plate wells. Plates were incubated statically at 24-h and 48-h. Methodologies for plate washing, staining, and optical density (OD) measurements of 0.1% w/v crystal violet (CV) can be followed in **section 2.8.4.** This assay generated 3 biological and 6 technical replicates.

To investigate the ratio of pathogen DNA present after 24-h or also at 48-h where significant differences in OD measurements occurred, 10 µl of mixed culture inoculum was transferred for Chelex® DNA extraction from separate 96 well plates (modified from Collins 2019; **see section 2.3.1**) and the resulting gDNA was RISA PCR amplified (**see section 2.4; see section 2.8.7**). Three biological and 3 technical replicates were generated where 3 replicates were chosen at random to be visualised using the Qiagen QIAxcel HS DNA cartridge (**see 2.5.3**). Pure culture controls were also generated to predict mixed sample ITS profiles.

6.2.11 Protease production

Protease production was examined in 6 clinical *Achromobacter* species and 6 CF pathogens (Table 2.2; Table 2.3; Ash and Ramteke 2018; **see section 2.8.5**). 10 µl of ~10^7 CFU/ml inoculum was spotted onto 1% skimmed milk agar as pure culture isolates or as a 1:1 mixed pathogen ratio. Plates were incubated at 37°C for 24-h and 48-h. The resulting zones, if any, were measured across 4 radii.

6.2.12 Data analysis and figure generation

Growth rates (μ), in addition to carrying capacity (k), doubling time (g), and maximum OD was estimated through the R Growthcurver package (Sprouffske and Wagner 2016), and lag time (λ) was determined in Grofit (Kahm et al. 2010) (**see section 2.8.2**). Predicted maximum OD, root mean square error (RMSE), and R^2 were also calculated to reflect goodness of fit for each logistic regression model. μ rates were statistically analysed with a Kruskal-Wallice test which was analysed *post-hoc* using the Dunn test and an aligned Rank Transform ANOVA with Benjamini-Hochberg correction for multiple comparisons.

Linear mixed effect regression models were constructed through R package nlme (Pinheiro et al. 2013) to compare clinical *Achromobacter* spp. and reference CF microorganism biofilm forming capacity, swimming motility, swarming motility, in addition to protease production. For each model, sample replicate and time were used as random terms while a random error term was included in both motility and protease regression models to account for repeated measures. Mixed culture motility was compared to single culture readings through regression, and R code for these motility regression models was kindly provided by Dr Sarah Christofides. A Kruskal-Wallice test using a Dunn *post hoc* with Benjamini-Hochberg correction was used to compare single culture swimming motility to each other while an ANOVA with a Tukey *post hoc* was employed to compare single culture swarming motility. These comparisons, and the comparisons between 24-h and 48-h reads had to be dropped from each motility regression model after violating the assumptions of linear regression. Biofilm regression model used the nlme constant variance function varIdent (Pinheiro et al. 2022) to consider unequal variances in this regression models affecting bacteria strain and time. A Dunnett's *post-hoc* test was used for the protease model while Tukey's test was used for biofilms model. All figures were generated in R ggplot2 package (R-Core-Team 2020; Wickham 2016) unless stated otherwise.

6.3 Results

6.3.1 Genome summary

The *Achromobacter* genomes ranged in size from 6106634 bp to 7034037 bp with a mean of 6531215 bp. The average GC% content ranged from 67.36% and 68.36%, with a mean of 67.9%. *A. insuavis* generally exhibited larger genomes and higher GC% content than the other 2 species (SI Table E1). The number of predicted coding sequences (CDS) ranged from 5483 to 6305 with a mean of 5893. The number of predicted RNA (including rRNA, tmRNA, and tRNA) genes ranged from 57 to 79 with a mean of 69. *A. insuavis also* possessed comparatively more RNA genes (SI Table E1).

6.3.2 CF *Achromobacter* pangenome and average nucleotide identity

All CF *Achromobacter* isolates shared a small core genome of 1874 core genes (shared between 99-100% isolates) and accounting for 11.9% of the pangenome (15778 genes). 6862 genes were shared between 15-95% of isolates and the remaining 6536 of the total were shared between <15% of isolates. The diversity of the *Achromobacter* isolates were further emphasised through initial phylogenetic construction of its core genome. The resulting phylogeny suggested the presence of 3 distinct *Achromobacter* species among a cohort of 16 patients (SI Figure E1). Furthermore, there appeared to be clade specific presence and absence of genes based on the complete pangenome dataset (SI Figure E1). ANI determined each strain sharing 91-92% sequence similarity (Figure 6.1). In contrast, MALDI-TOF failed to identify the CF *Achromobacter* isolates to the species level; 26/28 were confidently assigned to *A. xylosoxidans*, and the remaining 2 were assigned to *A. xylosoxidans* with lower confidence (SI Table E2).

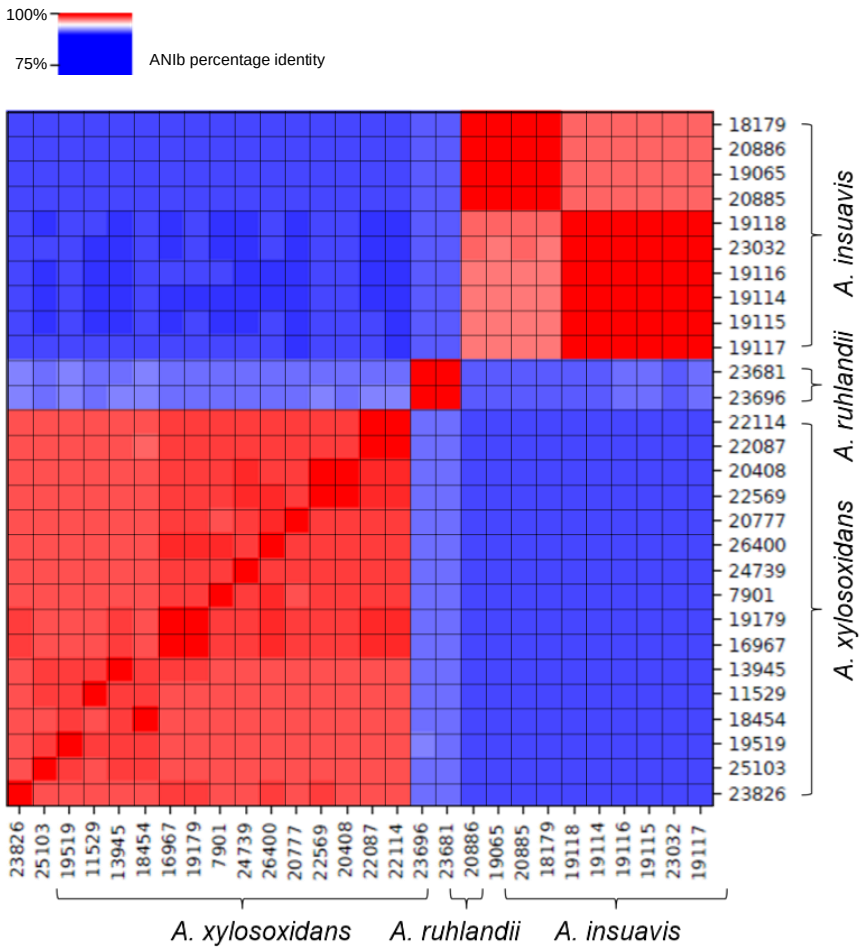


Figure 6.1 Average nucleotide identity analysis of CF *Achromobacter* isolates. Pairwise genome alignments were conducted between all clinical *Achromobacter* strains using the BLAST algorithm, and the percent nucleotide identity of all matching regions between each alignment is illustrated in each bordered cell where the x and y axes join. The legend bar denotes percent nucleotide identity where isolates sharing >95% sequence identity are shaded in red and are considered the same species, while blue cells suggest distinct species, sharing <95% sequence identity. Each axis is labelled to note strain number and species identity through WGS. Labels were edited in Inkscape and this heatmap was generated with the pyANI Python script (Pritchard 2018).

6.3.3 *Achromobacter* whole genome species assignation and multilocus sequence analysis

A whole core genome phylogeny was constructed from 425 core genes identified between the 28 clinical *Achromobacter* isolates and 16 reference *Achromobacter* species to achieve objective 1. As a result, the largest clade of 16 isolates clustered with *A. xylosoxidans* reference genomes, 10 isolates clustered with 2 distinct *A. insuavis* clades, and the remaining 2 *Achromobacter* isolates clustered with *A. ruhlandii* reference genomes (Figure 6.2). *A. xylosoxidans* was therefore concluded to be most frequently recovered in this study, followed

by *A. insuavis* then *A. ruhlandii.* Most STs provided partial matches to existing *Achromobacter* MLST references and were consequently unknown (Table 6.1). Nevertheless, isolates with identical allelic profiles were found in patients colonised with *A. insuavis* (patients B and C), *A. ruhlandii* (patient F) and *A. xylosoxidans* (patients A, D and E) which could suggest reinfection (Table 6.1; SI Table E3) This notion was further supported by all *A. insuavis* and *A. ruhlandii* isolates deriving from a single hospital, whereas *A. xylosoxidans* was more widespread (Table 6.1). Furthermore, MLSA profiles between these patients were all unique, suggesting no patient cross-infection. Patients from hospital 1 were colonised with 2 *A. xylosoxidans* isolates sharing the same sequence type/allelic profile suggesting a common source of infection (Table 6.1; SI Table E3).

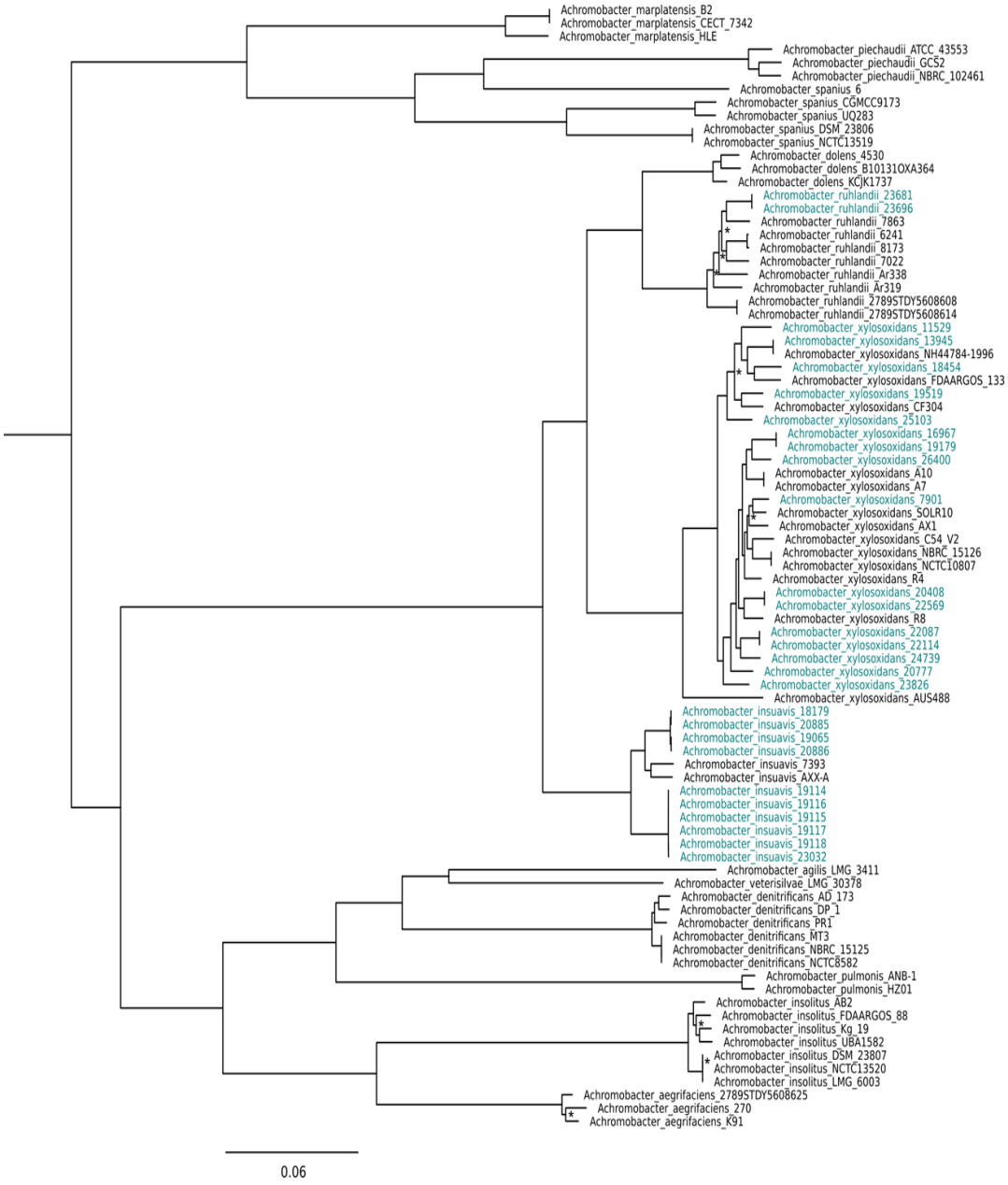


Figure 6.2 Population biology of the clinical and reference *Achromobacter* isolates. A core gene phylogenomic tree is presented to show the population biology represented in the collection. The scale bar represents the number of base substitutions per site. The root was moved to its midpoint in FigTree and node confidence (bootstrap) is shown as an Asterix where below 80. CF isolates from this study are highlighted in blue.

Table 6.1 *Achromobacter* spp. molecular epidemiology summary

Clinical ID	Genomic ID	Size (bp)	Location	Patient[a]	ST
18179	*A. insuavis*	6545660	Hospital 1	B	215~
19065	*A. insuavis*	6676555	Hospital 1	B	215~
19114	*A. insuavis*	6956129	Hospital 1	C	304~
19115	*A. insuavis*	7034037	Hospital 1	C	304~
19116	*A. insuavis*	6784170	Hospital 1	C	304~
19117	*A. insuavis*	6980565	Hospital 1	C	304~
19118	*A. insuavis*	6953259	Hospital 1	C	304~
20885	*A. insuavis*	6635970	Hospital 1	B	215~
20886	*A. insuavis*	6622380	Hospital 1	B	215~
23032	*A. insuavis*	6954773	Hospital 1	C	304~
23681	*A. ruhlandii*	6106634	Hospital 2	F	265~
23696	*A. ruhlandii*	6226266	Hospital 2	F	265~
7901	*A. xylosoxidans*	6305324	Hospital 5	U	180
11529	*A. xylosoxidans*	6213766	Study	U	24~
13945	*A. xylosoxidans*	6561758	Hospital 2	U	28~
16967	*A. xylosoxidans*	6178511	Hospital 1	A	184~
18454	*A. xylosoxidans*	6760908	Hospital 3	U	27
19179	*A. xylosoxidans*	6164203	Hospital 1	A	184~
19519	*A. xylosoxidans*	6261297	Hospital 1	U	176~
20408	*A. xylosoxidans*	6327736	Hospital 1	D	237~
20777	*A. xylosoxidans*	6510956	Hospital 1	U	184~
22087	*A. xylosoxidans*	6735159	Hospital 1	E	182
22114	*A. xylosoxidans*	6633523	Hospital 1	E	182
22569	*A. xylosoxidans*	6238863	Hospital 1	D	237~
23826	*A. xylosoxidans*	6287573	Hospital 4	U	27~
24739	*A. xylosoxidans*	6357181	Hospital 1	U	290
25103	*A. xylosoxidans*	6477026	Hospital 1	U	311~
26400	*A. xylosoxidans*	6383850	Hospital 2	U	12~

[a]Isolates recovered once from an individual patient is noted as unique "U".

6.3.4 *A. insuavis* mode of transmission

A higher frequency of culture-positive *Achromobacter* was observed in patients B and C compared to other patients. This was later identified as *A. insuavis* (Figure 6.2). SNP distances were compared between isolates belonging to different patients to understand whether transmission between patients or a single source of infection was involved (objective 2). In concordance to MLSA data, there was no evidence of patient-patient transmission due to the large number of SNPs (77302 SNPs) differing between the references recovered from patients B and C (Figure 6.3). Isolates from patient C were less diverse than those found in patient B, differing by a few hundred SNPs compared to those in the thousands (Figure 6.3). In both cases, the SNP distances were too large to suggest recrudescence, implicating acquisition from a common environmental source (Figure 6.3).

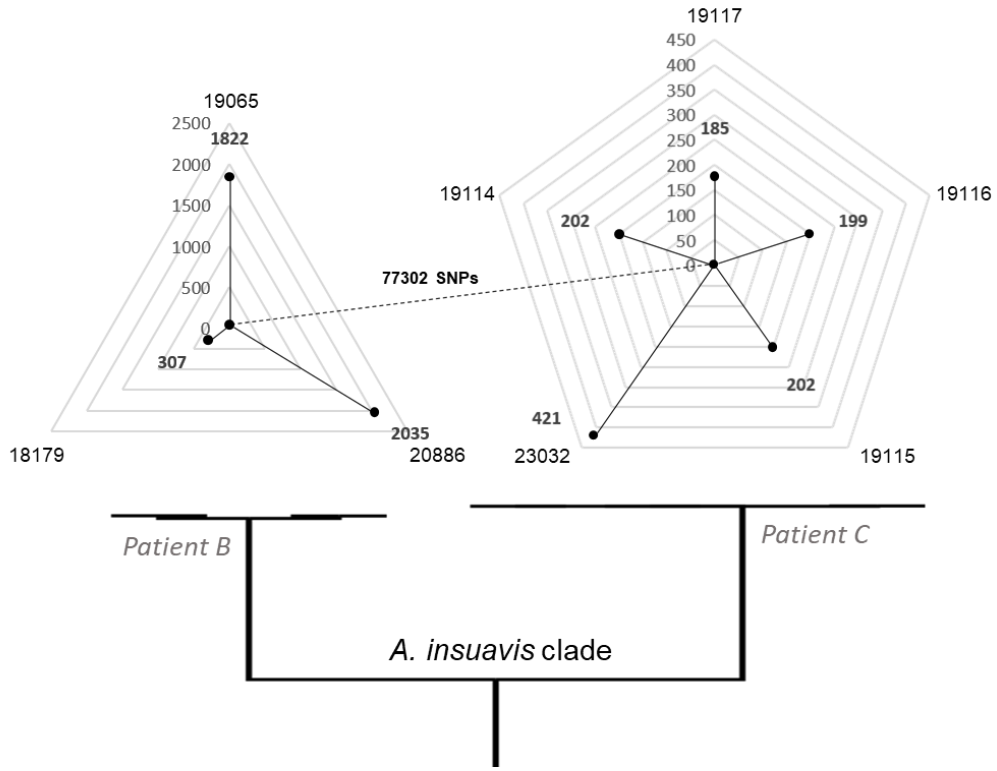


Figure 6.3 Transmission of *Achromobacter insuavis* from a common environmental source. SNP distances between and within the two *A. insuavis* groups are shown on this modified radar chart. Beneath the plots, a section of the *Achromobacter* core genome SNP-sites phylogeny has been included to mark each clade. Along the Y axis of each plot is a scale of the SNP distances relative to the in-group *A. insuavis* references. The exact SNP differences within each group have been labelled with the corresponding isolate labelled around the outside of the plot.

6.3.5 *Achromobacter* spp. antimicrobial resistance

Antibiotic MICs varied per antimicrobial per species; *A. xylosoxidans* possessed the highest recorded MICs for ceftazidime (CAZ), imipenem (IPM), meropenem (MEM) and piperacillin/tazobactam (TZP), despite intraspecific variability (Table 6.2). Few isolates belonging to *A. xylosoxidans* and *A. insuavis* held highest MICs for colistin (CST) and co-trimoxazole (TMP/SMX). *A. ruhlandii* held comparably high MICs for IPM and MEM, and lower MICs to CAZ, CST, TMP/SMX and TZP. *A. insuavis* generally displayed the lowest MICs except for CST, TZP, and 1 replicate of TMP/SMX. Overall, *Achromobacter* spp. isolates were most commonly resistant to carbapenems. MIC patterns did not appear to be clade specific (Table 6.2).

To gain further insight into AMR genes that could singularly and/or collectively give rise to these observed MIC patterns, AMR conferring genes were analysed firstly with ARIBA-CARD and afterwards with Roary to achieve objective 3. All isolates possessed the complete *axyXY-oprZ* efflux pump genes. Sixteen *A. xylosoxidans* isolates possessed the $bla_{OXA\text{-}114a}$ *gene* intrinsic to this species, 6 *A. insuavis* isolates belonging to one clade possessed the $bla_{OXA\text{-}243}$ variant and the remaining four belonging to the second clade possessed the $bla_{OXA\text{-}457}$. A. *ruhlandii* possessed its species-specific $bla_{OXA\text{-}258}$ variant. An amino acid alignment based on the CF *Achromobacter* oxacillinase amino acid sequences and reference *Achromobacter* species reflected the results of ARIBA-CARD. This phylogeny enabled greater resolution in identifying subclasses of *Achromobacter* spp. species specific OXA variants which were most diverse in *A. xylosoxidans* (Figure 6.4). While Roary did not identify *Achromobacter*-specific multidrug efflux pump genes or specific oxacillinase variants, it did report various β-lactamase genes. This included *ampC*, *blm* and *blh*, which were widespread in all isolates. *blaP* was detected in 8 *A. xylosoxidans* isolates and both *A. ruhlandii* isolates, and the metallo β-lactamase *ccrA* was specific to 12 *A. xylosoxidans* isolates. Genes for tetracycline resistance (*tetR/tetA*) and macrolide resistance (*macAB*) were also widespread. Genes conforming resistance to sulphonamides (*sul2*), bleomycin (*ble*), and bicyclomycin (*bcr*) were less prevalent (SI Table E4).

Table 6.2 CF *Achromobacter* spp. clinical MICs in µg/ml paired with its resistance profile following CLSI clinical breakpoint data of *P. aeruginosa* and "Other non-*Enterobacteriaceae*"

Isolate	ID	CAZ µg/ml	Resistance[a]	CST µg/ml	Resistance[a]	TMP/SMX µg/ml	Resistance[a]	IPM µg/ml	Resistance[a]	MEM µg/ml	Resistance[a]	TZP µg/ml	Resistance[a]
18179	*A. insuavis*	3	S	>128	R	0.06	S	2	S	0.25	S	2	S
19065	*A. insuavis*	4	S	>4	R	<0.5	S	2	S	3	S	16	S
19114	*A. insuavis*	8	S	1	S	>4	R	≤1	S	≤1	S	≤8	S
19115	*A. insuavis*	8	S	1	S	>16	R	≤1	S	≤1	S	≤8	S
19116	*A. insuavis*	16	I	2	S	4	R	≤1	S	≤1	S	64	I
19117	*A. insuavis*	4	S	1	S	≤0.5	S	≤1	S	≤1	S	≤8	S
19118	*A. insuavis*	8	S	1	S	≤0.5	S	≤1	S	≤1	S	≤8	S
20885	*A. insuavis*	4	S	2	S	0.023	S	1.5	S	0.75	S	3	S
20886	*A. insuavis*	6	S	>128	R	1.5	S	1	S	3	S	16	S
23681	*A. ruhlandii*	8	S	1	S	0.125	S	32	R	>32	R	24	S
23696	*A. ruhlandii*	8	S	1.5	S	0.064	S	>32	R	>32	R	48	S
7901	*A. xylosoxidans*	16	I	2	S	2	S	2	S	>32	R	3	S
15945	*A. xylosoxidans*	16	I	0.5	S	>32	R	>32	R	>32	R	32	S
16967	*A. xylosoxidans*	>256	R	0.25	S	0.032	S	>32	R	>32	R	>256	R
18454	*A. xylosoxidans*	12	S[b]	2	S	>32	R	>32	R	0.5	S	0.75	S
19179	*A. xylosoxidans*	16	I	0.5	S	0.064	S	>32	R	>32	R	>256	R
19519	*A. xylosoxidans*	>16	I[c]	>4	R	≤0.5	S	2	S	0.25	S	4	S
20408	*A. xylosoxidans*	>256	R	>128	R	0.5	S	6	S	>32	R	2	S
20777	*A. xylosoxidans*	6	S	1	S	0.047	S	1.5	S	0.094	S	1.5	S
22087	*A. xylosoxidans*	1.5	S	2	S	0.023	S	>32	R	0.19	S	0.5	S
22114	*A. xylosoxidans*	2	S	2	S	0.023	S	>32	R	0.19	S	0.5	S
22569	*A. xylosoxidans*	>256	R	>128	R	0.38	S	>32	R	>32	R	>256	R
23826	*A. xylosoxidans*	6	S	1	S	>32	R	1	S	6	S	1	S
24739	*A. xylosoxidans*	12	S[b]	1	S	0.047	S	0.75	S	0.125	S	0.5	S
25103	*A. xylosoxidans*	8	S	2	S	0.016	S	1.5	S	0.125	S	0.75	S

[a] Clinical breakpoints validate the concentration of antimicrobials used for antimicrobial susceptibility testing to determine antimicrobial sensitive (S), intermediate (I), and resistant (R) isolates. Caution in its interpretation is needed since no clinical breakpoint data is available for *Achromobacter* spp.

[b] Values being interpreted as sensitive for being below the guideline threshold for intermediate resistance.

[c] Value being interpreted as intermediate resistance as it is over the guideline threshold for intermediate resistance but cannot determine if over the guideline threshold for resistant.

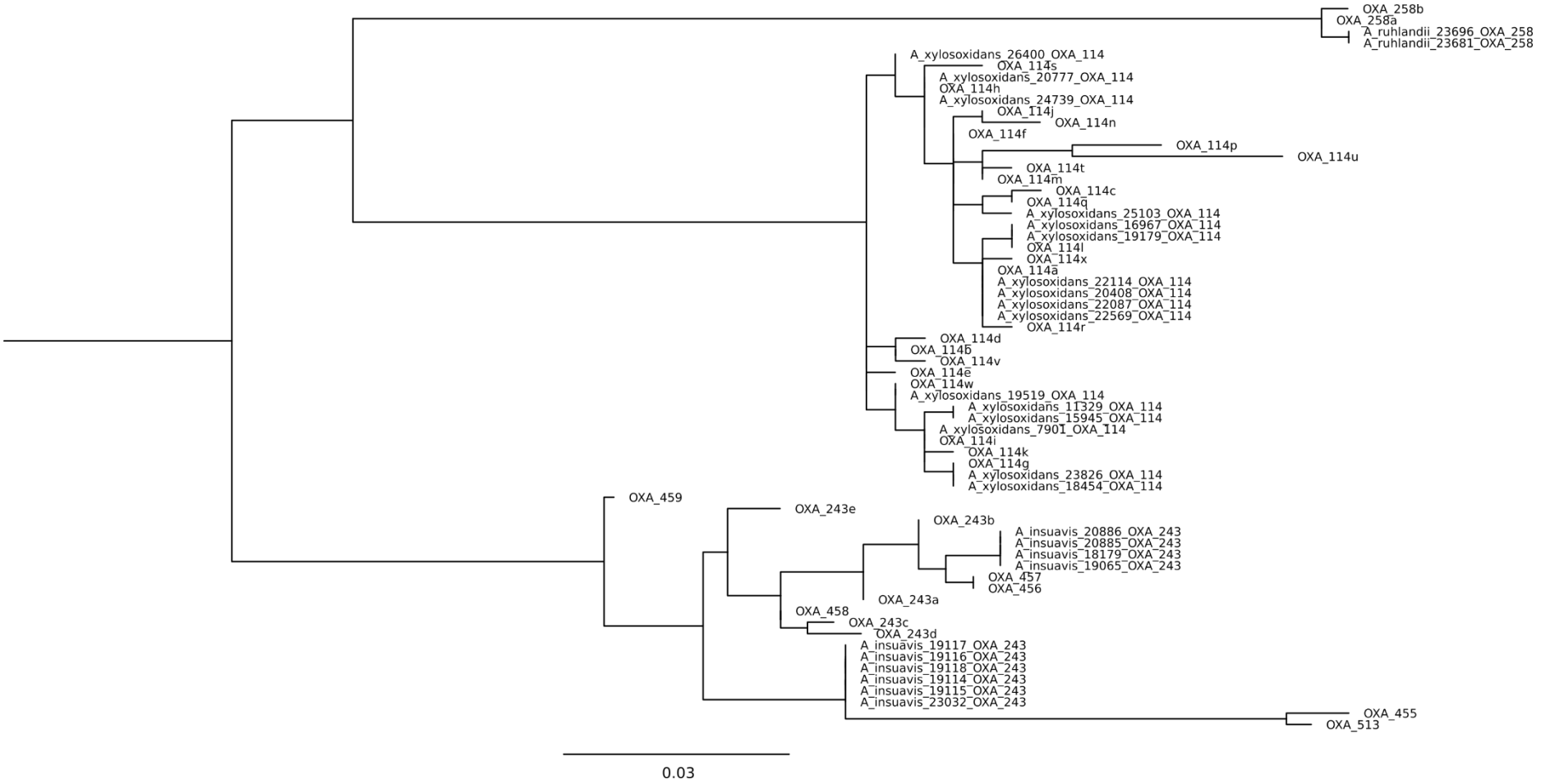


Figure 6.4 CF *Achromobacter* species-specific OXA variants. This maximum likelihood phylogenetic tree was constructed based on oxacillinase variation and was edited in FigTree and Inkscape. The tree was re-rooted by its midpoint and its root lengthened to facilitate viewing. The scale bar indicates the number of substitutions per site. All study-derived strains have been labelled with their sample ID and corresponding oxacillinase variant according to ARIBA-CARD.

6.3.6 Single and mixed CF pathogen growth rates

Pure culture and mixed pathogen (*S. aureus, P. aeruginosa* and *B. multivorans*) growth rates were assessed to infer *Achromobacter* adaptation to the environment and evaluate competition between interacting pathogens (objective 4). All clinical *Achromobacter* spp. exhibited growth at both 20°C and 37°C, while all but *A. xylosoxidans* were inhibited at 41°C (Table 6.3; Figure 6.5; SI Figure E2; SI Table E5). Growth rates (μ) significantly differed between each strain (R^2 0.95, $F_{5,36}$=78.9, p<0.0001) and temperature ($F_{2,36}$=336.1, P<0.0001), typically being lowest at 20°C (SI Table E6). Unusually high μ rates were observed in 1 *A. insuavis* and 1 *A. ruhlandii* strain at 41°C but low maximum OD values suggested sudden growth inhibition (SI Table E5).

At 37°C, μ was comparable across all pure culture clinical *Achromobacter* spp. isolates, *A. xylosoxidans* reference, *P. aeruginosa* PA14, and *P. aeruginosa* LES-B58 (Table 6.3; Figure 6.5). Both *S. aureus* and *B. multivorans* baseline growth rates were notably higher than all other isolates (Table 6.3; Figure 6.5*). μ* rates markedly increased when *B. multivorans* was in coculture with all clinical *Achromobacter isolates* (range 0.5-1.1), and when *S. aureus* was in coculture with *insuavis* and *A. ruhlandii* (range 0.9-1; Table 6.4; Figure 6.5*). μ* of other coculture combinations were comparable to both clinical *Achromobacter* spp. and CF pathogen pure culture references (Table 6.3; Table 6.4; Table 6.5). Although there was a significant effect overall (X^2=74.7, df=40, P=0.0007), a *post hoc* analysis found no significant pairwise differences in these growth rates. Notably, *A. insuavis* reached lower cell densities than *A. ruhlandii* and *A. xylosoxidans* at 48-h (Table 6.3). Bacteria lag phases (λ) were significantly affected by bacteria strain and interaction (R^2 0.6, $F_{40,82}$=3.1, p<0.0001) but again the pairwise comparisons were insignificant.

Table 6.3 Pure culture CF *Achromobacter* and CF pathogen panel 37°C growth parameters

Species	Strain	Lag phase (λ) in hours	Carrying capacity (k)	Growth rate (μ) in hours	Doubling time (g) in hours	Maximum predicted OD	Observed maximum OD	RMSE	R^2
A. insuavis	19065	7..6	0.5	0.1	0.2	0.4	0.4	0.01	0.99
A. insuavis	23032	5.7	0.5	0.2	0.2	0.5	0.6	0.03	0.99
A. ruhlandii	23681	2.2	0.9	0.2	0.2	0.9	0.8	0.03	0.99
A. ruhlandii	23696	0.6	1	0.1	0.2	1	1.1	0.04	0.98
A. xylosoxidans	18454	3	1.3	0.2	0.2	1.3	1.4	0.05	0.99
A. xylosoxidans	20777	3.4	1.1	0.2	0.2	1.1	1.1	0.05	0.98
A. xylosoxidans	LMG 1863	7	1.3	0.1	4.9	1.3	1.3	0.03	0.99
B. multivorans	ATCC 17616	5.7	1.2	0.6	1.2	1.2	1.2	0.05	0.98
P. aeruginosa	PA14	3.5	1	0.3	2.5	1	1.1	0.07	0.97
P. aeruginosa	LES-B58	4.3	1.1	0.2	4	1.1	1.1	0.03	1
S. aureus	NCTC 12981	1.2	0.9	0.4	1.6	0.9	1.1	0.1	0.95

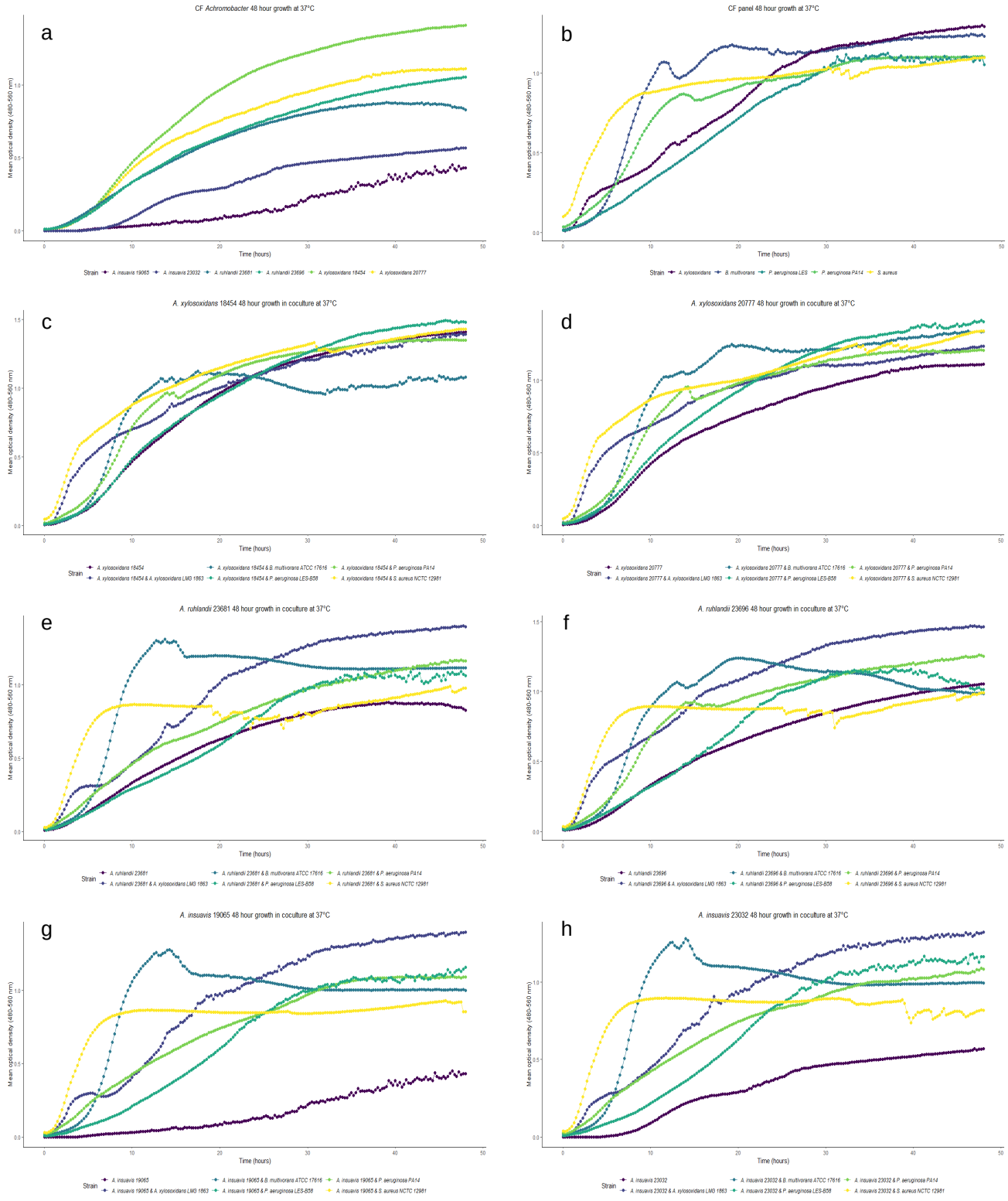


Figure 6.5 Single and coculture clinical *Achromobacter* spp. and secondary pathogen growth curves. Sample OD (480 nm-560 nm) was plotted at every 15-minute interval over 48 hours for pure culture clinical *Achromobacter* and reference CF isolates (a, b), clinical *A. xylosoxidans* in coculture (c, d), clinical *A. ruhlandii* in coculture (e, f), and *A. insuavis* in coculture (g, h).

Table 6.4 Growth parameters for the mixed culture CF *Achromobacter* and CF pathogen panel

Strain	Interaction	Lag time (*λ*) in hours	Carrying capacity (*k*)[a]	Growth rate (*μ*) in hours	Doubling time (*g*) in hours	Maximum predicted OD	Observed maximum OD	RMSE	R^2
A. insuavis 19065	*A. xylosoxidans* LMG1863	6	1.4	0.2	4.3	1.3	1.4	0.04	0.99
A. insuavis 19065	*B. multivorans* ATCC 17616	5.6	1.1	1	0.7	1.1	1	0.07	0.96
A. insuavis 19065	*P. aeruginosa* PA14	1.4	1.1	0.1	4.9	1.1	1.1	0.04	0.99
A. insuavis 19065	*P. aeruginosa* LES-B58	7.5	1.1	0.2	3.8	1.1	1.2	0.02	1
A. insuavis 19065	*S. aureus* NCTC 12981	1.2	0.8	0.9	0.8	0.8	0.9	0.04	0.98
A. insuavis 23032	*A. xylosoxidans* LMG1863	5.1	1.3	0.2	4.3	1.3	1.3	0.03	0.99
A. insuavis 23032	*B. multivorans* ATCC 17616	5.5	1	1.1	0.7	1	1	0.08	0.95
A. insuavis 23032	*P. aeruginosa* PA14	1.6	1	0.1	4.6	1	1.1	0.04	0.99
A. insuavis 23032	*P. aeruginosa* LES-B58	7.1	1.1	0.2	3.9	1.1	1.2	0.02	1
A. insuavis 23032	*S. aureus* NCTC 12981	1.1	0.8	1.0	0.7	0.8	0.8	0.04	0.97
A. ruhlandii 23681	*A. xylosoxidans* LMG1863	4.9	1.4	0.2	4.3	1.4	1.4	0.03	1
A. ruhlandii 23681	*B. multivorans* ATCC 17616	5.3	1.1	0.9	0.8	1.1	1.1	0.06	0.98
A. ruhlandii 23681	*P. aeruginosa* PA14	0.9	1.1	0.1	5.3	1.1	1.2	0.05	0.98
A. ruhlandii 23681	*P. aeruginosa* LES-B58	6.5	1.1	0.2	4.4	1.1	1.1	0.03	0.99
A. ruhlandii 23681	*S. aureus* NCTC 12981	1.1	0.8	1.0	0.7	0.8	1	0.06	0.91
A. ruhlandii 23696	*A. xylosoxidans* LMG1863	3.1	1.4	0.1	4.9	1.4	1.5	0.06	0.98
A. ruhlandii 23696	*B. multivorans* ATCC 17616	6.9	1.1	0.6	1.2	1.1	1	0.07	0.96
A. ruhlandii 23696	*P. aeruginosa* PA14	3.4	1.1	0.2	2.9	1.1	1.3	0.08	0.96
A. ruhlandii 23696	*P. aeruginosa* LES-B58	7.5	1.1	0.2	3.5	1.1	1	0.04	0.99
A. ruhlandii 23696	*S. aureus* NCTC 12981	1.1	0.9	0.9	0.7	1	1	0.05	0.96
A. xylosoxidans 18454	*A. xylosoxidans* LMG1863	2.5	1.3	0.1	5.2	1.3	1.4	0.07	0.97
A. xylosoxidans 18454	*B. multivorans* ATCC 17616	4.5	1	0.7	1	1	1.1	0.04	0.99
A. xylosoxidans 18454	*P. aeruginosa* PA14	3.4	1.3	0.2	2.9	1.3	1.3	0.06	0.98
A. xylosoxidans 18454	*P. aeruginosa* LES-B58	2.4	1.4	0.2	4.4	1.4	1.5	0.06	0.99
A. xylosoxidans 18454	*S. aureus* NCTC 12981	0.6	1.3	0.2	4.1	1.3	1.4	0.09	0.95
A. xylosoxidans 20777	*A. xylosoxidans* LMG1863	2.5	1.1	0.2	4.3	1.1	1.2	0.06	0.96
A. xylosoxidans 20777	*B. multivorans* ATCC 17616	6.6	1.2	0.5	1.5	1.2	1.3	0.05	0.98
A. xylosoxidans 20777	*P. aeruginosa* PA14	3.1	1.1	0.3	2.6	1.1	1.2	0.07	0.97
A. xylosoxidans 20777	*P. aeruginosa* LES-B58	2.4	1.4	0.2	4.2	1.3	1.4	0.05	0.99
A. xylosoxidans 20777	*S. aureus* NCTC 12981	0.9	1.2	0.1	4.7	1.2	1.3	0.1	0.91

[a]Carrying capacity units are related to optical density units.

6.3.7 CF pathogen swimming and swarming motility

Six *Achromobacter* isolates and 6 reference CF microorganisms were assayed for swimming and swarming motility in pure culture and coculture to estimate clinical *Achromobacter* spp. competitive fitness and observe antagonism in a low-viscosity culture environment. Motility was also measured to better understand *Achromobacter* spp. adaptation to the lung environment (objective 4). Complete pure culture motility assays and resulting motile mixed pathogen cultures are shown in SI Figure E3 and SI Figure E4. Notable examples of motility have been included in Figure 6.6 and Figure 6.7. Diameters of bacterial growth are shown in SI Table E7 and significant differences between relevant mixed culture interactions and pure culture readings are shown in SI Table E8.

Clinical *Achromobacter* spp. strains exhibited variable swimming and swarming patterns: all *A. insuavis* were nonmotile, *A. xylosoxidans* was motile and *A. ruhlandii* exhibited both motile and non-motile phenotypes after 48-h (SI Figure E3). Overall, swimming (X^2 498.2, df=11, P<0.0001) and swarming (R^2 0.7, $F_{11,564}$=130.3, P<0.0001) motility significantly differed between strains. Both *A. xylosoxidans* strains demonstrated significantly higher swimming motility than the swimming-positive *A. ruhlandii* isolate (P<0.0001). *B. multivorans* demonstrated swimming motility comparable to clinical *A. xylosoxidans* strains while *P. aeruginosa* PA14 exhibited the highest degree of motility above all other strains (P<0.0001; Figure 6.6). Only *B. multivorans* and *P. aeruginosa* PA14 were capable of true swarming motility; in particular, *P. aeruginosa* PA14 showed a significantly higher degree of swarming compared to *B. multivorans* (P<0.0001) and generated distinct finger-like projections on LB agar (Figure 6.6). Although the diameter of *A. ruhlandii* colonies was classified as motile, true swarming motility was considered in this case, but instead extensive spread due to its mucoid phenotype.

Overall, swimming (R^2 0.9, X^2 517.8, df=25, P<0.0001) and swarming (R^2 0.9, X^2 814.3, df=25, P<0.0001) motility significantly differed between mixed pathogen interactions compared to clinical *Achromobacter* spp. pure culture motility. All *Achromobacter* spp. swimming and swarming interactions with *P. aeruginosa* PA14 resulted in significantly higher motility compared to clinical *Achromobacter* spp. pure culture (P<0.0001; SI Table E8; SI Figure E4). Only one *A. xylosoxidans* strain co-migrated with swarming *P. aeruginosa* PA14 cultures (SI Table E10). Two ITS amplicons were amplified from *A. xylosoxidans* and *P. aeruginosa* PA14 swimming and the latter swarming coculture which matched to the 2 species because the ratio of *A. xylosoxidans* and *P. aeruginosa* PA14 amplified DNA was relatively even, this suggested neither species inhibited the growth of the other during the motility interactions (SI Figure E5; SI Figure E6; SI Table E9; SI Table E10).

B. multivorans and *Achromobacter* spp. mixed culture *s*wimming and swarming motility was generally significantly higher than *Achromobacter* spp. pure culture values. *A. ruhlandii* appeared to have delayed the growth of *B. multivorans* though *A. ruhlandii* was not detected through RISA PCR (SI Figure E6). All remaining combinations of *A. xylosoxidans* mixed cultures with few exceptions yielded results similar to *A. xylosoxidans* pure culture motility (SI Table E7). Notably, 1 *A. xylosoxidans* strain exhibited significantly higher swimming and swarming motility with *B. multivorans* (P=0.01) and *C. albicans* (P=0.0004) akin to *P. aeruginosa* PA14 motility in pure culture (SI Table E8). In this instance, atypical motility was observed with *C. albicans* mixed cultures by generating irregular projections similar to *P. aeruginosa* PA14 (Figure 6.6). While *A. xylosoxidans* RISA amplified DNA was detected with *C. albicans* mixed swimming and swarming cultures, it was not detected with *B. multivorans* cocultures (SI Figure E5; SI Figure E6; SI Table E9; SI Table E10). *A. xylosoxidans* swarming interactions also significantly increased in *S. aureus* and *B. multivorans* coculture compared to *A. xylosoxidans* pure culture readings (SI Table E8; Table E7).

Swimming and swarming motility patterns of non-motile *Achromobacter* species in coculture suggested these plates primarily consisted of secondary CF pathogens they were mixed with, especially when these differences were significant to pure culture readings (SI Table E7). Moreover, *A. insuavis* and *A. ruhlandii* specific ITS fragments did not amplify at the growth edge, supporting this notion. (SI Figure E5; SI Figure E6; SI Table E9; SI Table E10).

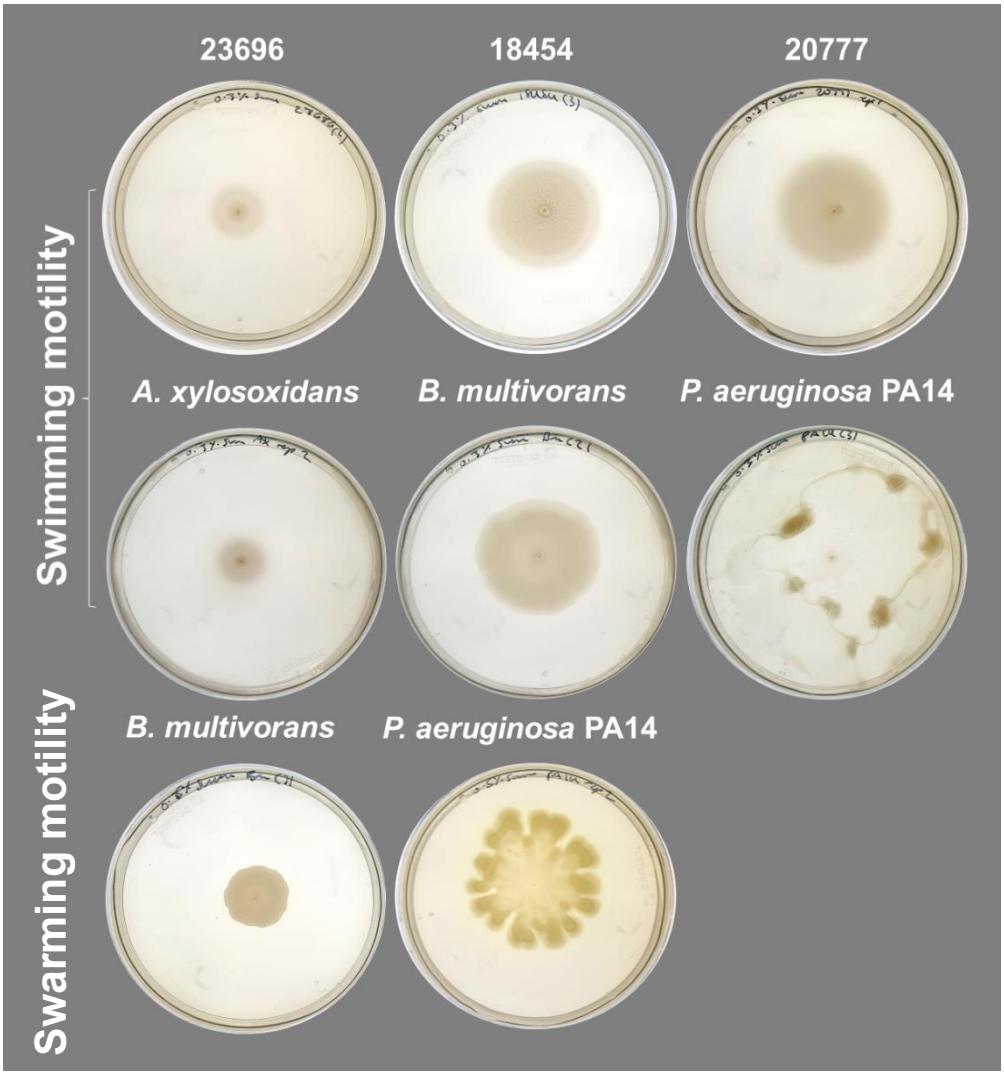


Figure 6.6 Swimming and swarming motility in clinical *Achromobacter* species and CF pathogens in pure axenic culture. Images have been cropped and corrections have been made based on image brightness and saturation to facilitate viewing. Sample labels that are numbered only corresponds to clinical *Achromobacter* isolates.

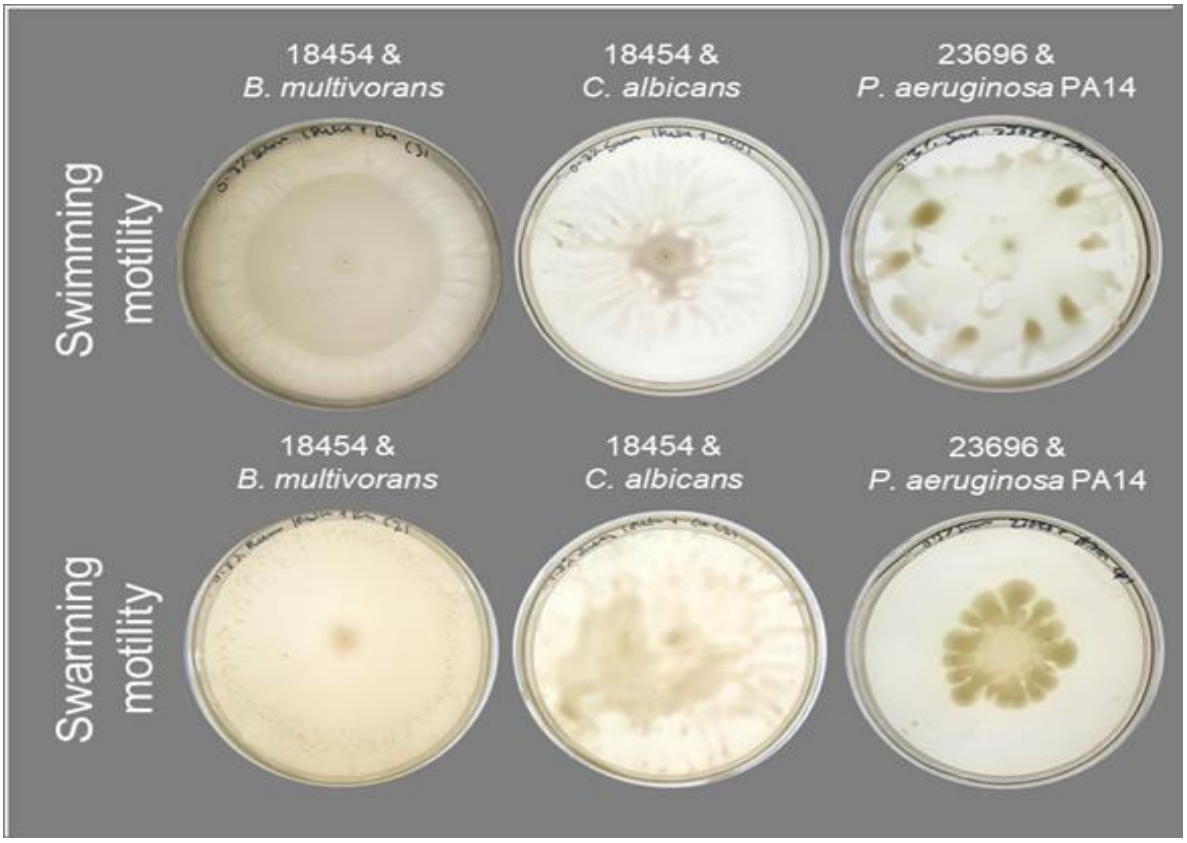


Figure 6.7 Examples of motile *Achromobacter* mixed culture assays. Images have been cropped and corrections have been made based on image brightness and saturation to facilitate viewing.

6.3.8 Pure and mixed culture biofilm production

For its role in persistent lung infection and AMR, *Achromobacter* biofilms were measured and compared to major CF pathogens, and interactions were noted (objective 4). Biofilm production observed via crystal violet staining significantly differed between bacterial strains (R^2 0.98, $F_{41,168}$=37.8, P<0.0001). At 24-h, *A. insuavis* and *A. ruhlandii* produced minimal biomass, comparable to *S. aureus* (Figure 6.8). Optical density readings of *C. albicans* and *P. aeruginosa* LES-B58 were significantly higher than *A. insuavis* and *A. ruhlandii* (SI Table E11). OD measurements were more variable in *A. xylosoxidans* isolates. The 2 clinical *A. xylosoxidans* strains exhibited a higher biofilm forming capacity than *A. insuavis* and *A. ruhlandii,* comparable to both the taxonomic reference *A. xylosoxidans* LMG 1863 and *B. multivorans* isolates (Figure 6.8). Biofilm production was significantly higher in *P. aeruginosa* PA14 compared to any other isolate (P<0.0001; SI Table E11).

At 24-h, all combinations of *A. insuavis* and *A. ruhlandii* mixed cultures yielded similar results to one another (Figure 6.9). These mixed cultures with low biofilm formers reflected low biomass pure culture OD readings while high biofilm formers reflected CF pathogen baseline

results (Figure 6.8; Figure 6.9). Neither *C. albicans* nor *P. aeruginosa* LES-B58 inhibited *A. insuavis* or *A. ruhlandii* growth based on the high proportion of RISA PCR amplified DNA specific to *Achromobacter* spp. in these mixed cultures (SI Figure E7; SI Table E12). Clinical *A. xylosoxidans* mixed culture biofilm production matched its baseline values (Figure 6.8; Figure 6.9). Despite of this result, *B. multivorans* consistently outgrew these *Achromobacter* isolates based on the sole amplification of a *B. multivorans*-specific ITS fragment from these samples (SI Figure E7; SI Table E12).

Biofilm formation also significantly differed over time ($F_{41,168}$=5.7, P<0.0001) due to *P. aeruginosa* PA14 (Figure 6.9; SI Figure E8). Most notably, biofilm quantitation of *P. aeruginosa* PA14 mixed cultures markedly increased after 48-h (P<0.0001) reflecting *P. aeruginosa* PA14 48-h value which also increased significantly from 24-h readings (P<0.0001) (Figure 6.8; SI Figure E8). Despite these results, amplification of both *Achromobacter* spp. and *P. aeruginosa* mixed coculture ITS profiles at 48-h suggested synergism between the (SI Figure E7; SI Table E12).

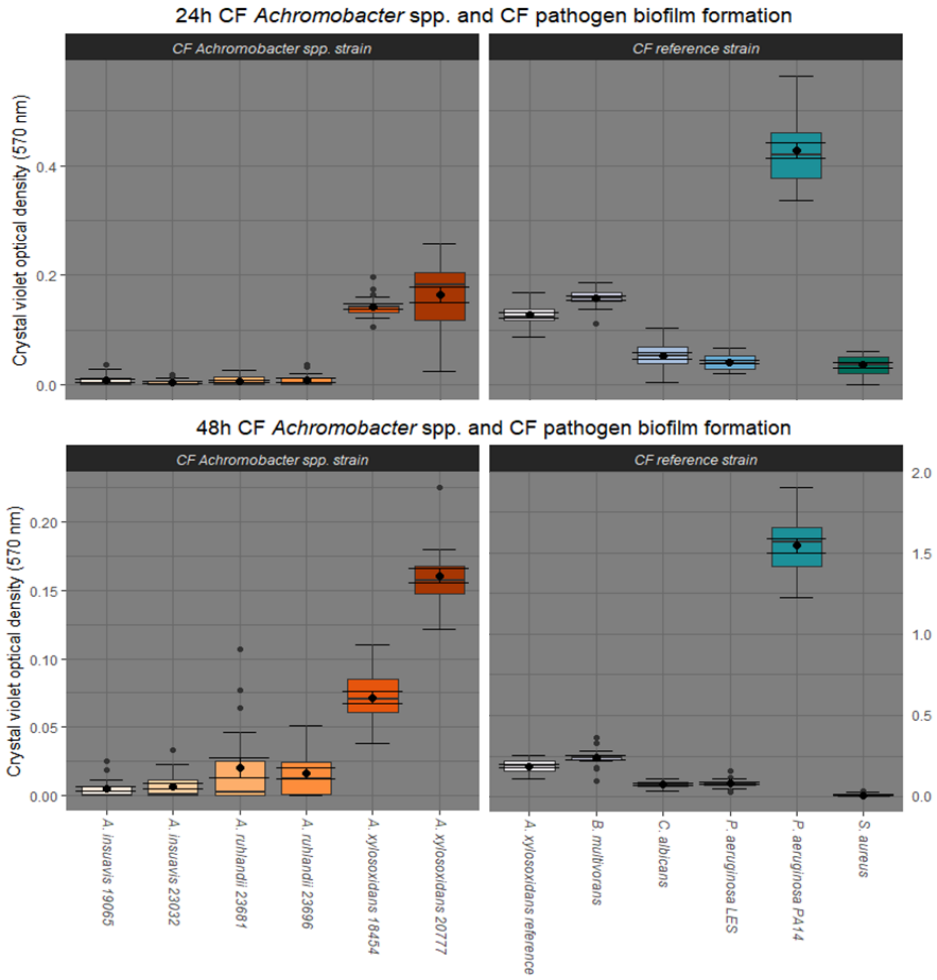


Figure 6.8 Biofilm production in pure *Achromobacter* spp. and secondary CF pathogen cultures. Crystal violet (CV) OD was plotted for 24-h and 48-h data. Standard error bars for each sample have also been included from the centre of each plot. Clinical vs, reference strains were coloured separately using R Color Brewer (Neuwirth 2015). Colours simply define clinical *Achromobacter* strains (oranges) from reference CF strains (blues). The *A. xylosoxidans* reference refers to *A. xylosoxidans* LMG 1863.

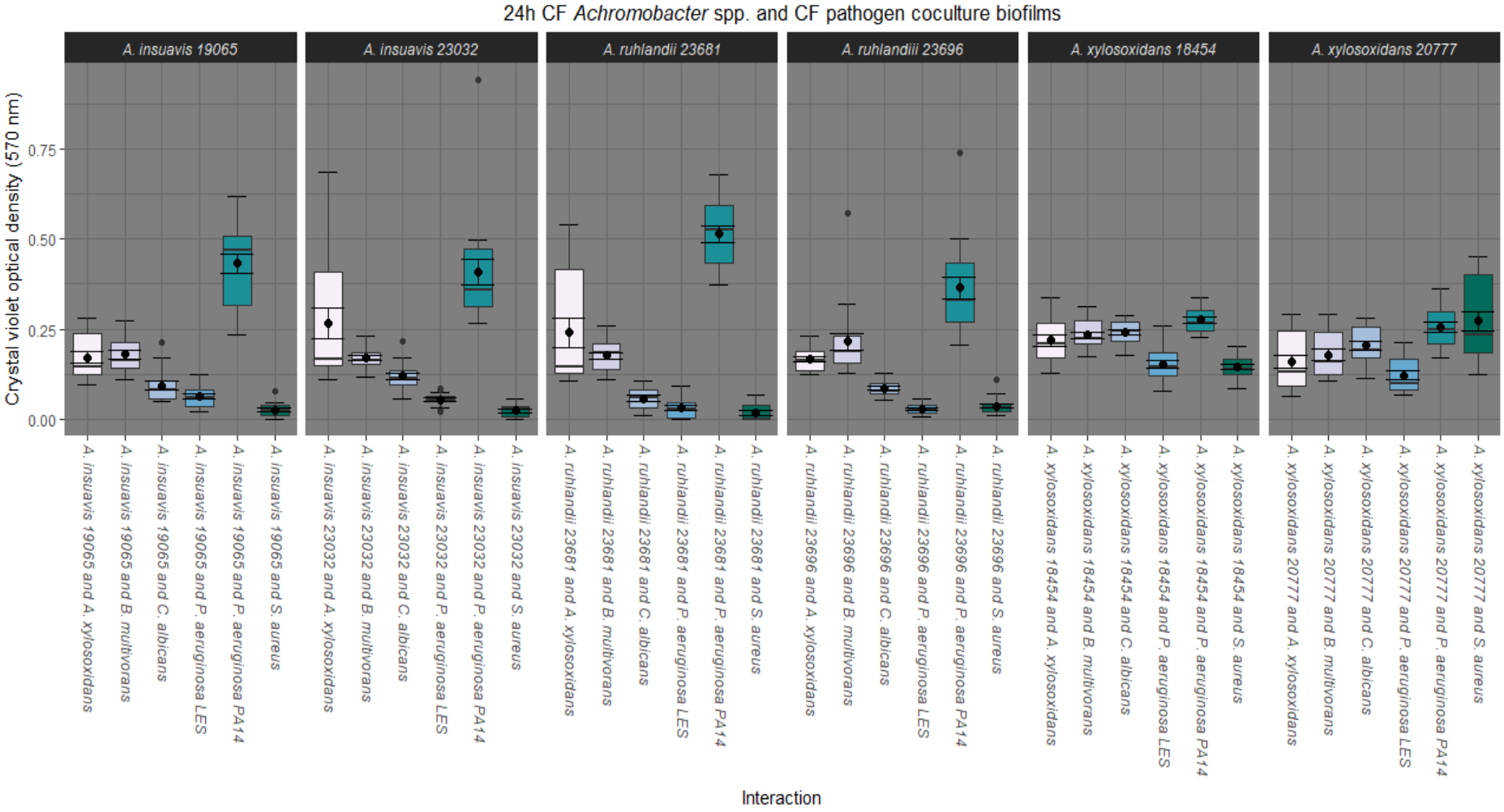


Figure 6.9 24-hour biofilm production in clinical *Achromobacter* spp. and CF pathogen mixed cultures. Standard error bars for each sample have been included from the centre of each plot. Each interaction was coloured according to secondary pathogens which was accomplished in R Color Brewer (Neuwirth 2015). The *A. xylosoxidans* reference refers to *A. xylosoxidans* LMG 1863 being paired with each clinical isolate.

6.3.9 Effect of mixed pathogen cultures on protease production

Achromobacter strains were inoculated with a secondary CF pathogen to understand whether *Achromobacter* spp. could modulate the production of protease (objective 4). Protease production significantly differed between strain (R^2 0.9, $F_{20,21}$=5.6, P=0.0001) and time ($F_{1,189}$=502.4, P<0.0001; SI Table E13). *Pseudomonas* LES-B58 and PA14 exhibited strong protease activity (21.4 mm and 28.8 mm diameter, respectively). Measurable zones surrounding *S. aureus* cultures were smaller (~10 mm diameter after 24-h) than the zones of either *Pseudomonas* strain, however these differences were insignificant (SI Figure E9). Following 48-h, the resulting zones surrounding *Pseudomonas* LES-B58, *P. aeruginosa* PA14, and *S. aureus* increased and resulting zones were comparable to one other (33.8 mm, 37.8 mm, and 19.1 mm diameter respectively; SI Figure E9). *P. aeruginosa* LES-B58 and *P. aeruginosa* PA14 protease production did however significantly differ from 24-h to 48-h timepoints (SI Table E13). All clinical *Achromobacter* isolates were protease-negative, and despite mixed culture CF pathogen protease production increasing after 48-h, these differences were not considered to have been influenced by any clinical *Achromobacter* strain (SI Table E13).

6.4 Discussion

6.4.1 Whole genome sequencing implicated *Achromobacter* species missed by conventional laboratory diagnostics

In the last 2 decades, reported prevalence rates of *Achromobacter* spp. have increased partly due to improved molecular diagnostic tools, and potentially also AMR selective pressures (De Baets et al. 2007; Lambiase et al. 2011; Pereira et al. 2011; Amoureux et al. 2013; Pereira et al. 2017). WGS is becoming increasingly commonplace and standardised in clinical microbiology for its application in pathogen and transmission surveillance, infection control, taxonomy, metagenomics, and rapid AMR/virulence detection (Deurenberg et al. 2017). Here, the power of WGS for CF pathogen diagnostics was clearly demonstrated, further emphasising its application in the rapid and accurate identification of *Achromobacter* species, transmission surveillance and AMR within the context of a Welsh CF population.

In support of hypothesis 1 (WGS significantly outperforms routine MALDI-TOF in the accurate identification of Achromobacter species), WGS uncovered additional *Achromobacter* species where MALDI-TOF was unable to differentiate *A. xylosoxidans* species from *A. insuavis* or *A. ruhlandii*. This was also in concordance with the findings of Blondiaux (2010), Gomila et al. (2014), and Almuzara et al. (2015). The stated prevalence of multicentre *Achromobacter* species varies, yet *A. xylosoxidans* is often the most prevalent reported species in CF and non-respiratory samples, usually followed by *A. insuavis*, *A. dolens*, and *A. ruhlandii* (Ridderberg et al. 2012; Spilker et al. 2013; Amoureux et al. 2016; Coward et al. 2016; Firmida et al. 2016; Filipic 2017). Similarly, *A xylosoxidans* was most frequently recovered from this study followed by *A. insuavis* and *A. ruhlandii*. Although MALDI-TOF is a rapid and technically undemanding diagnostic tool, the potential misdiagnosis of *Achromobacter* spp. infections may lead to worsened patient outcomes as a result of inappropriate antimicrobial therapies and potential differences in virulence.

6.4.2 Whole genome species assignation and multilocus sequence analysis suggested *A. insuavis* reinfection

The transmissibility of *Achromobacter* has remained controversial due to a lack of cohesion among past and recent identification methods (Kanellopoulou et al., 2004, Edwards et al., 2017, Coward et al., 2016). *A. insuavis* transmission was investigated since isolates belonging to patients B and C shared the same ST. As MLST lacks the resolution to determine mode of transmission, it was necessary to compare SNPs to determine whether the single ST of *A. insuavis* was due to reinfection from an environmental source or recrudescence of the isolate during reinfection. This study found no evidence of patient cross-infection, which was already

highly unlikely due to modern patient segregation measures (Vonberg and Gastmeier 2005). Instead, the principal mode of *Achromobacter* transmission was suggested to originate from a common environmental source and is arguably one reason for increased incidence in *Achromobacter* spp. prevalence (Ridderberg et al. 2011). Recombination events leading to diversification of chronically-infecting *Achromobacter* spp. should however not be overlooked; Darch et al. (2015) demonstrated the importance of recombination in the continual evolution and adaptation of *P. aeruginosa* in the CF lung, warranting future investigation for *Achromobacter* species.

6.4.3 *Achromobacter* spp. are multi-drug resistant, especially towards "last-line" antimicrobial agents

Achromobacter infections are routinely treated with piperacillin-tazobactam, imipenem, meropenem and trimethoprim-sulfamethoxazole due to *Achromobacter*'s frequent AMR to aminoglycosides, aztreonam, tetracyclines, penicillins and cephalosporins. Data on non-*A. xylosoxidans* AMR and resistance mechanisms are limited (Abbott and Peleg 2015; Swenson and Sadikot 2015; Filipic 2017; Fleurbaaij et al. 2018). Meropenem is commonly used as carbapenem resistance is not thought to be widespread, although carbapenem-resistant *A. xylosoxidans* have emerged (Liu et al., 2016, Amoureux et al., 2013b). Importantly, carbapenem-resistant *Achromobacter* spp. strains were the most frequently recovered AMR phenotype seen in this study.

A. xylosoxidans is innately multidrug resistant, and harbours resistance-nodulation-division (RND) efflux pumps. These include AxyAB-OprM, AxyXY-OprZ, and AxyEF-OprN which give rise to distinct resistance profiles (Bador et al. 2011; Bador et al. 2013; Nielsen et al. 2019). AxyXY-OprZ is responsible for resistance to aminoglycosides, some fluoroquinolones, erythromycin, cefepime, tetracyclines and carbapenems, and is widespread among the 3 CF *Achromobacter* species (Bador et al. 2013). This multidrug efflux pump could result in high carbapenem cephalosporin and tetracycline MICs and is of particular concern as its substrate range includes major compounds used to treat CF pulmonary infection (Bador et al. 2017). Neither AxyAB-OprM nor AxyEF-OprN, or its homologs was detected in the present study.

Numerous acquired β-lactamase genes have been characterised in *A. xylosoxidans*. These include class A β-lactamases bla_{VEB-1} and *axc* (Neuwirth et al. 2006; Fleurbaaij et al. 2018), and class B metallo-β-lactamases comprising of IMP-type bla_{IMP-1}, bla_{IMP-10}, and bla_{IMP-19} (Iyobe et al. 2002; Yamamoto et al. 2012). These β-lactamases demonstrate low level carbapenemase activity, along with VIM-type bla_{VIM-1} and bla_{VIM-2} (Riccio et al. 2001; Shin et al. 2005), and bla_{TMB-1} (El Salabi et al. 2012). While acquired AMR genes were postulated to have contributed to clinical *Achromobacter* spp. carbapenem resistance, not one of these

known acquired AMR genes were detected. Instead, the metallo-β-lactamase gene *ccrA* was postulated to have influenced higher IPM and MEM MICs in *A. ruhlandii* and *A. xylosoxidans*. Additionally, all *A. xylosoxidans* isolates reported with *ccrA* were resistant to at least one carbapenem, although this gene was not present in all carbapenem resistant isolates. The presence of *ccrA* did not also appear to correlate with elevated CAZ or TZP MICs. The metallo-β-lactamase *blm* was not believed to have significantly altered carbapenem resistance as *blm* was widespread including in carbapenem-sensitive *A. insuavis.*

Achromobacter spp. also harbour intrinsic class D oxacillinases that have narrow-spectrum hydrolysis profiles but are capable of extended activity against expanded-spectrum cephalosporins (Doi et al. 2008; Antunes et al. 2014). *A. xylosoxidans* possesses the constitutively expressed *bla*OXA_{-114} gene variant, known to confer resistance to cefotaxime, cefoxitin, cefepime and aztreonam (Doi et al., 2008). Similarly, weak β-lactam hydrolysis has been observed from *A. ruhlandii* OXA-258 (Papalia et al. 2013). Oxacillinases were present among the study-derived *A. xylosoxidans* and in the *A. insuavis* and *A. ruhlandii* species-specific variants (Figure 6.4; Papalia et al., 2013, Traglia et al., 2014). Interestingly, *A. xylosoxidans* that clustered with known OXA variants demonstrated variable MICs to carbapenems, suggesting the oxacillinase did not strongly influence resistance to IPM or MEM, CAZ, or TZP. Altogether, the specific presence and absence of AMR genes did not all correlate with MICs, disproving hypothesis 2 (WGS provides the resolution required to guide appropriate antimicrobial therapies). In the case of *Achromobacter* species, better AMR gene references are required.

Inconsistencies between *in silico* and MIC data must be considered due to underreporting of, or incorrect, AMR gene identification due to the lack of *Achromobacter*-specific references in existing AMR databases. For example, *sul2* is related to TMP/SMX resistance but *in silico* detection did not match MIC data. Additionally, no colistin resistance genes were detected despite 6 isolates being colistin resistant. The clinical relevance of these MIC results must also be scrutinised in the context of a lack of existing *Achromobacter* specific clinical breakpoint data and instead relies on *P. aeruginosa* clinical breakpoint data.

6.4.4 *Achromobacter* spp. pathogenic potential

Despite increased interest in studying novel respiratory *Achromobacter* species, little is known regarding their virulence. Mechanisms involving their ability to adhere, colonise and infect the lung must be investigated further especially as the incidence of *Achromobacter* infections continue to rise (Filipic et al. 2017). This study aimed to analyse *Achromobacter* spp. pathogenic potential through basic phenotypic traits such as growth rates, biofilm production, motility, and protease production as these features are associated with *in vivo* colonisation,

infection, progression of bronchiectasis and lung function decline (Wiser and Lenski 2015; Filipic 2017; McKelvey et al. 2020). These traits can also indicate competitive fitness between one or more pathogens which could have significant implications on the modulation of clinically relevant phenotypes in CF (Jean-Pierre et al. 2021). Furthermore, literature surrounding pathogen interactions are mostly limited to *P. aeruginosa* and *S. aureus*, often neglecting other clinically relevant respiratory species (Menetrey et al. 2020). As none of the clinical *Achromobacter* isolates altered protease production, this result has been omitted from subsequent sections.

As a proxy of fitness, *Achromobacter* growth rates were most similar to P. *aeruginosa* LES-B58 and PA14*.* The relationship between *Achromobacter* spp. and *P. aeruginosa* did not appear antagonistic due to unaltered mixed culture growth rates. Based on the ability of *P. aeruginosa* to outcompete the growth and metabolism of *S. aureus* (Camus et al. 2021), a comparable level of competitive fitness was reflected in *A. xylosoxidans* and *S. aureus* interactions where their combined growth rates mirrored *A. xylosoxidans* pure culture readings, however these differences were not significant. This insignificant result may have arisen from a reduction in statistical power when comparing a large number of factors (i.e., the number of strain identifiers being compared). This result also suggests the competitive fitness of *A. xylosoxidans* was similar to *P. aeruginosa,* supporting hypothesis 3 (*Achromobacter xylosoxidans* competitive fitness as inferred from phenotypic traits is demonstrably similar to major CF Gram-negative species). The competitive fitness of *A. insuavis* and *A ruhlandii* was inferred to be lesser compared to *A. xylosoxidans* since non-*A. xylosoxidans S. aureus* mixed cultures reflected *S. aureus* baseline growth, and μ rates in the *B. multivorans* cocultures were substantially higher than *A. xylosoxidans* mixed cultures. Additionally, *A. xylosoxidans* was demonstrated to be more readily adapted to a temperate environment by being able to reach full turbidity at 20°C and 41°C, though any advantage this may provide *A. xylosoxidans* over the other *Achromobacter* species in terms of competitive advantage or persistence in the environment or host is so far unknown.

B. multivorans solely appeared to have outcompeted the growth of all *Achromobacter* isolates where their combined growth rates most similarly reflected *B. multivorans* pure culture μ. This did not necessarily imply that *B. multivorans* demonstrated higher fitness above all other species investigated. For example, *P. aeruginosa* can dominate *Burkholderia* species during lung infection, and is postulated to demonstrate a similar level of competitive fitness as *A. xylosoxidans*. Furthermore, there were no significant differences in the growth rates of mixed culture assays compared to pure culture. Most importantly, phenotypic characterisation is also not a direct measure of fitness within the context of disease.

Biofilm production is associated with the ability of bacteria to initiate and persist in chronic infections which significantly contributes to AMR and resistance to antiseptics (Jakobsen et al. 2013; Filipic et al. 2017). As a result, eradication of biofilm infections often fails despite aggressive antibiotic treatment (Nielsen et al. 2016). The capacity of *Achromobacter* spp. to form biofilms has already been demonstrated where very strong biofilm producers have been isolated from CF centres (Jakobsen et al. 2013; Trancassini et al. 2014; Konstantinović et al. 2017; Pereira et al. 2017). Likewise, strong biofilm production was observed in clinical *A. xylosoxidans* strains and *A. xylosoxidans* mixed culture biofilms largely resembled *A. xylosoxidans* pure culture readings (supporting hypothesis 3). This study did however highlight weaker non-*A. xylosoxidans* biofilm formers. Despite biofilm production being comparable between clinical *A. xylosoxidans* strains and *B. multivorans,* RISA analysis also suggested *B. multivorans* outcompeted and outgrew *A. xylosoxidans* in coculture based on the sole amplification of a *B. multivorans* amplicon.

After 24 hours, competition between *A. xylosoxidans* and *P. aeruginosa* PA14 was also observed, where the OD of *P. aeruginosa* PA14 stained biomass was drastically reduced, comparable to *A. xylosoxidans* pure culture OD measurements. This result reflected the findings of Menetrey et al. (2020) who observed *A. xylosoxidans* inhibit the growth, pigmentation, and biofilm formation of *P. aeruginosa.* While *P. aeruginosa* PA14 appeared inhibited at 24 hours, no change in *P. aeruginosa* PA14 pigment was observed, and its biofilm production significantly increased after 48 hours. In contrast, the ratio of *A. xylosoxidans* RISA amplified DNA remained consistently higher than *P. aeruginosa* in these samples suggesting the potential of *A. xylosoxidans* to become dominant over time (Collins 2019). This warrants further investigation into the mechanisms in which *A. xylosoxidans* could prevent the outgrowth of *Pseudomonas* spp., and how this knowledge could be applied to preventing the proliferation of major respiratory pathogens. There was also a surprisingly large amount of *A. insuavis* and *A. ruhlandii* RISA-amplified DNA detected with *P. aeruginosa* PA14. This result changed the original assumption that *P. aeruginosa* outgrew these species according to mixed culture optical density readings and their competitive fitness in culture may be underestimated.

The importance of flagellar motility in biofilm formation has also been demonstrated for other major respiratory pathogens such as *P. aeruginosa* and *B. cenocepacia and* shares a reciprocal relationship with biofilm formation (Richter et al. 2018; Nielsen et al. 2019; Schulze et al. 2021). Swimming motility was examined for flagella-mediated movement through liquids, and swarming motility was also examined for being a quorum-sensing regulated phenotype that has been proposed to contribute to AMR and bacterial virulence. Swarming facilitates bacterial movement through mucosal mucus coating of the epithelium lining of the respiratory

tract and plays a role in early biofilm development (Pereira et al. 2017). *A. xylosoxidans* exhibited the highest degree of swimming motility of all the clinical *Achromobacter* strains. This result could suggest a higher competitive fitness for lung infections, in *A. xylosoxidans* being able to compete for resources or limit potential competitors compared to the other *Achromobacter* species. Furthermore, swimming motility of both clinical *A. xylosoxidans* strains and *B. multivorans* were comparable, supporting hypothesis 3.

In contrast, *A. xylosoxidans* swimming motility was significantly weaker than *P. aeruginosa* PA14, suggesting a lower competitive fitness compared to this bacterium. *Achromobacter* spp. failed to exhibit a swarming phenotype similar to the findings of Trancassini et al. (2014) and Khademi et al. (2021) which is in contrast to the works of Filipic et al. (2017) and Pereira et al. (2017) who demonstrated *A. xylosoxidans* and *A. ruhlandii* motility was strain dependent. It is important to acknowledge that *A. ruhlandii* appeared positive for swarming after 48 hours, but this result was most likely as a consequence of its mucoid phenotype showing colony spread across the agar rather than true swarming. Nevertheless, it is important to acknowledge that this mucoid phenotype is a hallmark of chronic infection in *P. aeruginosa* and may also apply to *A. ruhlandii* (Nielsen et al. 2019), and further research is needed to investigate whether mucoid *Achromobacter* species is capable of delaying the swarming of *B. multivorans*. Compared to non-CF isolates, reduced motility might act as a competitive advantage in establishing airway infections in a host as was shown for other Gram-negative non-fermenting bacteria (Filipic et al. 2017). This rationale could help explain the persistence of *A. insuavis* colonisation in patients B and C.

Notable motile interactions implicated the potential synergism between *A. xylosoxidans* and *P. aeruginosa* PA14 after both species were RISA PCR amplified, suggesting co-migration. Increased motility was also observed between *A. xylosoxidans* and *C. albicans* compared to either baseline result, however *C. albicans* could not be identified through RISA. While motility increased from *A. xylosoxidans* pure culture rates when in coculture with *B. multivorans, Achromobacter* spp. could not be RISA PCR amplified from these mixed cultures, suggesting *B. multivorans* antagonised *A. xylosoxidans* and prevented its outgrowth or motility. Synergistic microbial interactions have been known to modulate the pathogenicity of major pathogens in the lung (Flynn et al. 2016; Li et al. 2021), and with this rationale, this potential synergism between *A. xylosoxidans* and *P. aeruginosa* warrants further investigation into the potential modulation of *P. aeruginosa* virulence *in vitro* and vice versa.

6.4.5 Study pitfalls, considerations, and future direction

A few issues arose from this study potentially impacting measurements and interpretation of results. This included OD readings of crystal violet pertaining to *A. insuavis* isolates. These isolates consistently produced insoluble aggregates in liquid culture. This result can be explained by different biofilm morphology between *A. xylosoxidans* and *A. insuavis. A. xylosoxidans* is reported to form surface-attached biofilms while *A. insuavis* forms loosely attached or unattached suspended aggregates under static conditions (Nielsen et al. 2016). Consequently, growth and biomass measurements may be misleading and underestimate its fitness in culture. The results of both this study and Nielsen et al. (2016) does however support the notion that surface-attached biofilms may not be required for *Achromobacter* spp. biofilm formation where unattached aggregates may possibly be an occurring mode of growth *in vivo* (Nielsen et al. 2016). Finally, it important to acknowledge these phenotypic observations on abiotic surfaces do not necessarily correlate to *in vivo* biofilm formation. Also, the crystal violet assay is a proxy for microbial biomass and not microbial viability/total numbers in a biofilm and hence further methods designed to enumerate biofilm cell numbers should be carried out.

Other drawbacks in methodology arose from using direct colony (Chelex) gDNA extractions and RISA PCR amplification to measure interactions. gDNA extraction with *Achromobacter* species and *S. aureus* consistently failed despite numerous freeze-thaw cycles. RISA PCR may also fail to amplify clinical *Achromobacter* species or CF reference pathogens in low abundance, and PCR amplification bias inevitably influences the ratios of pathogen DNA amplified. It was also possible that *A. xylosoxidans* was amplified with *B. multivorans,* but the QIAxcel may have lacked the resolution to separate these RISA fragments. Additionally, the growth rates of *C. albicans* could not be recorded because of its different absorbance spectrum to bacteria, nor could it have been detected through bacterial RISA PCR.

Compared to many other Gram-negative non-fermenting rods, research using comparative genomics to study *Achromobacter* spp. taxonomy, virulence, and AMR is lacking to address key questions about their clinical relevance and how these species interact in a polymicrobial community. In order to begin addressing these gaps in knowledge, further studies are required to characterise *Achromobacter* spp. virulence factors and AMR mechanisms while updating *in silico* genomic databases with high-quality *Achromobacter* whole-genome references that can be achieved through long-read sequencing. Mechanisms accounting for increasing *Achromobacter* prevalence could be then addressed along with how these species contribute to lung disease and failure of antimicrobial therapy (Filipic et al. 2017).

WGS of longitudinally collected isolates from patients with chronic infections could be conducted to find the genetic changes responsible for adaptive phenotypes associated with establishing chronic infection (Khademi et al. 2021). Functional metagenomic analysis of quorum sensing signalling could prove beneficial in elucidating its role in competition and growth inhibition of major pathogens (Menetrey et al. 2020). Another approach to documenting pathogen interactions could enumerate live vs. dead proportion of bacterial cells to investigate antagonisms that is undertaken through fluorometric and chromogenic screening (Menetrey et al. 2020). A more complete analysis of interacting pathogens may also aid in further understanding the dynamics of bacterial colonisation of the respiratory tract and how interacting pathogens may cause harm to the host.

6.4.6 Conclusions

In summary, this study analysed *Achromobacter* taxonomy, transmission, AMR, and pathogenic potential through phenotypic assays. *A. xylosoxidans* was the most frequently isolated in the Welsh CF population assessed. *A. insuavis* and *A. ruhlandii* were identified accurately by WGS, but encountered less frequently in the study population. MALDI-TOF failed to recognise the latter two species supporting hypothesis 1. WGS was also suitable for epidemiological surveillance showing that the environment was a major source of acquisition for *A. insuavis.* Although a wealth of AMR data was generated through *in silico* AMR gene prediction, the lack of deposited *Achromobacter* spp. AMR gene references impacted differences in MIC and *in silico* data, thus disproving hypothesis 2. The competitive fitness and pathogenic potential of all 3 *Achromobacter* species was shown to be different from one another across a range of phenotypic assays. The competitive fitness of *A. xylosoxidans* was demonstrated to be comparable to *P. aeruginosa* LES-B58, *P. aeruginosa* PA14, and *B. multivorans,* accepting hypothesis 3. In contrast, *A. insuavis* aggregate formation may have underestimated its competitive fitness. Further research is required to understand how these different *Achromobacter* species are equipped for chronic persistence in the lung and contribute to disease progression either directly or indirectly when forming interactions.

7 General discussion

7.1 Overview of key research questions

Culture-independent respiratory microbiome research has revealed complex and highly individual polymicrobial communities across chronic lung diseases. Because of this high individuality, it is vital to characterise these polymicrobial communities and to not adhere to the outdated presumption that few traditional pathogens cause infection. 16S rRNA gene next generation sequencing (NGS) is the most commonly used method to gain insight into bacterial community composition and diversity (Watson et al. 2019). It is well established that bacterial polymicrobial communities can drive the progression of chronic lung diseases, particularly cystic fibrosis (CF) microbiota. Whole genome sequencing (WGS) has been instrumental in targeting emerging respiratory pathogens, where a higher resolution of species identification is required in addition to determining transmission, antimicrobial resistance (AMR), and virulence traits. Because these high-throughput NGS approaches are still primarily research tools, key clinical and research questions were addressed: 1) What type of PCR, sequencing, and gDNA extraction methods is most suitable for routine analysis respiratory microbiota in the Public Health Wales? 2) Can 16S rRNA hypervariable gene sequencing identify a causal link between altered airway microbiota diversity and disease progression? 3) Can WGS better inform clinicians to treat and survey *Achromobacter* spp. infection? The clinical relevance and competitive fitness of *Achromobacter* species was also explored through phenotypic testing, leading to 4) To what extent can polymicrobial interactions be documented and what are the resulting implications on disease progression and antimicrobial treatment?

7.2 Synthesis of key results

Because there is an increased incentive to find a causal relationship between lung microbiota composition, diversity, and disease progression, researchers have attempted to standardise NGS protocols for comparability between studies. However, inconsistencies in lung microbiome research techniques have arisen from different choices of gDNA extraction methodologies, PCR primer, 16S rRNA hypervariable gene region, PCR cycling conditions, and bioinformatic pipeline to name a few (Fouhy et al. 2016). To address some of these inconsistencies, **chapter 3** explored the suitability of a modified 16S rRNA gene sequencing protocol targeting hypervariable region 4 (Kozich et al. 2013) for respiratory microbiota profiling targeting. This methodology was also extended to **chapter 4** where a bacterial gDNA extraction protocol was validated for sputum bacterial populations, and in **chapter 5** for the profiling of lung microbiota in bronchiectasis, cystic fibrosis, and chronic obstructive pulmonary

disease (COPD). A standardised sequencing approach was achieved using PHW reagents and equipment, permitting highly translatable research, and promoting routinisation by fitting into current clinical framework. This V4 16S rRNA gene PCR was specific to all reference respiratory bacteria and they could be detected in low abundance except for *C. acnes.* Abundance and diversity data illustrated distinct community types that simulated different disease states and were distinguishable on an OTU basis. Here, Shannon and inverse Simpson diversity measurements best reflected predicted community diversity. On the other hand, community composition substantially changed in sequenced samples: Gram-positive species were generally preferentially amplified, and *P. aeruginosa* and *B. cepacia* were severely underrepresented. Primer degeneracy and 16S rRNA gene copy were originally considered to have given rise to this result however **chapter 4** and **chapter 5** highlighted potential erroneous preparation of these mock communities. This notion was inferred by the even representation of amplified *S. aureus* and *B. cepacia* DNA that was extracted using the ZymoBIOMICS™ kit (**chapter 4**) in addition to observing highly comparable 16S rRNA gene sequence data and RISA data (**chapters 4 and 5**) which was not the case in **chapter 3**.

While 16S rRNA gene sequencing has been described as a valuable supplementary test in everyday clinical practice (Deurenberg et al. 2017), this method does not always identify bacteria to the genus level. To accompany 16S rRNA gene-based NGS, **chapter 3** also evaluated the usefulness of a RISA PCR to be used alongside sequencing for bacterial species-level resolution and to roughly estimate mock community diversity and dominant species. This method was advantageous for its rapidity and simple means of profiling, however community diversity through RISA was less accurate than 16S rRNA gene sequencing. At this stage, RISA appeared to have performed better to predict species identification rather than being a definitive indicator of bacterial species, mainly due to variation in sample DNA molecular weight that was analysed through the TapeStation and QIAxcel. The resolution of RISA PCR was best when visualising specific abundant species which did not produce ITS amplicons that overlapped closely to other species and were fairly simple. This protocol did highlight a better representation of Gram-negative species, appearing to have outperformed the performance of V4 primers, however it was clear that the resolution of this PCR method was mostly limited to the genus level in mock communities.

Chapter 4 extended the use of optimised RISA PCR and 16S rRNA gene sequencing protocols to validate a bacterial gDNA extraction protocol capable of evenly representing respiratory microbiota and degrading sputum components while limiting the shearing impact on Gram-negative gDNA. An important challenge gDNA extraction poses is that among a wealth of available extraction methods, few kits are specifically designed to extract complex

polymicrobial communities and for metagenomics. The ZymoBIOMICS™ DNA Miniprep kit was the only kit validated for microbiome analysis, being capable of processing both high biomass and low biomass samples. This kit unsurprisingly gave rise to highest gDNA extraction yields and most even representation between *S. aureus* and *B. cepacia*, based on PCR detection limit studies. This kit was highly appropriate for the routine extraction of sputum, and even though this kit required more manual processing than the EMAG, the time taken for pre-processing samples for the EMAG took substantially longer to complete. Further limitations to the EMAG were evident where bead-beating coupled with EMAG chemistry was either too aggressive or too weak, resulting in highly uneven bacterial lysis efficiencies.

Enzymatic lysis did provide the EMAG with more even extractions between both *S. aureus* and *B. cepacia,* and resulted in higher quality gDNA than from bead-beating: however, it was apparent that sputum interfered with gDNA extraction process, resulting in noticeably different results from each extracted replicate. Dithiothreitol (DTT) may have been the most likely inhibitory compound based on successful gDNA extraction and PCR amplification of extracted sputum when sputum was pelleted, and DTT was discarded. Nevertheless, DNA extraction inhibition was believed to be multifactorial. It is noteworthy to mention that MetaPolyzyme was originally designed for soil where the bacterial load is potentially higher, and sputum may have contained inhibiting compounds. As NGS technologies continue to grow and become more feasible in clinical laboratories, a systematic revaluation of current gDNA extraction platforms is needed to study complex polymicrobial communities, metagenomics, and for methods requiring high molecular weight gDNA. It is worth revisiting the application of MetaPolyzyme in different clinical sample types processed by PHW laboratories for microbiota analysis. Alternatively, it is worth determining what components of sputum inhibits various lytic enzymes in order to achieve higher quality gDNA extracts. This can be achieved by creating artificial sputum environments simulating a viscous and proteinaceous material that is controlled by known quantities of microorganisms.

Chapter 5 applied the optimised gDNA extraction protocol (ZymoBIOMICS™ kit), V4 16S rRNA gene sequencing, and RISA PCR to sputum to compare against routine culture and MALDI-TOF. This study encompassed a small cohort of patients who suffer from BR, CF, and COPD, and has contributed to the growing pool of knowledge that respiratory microbiota is highly individual in each patient and there is not always a clear distinction between these disease-specific communities as the majority of highly abundant OTUs were found in each disease category. This does not however reflect the failure of DNA-based methods in differentiating these communities based on composition and diversity. This notion was further supported through greater cohesion between both 16S rRNA gene sequencing and RISA data

as seen in **chapter 4 (and chapter 5**) compared to **chapter 3**. This study highlighted a significant loss of diversity when traditional pathogens dominated respiratory samples in accordance with Flight et al. (2015), Richardson et al. (2019), and Cuthertson et al. (2020). No causal relationships could be determined between loss of bacterial diversity and lung function decline, due to the highly individual microbiota and potential presence of non-bacterial microbiota which could be driving disease progression. Microbe-microbe and host-microbe interactions influencing this change remain poorly understood.

Chapter 5 demonstrates the richness of data that can be generated using relatively basic approaches where both 16S rRNA gene sequencing and RISA can be deployed at a low cost and be integrated into existing PHW laboratory framework. As demonstrated here and in **chapter 3**, RISA is more appropriate as a tool that predicts ITS profiles than it is to provide definitive species identification. This predictive tool can be used to infer virtually any bacterial species which is matched to a user-specified reference database of microorganisms. In complex polymicrobial communities, RISA is realistically limited to genus-level prediction, however in special circumstances, species-level predictions can be inferred with one key example being *P. monteilii* that was missed through culture. These circumstances pertain to heavily dominated and/or simple ITS profiles. Nevertheless, RISA in combination with culture and MALDI-TOF could be a valuable tool to provide additional resolution where 16S rRNA gene sequencing resolves dominant microbiota to the phylum level. For V4 16S rRNA gene sequencing, this lack of resolution seems to be influenced most by *Enterobacteriaceae*. This research demonstrated the requirement of culture-based tools to supplement poorly resolved sequence data along with RISA PCR however it is apparent that culture-based data should not guide patient therapy; this was emphasised in culture positive and sequence negative samples in addition to PPMs that were identified where culture yielded no microbial growth.

Chapter 6 evaluated the use of whole genome sequencing and was at the time a novel clinical research tool to better characterise an understudied genus of unknown clinical relevance in context of lung infection. At the time of completion, this was also the first known study contributing towards understanding the mode of *A. insuavis* acquisition from a common environmental source, in support of more recent literature (Pereira et al., 2011, Ridderberg et al., 2011). This study was also instrumental in determining the multidrug resistance of *A. xylosoxidans, A. insuavis,* and *A. ruhlandii.* While *A. xylosoxidans* AMR is more recognised (Fleurbaaij et al., 2018), metallo-beta-lactamase CcrA was newly reported to *A. xylosoxidans* and could have contributed to greater carbapenem resistance and is worth testing for its potential dissemination among a larger Welsh population. Carbapenem resistance was

believed to be multifactorial as carbapenem resistance was most common in *A. xylosoxidans* and *A. ruhlandii*.

It is important to acknowledge due to a lack of deposited *Achromobacter* spp. whole genome references, *in silico* AMR predictions were made based on the closest related reference available being *P. aeruginosa.* The relevance of interpreting these results must questioned however as no clinical breakpoint data exists for *Achromobacter* species and clinicians cannot rely on phenotypic and *in silico* AMR data until these reference databases have been updated to include high quality and relevant references, and for more relevant clinical breakpoint data to be published. For the time being, WGS is still a research tool for the NHS as this technology is dependent on funding, resource allocation, and additional laboratory personnel.

Chapter 6 provided insight into *Achromobacter* spp. AMR mechanisms and virulence traits. Little is known regarding mechanisms allowing these species to adhere, colonise, and infect the lung since genomic and phenotypic data is limited, especially for non-*A. xylosoxidans* species (Filipic et al. 2017; Dobson et al. 2018, Fleurbaajj et al. 2018). *A. xylosoxidans* was inferred to have exhibited a higher pathogenic potential over other *Achromobacter* strains, however it is important to note this was *in vitro* rather than *in situ.* Because little is known regarding the competitive fitness of *Achromobacter* spp. paired with respiratory pathogens, these phenotypic traits were also recorded in coculture with traditional CF pathogens. Competition was strain dependent, in accord with the findings of Menetrey et al. (2020). Antagonism was observed in *B. multivorans* interactions which originated from the soil while interactions between *Achromobacter* spp. and the clinical *P. aeruginosa* PA14 appeared synergistic rather than antagonistic. This synergism has also been recorded between different clinical *P. aeruginosa* and *A. xylosoxidans* strains (Collins 2019). This is importantly as synergistic interactions modulate the pathogenicity of *P. aeruginosa* (Flynn et al. 2016; Li et al. 2021) and potentially *A. xylosoxidans,* warranting further investigation in these interactions. This research also emphasised the complexity of interacting polymicrobial hierarchies; intransitive hierarchies can have unpredictable implications on modulating pathogenic species without prior understanding of competition outcomes.

7.3 Proposal of routine 16S rRNA gene NGS and RISA PCR for PHW bacteriology

The DNA-based methods offered an increase in taxonomic resolution compared to culture alone, which has the potential to benefit vulnerable patient groups should these technologies be adopted by PHW. 16S rRNA gene library preparation was rapid, requiring ~1 hour inside a thermal cycler. This short time frame should maximise the number of sequencing libraries prepared in a single day, and automated library preparation in the PHW could further speed up this process while reducing contamination risk. A reduction in primer concentration also resulted in highly concentrated libraries that were consistently free from library preparation contaminants, facilitating purification, optimal Illumina clustering efficiency, and the production of high-quality sequence data. The SILVA 16S rRNA gene reference taxonomic database is recommended for the identification of respiratory taxa by resolving the highest number of unclassified sequences and providing greatest discrimination power.
RISA PCR shows promise for being a rapid and cost-effective tool that can be viewed under various high-resolution electrophoretic methods. RISA PCR can estimate dominant pathogens and diversity profiles in the same day after gDNA extraction, but it should not be used as definitive means of identification.

7.4 Project limitations

Project limitations that were considered to have impacted more than one results chapter and/or have wider implications are noted in this section:

In future, human gDNA depletion may lead to a more accurate estimation of sputum spike-in PCR detection limits (**chapter 4**). Although this was not an essential step of sequencing or PCR optimisation, the vast ratio of human to microbial DNA may have led to little amplification of target DNA. Such large quantities of human gDNA would also not permit functional analyses of sputum should these methodologies be adapted for this purpose.

Another major limitation involved the DNA-based analysis of patient sputa (**chapter 5**). One of the many challenges brought on by the COVID-19 pandemic led to cessation of sputum processing, extraction, sequencing, and analysis in the PHW. While the direction of this project was modified in response, only a fraction of the intended sample size could be analysed. This small sample size may have contributed to the lack of significant associations between microbial diversity and lung function decline or among other patient clinical parameters.

Due to the number of optimisation stages and treatments applied to **chapter 3** and **chapter 5**, only a small panel of bacterial species could be tested. It is unclear what would be the total extent of respiratory microbiota undetected through 16S rRNA hypervariable gene sequencing

(i.e., endospores, yeasts, other very robust Gram-positive bacteria). Importantly, this project has focused solely on the bacterial component, ignoring non-bacterial components that can contribute towards disease progression. However, this project was too constrained by time to investigate these avenues.

7.5 Future direction

With an ever-increasing amount of 16S rRNA gene sequence-based data capturing respiratory microbiota composition and diversity, it is now essential to study microbe-microbe and host-microbe interactions in more detail to explain associations between infection and disease progression. A greater understanding of functional response diversity will help in understanding respiratory microbiota composition and diversity changes that are associated with pulmonary exacerbations and states of stability (Filkins and O'Toole 2015). Metagenomic sequencing captures all bacteria, viruses, fungi, and eukaryotes and their functional capabilities. Changes in microbial gene expression relevant to persistence in the lung, increased resistance, and virulence can be identified through metatranscriptomics. Metabolites associated with exacerbations can also be detected by metabolomics (Toulousse et al 2021). These technologies are still very expensive, error prone, and not feasible for smaller laboratories. Not only will these methods require standardisation, but these methods are also complicated by the analysis of low-biomass samples risking significant contamination from the oropharynx.

Because of these challenges, it is not uncommon for clinical laboratories to adopt amplicon-based sequencing when dealing with low-biomass specimens such as sputum and blood; in context of PHW molecular diagnostics for lung infection, additional research should be carried out such as inclusion of human DNA depletion step to assess the extent of microbial abundance changes from untreated samples and to assess its suitability for future metagenomic work. To help form better associations between past and current infection, the exclusion of dead proportion of bacteria should be considered. If possible, PCR and gDNA extraction automation should be considered to limit sample contamination and to free up resources. Inhibition of gDNA extraction and PCR amplification from sputum should also be investigated to determine whether further optimisation can be applied to these samples to maximise yields.

WGS demonstrated potential for re-examining samples that were previously culture-positive for emerging or unusual respiratory pathogens. In context of *Achromobacter* species, long

read sequencing can help generate a more relevant references for the detection of AMR and virulence genes. In situations where AMR genes have been identified but cannot explain resistance phenotypes, genome-wide association studies may address this problem (Gabrielaite 2021). Longitudinal studies should find pathoadaptive features associated with establishing chronic infection. Moreover, efflux pump components which were widespread in all *Achromobacter* isolates should also be investigated as they are implicated in bacterial virulence (Alcalde-Rico et al. 2016). Further studies investigating the expression of *Achromobacter* spp. virulence factors should be conducted to elucidate mechanisms involved in persistent infection and lung inflammation. A broader investigation into *Achromobacter* spp. and traditional and emerging respiratory pathogens is required to search for specific competition that could be established in the lung (Menetrey et al. 2020) and understand the relevance of these interactions, and whether these interactions can be exploited for the growth inhibition of major pathogens.

7.6 Conclusions

This project was mostly dedicated to developing 16S rRNA gene NGS and RISA methodologies to overcome the main disadvantages of culture-based tools in context of diagnosing lung infection. From this, standardised sequencing and PCR approaches were achieved using PHW resources, representing the first step towards routinisation. For respiratory microbiota analysis, sequencing of 16S rRNA hypervariable gene region 4 and the SILVA taxonomic database provided basic information regarding bacterial community composition and diversity. Throughout, RISA has shown its use in predicting species identification but should not be used as formal identification. A bead-based gDNA extraction kit (ZymoBIOMICS™) gave rise to highly even and concentrated gDNA extracts, partly mitigating cell lysis bias. These DNA-based tools ultimately complemented routine culture and MALDI-TOF identification in sputum, enabling the detection of potentially pathogenic microorganisms missed in culture, and best-resolved poor sequence identification in a minority of samples. Despite this wealth of data, no correlations between loss of bacterial diversity and lower lung function were drawn. The discriminatory power of whole genome sequencing was demonstrated in the characterisation of multidrug resistant CF *Achromobacter* species and demonstrated major limitations in current routine identification of these species. The significance of these findings in combination with transmission data points towards these emerging species as clinically relevant though it is uncertain as to what extent. Phenotypic traits suggested the pathogenic potential of *A. xylosoxidans* was greater than the other species, nevertheless, further research is warranted to understand the pathogenicity of all three species.

References

Abbott, I. J. and Peleg, A. Y. 2015. *Stenotrophomonas*, *Achromobacter*, and nonmelioid *Burkholderia* species: antimicrobial resistance and therapeutic strategies. *Semin Respir Crit Care Med* 36(1), pp. 99-110. doi: 10.1055/s-0034-1396929

AbdulWahab, A., Taj-Aldeen, S. J., Ibrahim, E. B., Talaq, E., Abu-Madi, M. and Fotedar, R. 2015. Discrepancy in MALDI-TOF MS identification of uncommon Gram-negative bacteria from lower respiratory secretions in patients with cystic fibrosis. *Infect Drug Resist* 8, pp. 83-88. doi: 10.2147/IDR.S80341

Acosta, N., Whelan, F. J., Somayaji, R., Poonja, A., Surette, M. G., Rabin, H. R. and Parkins, M. D. 2017. The Evolving Cystic Fibrosis Microbiome: A Comparative Cohort Study Spanning 16 Years. *Ann Am Thorac Soc* 14(8), pp. 1288-1297. doi: 10.1513/AnnalsATS.201609-668OC

Aditi, A., Shariff, M. and Beri, K. 2017. Exacerbation of bronchiectasis by *Pseudomonas monteilii*: a case report. *BMC Infect Dis* 17(1), p. 511. doi: 10.1186/s12879-017-2600-9

Agarwal, A. K., Raja, A. and Brown, B. D. 2021. *Chronic Obstructive Pulmonary Disease*. Tampa, Florida:StatPearls Publishing.

Al Ahmar, R., Kirby, B. D. and Yu, H. D. 2020. Culture of Small Colony Variant of *Pseudomonas aeruginosa* and Quantitation of its Alginate. *J Vis Exp* (156), doi: 10.3791/60466

Alby, K., Gilligan, P. H. and Miller, M. B. 2013. Comparison of matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry platforms for the identification of Gram-negative rods from patients with cystic fibrosis. *J Clin Microbiol* 51(11), pp. 3852-3854. doi: 10.1128/JCM.01618-13

Ali, N., Rampazzo, R. C. P., Costa, A. D. T. and Krieger, M. A. 2017. Current Nucleic Acid Extraction Methods and Their Implications to Point-of-Care Diagnostics. *Biomed Res Int* 2017, p. 9306564. doi: 10.1155/2017/9306564

Amati, F., Simonetta, E., Gramegna, A., Tarsia, P., Contarini, M., Blasi, F. and Aliberti, S. 2019. The biology of pulmonary exacerbations in bronchiectasis. *Eur Respir Rev* 28(154), doi: 10.1183/16000617.0055-2019

Amoureux, L., Bador, J., Siebor, E., Taillefumier, N., Fanton, A. and Neuwirth, C. 2013. Epidemiology and resistance of *Achromobacter xylosoxidans* from cystic fibrosis patients in Dijon, Burgundy: first French data. *J Cyst Fibros* 12(2), pp. 170-176. doi: 10.1016/j.jcf.2012.08.005

Amoureux, L., Bador, J., Verrier, T., Mjahed, H., de Curraize, C. and Neuwirth, C. 2016. *Achromobacter xylosoxidans* is the predominant *Achromobacter* species isolated from diverse non-respiratory samples. *Epidemiol Infect* 144(16), pp. 3527-3530. doi: 10.1017/s0950268816001564

Amoureux, L. Bador, J., Zouak, F. B., Chapouis, A., de Curraize, C., and Neuwirth, C. 2016. Distribution of the species of *Achromobacter* in a French Cystic Fibrosis Centre and multilocus sequence typing analysis reveal the predominance of *A. xylosoxidans* and clonal relationships between some clinical and environmental isolates. *J Cyst Fibros* 15(4), pp. 486-494. doi: doi:10.1016/j.jcf.2015.12.009

Anabtawi, A., Le, T., Putman, M., Tangpricha, V. and Bianchi, M. L. 2019. Cystic fibrosis bone disease: Pathophysiology, assessment and prognostic implications. *J Cyst Fibros* 18 Suppl 2, pp. S48-S55. doi: 10.1016/j.jcf.2019.08.018

Anderson, M. J. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 26, pp. 32–46. doi: 10.1111/j.1442-9993.2001.01070.pp.x

Andrews, S. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc [Accessed: 8 February 2018].

Antunes, N. T., Lamoureaux, T. L., Toth, M., Stewart, N. K., Frase, H. and Vakulenko, S. B. 2014. Class D β-Lactamases: Are They All Carbapenemases? *Antimicrob Agents Chemother* 58(4), pp. 2119-2125. doi: 10.1128/AAC.02522-13

Ash, K. S. and Ramteke, P. W. 2018. Optimization of Extracellular Alkaline Protease Production From *Pseudomonas aeruginosa* Isolated from Soil Samples. *IJAEB* 11(1), pp. 187-194.

Babu, K. N., Sheeja, T. E., Minoo, D., Rajesh, M. K., Samsudeen, K., Suraby, E. J. and Kumar, I. P. V. 2021. Random Amplified Polymorphic DNA (RAPD) and Derived Techniques. *Methods Mol Biol* 2222, pp. 219-247. doi: 10.1007/978-1-0716-0997-2_13

Bador, J., Amoureux, L., Blanc, E. and Neuwirth, C. 2013. Innate aminoglycoside resistance of *Achromobacter xylosoxidans* is due to AxyXY-OprZ, an RND-type multidrug efflux pump. *Antimicrob Agents Chemother* 57(1), pp. 603-605. doi: 10.1128/aac.01243-12

Bador, J., Amoureux, L., Duez, J. M., Drabowicz, A., Siebor, E., Llanes, C. and Neuwirth, C. 2011. First Description of an RND-Type Multidrug Efflux Pump in *Achromobacter xylosoxidans*, AxyABM. *Antimicrob Agents Chemother* 55(10), pp. 4912-4914. doi: 10.1128/aac.00341-11

Bador, J. et al. 2017. Role of AxyZ Transcriptional Regulator in Overproduction of AxyXY-OprZ Multidrug Efflux System in *Achromobacter* Species Mutants Selected by Tobramycin. *Antimicrob Agents Chemother* 61(8), doi: 10.1128/aac.00290-17

Baldwin, A. et al. 2005. Multilocus sequence typing scheme that provides both species and strain differentiation for the *Burkholderia cepacia* complex. *J Clin Microbiol* 43(9), pp. 4665-4673. doi: 10.1128/JCM.43.9.4665-4673.2005

Ballarini, A., Scalet, G., Kos, M., Cramer, N., Wiehlmann, L. and Jousson, O. 2012. Molecular typing and epidemiological investigation of clinical populations of *Pseudomonas*

aeruginosa using an oligonucleotide-microarray. *BMC Microbiol* 12, p. 152. doi: 10.1186/1471-2180-12-152

Barnes, P. J. 2016. Sex Differences in Chronic Obstructive Pulmonary Disease Mechanisms. *Am J Respir Crit Care Med* 193(8), pp. 813-814. doi: 10.1164/rccm.201512-2379ED

Beasley, V., Joshi, P. V., Singanayagam, A., Molyneaux, P. L., Johnston, S. L. and Mallia, P. 2012. Lung microbiology and exacerbations in COPD. *Int J Chron Obstruct Pulmon Dis* 7, pp. 555-569. doi: 10.2147/COPD.S28286

Bellali, S. et al. 2021. Running after ghosts: are dead bacteria the dark matter of the human gut microbiota? *Gut Microbes* 13(1), pp. 1-12. doi: 10.1080/19490976.2021.1897208

Bellali, S., Lagier, J. C., Raoult, D. and Bou Khalil, J. 2019. Among Live and Dead Bacteria, the Optimization of Sample Collection and Processing Remains Essential in Recovering Gut Microbiota Components. *Front Microbiol* 10, p. 1606. doi: 10.3389/fmicb.2019.01606

Ben Dekhil, S. M., Peel, M. M., Lennox, V. A., Stackebrandt, E. and Sly, L. I. 1997. Isolation of *Lautropia mirabilis* from sputa of a cystic fibrosis patient. *J Clin Microbiol* 35(4), pp. 1024-1026. doi: 10.1128/jcm.35.4.1024-1026.1997

Bernal, E. 2014. Limit of Detection and Limit of Quantification Determination in Gas Chromatography. 1st ed. Guo, X (1st ed). *Advances in Gas Chromatography*. London: IntechOpen.

Bevivino, A., Bacci, G., Drevinek, P., Nelson, M. T., Hoffman, L. and Mengoni, A. 2019. Deciphering the Ecology of Cystic Fibrosis Bacterial Communities: Towards Systems-Level Integration. *Trends Mol Med* 25(12), pp. 1110-1122. doi: 10.1016/j.molmed.2019.07.008

Bierlaagh, M. C., Muilwijk, D., Beekman, J. M. and van der Ent, C. K. 2021. A new era for people with cystic fibrosis. *Eur J Pediatr* 180(9), pp. 2731-2739. doi: 10.1007/s00431-021-04168-y

Bikandi, J., San Millán, R., Rementeria, A. and Garaizar, J. 2004. *In silico* analysis of complete bacterial genomes: PCR, AFLP-PCR and endonuclease restriction. *Bioinformatics* 20(5), pp. 798-799. doi: 10.1093/bioinformatics/btg491

Bittar, F. and Rolain, J. M. 2010. Detection and accurate identification of new or emerging bacteria in cystic fibrosis patients. *Clin Microbiol Infect* 16(7), pp. 809-820. doi: 10.1111/j.1469-0691.2010.03236.x

Blackburn, L., Brownlee, K., Conway, S. and Denton, M. 2004. 'Cepacia syndrome' with *Burkholderia multivorans*, 9 years after initial colonization. *J Cyst Fibros* 3(2), pp. 133-134. doi: 10.1016/j.jcf.2004.03.007

Bopaka, R. G., El Khattabi, W., Janah, H., Jabri, H. and Afif, H. 2015. Bronchiectasis: a bacteriological profile. *Pan Afr Med J* 22, p. 378. doi: 10.11604/pamj.2015.22.378.7775

Borneman, J. and Triplett, E. W. 1997. Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. *Appl Environ Microbiol* 63(7), pp. 2647-2653. doi: 10.1128/aem.63.7.2647-2653.1997

Bouquet, J. et al. 2020. Microbial burden and viral exacerbations in a longitudinal multicenter COPD cohort. *Respir Res* 21(1), p. 77. doi: 10.1186/s12931-020-01340-0

Brode, S. K., Ling, S. C. and Chapman, K. R. 2012. Alpha-1 antitrypsin deficiency: a commonly overlooked cause of lung disease. *CMAJ* 184(12), pp. 1365-1371. doi: 10.1503/cmaj.111749

Burden, D. W. 2008. Guide to the Homogenization of Biological Samples. *Random Primers* (7), pp. 1-14.

Burki, T. K. 2017. The economic cost of respiratory disease in the UK. *Lancet Respir Med* 5(5), p. 381. doi: 10.1016/S2213-2600(17)30108-X

Burns, J. L. and Rolain, J. M. 2014. Culture-based diagnostic microbiology in cystic fibrosis: can we simplify the complexity? *J Cyst Fibros* 13(1), pp. 1-9. doi: 10.1016/j.jcf.2013.09.004

Caballero, J. D. et al. 2014. Microbiological diagnostic procedures for respiratory cystic fibrosis samples in Spain: towards standard of care practices. *BMC Microbiol* 14, p. 335. doi: 10.1186/s12866-014-0335-y

Cabrera-Rubio, R., Garcia-Núñez, M., Setó, L., Antó, J. M., Moya, A., Monsó, E. and Mira, A. 2012. Microbiome diversity in the bronchial tracts of patients with chronic obstructive pulmonary disease. *J Clin Microbiol* 50(11), pp. 3562-3568. doi: 10.1128/JCM.00767-12

Camus, L., Briaud, P., Vandenesch, F. and Moreau, K. 2021. How Bacterial Adaptation to Cystic Fibrosis Environment Shapes Interactions Between. *Front Microbiol* 12, p. 617784. doi: 10.3389/fmicb.2021.617784

Capizzani, C. P. D. C., Caçador, N. C., Marques, E. A., Levy, C. E., Tonani, L., Torres, L. A. G. M. and Darini, A. L. D. C. 2018. A practical molecular identification of nonfermenting Gram-negative bacteria from cystic fibrosis. *Braz J Microbiol* 49(2), pp. 422-428. doi: 10.1016/j.bjm.2017.07.002

Caporaso, J. G. et al. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A* 108 Suppl 1, pp. 4516-4522. doi: 10.1073/pnas.1000080107

Carreto, L. et al. 2020. Utility of routine screening for alpha-1 antitrypsin deficiency in patients with bronchiectasis. *Thorax* 75(7), pp. 592-593. doi: 10.1136/thoraxjnl-2019-214195

Cavallazzi, R. and Ramirez, J. 2020. Community-acquired pneumonia in chronic obstructive pulmonary disease. *Curr Opin Infect Dis* 33(2), pp. 173-181. doi: 10.1097/QCO.0000000000000639

Caverly, L. J. and LiPuma, J. J. 2018. Cystic fibrosis respiratory microbiota: unraveling complexity to inform clinical practice. *Expert Rev Respir Med* 12(10), pp. 857-865. doi: 10.1080/17476348.2018.1513331

Caverly, L. J., Zhao, J. and LiPuma, J. J. 2015. Cystic fibrosis lung microbiome: opportunities to reconsider management of airway infection. *Pediatr Pulmonol* 50 Suppl 40, pp. S31-38. doi: 10.1002/ppul.23243

Chandrasekaran, R., Mac Aogáin, M., Chalmers, J. D., Elborn, S. J. and Chotirmall, S. H. 2018. Geographic variation in the aetiology, epidemiology and microbiology of bronchiectasis. *BMC Pulm Med* 18(1), p. 83. doi: 10.1186/s12890-018-0638-0

Chen, C. L. et al. 2020. The Roles of Bacteria and Viruses in Bronchiectasis Exacerbation: A Prospective Study. *Arch Bronconeumol (Engl Ed)* 56(10), pp. 621-629. doi: 10.1016/j.arbr.2019.12.014

Chen, Y., Liu, Y. and Wang, X. 2017. Spatiotemporal variation of bacterial and archaeal communities in sediments of a drinking reservoir, Beijing, China. *Appl Microbiol Biotechnol* 101(8), pp. 3379-3391. doi: 10.1007/s00253-016-8019-1

Chiaradia, L. et al. 2017. Dissecting the mycobacterial cell envelope and defining the composition of the native mycomembrane. *Sci Rep* 7(1), p. 12807. doi: 10.1038/s41598-017-12718-4

Cho, E. J., Sung, H., Park, S. J., Kim, M. N. and Lee, S. O. 2013. *Rothia mucilaginosa* pneumonia diagnosed by quantitative cultures and intracellular organisms of bronchoalveolar lavage in a lymphoma patient. *Ann Lab Med* 33(2), pp. 145-149. doi: 10.3343/alm.2013.33.2.145

Chomczynski, P. and Sacchi, N. 2006. The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nat Protoc* 1(2), pp. 581-585. doi: 10.1038/nprot.2006.83

Chuang, L. Y., Cheng, Y. H. and Yang, C. H. 2013. Specific primer design for the polymerase chain reaction. *Biotechnol Lett* 35(10), pp. 1541-1549. doi: 10.1007/s10529-013-1249-8

Claassen-Weitz, S., Gardner-Lubbe, S., Mwaikono, K. S., du Toit, E., Zar, H. J. and Nicol, M. P. 2020. Optimizing 16S rRNA gene profile analysis from low biomass nasopharyngeal and induced sputum specimens. *BMC Microbiol* 20(1), p. 113. doi: 10.1186/s12866-020-01795-7

Cobo, F., Borrego, J., Rodríguez-Granger, J., Sampedro, A. and Navarro-Marí, J. M. 2018. A rare case of pleural infection due to *Propionibacterium acnes* (*Cutibacterium acnes*). *Rev Esp Quimioter* 31(2), pp. 173-174.

Cohen-Cymberknoh, M., Kerem, E., Ferkol, T. and Elizur, A. 2013. Airway inflammation in cystic fibrosis: molecular mechanisms and clinical implications. *Thorax* 68(12), pp. 1157-1162. doi: 10.1136/thoraxjnl-2013-203204

Cole, J. R. et al. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* 37(11), pp. D141-145. doi: 10.1093/nar/gkn879

Collins, A. 2019. *Ex vivo modelling of bacterial communities in sputum from cystic fibrosis pulmonary infection.* Cardiff University.

Conese, M., Castellani, S., D'Oria, S., Di Gioia, S. and Montemurro, P. 2017. Role of Neutrophils in Cystic Fibrosis Lung Disease.1st ed. Khajah, M. (1st ed). In: *Role of Neutrophils in Disease Pathogenesis*. p. 180, Vienna, Austria:IntechOpen.

Coward, A., Kenna, D. T., Perry, C., Martin, K., Doumith, M. and Turton, J. F. 2016. Use of *nrdA* gene sequence clustering to estimate the prevalence of different *Achromobacter* species among Cystic Fibrosis patients in the UK. *J Cyst Fibros* 15(4), pp. 479-485. doi: 10.1016/j.jcf.2015.09.005

Cox, M. J. et al. 2017. Longitudinal assessment of sputum microbiome by sequencing of the 16S rRNA gene in non-cystic fibrosis bronchiectasis patients. *PLoS One* 12(2), p. e0170622. doi: 10.1371/journal.pone.0170622

Cuthbertson, L. et al. 2020. Lung function and microbiota diversity in cystic fibrosis. *Microbiome* 8(1), p. 45. doi: 10.1186/s40168-020-00810-3

Darch, S. E. et al. 2015. Recombination is a key driver of genomic and phenotypic diversity in a *Pseudomonas aeruginosa* population during cystic fibrosis infection. Sci Rep 5, p. 7649. doi: 10.1038/srep07649

De Baets, F., Schelstraete, P., Van Daele, S., Haerynck, F. and Vaneechoutte, M. 2007. *Achromobacter xylosoxidans* in cystic fibrosis: prevalence and clinical relevance. *J Cyst Fibros* 6(1), pp. 75-78. doi: 10.1016/j.jcf.2006.05.011

De Soyza, A. and Aliberti, S. 2017. Bronchiectasis and *Aspergillus*: How are they linked? *Med Mycol* 55(1), pp. 69-81. doi: 10.1093/mmy/myw109

Degiacomi, G., Sammartino, J. C., Chiarelli, L. R., Riabova, O., Makarov, V. and Pasca, M. R. 2019. *Mycobacterium abscessus*, an Emerging and Worrisome Pathogen among Cystic Fibrosis Patients. *Int J Mol Sci* 20(23), doi: 10.3390/ijms20235868

Delhaes, L. et al. 2012. The airway microbiota in cystic fibrosis: a complex fungal and bacterial community-implications for therapeutic management. *PLoS One* 7(4), p. e36313. doi: 10.1371/journal.pone.0036313

DeSantis, T. Z. et al. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72(7), pp. 5069-5072. doi: 10.1128/AEM.03006-05

Deurenberg, R. H. et al. 2017. Application of next generation sequencing in clinical microbiology and infection prevention. *J Biotechnol* 243, pp. 16-24. doi: 10.1016/j.jbiotec.2016.12.022

Dey, S., Eapen, M. S., Chia, C., Gaikwad, A. V., Wark, P. A. B. and Sohal, S. S. 2022. Pathogenesis, clinical features of asthma COPD overlap, and therapeutic modalities. *Am J Physiol Lung Cell Mol Physiol* 322(1), pp. L64-L83. doi: 10.1152/ajplung.00121.2021

Diancourt, L., Passet, V., Verhoef, J., Grimont, P. A. and Brisse, S. 2005. Multilocus sequence typing of *Klebsiella pneumoniae* nosocomial isolates. *J Clin Microbiol* 43(8), pp. 4178-4182. doi: 10.1128/JCM.43.8.4178-4182.2005

Dicker, A. J. et al. 2021. The sputum microbiome and clinical outcomes in patients with bronchiectasis: a prospective observational study. *Lancet Respir Med* 9(8), pp. 885-896. doi: 10.1016/S2213-2600(20)30557-9

Dickson, R. P., Erb-Downward, J. R. and Huffnagle, G. B. 2013. The role of the bacterial microbiome in lung disease. *Expert Rev Respir Med* 7(3), pp. 245-257. doi: 10.1586/ers.13.24

Dickson, R. P., Erb-Downward, J. R., Martinez, F. J. and Huffnagle, G. B. 2016. The Microbiome and the Respiratory Tract. *Annu Rev Physiol* 78, pp. 481-504. doi: 10.1146/annurev-physiol-021115-105238

Ditz, B., Christenson, S., Rossen, J., Brightling, C., Kerstjens, H. A. M., van den Berge, M. and Faiz, A. 2020. Sputum microbiome profiling in COPD: beyond singular pathogen detection. *Thorax* 75(4), pp. 338-344. doi: 10.1136/thoraxjnl-2019-214168

Doi, Y., Poirel, L., Paterson, D. L. and Nordmann, P. 2008. Characterization of a naturally occurring class D beta-lactamase from *Achromobacter xylosoxidans*. *Antimicrob Agents Chemother* 52(6), pp. 1952-1956. doi: 10.1128/aac.01463-07

Dong, D. et al. 2015. Survey and rapid detection of *Klebsiella pneumoniae* in clinical samples targeting the *rcsA* gene in Beijing, China. *Front Microbiol* 6, p. 519. doi: 10.3389/fmicb.2015.00519

Doud, M. S., Light, M., Gonzalez, G., Narasimhan, G. and Mathee, K. 2010. Combination of 16S rRNA variable regions provides a detailed analysis of bacterial community dynamics in the lungs of cystic fibrosis patients. *Hum Genomics* 4(3), pp. 147-169. doi: 10.1186/1479-7364-4-3-147

Drengenes, C., Wiker, H. G., Kalananthan, T., Nordeide, E., Eagan, T. M. L. and Nielsen, R. 2019. Laboratory contamination in airway microbiome studies. *BMC Microbiol* 19(1), p. 187. doi: 10.1186/s12866-019-1560-1

Duan, K., Dammel, K., Stein, J., Rabin, H. and G., S. M. 2003. Modulation of *Pseudomonas aeruginosa* gene expression by host microflora through interspecies communication. *Mol Microbiol* 50(5), pp. 1477-1491. doi: 10.1046/j.1365-2958.2003.03803.x

Dupont, C., Michon, A. L., Jumas-Bilak, E., Nørskov-Lauritsen, N., Chiron, R. and Marchandin, H. 2015. Intrapatient diversity of *Achromobacter* spp. involved in chronic colonization of Cystic Fibrosis airways. *Infect Genet Evol* 32, pp. 214-223. doi: 10.1016/j.meegid.2015.03.012

Díez-Aguilar, M. et al. 2019. Antimicrobial susceptibility of non-fermenting Gram-negative pathogens isolated from cystic fibrosis patients. *Int J Antimicrob Agents* 53(1), pp. 84-88. doi: 10.1016/j.ijantimicag.2018.09.001

Döşkaya, M. et al. 2011. Degree and frequency of inhibition in a routine real-time PCR detecting *Pneumocystis jirovecii* for the diagnosis of *Pneumocystis* pneumonia in Turkey. *J med microbiol* 60(7), doi: 10.1099/jmm.0.030775-0

Edgar, R. C. 2017. UNBIAS: An attempt to correct abundance bias in 16S sequencing, with limited success. *bioRxiv* 2017:124149. doi: 10.1101/124149

Edwards, B. D. et al. 2017. Prevalence and Outcomes of *Achromobacter* Species Infections in Adults with Cystic Fibrosis: a North American Cohort Study. *J Clin Microbiol* 55(7), pp. 2074-2085. doi: 10.1128/JCM.02556-16

Einarsson, G. G., Comer, D. M., McIlreavey, L., Parkhill, J., Ennis, M., Tunney, M. M. and Elborn, J. S. 2016. Community dynamics and the lower airway microbiota in stable chronic obstructive pulmonary disease, smokers and healthy non-smokers. *Thorax* 71(9), pp. 795-803. doi: 10.1136/thoraxjnl-2015-207235

El Salabi, A., Borra, P. S., Toleman, M. A., Samuelsen, O. and Walsh, T. R. 2012. Genetic and biochemical characterization of a novel metallo-beta-lactamase, TMB-1, from an *Achromobacter xylosoxidans* strain isolated in Tripoli, Libya. *Antimicrob Agents Chemother* 56(5), pp. 2241-2245. doi: 10.1128/aac.05640-11

El-Korashy, R. I. M. and El-Sherif, R. H. 2014. Gram negative organisms as a cause of acute exacerbation of COPD. *Egypt J Chest Dis Tuberc* 63(2), pp. 345-349.

Elborn, J. S. 2008. Identification and management of unusual pathogens in cystic fibrosis. *J R Soc Med* 101 Suppl 1, pp. S2-5. doi: 10.1258/jrsm.2008.s18002

Elnifro, E. M., Ashshi, A. M., Cooper, R. J. and Klapper, P. E. 2000. Multiplex PCR: optimization and application in diagnostic virology. *Clin Microbiol Rev* 13(4), pp. 559-570. doi: 10.1128/CMR.13.4.559

Enaud, R., Prevel, R., Ciarlo, E., Beaufils, F., Wieërs, G., Guery, B. and Delhaes, L. 2020. The Gut-Lung Axis in Health and Respiratory Diseases: A Place for Inter-Organ and Inter-Kingdom Crosstalks. *Front Cell Infect Microbiol* 10, p. 9. doi: 10.3389/fcimb.2020.00009

Erb-Downward, J. R. et al. 2011. Analysis of the lung microbiome in the "healthy" smoker and in COPD. *PLoS One* 6(2), p. e16384. doi: 10.1371/journal.pone.0016384

Ercolini, D. and Cocolin, L. 2014. IDENTIFICATION METHODS | Culture-Independent Techniques.*Encyclopedia of Food Microbiology*. 2 ed. Amsterdam:Elsevier.

Erwin, A. L. and Smith, A. L. 2007. Nontypeable *Haemophilus influenzae*: understanding virulence and commensal behavior. *Trends Microbiol* 15(8), pp. 355-362. doi: 10.1016/j.tim.2007.06.004

Escalas, A., Hale, L., Voordeckers, J. W., Yang, Y., Firestone, M. K., Alvarez-Cohen, L. and Zhou, J. 2019. Microbial functional diversity: From concepts to applications. *Ecol Evol* 9(20), pp. 12000-12016. doi: 10.1002/ece3.5670

Escobar-Zepeda, A., Vera-Ponce de León, A. and Sanchez-Flores, A. 2015. The Road to Metagenomics: From Microbiology to DNA Sequencing Technologies and Bioinformatics. *Front Genet* 6, p. 348. doi: 10.3389/fgene.2015.00348

Evans, J., Sheneman, L. and Foster, J. 2006. Relaxed neighbor joining: a fast distance-based phylogenetic tree construction method. *J Mol Evol* 62(6), pp. 785-792. doi: 10.1007/s00239-005-0176-2

Falagas, M. E., Kasiakou, S. K. and Michalopoulos, A. 2005. Treatment of multidrug-resistant *Pseudomonas aeruginosa* and *Acinetobacter baumannii* pneumonia. *J Cyst Fibros* 4(2), pp. 149-150. doi: 10.1016/j.jcf.2005.02.003

Faner, R. et al. 2017. The microbiome in respiratory medicine: current challenges and future perspectives. *Eur Respir J* 49(4), doi: 10.1183/13993003.02086-2016

Faraji, M., Yamini, Y. and Gholami, M. 2019. Recent Advances and Trends in Applications of Solid-Phase Extraction Techniques in Food and Environmental Analysis. *Chromatographia* 82, pp. 1207–1249. doi: 10.1007/s10337-019-03726-9

Farrell Jr, R. E. 2017. *Chapter 3 - RNA Isolation Strategies*. 5th ed. Cambridge, Massachusetts: Academic Press.

Fauroux, B. et al. 2004. *Burkholderia cepacia* is associated with pulmonary hypertension and increased mortality among cystic fibrosis patients. *J Clin Microbiol* 42(12), pp. 5537-5541. doi: 10.1128/JCM.42.12.5537-5541.2004

Fenker, D. E. et al. 2018. A Comparison between Two Pathophysiologically Different yet Microbiologically Similar Lung Diseases: Cystic Fibrosis and Chronic Obstructive Pulmonary Disease. *Int J Respir Pulm Med* 5(2), doi: 10.23937/2378-3516/1410098

Filipic, B. M., Malesevic, M., Vasiljevic, Z., Lukic, J., Novovkic, K., Kojic, M. and Jovkic, B. 2017. Uncovering Differences in Virulence Markers Associated with *Achromobacter* Species of CF and Non-CF Origin. *Front Cellular Infect Microbiol* 7(224), doi: doi:10.3389/fcimb.2017.00224

Filkins, L. M. and O'Toole, G. A. 2015. Cystic Fibrosis Lung Infections: Polymicrobial, Complex, and Hard to Treat. *PLoS Pathog* 11(12), p. e1005258. doi: 10.1371/journal.ppat.1005258

Finch, S. et al. 2018. Sex differences in bronchiectasis patient characteristics: an analysis of the EMBARC cohort. *Eur Respir J* 52, p. 2282. doi: 10.1183/13993003.congress-2018.PA2282

Firmida, M. C., Pereira, R. H., Silva, E. A., Marques, E. A. and Lopes, A. J. 2016. Clinical impact of *Achromobacter xylosoxidans* colonization/infection in patients with cystic fibrosis. *Braz J Med Biol Res* 49(4), p. e5097. doi: 10.1590/1414-431x20155097

Fleurbaaij, F. et al. 2018. Proteomic identification of Axc, a novel beta-lactamase with carbapenemase activity in a meropenem-resistant clinical isolate of *Achromobacter xylosoxidans*. *Sci Rep* 8(1), p. 8181. doi: doi:10.1038/s41598-018-26079-z

Flight, W. G., Hough, N. E. and Chapman, S. J. 2020. Outcomes of pulmonary *mycobacterium abscessus* infection. *Int J Mycobacteriol* 9(1), pp. 48-52. doi: 10.4103/ijmy.ijmy_3_20

Flight, W. G. et al. 2015. Rapid Detection of Emerging Pathogens and Loss of Microbial Diversity Associated with Severe Lung Disease in Cystic Fibrosis. *J Clin Microbiol* 53(7), pp. 2022-2029. doi: 10.1128/JCM.00432-15

Florio, W., Cappellini, S., Giordano, C., Vecchione, A., Ghelardi, E. and Lupetti, A. 2019. A new culture-based method for rapid identification of microorganisms in polymicrobial blood cultures by MALDI-TOF MS. *BMC Microbiol* 19(1), p. 267. doi: 10.1186/s12866-019-1641-1

Flynn, J. M., Niccum, D., Dunitz, J. M. and Hunter, R. C. 2016. Evidence and Role for Bacterial Mucin Degradation in Cystic Fibrosis Airway Disease. *PLoS Pathog* 12(8), p. e1005846. doi: 10.1371/journal.ppat.1005846

Fouhy, F., Clooney, A. G., Stanton, C., Claesson, M. J. and Cotter, P. D. 2016. 16S rRNA gene sequencing of mock microbial populations- impact of DNA extraction method, primer choice and sequencing platform. *BMC Microbiol* 16(1), p. 123. doi: 10.1186/s12866-016-0738-z

Frank, J. A., Reich, C. I., Sharma, S., Weisbaum, J. S., Wilson, B. A. and Olsen, G. J. 2008. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl Environ Microbiol* 74(8), pp. 2461-2470. doi: 10.1128/AEM.02272-07

Françoise, A. and Héry-Arnaud, G. 2020. The Microbiome in Cystic Fibrosis Pulmonary Disease. *Genes* 11(5), doi: 10.3390/genes11050536

Gade, S. S., Norskov-Lauritsen, N. and Ridderberg, W. 2017. Prevalence and species distribution of *Achromobacter* sp. cultured from cystic fibrosis patients attending the Aarhus centre in Denmark. *J Med Microbiol* 66(5), pp. 686-689. doi: 10.1099/jmm.0.000499

Garcia, M. et al. 2017. Comparison of eMAG™ versus NucliSENS® EasyMAG® performance on clinical specimens. *J Clinical Virol* 88, pp. 52-57. doi: 10.1016/j.jcv.2017.01.004

Garcia-Nuñez, M. et al. 2020. The Respiratory Microbiome in Cystic Fibrosis: Compartment Patterns and Clinical Relationships in Early Stage Disease. *Front Microbiol* 11, p. 1463. doi: 10.3389/fmicb.2020.01463

Garcia-Nuñez, M. et al. 2014. Severity-related changes of bronchial microbiome in chronic obstructive pulmonary disease. *J Clin Microbiol* 52(12), pp. 4217-4223. doi: 10.1128/JCM.01967-14

Gargis, A. S., Kalman, L. and Lubin, I. M. 2016. Assuring the Quality of Next-Generation Sequencing in Clinical Microbiology and Public Health Laboratories. *J Clin Microbiol* 54(12), pp. 2857-2865. doi: 10.1128/JCM.00949-16

Garnier, S., Noam, R., Robert, R., Camargo, P. A., Sciaini, M. and Scherer, C. 2021. Viridis-Colorblind Friendly Color Maps for R. Available at: https://sjmgarnier.github.io/viridis/ [Accessed: 1 March 2020].

Garrigos, T. et al. 2021. Development of a database for the rapid and accurate routine identification of *Achromobacter* species by matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF MS). *Clin Microbiol Infect* 27(1), pp. 126.e121-126.e125. doi: 10.1016/j.cmi.2020.03.031

Gentzsch, M. and Mall, M. A. 2018. Ion Channel Modulators in Cystic Fibrosis. *Chest* 154(2), pp. 383-393. doi: 10.1016/j.chest.2018.04.036

Ghyselinck, J., Pfeiffer, S., Heylen, K., Sessitsch, A. and De Vos, P. 2013. The effect of primer choice and short read sequences on the outcome of 16S rRNA gene based diversity studies. *PLoS One* 8(8), p. e71360. doi: 10.1371/journal.pone.0071360

Gill, C., van de Wijgert, J. H., Blow, F. and Darby, A. C. 2016. Evaluation of Lysis Methods for the Extraction of Bacterial DNA for Analysis of the Vaginal Microbiota. *PLoS One* 11(9), p. e0163148. doi: 10.1371/journal.pone.0163148

Giuliano, C., Patel, C. R. and Kale-Pradhan, P. B. 2019. A Guide to Bacterial Culture Identification And Results Interpretation. *P&T* 44(4), pp. 192-200.

Goddard, A. F. et al. 2012. Direct sampling of cystic fibrosis lungs indicates that DNA-based analyses of upper-airway specimens can misrepresent lung microbiota. *Proc Natl Acad Sci USA* 109(34), pp. 13769-13774. doi: 10.1073/pnas.1107435109

Gosse, L., Amrane, S., Mailhe, M., Dubourg, G. and Lagier, J. C. 2019. : An unusual cause of community-acquired pneumonia. *IDCases* 17, p. e00572. doi: 10.1016/j.idcr.2019.e00572

Greathouse, K. L., Sinha, R. and Vogtmann, E. 2019. DNA extraction for human microbiome studies: the issue of standardization. *Genome Biol* 20(1), p. 212. doi: 10.1186/s13059-019-1843-8

Grinwis, M. E., Sibley, C. D., Parkins, M. D., Eshaghurshan, C. S., Rabin, H. R. and Surette, M. G. 2010. Characterization of *Streptococcus milleri* group isolates from expectorated sputum of adult patients with cystic fibrosis. *J Clin Microbiol* 48(2), pp. 395-401. doi: 10.1128/JCM.01807-09

Grochowalska, A., Kozioł-Montewka, M. and Sobieszczańska, A. 2017. Analysis of *Acinetobacter baumannii* resistance patterns in patients with chronic obstructive pulmonary disease (COPD) in terms of choice of effective empiric antibiotic therapy. *Ann Agric Environ Med* 24(2), pp. 307-311. doi: 10.26444/aaem/74710

Gupta, V., Sharma, S., Singhal, L., Soni, R. and Chander, J. 2018. *Pseudomonas Monteilii* an Emerging Pathogen in Meningoencephalitis. *JCDR* 12, doi: 10.7860/JCDR/2018/32670.11437

Gurevich, A., Saveliev, V., Vyahhi, N. and Tesler, G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29(8), pp. 1072-1075. doi: 10.1093/bioinformatics/btt086

Haas, B. J. et al. 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* 21(3), pp. 494-504. doi: 10.1101/gr.112730.110

Haiko, J., Saeedi, B., Bagger, G., Karpati, F. and Özenci, V. 2019. Coexistence of *Candida* species and bacteria in patients with cystic fibrosis. *Eur J Clin Microbiol Infect Dis* 38(6), pp. 1071-1077. doi: 10.1007/s10096-019-03493-3

Haldar, K. et al. 2020. The sputum microbiome is distinct between COPD and health, independent of smoking history. *Respir Res* 21(1), p. 183. doi: 10.1186/s12931-020-01448-3

Hall, A. T., Zovanyi, A. K., Christensen, D. R., Koehler, J. W. and Minogue, T. D. 2013. Evaluation of inhibitor-resistant real-time PCR methods for diagnostics in clinical and environmental samples. *PLoS One* 8(9), p. e73845. doi: 10.1371/journal.pone.0073845

Hallmaier-Wacker, L. K., Lueert, S., Roos, C. and Knauf, S. 2018. The impact of storage buffer, DNA extraction method, and polymerase on microbial analysis. *Sci Rep* 8(1), p. 6292. doi: 10.1038/s41598-018-24573-y

Harness-Brumley, C. L., Elliott, A. C., Rosenbluth, D. B., Raghavan, D. and Jain, R. 2014. Gender differences in outcomes of patients with cystic fibrosis. *J Womens Health (Larchmt)* 23(12), pp. 1012-1020. doi: 10.1089/jwh.2014.4985

Hatziagorou, E. et al. 2020. Changing epidemiology of the respiratory bacteriology of patients with cystic fibrosis-data from the European cystic fibrosis society patient registry. *J Cyst Fibros* 19(3), pp. 376-383. doi: 10.1016/j.jcf.2019.08.006

Hauser, P. M., Bernard, T., Greub, G., Jaton, K., Pagni, M. and Hafen, G. M. 2014. Microbiota present in cystic fibrosis lungs as revealed by whole genome sequencing. *PLoS One* 9(3), p. e90934. doi: 10.1371/journal.pone.0090934

Head, M. G., Fitchett, J. R., Cooke, M. K., Wurie, F. B., Hayward, A. C., Lipman, M. C. and Atun, R. 2014. Investments in respiratory infectious disease research 1997-2010: a systematic analysis of UK funding. *BMJ Open* 4(3), p. e004600. doi: 10.1136/bmjopen-2013-004600

Heras, J., Domínguez, C., Mata, E., Pascual, V., Lozano, C., Torres, C. and Zarazaga, M. 2015. GelJ--a tool for analyzing DNA fingerprint gel images. *BMC Bioinform* 16, p. 270. doi: 10.1186/s12859-015-0703-0

Higashi, Y. et al. 2017. Pulmonary Actinomycosis Mimicking Pulmonary Aspergilloma and a Brief Review of the Literature. *Intern Med* 56(4), pp. 449-453. doi: 10.2169/internalmedicine.56.7620

Hilty, M. et al. 2010. Disordered microbial communities in asthmatic airways. *PLoS One* 5(1), p. e8578. doi: 10.1371/journal.pone.0008578

Hindiyeh, M. et al. 2019. Comparison of the new fully automated extraction platform eMAG to the MagNA PURE 96 and the well-established easyMAG for detection of common human respiratory viruses. *PLoS One* 14(2), p. e0211079. doi: 10.1371/journal.pone.0211079

Holgate, S. T. 2007. Priorities for respiratory research in the UK. *Thorax* 62(1), pp. 5-7. doi: 10.1136/thx.2006.073882

Hosomi, K. et al. 2017. Method for preparing DNA from feces in guanidine thiocyanate solution affects 16S rRNA-based profiling of human microbiota diversity. *Sci Rep* 7(1), p. 4339. doi: 10.1038/s41598-017-04511-0

Hu, Y. et al. 2015. Genomic insights into intrinsic and acquired drug resistance mechanisms in *Achromobacter xylosoxidans*. *Antimicrob Agents Chemother* 59(2), pp. 1152-1161. doi: 10.1128/aac.04260-14

Humbert, M. V. and Christodoulides, M. 2019. Atypical, Yet Not Infrequent, Infections with *Neisseria* Species. *Pathogens* 9(1), pp. 1-27. doi: 10.3390/pathogens9010010

Hurley, M. N. 2018. *Staphylococcus aureus* in cystic fibrosis: problem bug or an innocent bystander? *Breathe* 14(2), pp. 87-90. doi: 10.1183/20734735.014718

Imai, K., Iinuma, T. and Sato, S. 2021. Relationship between the oral cavity and respiratory diseases: Aspiration of oral bacteria possibly contributes to the progression of lower airway inflammation. *Jpn Dent Sci Rev* 57, pp. 224-230. doi: 10.1016/j.jdsr.2021.10.003

Inchingolo, R., Pierandrei, C., Montemurro, G., Smargiassi, A., Lohmeyer, F. M. and Rizzi, A. 2021. Antimicrobial Resistance in Common Respiratory Pathogens of Chronic Bronchiectasis Patients: A Literature Review. *Antibiotics* 10(3), doi: 10.3390/antibiotics10030326

Islam, M. S., Aryasomayajula, A. and Selvaganapathy, P. R. 2017. A Review on Macroscale and Microscale Cell Lysis Methods. *Micromachines* 8(3), doi: 10.3390/mi8030083

Isler, B., Kidd, T. J., Stewart, A. G., Harris, P. and Paterson, D. L. 2020. Infections and Treatment Options. *Antimicrob Agents Chemother* 64(11), doi: 10.1128/AAC.01025-20

Iyobe, S., Kusadokoro, H., Takahashi, A., Yomoda, S., Okubo, T., Nakamura, A. and O'Hara, K. 2002. Detection of a variant metallo-beta-lactamase, IMP-10, from two unrelated strains of *Pseudomonas aeruginosa* and an *alcaligenes xylosoxidans* strain. *Antimicrob Agents Chemother* 46(6), pp. 2014-2016. doi: 10.1128/aac.46.6.2014-2016.2002

Jakobsen, T. H. et al. 2013. Complete Genome Sequence of the Cystic Fibrosis Pathogen *Achromobacter xylosoxidans* NH44784-1996 Complies with Important Pathogenic Phenotypes. *PLoS One* 8(7), doi: 10.1371/journal.pone.0068484

Jean-Pierre, F., Vyas, A., Hampton, T. H., Henson, M. A. and O'Toole, G. A. 2021. One versus Many: Polymicrobial Communities and the Cystic Fibrosis Airway. *mBio* 12(2), doi: 10.1128/mBio.00006-21

Jeukens, J., Freschi, L., Vincent, A. T., Emond-Rheault, J. G., Kukavica-Ibrulj, I., Charette, S. J. and Levesque, R. C. 2017. A Pan-Genomic Approach to Understand the Basis of Host Adaptation in *Achromobacter*. *Genome Biol Evol* 9(4), pp. 1030-1046. doi: 10.1093/gbe/evx061

Jia, B. et al. 2017. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res* 45(D1), pp. D566-D573. doi: 10.1093/nar/gkw1004

Jorth, P. et al. 2019. Direct Lung Sampling Indicates That Established Pathogens Dominate Early Infections in Children with Cystic Fibrosis. *Cell Rep* 27(4), pp. 1190-1204.e1193. doi: 10.1016/j.celrep.2019.03.086

Kanellopoulou, M., Pournaras, S., Iglezos, H., Skarmoutsou, N., Papafrangas, E. and Maniatis, A. N. 2004. Persistent colonization of nine cystic fibrosis patients with an *Achromobacter* (*Alcaligenes*) *xylosoxidans* clone. *Eur J Clin Microbiol Infect Dis* 23(4), pp. 336-339. doi: 10.1007/s10096-004-1105-9

Karstens, L., Siddiqui, N. Y., Zaza, T., Barstad, A., Amundsen, C. L. and Sysoeva, T. A. 2021. Benchmarking DNA isolation kits used in analyses of the urinary microbiome. *Sci Rep* 11(1), p. 6186. doi: 10.1038/s41598-021-85482-1

Kayani, K., Mohammed, R. and Mohiaddin, H. 2018. Cystic Fibrosis-Related Diabetes. *Front Endocrinol (Lausanne)* 9, p. 20. doi: 10.3389/fendo.2018.00020

Kearns, D. B. 2010. A field guide to bacterial swarming motility. *Nat Rev Microbiol* 8(9), pp. 634-644. doi: 10.1038/nrmicro2405

Kelly, M. M., Efthimiadis, A. and Hargreave, F. E. 2001. Induced Sputum. In: Rogers, D.F. and Donnelly, L.E. *Human Airway Inflammation*. 56th ed. pp.77-91, New York:SpringerLink.

Kelly, N. et al. 2021. Periodontal status and chronic obstructive pulmonary disease (COPD) exacerbations: a systematic review. *BMC Oral Health* 21(1), p. 425. doi: 10.1186/s12903-021-01757-z

Khademi, S. M. H., Gabrielaite, M., Paulsson, M., Knulst, M., Touriki, E., Marvig, R. L. and Påhlman, L. I. 2021. Genomic and Phenotypic Evolution of *Achromobacter xylosoxidans* during Chronic Airway Infections of Patients with Cystic Fibrosis. *mSystems*, p. e0052321. doi: 10.1128/mSystems.00523-21

Kim, B. R. et al. 2017. Deciphering Diversity Indices for a Better Understanding of Microbial Communities. *J Microbiol Biotechnol* 27(12), pp. 2089-2093. doi: 10.4014/jmb.1709.09027

Kim, M. and Chun, J. 2014. *Chapter 4 - 16S rRNA Gene-Based Identification of Bacteria and Archaea using the EzTaxon Server*. 1st ed. Cambridge, Massachusetts:Academic Press.

King, P. T. 2015. Inflammation in chronic obstructive pulmonary disease and its role in cardiovascular disease and lung cancer. *Clin Transl Med* 4(1), p. 68. doi: 10.1186/s40169-015-0068-z

King, P. T., Holdsworth, S. R., Freezer, N. J., Villanueva, E., Farmer, M. W., Guy, P. and Holmes, P. W. 2010. Lung diffusing capacity in adult bronchiectasis: a longitudinal study. *Respir Care* 55(12), pp. 1686-1692.

Kitsios, G. D. et al. 2018. Microbiome in lung explants of idiopathic pulmonary fibrosis: a case-control study in patients with end-stage fibrosis. *Thorax* 73(5), pp. 481-484. doi: 10.1136/thoraxjnl-2017-210537

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M. and Glöckner, F. O. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 41(1), p. e1. doi: 10.1093/nar/gks808

Kolia-Diafouka, P., Godreuil, S., Bourdin, A., Carrère-Kremer, S., Kremer, L., Van de Perre, P. and Tuaillon, E. 2018. Optimized Lysis-Extraction Method Combined With. *Front Microbiol* 9, p. 2224. doi: 10.3389/fmicb.2018.02224

Konstantinović, N., Ćirković, I., Đukić, S., Marić, V. and Božić, D. D. 2017. Biofilm formation of *Achromobacter xylosoxidans* on contact lens. *Acta Microbiol Immunol Hung* 64(3), pp. 293-300. doi: 10.1556/030.64.2017.005

Kovach, K. et al. 2017. Evolutionary adaptations of biofilms infecting cystic fibrosis lungs promote mechanical toughness by adjusting polysaccharide production. *NPJ Biofilms Microbiomes* 3, p. 1. doi: 10.1038/s41522-016-0007-9

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. and Schloss, P. D. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 79(17), pp. 5112-5120. doi: 10.1128/AEM.01043-13

Krueger, F. 2021. *Trim Galore!* Available at: https://github.com/FelixKrueger/TrimGalore [Accessed: 8 February 2018].

Kumar, S., Stecher, G. and Tamura, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 33(7), pp. 1870-1874. doi: 10.1093/molbev/msw054

Labombarda, F., Saloux, E., Brouard, J., Bergot, E. and Milliez, P. 2016. Heart involvement in cystic fibrosis: A specific cystic fibrosis-related myocardial changes? *Respir Med* 118, pp. 31-38. doi: 10.1016/j.rmed.2016.07.011

Lambiase, A., Catania, M. R., Del Pezzo, M., Rossano, F., Terlizzi, V., Sepe, A. and Raia, V. 2011. *Achromobacter xylosoxidans* respiratory tract infection in cystic fibrosis patients. *Eur J Clin Microbiol Infect Dis* 30(8), pp. 973-980. doi: 10.1007/s10096-011-1182-5

Laupland, K. B. and Valiquette, L. 2013. The changing culture of the microbiology laboratory. *Can J Infect Dis Med Microbiol* 24(3), pp. 125-128. doi: 10.1155/2013/101630

Lazarus, A., Myers, J. and Fuhrer, G. 2008. Bronchiectasis in adults: a review. *Postgrad Med* 120(3), pp. 113-121. doi: 10.3810/pgm.2008.09.1912

Lee, S. H., Lee, Y., Park, J. S., Cho, Y. J., Yoon, H. I., Lee, C. T. and Lee, J. H. 2018. Characterization of Microbiota in Bronchiectasis Patients with Different Disease Severities. *J Clin Med* 7(11), doi: 10.3390/jcm7110429

Leslie, P. and McCabe, J. T. 2013. Response Diversity and Resilience in Social-Ecological Systems. *Curr Anthropol* 54(2), pp. 114-143. doi: 10.1086/669563

Lewis, T., Loman, N. J., Bingle, L., Jumaa, P., Weinstock, G. M., Mortiboy, D. and Pallen, M. J. 2010. High-throughput whole-genome sequencing to dissect the epidemiology of

Acinetobacter baumannii isolates from a hospital outbreak. *J Hosp Infect* 75(1), pp. 37-41. doi: 10.1016/j.jhin.2010.01.012

Li, Q., Wang, H., Tan, L., Zhang, S., Lin, L., Tang, X. and Pan, Y. 2021. Oral Pathogen. *Front Cell Infect Microbiol* 11, p. 643913. doi: 10.3389/fcimb.2021.643913

Li, X., Bosch-Tijhof, C. J., Wei, X., de Soet, J. J., Crielaard, W., Loveren, C. V. and Deng, D. M. 2020. Efficiency of chemical versus mechanical disruption methods of DNA extraction for the identification of oral Gram-positive and Gram-negative bacteria. *J Int Med Res* 48(5), p. 300060520925594. doi: 10.1177/0300060520925594

Li, Y. J., Pan, C. Z., Fang, C. Q., Zhao, Z. X., Chen, H. L., Guo, P. H. and Zhao, Z. W. 2017. Pneumonia caused by extensive drug-resistant *Acinetobacter baumannii* among hospitalized patients: genetic relationships, risk factors and mortality. *BMC Infect Dis* 17(1), p. 371. doi: 10.1186/s12879-017-2471-0

Lieberman, D. 2003. Pseudomonal infections in patients with COPD: epidemiology and management. *Am J Respir Med* 2(6), pp. 459-468. doi: 10.1007/BF03256673

Lin, J., He, S. S., Xu, Y. Z., Li, H. Y., Wu, X. M. and Feng, J. X. 2019. Bacterial etiology in early re-admission patients with acute exacerbation of chronic obstructive pulmonary disease. *Afr Health Sci* 19(2), pp. 2073-2081. doi: 10.4314/ahs.v19i2.31

LiPuma, J. J. 2010. The Changing Microbial Epidemiology in Cystic Fibrosis.*Clin Microbiol Rev*. Vol. 23. pp. 299-323.

Liu, C., Guo, J., Yan, W., Jin, Y., Pan, F., Fang, X. and Qin, L. 2017. Hospital-acquired pneumonia due to *Achromobacter xylosoxidans* in the elderly: A single-center retrospective study in Beijing. *J Infect Dev Ctries* 11(1), pp. 10-18. doi: 10.3855/jidc.8747

Liu, L. et al. 2012. Comparison of next-generation sequencing systems. *J Biomed Biotechnol* 2012, p. 251364. doi: 10.1155/2012/251364

Lo, T. S. 2012. A cavitary pneumonia caused by *Leptotrichia* species in an immunocompetent patient. *Infect Dis Rep* 4(1), p. e24. doi: 10.4081/idr.2012.e24

Lopes, S. P., Azevedo, N. F. and Pereira, M. O. 2018. Quantitative assessment of individual populations within polymicrobial biofilms. *Sci Rep* 8(1), pp. 1-13. doi: doi:10.1038/s41598-018-27497-9

Lozupone, C. A., Hamady, M., Kelley, S. T. and Knight, R. 2007. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol* 73(5), pp. 1576-1585. doi: 10.1128/AEM.01996-06

Lucas, R. et al. 2020. Impact of Bacterial Toxins in the Lungs. *Toxins* 12(4), doi: 10.3390/toxins12040223

Lucas, S. K., Yang, R., Dunitz, J. M., Boyer, H. C. and Hunter, R. C. 2018. 16S rRNA gene sequencing reveals site-specific signatures of the upper and lower airways of cystic fibrosis patients. *J Cyst Fibros* 17(2), pp. 204-212. doi: 10.1016/j.jcf.2017.08.007

Mac Aogáin, M. et al. 2021. Integrative microbiomics in bronchiectasis exacerbations. *Nat Med* 27(4), pp. 688-699. doi: 10.1038/s41591-021-01289-7

Mahboubi, M. A., Carmody, L. A., Foster, B. K., Kalikin, L. M., VanDevanter, D. R. and LiPuma, J. J. 2016. Culture-Based and Culture-Independent Bacteriologic Analysis of Cystic Fibrosis Respiratory Specimens. *J Clin Microbiol* 54(3), pp. 613-619. doi: 10.1128/JCM.02299-15

Mahenthiralingam, E. 2014. Emerging cystic fibrosis pathogens and the microbiome. *Paediatr Respir Rev* 15 Suppl 1, pp. 13-15. doi: 10.1016/j.prrv.2014.04.006

Mahenthiralingam, E., Campbell, M. E., Foster, J., Lam, J. S. and Speert, D. P. 1996a. Random amplified polymorphic DNA typing of *Pseudomonas aeruginosa* isolates recovered from patients with cystic fibrosis. *J Clin Microbiol* 34(5), pp. 1129-1135. doi: 10.1128/jcm.34.5.1129-1135.1996

Mahenthiralingam, E., Campbell, M. E., Henry, D. A. and Speert, D. P. 1996b. Epidemiology of *Burkholderia cepacia infection* in patients with cystic fibrosis: analysis by randomly amplified polymorphic DNA fingerprinting. *J Clin Microbiol* 34(12), pp. 2914-2920. doi: 10.1128/jcm.34.12.2914-2920.1996

Mallott, E. K., Malhi, R. S. and Amato, K. R. 2019. Assessing the comparability of different DNA extraction and amplification methods in gut microbial community profiling. *Access Microbiol* 1(7), p. e000060. doi: 10.1099/acmi.0.000060

Marion-Sanchez, K., Pailla, K., Cesarine, M., Platon, M. G., Derancourt, C. and Olive, C. 2019. *Achromobacter xylosoxidans* resistance to antiseptics and disinfectants is far from obvious. *Trans R Soc Trop Med Hyg* 113(6), pp. 356-358. doi: 10.1093/trstmh/trz016

Martin-Loeches, I. et al. 2020. The importance of airway and lung microbiome in the critically ill. *Crit Care* 24(1), p. 537. doi: 10.1186/s13054-020-03219-4

Martina, P. F. et al. 2017. First time identification of *Pandoraea sputorum* from a patient with cystic fibrosis in Argentina: a case report. *BMC Pulm Med* 17(1), p. 33. doi: 10.1186/s12890-017-0373-y

Martinucci, M., Roscetto, E., Iula, V. D., Votsi, A., Catania, M. R. and De Gregorio, E. 2016. Accurate identification of members of the *Burkholderia cepacia* complex in cystic fibrosis sputum. *Lett Appl Microbiol* 62(3), pp. 221-229. doi: 10.1111/lam.12537

Mathieu, E. et al. 2018. Paradigms of Lung Microbiota Functions in Health and Disease, Particularly, in Asthma. *Front Physiol* 9, p. 1168. doi: 10.3389/fphys.2018.01168

Matkovic, Z. and Miravitlles, M. 2013. Chronic bronchial infection in COPD. Is there an infective phenotype? *Respir Med* 107(1), pp. 10-22. doi: 10.1016/j.rmed.2012.10.024

Matthaiou, D. K., Chasou, E., Atmatzidis, S. and Tsolkas, P. 2011. A case of bacteremia due to *Burkholderia cepacia* in a patient without cystic fibrosis. *Respir Med CME* 4(3), pp. 144-145. doi: 10.1016/j.rmedc.2010.11.002

Mayhew, D. et al. 2018. Longitudinal profiling of the lung microbiome in the AERIS study demonstrates repeatability of bacterial and eosinophilic COPD exacerbations. *Thorax* 73(5), pp. 422-430. doi: 10.1136/thoraxjnl-2017-210408

McCallum, G. B. and Binks, M. J. 2017. The Epidemiology of Chronic Suppurative Lung Disease and Bronchiectasis in Children and Adolescents. *Front Pediatr* 5, p. 27. doi: 10.3389/fped.2017.00027

McKelvey, M. C., Weldon, S., McAuley, D. F., Mall, M. A. and Taggart, C. C. 2020. Targeting Proteases in Cystic Fibrosis Lung Disease. Paradigms, Progress, and Potential. *Am J Respir Crit Care Med* 201(2), pp. 141-147. doi: 10.1164/rccm.201906-1190PP

Menetrey, Q., Dupont, C., Chiron, R., Jumas-Bilak, E. and Marchandin, H. 2020. High Occurrence of Bacterial Competition Among Clinically Documented Opportunistic Pathogens Including. *Front Microbiol* 11, p. 558160. doi: 10.3389/fmicb.2020.558160

Mitchelmore, P. et al. 2020. Immunodeficiency associated bronchiectasis in the European Bronchiectasis Registry (EMBARC). *European Respiratory Journal* 56, p. 2057. doi: 10.1183/13993003.congress-2020.2057

Moffatt, M. F. and Cookson, W. O. 2017. The lung microbiome in health and disease. *Clin Med* 17(6), pp. 525-529. doi: 10.7861/clinmedicine.17-6-525

Monsó, E. 2017. Microbiome in chronic obstructive pulmonary disease. *Ann Transl Med* 5(12), p. 251. doi: 10.21037/atm.2017.04.20

Monteiro, J. M. et al. 2015. Cell shape dynamics during the staphylococcal cell cycle. *Nat Commun* 6, p. 8055. doi: 10.1038/ncomms9055

Moran Losada, P. et al. 2016. The cystic fibrosis lower airways microbial metagenome. *ERJ Open Res* 2(2), doi: 10.1183/23120541.00096-2015

Muhlebach, M. S. et al. 2018. Anaerobic bacteria cultured from cystic fibrosis airways correlate to milder disease: a multisite study. *Eur Respir J* 52(1), doi: 10.1183/13993003.00242-2018

Murphy, E. C. and Frick, I. M. 2013. Gram-positive anaerobic cocci--commensals and opportunistic pathogens. *FEMS Microbiol Rev* 37(4), pp. 520-553. doi: 10.1111/1574-6976.12005

Humbert, M. V. and Christodoulides, M. 2019. Atypical, Yet Not Infrequent, Infections with *Neisseria* Species. *Pathogens* 9(1), pp. 1-27. doi: 10.3390/pathogens9010010

Myer, P. R. et al. 2020. Classification of 16S rRNA reads is improved using a niche-specific database constructed by near-full length sequencing. *PLoS One* 15(7), p. e0235498. doi: 10.1371/journal.pone.0235498

Nagaoka, K. et al. 2014. *Prevotella intermedia* induces severe bacteremic pneumococcal pneumonia in mice with upregulated platelet-activating factor receptor expression. *Infect Immun* 82(2), pp. 587-593. doi: 10.1128/IAI.00943-13

Nagendra, S., Bourbeau, P., Brecher, S., Dunne, M., LaRocco, M. and Doern, G. 2001. Sampling variability in the microbiological evaluation of expectorated sputa and endotracheal aspirates. *J Clin Microbiol* 39(6), pp. 2344-2347. doi: 10.1128/JCM.39.6.2344-2347.2001

Nagler, M., Insam, H., Pietramellara, G. and Ascher-Jenull, J. 2018. Extracellular DNA in natural environments: features, relevance and applications. *Appl Microbiol Biotechnol* 102(15), pp. 6343-6356. doi: doi:10.1007/s00253-018-9120-4

Nair, B., Stapp, J., Stapp, L., Bugni, L., Van Dalfsen, J. and Burns, J. L. 2002. Utility of gram staining for evaluation of the quality of cystic fibrosis sputum samples. *J Clin Microbiol* 40(8), pp. 2791-2794. doi: 10.1128/JCM.40.8.2791-2794.2002

Naser, A. Y. et al. 2021. Hospital admission trends due to respiratory diseases in England and Wales between 1999 and 2019: an ecologic study. *BMC Pulm Med* 21(1), p. 356. doi: 10.1186/s12890-021-01736-8

Navratilova, L. et al. 2016. The *Streptococcus milleri* group in chronic obstructive pulmonary disease. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* 160(3), pp. 378-384. doi: 10.5507/bp.2016.017

Neuwirth, C. et al. 2006. VEB-1 in *Achromobacter xylosoxidans* from Cystic Fibrosis Patient, France. *Emerg Infect Dis* 12(11), pp. 1737-1739. doi: 10.3201/eid1211.060143

Nguyen, L. T., Schmidt, H. A., von Haeseler, A. and Minh, B. Q. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32(1), pp. 268-274. doi: 10.1093/molbev/msu300

Nielsen, S. M., Nørskov-Lauritsen, N., Bjarnsholt, T. and Meyer, R. L. 2016. *Achromobacter* Species Isolated from Cystic Fibrosis Patients Reveal Distinctly Different Biofilm Morphotypes. *Microorganisms* 4(3), doi: 10.3390/microorganisms4030033

Nielsen, S. M., Penstoft, L. N. and Nørskov-Lauritsen, N. 2019. Motility, Biofilm Formation and Antimicrobial Efflux of Sessile and Planktonic Cells of *Achromobacter xylosoxidans.* *Pathogens* 8(1), p. 14. doi: 10.3390/pathogens8010014

O'Donnell, A. E. 2017. *The Role of Other Bacteria, Fungi, and Viruses in Bronchiectasis*. New York:SpringerLink.

O'Dwyer, D. N., Dickson, R. P. and Moore, B. B. 2016. The Lung Microbiome, Immunity, and the Pathogenesis of Chronic Lung Disease. *J Immunol* 196(12), pp. 4839-4847. doi: 10.4049/jimmunol.1600279

O'Neal, W. K. and Knowles, M. R. 2018. Cystic Fibrosis Disease Modifiers: Complex Genetics Defines the Phenotypic Diversity in a Monogenic Disease. *Annu Rev Genomics Hum Genet* 19, pp. 201-222. doi: 10.1146/annurev-genom-083117-021329

O'Neill, K. et al. 2015. Reduced bacterial colony count of anaerobic bacteria is associated with a worsening in lung clearance index and inflammation in cystic fibrosis. *PLoS One* 10(5), p. e0126980. doi: 10.1371/journal.pone.0126980

O'Toole, G. and Kolter, R. 1998. Initiation of biofilm formation in *Pseudomonas fluorescens* WCS365 proceeds via multiple, convergent signalling pathways: a genetic analysis. *Mol Microbiol* 28(3), pp. 449-461. doi: 10.1046/j.1365-2958.1998.00797.x

Ogawa, H., Kitsios, G. D., Iwata, M. and Terasawa, T. 2020. Sputum Gram Stain for Bacterial Pathogen Diagnosis in Community-acquired Pneumonia: A Systematic Review and Bayesian Meta-analysis of Diagnostic Accuracy and Yield. *Clin Infect Dis* 71(3), pp. 499-513. doi: 10.1093/cid/ciz876

Oksanen, J. et al. 2019. vegan: Community Ecology Package. Available at: https://CRAN.R-project.org/package=vegan [Accessed: 9 September 2020]

Oladunjoye, O. O., Oladunjoye, A. O., Oladiran, O. and Donato, A. A. 2020. *Stenotrophomonas maltophilia* Infection in a Patient with Acute Exacerbation of Chronic Obstructive Pulmonary Disease (COPD): A Colonizer or True Infection? *Am J Case Rep* 21, p. e924577. doi: 10.12659/AJCR.924577

Ooi, C. Y. and Durie, P. R. 2016. Cystic fibrosis from the gastroenterologist's perspective. *Nat Rev Gastroenterol Hepatol* 13(3), pp. 175-185. doi: 10.1038/nrgastro.2015.226

Oriano, M. et al. 2019. Comparison of different conditions for DNA extraction in sputum - a pilot study. *Multidiscip Respir Med* 14, p. 6. doi: 10.1186/s40248-018-0166-z

Osorio, C. R., Collins, M. D., Romalde, J. L. and Toranzo, A. E. 2005. Variation in 16S-23S rRNA intergenic spacer regions in *Photobacterium damselae*: a mosaic-like structure. *Appl Environ Microbiol* 71(2), pp. 636-645. doi: 10.1128/AEM.71.2.636-645.2005

Paganin, P. et al. 2015. Changes in cystic fibrosis airway microbial community associated with a severe decline in lung function. *PLoS One* 10(4), p. e0124348. doi: 10.1371/journal.pone.0124348

Page, A. J. et al. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31(22), pp. 3691-3693. doi: 10.1093/bioinformatics/btv421

Page, A. J., Taylor, B., Delaney, A. J., Soares, J., Seemann, T., Keane, J. A. and Harris, S. R. 2016. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genom* 2(4), p. e000056. doi: 10.1099/mgen.0.000056

Papalia, M. et al. 2013. OXA-258 from *Achromobacter ruhlandii*: a species-specific marker. *J Clin Microbiol* 51(5), pp. 1602-1605. doi: 10.1128/jcm.03043-12

Parkins, M. D. and Floto, R. A. 2015. Emerging bacterial pathogens and changing concepts of bacterial pathogenesis in cystic fibrosis. *J Cyst Fibros* 14(3), pp. 293-304. doi: 10.1016/j.jcf.2015.03.012

Parte, A. C., Carbasse, J. S., Meier-Kolthoff, J. P., Reimer, L. C. and Göker, M. 2020. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int J Syst Evol* 70(11), pp. 5607-5612. doi: doi:10.1099/ijsem.0.004332

Passos, F., Uggetti, E., Carrere, H. and Ferrer, I. 2015. Algal Biomass: Physical Pretreatments. 1st ed. Pandey, E., Negi, S., Pinod, P. and Larroche (1st ed). In: *Pretreatment of Biomass: Processes and Technologies*. pp. 195-226, Amsterdam:Elsevier.

Payne, G. W. et al. 2005. Development of a *recA* gene-based identification approach for the entire *Burkholderia* genus. *Appl Environ Microbiol* 71(7), pp. 3917-3927. doi: 10.1128/AEM.71.7.3917-3927.2005

Pereira, R. H. et al. 2017. Patterns of virulence factor expression and antimicrobial resistance in *Achromobacter xylosoxidans* and *Achromobacter ruhlandii* isolates from patients with cystic fibrosis. *Epidemiol Infect* 145(3), pp. 600-606. doi: 10.1017/S0950268816002624

Pereira, R. H. V., Carvalho-Assef, A. P., Albano, R. M., Folescu, T. W., Jones, M., Leão, R. S. and Marques, E. A. 2011. *Achromobacter xylosoxidans*: Characterization of Strains in Brazilian Cystic Fibrosis Patients. *J Clin Microbiol* 49(10), pp. 3649-3651. doi: 10.1128/jcm.05283-11

Pérez-Brocal, V. et al. 2020. Optimized DNA extraction and purification method for characterization of bacterial and fungal communities in lung tissue samples. *Sci Rep* 10(1), p. 17377. doi: 10.1038/s41598-020-74137-2

Pérez-Frías, J., Pérez Ruiz, E., Castillo, M. C. L. and Pilar Caro García, P. C. 2019. The History of Cystic Fibrosis. *Open J Pediatr* 4(1), pp. 1-6. doi: doi:10.17352/ojpch.000015

Pezzulo, A. A. et al. 2012. Reduced airway surface pH impairs bacterial killing in the porcine cystic fibrosis lung. *Nature* 487(7405), pp. 109-113. doi: 10.1038/nature11130

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. and Team., R. C. 2020. nlme: Linear and Nonlinear Mixed Effects Models. Available at: https://CRAN.R-project.org/package=nlme [Accessed:3 December 2020].

Poeckh, T., Lopez, S., Fuller, A. O., Solomon, M. J. and Larson, R. G. 2008. Adsorption and elution characteristics of nucleic acids on silica surfaces and their use in designing a miniaturized purification unit. *Anal Biochem* 373(2), pp. 253-262. doi: 10.1016/j.ab.2007.10.026

Pollock, J., Glendinning, L., Wisedchanwet, T. and Watson, M. 2018. The Madness of Microbiome: Attempting To Find Consensus "Best Practice" for 16S Microbiome Studies. *Appl Environ Microbiol* 84(7), doi: 10.1128/AEM.02627-17

Poppleton, D. I. et al. 2017. Outer Membrane Proteome of. *Front Microbiol* 8, p. 1215. doi: 10.3389/fmicb.2017.01215

Pragman, A. A., Berger, J. P. and Williams, B. J. 2016. Understanding persistent bacterial lung infections: clinical implications informed by the biology of the microbiota and biofilms. *Clin Pulm Med* 23(2), pp. 57-66. doi: 10.1097/cpm.0000000000000108

Pragman, A. A., Knutson, K. A., Gould, T. J., Isaacson, R. E., Reilly, C. S. and Wendt, C. H. 2019. Chronic obstructive pulmonary disease upper airway microbiota alpha diversity is associated with exacerbation phenotype: a case-control observational study. *Respir Res* 20(1), p. 114. doi: 10.1186/s12931-019-1080-4

Prevaes, S. M. et al. 2017. Concordance between upper and lower airway microbiota in infants with cystic fibrosis. *Eur Respir J* 49(3), doi: 10.1183/13993003.02235-2016

Pritchard. 2016. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal. Methods* 8, pp. 12-24.

Proctor, P., Soldat, S. M., Easparro, B., Nash, R. and Atwood, J. 2019. Evaluating the impact of bead media diameter and material composition on bacterial cell lysis and genomic DNA extraction. *FASEB J* 33, p. 648.646. doi: 10.1096/fasebj.2019.33.1_supplement.648.6

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J. and Glöckner, F. O. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 35(21), pp. 7188-7196. doi: 10.1093/nar/gkm864

Qian, Y. Y. et al. 2020. Improving Pulmonary Infection Diagnosis with Metagenomic Next Generation Sequencing. *Front Cell Infect Microbiol* 10, p. 567615. doi: 10.3389/fcimb.2020.567615

Quaderi, S. A. and Hurst, J. R. 2018. The unmet global burden of COPD. *Glob Health Epidemiol Genom* 3, p. e4. doi: 10.1017/gheg.2018.1

Quon, B. S., Mayer-Hamblett, N., Aitken, M. L., Smyth, A. R. and Goss, C. H. 2011. Risk factors for chronic kidney disease in adults with cystic fibrosis. *Am J Respir Crit Care Med* 184(10), pp. 1147-1152. doi: 10.1164/rccm.201105-0932OC

Rabilloud, T., Luche, S., Santoni, V. and Chevallet, M. 2007. Detergents and chaotropes for protein solubilization before two-dimensional electrophoresis. *Methods mol biol* 355, pp. 111-119. doi: 10.1385/1-59745-227-0:111

Raidt, L., Idelevich, E. A., Dübbers, A., Küster, P., Drevinek, P., Peters, G. and Kahl, B. C. 2015. Increased Prevalence and Resistance of Important Pathogens Recovered from Respiratory Specimens of Cystic Fibrosis Patients During a Decade. *Pediatr Infect Dis J* 34(7), pp. 700-705. doi: 10.1097/INF.0000000000000714

Ramos, A. T., Figueirêdo, M. M., Aguiar, A. P., Almeida, C. e. G., Mendes, P. S. and Souza, E. L. 2016. Celiac Disease and Cystic Fibrosis: Challenges to Differential Diagnosis. *Folia Med* 58(2), pp. 141-147. doi: 10.1515/folmed-2016-0020

Ramsheh, M. Y., Haldar, K., Esteve-Codina, A., Purser, L. F., Richardson, M. and Muller-Quernheim, J. 2021. Lung microbiome composition and bronchial epithelial gene expression in patients with COPD versus healthy individuals: a bacterial 16S rRNA gene sequencing and host transcriptomic analysis. *Lancet Microbe* 2(7), pp. E300-E310. doi:10.1016/S2666-5247(21)00035-5

Rang, C. and Wilson, J. 2019. Remarkable progress in cystic fibrosis-But why? *Respirology* 24(1), pp. 17-18. doi: 10.1111/resp.13440

Rashid, M. H. and Kornberg, A. 2000. Inorganic polyphosphate is needed for swimming, swarming, and twitching motilities of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A* 97(9), pp. 4885-4890. doi: 10.1073/pnas.060030097

Raso, T., Bianco, O., Grosso, B., Zucca, M. and Savoia, D. 2008. *Achromobacter xylosoxidans* respiratory tract infections in cystic fibrosis patients. *Apmis* 116(9), pp. 837-841.

Ratjen, F., Bell, S. C., Rowe, S. M., Goss, C. H., Quittner, A. L. and Bush, A. 2015. Cystic fibrosis. *Nat Rev Dis Primers* 1, p. 15010. doi: 10.1038/nrdp.2015.10

Ratnatunga, C. N. et al. 2020. The Rise of Non-Tuberculosis Mycobacterial Lung Disease. *Front Immunol* 11, p. 303. doi: 10.3389/fimmu.2020.00303

Renom, F. et al. 2010. Prognosis of COPD patients requiring frequent hospitalization: Role of airway infection. *Respir Med* 104(6), pp. 840-848. doi: 10.1016/j.rmed.2009.12.010

Riccio, M. L., Pallecchi, L., Fontana, R. and Rossolini, G. M. 2001. In70 of Plasmid pAX22, a blaVIM-1-Containing Integron Carrying a New Aminoglycoside Phosphotransferase Gene Cassette. *Antimicrob Agents Chemother* 45(4), pp. 1249-1253. doi: 10.1128/aac.45.4.1249-1253.2001

Richardson, H., Dicker, A. J., Barclay, H. and Chalmers, J. D. 2019. The microbiome in bronchiectasis. *Eur Respir Rev* 28(153), doi: 10.1183/16000617.0048-2019

Richter, A. M. et al. 2018. Key Players and Individualists of Cyclic-di-GMP Signaling in. *Front Microbiol* 9, p. 3286. doi: 10.3389/fmicb.2018.03286

Ridderberg, W., Andersen, C., Væth, M., Bregnballe, V., Nørskov-Lauritsen, N. and Schiøtz, P. O. 2016. Lack of evidence of increased risk of bacterial transmission during cystic fibrosis educational programmes. *J Cyst Fibros* 15(1), pp. 109-115. doi: 10.1016/j.jcf.2015.04.007

Ridderberg, W., Bendstrup, K. E., Olesen, H. V., Jensen-Fangel, S. and Norskov-Lauritsen, N. 2011. Marked increase in incidence of *Achromobacter xylosoxidans* infections caused by sporadic acquisition from the environment. *J Cyst Fibros* 10(6), pp. 466-469. doi: 10.1016/j.jcf.2011.07.004

Ridderberg, W., Wang, M. and Nørskov-Lauritsen, N. 2012. Multilocus Sequence Analysis of Isolates of *Achromobacter* from Patients with Cystic Fibrosis Reveals Infecting Species Other than *Achromobacter xylosoxidans*. *J Clin Microbiol*. Vol. 50. pp. 2688-2694.

Ritari, J., Salojärvi, J., Lahti, L. and de Vos, W. M. 2015. Improved taxonomic assignment of human intestinal 16S rRNA sequences by a dedicated reference database. *BMC Genomics* 16, p. 1056. doi: 10.1186/s12864-015-2265-y

Rocchetti, T. T., Silbert, S., Gostnell, A., Kubasek, C., Jerris, R., Vong, J. and Widen, R. 2018. Rapid detection of four non-fermenting Gram-negative bacteria directly from cystic fibrosis patient's respiratory samples on the BD MAX™ system. *Pract Lab Med* 12, p. e00102. doi: 10.1016/j.plabm.2018.e00102

Rocha, G. A. et al. 2018. Species distribution, sequence types and antimicrobial resistance of *Acinetobacter* spp. from cystic fibrosis patients. *Epidemiol Infect* 146(4), pp. 524-530. doi: 10.1017/S0950268817002849

Rogers, G. B., Carroll, M. P., Hoffman, L. R., Walker, A. W., Fine, D. A. and Bruce, K. D. 2010. Comparing the microbiota of the cystic fibrosis lung and human gut. *Gut Microbes* 1(2), pp. 85-93. doi: 10.4161/gmic.1.2.11350

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 conventional and molecular microbiological approaches. *BMC Pulm Med* 9, p. 14. doi: 10.1186/1471-2466-9-14

Rogers, G. B., van der Gast, C. J., Cuthbertson, L., Thomson, S. K., Bruce, K. D., Martin, M. L. and Serisier, D. J. 2013. Clinical measures of disease in adult non-CF bronchiectasis correlate with airway microbiota composition. *Thorax* 68(8), pp. 731-737. doi: 10.1136/thoraxjnl-2012-203105

Rognes, T., Flouri, T., Nichols, B., Quince, C. and Mahé, F. 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4, p. e2584. doi: 10.7717/peerj.2584

Ronchetti, K. et al. 2018. The CF-Sputum Induction Trial (CF-SpIT) to assess lower airway bacterial sampling in young children with cystic fibrosis: a prospective internally controlled interventional trial. *Lancet Respir Med* 6(6), pp. 461-471. doi: 10.1016/S2213-2600(18)30171-1

Ross, A. A., Müller, K. M., Weese, J. S. and Neufeld, J. D. 2018. Comprehensive skin microbiome analysis reveals the uniqueness of human skin and evidence for phylosymbiosis within the class Mammalia. *Proc Natl Acad Sci U S A* 115(25), pp. E5786-E5795. doi: 10.1073/pnas.1801302115

Round, J. L. and Palm, N. W. 2018. Causal effects of the microbiota on immune-mediated diseases. *Sci Immunol* 3(20), doi: 10.1126/sciimmunol.aao1603

Ryan, K. and Byrd, T. F. 2018. *Mycobacterium abscessus*: Shapeshifter of the Mycobacterial World. *Front Microbiol* 9, p. 2642. doi: 10.3389/fmicb.2018.02642

Rychert, J. 2019. Benefits and Limitations of MALDI-TOF Mass Spectrometry for the Identification of Microorganisms. *J Infectiology & Epidemiol* 2(4), pp. 1-5.

Rytter, H., Jamet, A., Coureuil, M., Charbit, A. and Ramond, E. 2020. Which Current and Novel Diagnostic Avenues for Bacterial Respiratory Diseases? *Front Microbiol* 11, p. 616971. doi: 10.3389/fmicb.2020.616971

Salipante, S. J. et al. 2014. Performance comparison of Illumina and ion torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. *Appl Environ Microbiol* 80(24), pp. 7583-7591. doi: 10.1128/AEM.02206-14

Salter, S. J. et al. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 12, p. 87. doi: 10.1186/s12915-014-0087-z

Salunkhe, P. et al. 2005. A cystic fibrosis epidemic strain of *Pseudomonas aeruginosa* displays enhanced virulence and antimicrobial resistance. *J Bacteriol* 187(14), pp. 4908-4920. doi: 10.1128/JB.187.14.4908-4920.2005

Sayers, E. W., Cavanaugh, M., Clark, K., Pruitt, K. D., Schoch, C. L., Sherry, S. T. and Karsch-Mizrachi, I. 2021. GenBank. *Nucleic Acids Res* 49(D1), pp. D92-D96. doi: 10.1093/nar/gkaa1023

Schloss, P. D. 2010. The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. *PLoS Comput Biol* 6(7), p. e1000844. doi: 10.1371/journal.pcbi.1000844

Schrader, C., Schielke, A., Ellerbroek, L. and Johne, R. 2012. PCR inhibitors - occurrence, properties and removal. *J Appl Microbiol* 113(5), pp. 1014-1026. doi: 10.1111/j.1365-2672.2012.05384.x

Schulze, A., Mitterer, F., Pombo, J. P. and Schild, S. 2021. Biofilms by bacterial human pathogens: Clinical relevance - development, composition and regulation - therapeutical strategies. *Microb Cell* 8(2), pp. 28-56. doi: 10.15698/mic2021.02.741

Seemann, T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14), pp. 2068-2069. doi: 10.1093/bioinformatics/btu153

Selway, C. A., Eisenhofer, R. and Weyrich, L. S. 2020. Microbiome applications for pathology: challenges of low microbial biomass samples during diagnostic testing. *J Pathol Clin Res* 6(2), pp. 97-106. doi: 10.1002/cjp2.151

Sethi, S. 2004. Bacteria in exacerbations of chronic obstructive pulmonary disease: phenomenon or epiphenomenon? *Proc Am Thorac Soc* 1(2), pp. 109-114. doi: 10.1513/pats.2306029

Sethi, S. 2010. Infection as a comorbidity of COPD. *Eur Respir J* 35(6), pp. 1209-1215. doi: 10.1183/09031936.00081409

Sharan, H. 2015. Aerobic Bacteriological Study of Acute Exacerbations of Chronic Obstructive Pulmonary Disease. *J Clin Diagn Res* 9(8), pp. DC10-12. doi: 10.7860/JCDR/2015/14515.6367

Sharma-Kuinkel, B. K., Rude, T. H. and Fowler, V. G. 2016. Pulse Field Gel Electrophoresis. *Methods Mol Biol* 1373, pp. 117-130. doi: 10.1007/7651_2014_191

Sharma, M. V. 2019. Early diagnosis of viral pneumonia. *J Pulmon* 3(1), pp. 1-3.

Shen, C. H. 2019. *Chapter 6 - Extraction and Purification of Nucleic Acids and Proteins*. 1st ed. Cambridge, Massachusetts: Academic press.

Sherrard, L. J., Bell, S. C. and Tunney, M. M. 2016. The role of anaerobic bacteria in the cystic fibrosis airway. *Curr Opin Pulm Med* 22(6), pp. 637-643. doi: 10.1097/MCP.0000000000000299

Shimizu, K. et al. 2015. Pathogens in COPD exacerbations identified by comprehensive real-time PCR plus older methods. *Int J Chron Obstruct Pulmon Dis* 10, pp. 2009-2016. doi: 10.2147/COPD.S82752

Shin, K. S. et al. 2005. Imipenem-resistant *Achromobacter xylosoxidans* carrying *bla*$_{VIM-2}$-containing class 1 integron. *Diagn Microbiol Infect Dis* 53(3), pp. 215-220. doi: 10.1016/j.diagmicrobio.2005.06.018

Sibley, C. D., Sibley, K. A., Leong, T. A., Grinwis, M. E., Parkins, M. D., Rabin, H. R. and Surette, M. G. 2010. The *Streptococcus milleri* population of a cystic fibrosis clinic reveals patient specificity and intraspecies diversity. *J Clin Microbiol* 48(7), pp. 2592-2594. doi: 10.1128/JCM.00414-10

Sidstedt, M. et al. 2018. Inhibition mechanisms of hemoglobin, immunoglobulin G, and whole blood in digital and real-time PCR. *Anal Bioanal Chem* 410(10), pp. 2569-2583. doi: 10.1007/s00216-018-0931-z

Silhavy, T. J., Kahne, D. and Walker, S. 2010. The bacterial cell envelope. *Cold Spring Harb Perspect Biol* 2(5), p. a000414. doi: 10.1101/cshperspect.a000414

Silverman, J. D., Bloom, R. J., Jiang, S., Durand, H. K., Dallow, E., Mukherjee, S. and David, L. A. 2021. Measuring and mitigating PCR bias in microbiota datasets. *PLoS Comput Biol* 17(7), p. e1009113. doi: 10.1371/journal.pcbi.1009113

Smith, D. J., Badrick, A. C., Zakrzewski, M., Krause, L., Bell, S. C., Anderson, G. J. and Reid, D. W. 2014. Pyrosequencing reveals transient cystic fibrosis lung microbiome changes with intravenous antibiotics. *Eur Respir J* 44(4), pp. 922-930. doi: 10.1183/09031936.00203013

Smith, E. E. et al. 2006. Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *Proc Natl Acad Sci U S A* 103(22), pp. 8487-8492. doi: 10.1073/pnas.0602138103

Snell, N., Gibson, J., Jarrold, I. and Quint, J. K. 2019. Epidemiology of bronchiectasis in the UK: Findings from the British lung foundation's 'Respiratory health of the nation' project. *Respir Med* 158, pp. 21-23. doi: 10.1016/j.rmed.2019.09.012

Somayaji, R. and Chalmers, J. D. 2022. Just breathe: a review of sex and gender in chronic lung disease. *Eur Respir Rev* 31(163), doi: 10.1183/16000617.0111-2021

Somayaji, R., Stanojevic, S., Tullis, D. E., Stephenson, A. L., Ratjen, F. and Waters, V. 2017. Clinical Outcomes Associated with *Achromobacter* Species Infection in Patients with Cystic Fibrosis. *Ann Am Thorac Soc* 14(9), pp. 1412-1418. doi: 10.1513/AnnalsATS.201701-071OC

Spilker, T., Vandamme, P. and Lipuma, J. J. 2012. A multilocus sequence typing scheme implies population structure and reveals several putative novel *Achromobacter* species. *J Clin Microbiol* 50(9), pp. 3010-3015. doi: 10.1128/jcm.00814-12

Spilker, T., Vandamme, P. and Lipuma, J. J. 2013. Identification and distribution of *Achromobacter* species in cystic fibrosis. *J Cyst Fibros* 12(3), pp. 298-301. doi: 10.1016/j.jcf.2012.10.002

Sprouffske, K. and Wagner, A. 2016. Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinform* 17(1), pp. 1-4. doi: doi:10.1186/s12859-016-1016-7

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9), pp. 1312-1313. doi: 10.1093/bioinformatics/btu033

Stanton, N., Francis, N. A. and Butler, C. C. 2010. Reducing uncertainty in managing respiratory tract infections in primary care. *Br J Gen Pract* 60(581), pp. e466-475. doi: 10.3399/bjgp10X544104

Starke, R., Jehmlich, N., Alfaro, T., Dohnalkova, A., Capek, P., Bell, S. L. and Hofmockel, K. S. 2019. Incomplete cell disruption of resistant microbes. *Sci Rep* 9(1), p. 5618. doi: 10.1038/s41598-019-42188-9

Stepman, G., Dabb, K., Khan, I. A., Young, J. T. and Frunzi, J. 2020. Bilateral Pneumonia in a Patient with Chronic Bronchiectasis Caused by *Achromobacter xylosoxidans* Subspecies *denitrificans*. *Cureus* 12(3), p. e7381.

Stressmann, F. A. et al. 2012. Long-term cultivation-independent microbial diversity analysis demonstrates that bacterial communities infecting the adult cystic fibrosis lung show stability and resilience. *Thorax* 67(10), pp. 867-873. doi: 10.1136/thoraxjnl-2011-200932

Su, J. et al. 2015. Sputum Bacterial and Fungal Dynamics during Exacerbations of Severe COPD. *PLoS One* 10(7), p. e0130736. doi: 10.1371/journal.pone.0130736

Sui, H. Y. et al. 2020. Impact of DNA Extraction Method on Variation in Human and Built Environment Microbial Community and Functional Profiles Assessed by Shotgun Metagenomics Sequencing. *Front Microbiol* 11, p. 953. doi: 10.3389/fmicb.2020.00953

Sulaiman, I. et al. 2018. Evaluation of the airway microbiome in nontuberculous *mycobacteria* disease. *Eur Respir J* 52(4), doi: 10.1183/13993003.00810-2018

Swenson, C. E. and Sadikot, R. T. 2015. *Achromobacter* respiratory infections. *Ann Am Thorac Soc* 12(2), pp. 252-258. doi: 10.1513/AnnalsATS.201406-288FR

Sze, M. A. et al. 2012. The lung tissue microbiome in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 185(10), pp. 1073-1080. doi: 10.1164/rccm.201111-2075OC

Sze, M. A. et al. 2015. Host Response to the Lung Microbiome in Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* 192(4), pp. 438-445. doi: 10.1164/rccm.201502-0223OC

Tabatabaei, M., Dastbarsar, M. and Moslehi, M. A. 2019. Isolation and identification of *Pandoraea* spp. From bronchoalveolar lavage of cystic fibrosis patients in Iran. *Ital J Pediatr* 45(1), p. 118. doi: 10.1186/s13052-019-0687-x

Taccetti, G., Francalanci, M., Pizzamiglio, G., Messore, B., Carnovale, V., Cimino, G. and Cipolli, M. 2021. Cystic Fibrosis: Recent Insights into Inhaled Antibiotic Treatment and Future Perspectives. *Antibiotics* 10(3), doi: 10.3390/antibiotics10030338

Talbot, N. P. and Flight, W. G. 2016. Severe *Achromobacter xylosoxidans* infection and loss of sputum bacterial diversity in an adult patient with cystic fibrosis. *Paediatr Respir Rev* 20 Suppl, pp. 27-29. doi: 10.1016/j.prrv.2016.06.011

Tan, S. C. and Yiap, B. C. 2009. DNA, RNA, and protein extraction: the past and the present. *J Biomed Biotechnol* 2009, p. 574398. doi: 10.1155/2009/574398

Teng, F. et al. 2018. Impact of DNA extraction method and targeted 16S-rRNA hypervariable region on oral microbiota profiling. *Sci Rep* 8(1), p. 16321. doi: 10.1038/s41598-018-34294-x

Thomson, R. B. 2016. One Small Step for the Gram Stain, One Giant Leap for Clinical Microbiology. *J Clin Microbiol* 54(6), pp. 1416-1417. doi: 10.1128/JCM.00303-16

Tiew, P. Y., Jaggi, T. K., Chan, L. L. Y. and Chotirmall, S. H. 2021. The airway microbiome in COPD, bronchiectasis and bronchiectasis-COPD overlap. *Clin Respir J* 15(2), pp. 123-133. doi: 10.1111/crj.13294

Tourlousse, D. M. et al. 2021. Validation and standardization of DNA extraction and library construction methods for metagenomics-based human fecal microbiome measurements. *Microbiome* 9(1), pp. 1-19. doi: doi:10.1186/s40168-021-01048-3

Traglia, G. M. et al. 2012. *Achromobacter xylosoxidans*: An Emerging Pathogen Carrying Different Elements Involved in Horizontal Genetic Transfer.*Curr Microbiol*. Vol. 65. pp. 673-678.

Trancassini, M. et al. 2014. Outbreak of *Achromobacter xylosoxidans* in an Italian Cystic fibrosis center: genome variability, biofilm production, antibiotic resistance, and motility in isolated strains. *Front Microbiol* 5, p. 138. doi: 10.3389/fmicb.2014.00138

Tremblay, J. et al. 2015. Primer and platform effects on 16S rRNA tag sequencing. *Front Microbiol* 6, p. 771. doi: 10.3389/fmicb.2015.00771

Trombley Hall, A., McKay Zovanyi, A., Christensen, D. R., Koehler, J. W. and Devins Minogue, T. 2013. Evaluation of inhibitor-resistant real-time PCR methods for diagnostics in clinical and environmental samples. *PLoS One* 8(9), p. e73845. doi: 10.1371/journal.pone.0073845

Tunney, M. M. et al. 2013. Lung Microbiota and Bacterial Abundance in Patients with Bronchiectasis when Clinically Stable and during Exacerbation. *Am J Respir Crit Care Med* 187(10), pp. 1118-1126. doi: 10.1164/rccm.201210-1937OC

Van Buuren, S. and Groothuis-Oudshoorn, K. 2011. Mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw* 45(3), pp. 1-67. doi: 10.18637/jss.v045.i03

Van Daele, S. et al. 2005. Shared genotypes of *Achromobacter xylosoxidans* strains isolated from patients at a cystic fibrosis rehabilitation center. *J Clin Microbiol* 43(6), pp. 2998-3002. doi: 10.1128/jcm.43.6.2998-3002.2005

Van den Bossche, S., De Broe, E., Coenye, T., Van Braeckel, E. and Crabbé, A. 2021. The cystic fibrosis lung microenvironment alters antibiotic activity: causes and effects. *Eur Respir Rev* 30(161), doi: 10.1183/16000617.0055-2021

Vandamme, P. et al. 2013. *Achromobacter animicus* sp. nov., *Achromobacter mucicolens* sp. nov., *Achromobacter pulmonis* sp. nov. and *Achromobacter spiritinus* sp. nov., from human clinical samples. *Syst Appl Microbiol* 36(1), pp. 1-10. doi: 10.1016/j.syapm.2012.10.003

Vandeplassche, E., Tavernier, S., Coenye, T. and Crabbé, A. 2019. Influence of the lung microbiome on antibiotic susceptibility of cystic fibrosis pathogens. *Eur Respir Rev* 28(152), doi: 10.1183/16000617.0041-2019

Vandeventer, P. E. et al. 2011. Mechanical disruption of lysis-resistant bacterial cells by use of a miniature, low-power, disposable device. *J Clin Microbiol* 49(7), pp. 2533-2539. doi: 10.1128/JCM.02171-10

Venkataraman, A., Bassis, C. M., Beck, J. M., Young, V. B., Curtis, J. L., Huffnagle, G. B. and Schmidt, T. M. 2015. Application of a neutral community model to assess structuring of the human lung microbiome. *mBio* 6(1), doi: 10.1128/mBio.02284-14

Verduin, C. M., Hol, C., Fleer, A., van Dijk, H. and van Belkum, A. 2002. *Moraxella catarrhalis*: from emerging to established pathogen. *Clin Microbiol Rev* 15(1), pp. 125-144. doi: 10.1128/cmr.15.1.125-144.2002

Verregghen, M., Heijerman, H. G., Reijers, M., van Ingen, J. and van der Ent, C. K. 2012. Risk factors for *Mycobacterium abscessus* infection in cystic fibrosis patients; a case-control study. *J Cyst Fibros* 11(4), pp. 340-343. doi: 10.1016/j.jcf.2012.01.006

Veschetti, L., Sandri, A., Patuzzo, C., Melotti, P., Malerba, G. and Lleo, M. M. 2021. Genomic characterization of *Achromobacter* species isolates from chronic and occasional lung infection in cystic fibrosis patients. *Microb Genom* 7(7), doi: 10.1099/mgen.0.000606

Vilo, C. and Dong, Q. 2012. Evaluation of the RDP Classifier Accuracy Using 16S rRNA Gene Variable Regions. *Metagenomics*, pp. 1-5. doi: 10.4303/mg/235551

Vindhyvasni, R. P. D. and Gupta, A. 2020. Chronic obstructive pulmonary disease (COPD). *J Pharm Innov* 9(1), pp. 302-304.

von Wintzingerode, F., Göbel, U. B. and Stackebrandt, E. 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* 21(3), pp. 213-229. doi: 10.1111/j.1574-6976.1997.tb00351.x

Vonberg, R. P. and Gastmeier, P. 2005. Isolation of infectious cystic fibrosis patients: results of a systematic review. *Infect Control Hosp Epidemiol* 26(4), pp. 401-409. doi: 10.1086/502558

Voynow, J. A., Fischer, B. M. and Zheng, S. 2008. Proteases and cystic fibrosis. *Int J Biochem Cell Biol* 40(6-7), pp. 1238-1245. doi: 10.1016/j.biocel.2008.03.003

Waite, R. D., Qureshi, M. R. and Whiley, R. A. 2017. Modulation of behaviour and virulence of a high alginate expressing *Pseudomonas aeruginosa* strain from cystic fibrosis by oral commensal bacterium *Streptococcus anginosus*. *PLoS One* 12(3), p. e0173741. doi: 10.1371/journal.pone.0173741

Wang, H., Anthony, D., Selemidis, S., Vlahos, R. and Bozinovski, S. 2018. Resolving Viral-Induced Secondary Bacterial Infection in COPD: A Concise Review. *Front Immunol* 9, p. 2345. doi: 10.3389/fimmu.2018.02345

Wang, L., Hao, K., Yang, T. and Wang, C. 2017. Role of the Lung Microbiome in the Pathogenesis of Chronic Obstructive Pulmonary Disease. *Chin Med J (Engl)* 130(17), pp. 2107-2111. doi: 10.4103/0366-6999.211452

Wang, Q., Garrity, G. M., Tiedje, J. M. and Cole, J. R. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73(16), pp. 5261-5267. doi: 10.1128/AEM.00062-07

Wang, S. S., Liu, K. N. and Wang, B. W. 2010. Effects of dithiothreitol on the amyloid fibrillogenesis of hen egg-white lysozyme. *Eur Biophys J* 39(8), pp. 1229-1242. doi: 10.1007/s00249-010-0576-0

Wang, Y., Xu, L., Gu, Y. Q. and Coleman-Derr, D. 2016. MetaCoMET: a web platform for discovery and visualization of the core microbiome. *Bioinformatics* 32(22), pp. 3469-3470. doi: 10.1093/bioinformatics/btw507

Wang, Z. et al. 2020. A Refined View of Airway Microbiome in Chronic Obstructive Pulmonary Disease at Species and Strain-Levels. *Front Microbiol* 11, p. 1758. doi: 10.3389/fmicb.2020.01758

Waters, V. J. and Lipuma, J. 2020. Bacterial Infections and the Respiratory Microbiome. *Journal of Cystic Fibrosis*, pp. 73-92. doi: 10.1007/978-3-030-42382-7_5

Watson, R. L., de Koff, E. M. and Bogaert, D. 2019. Characterising the respiratory microbiome. *Eur Respir J* 53(2), doi: 10.1183/13993003.01711-2018

Watterson, W. J. et al. 2020. Droplet-based high-throughput cultivation for accurate screening of antibiotic resistant gut microbes. *Elife* 9, doi: 10.7554/eLife.56998

Wears, R. L. 2015. Standardisation and Its Discontents. *Cogn Technol Work* 17(1), pp. 89-94. doi: 10.1007/s10111-014-0299-6

Weiler, C. A. and Drumm, M. L. 2013. Genetic influences on cystic fibrosis lung disease severity. *Front Pharmacol* 4, p. 40. doi: 10.3389/fphar.2013.00040

Weiser, R. 2015. *The resistance of Pseudomonas aeruginosa to preservatives used in industrial formulations.* Cardiff University.

Weiser, R., Rye, P. D. and Mahenthiralingam, E. 2021. Implementation of microbiota analysis in clinical trials for cystic fibrosis lung infection: Experience from the OligoG phase 2b clinical trials. *J Microbiol Methods* 181, p. 106133. doi: 10.1016/j.mimet.2021.106133

Welp, A. L. and Bomberger, J. M. 2020. Bacterial Community Interactions During Chronic Respiratory Disease. *Front Cell Infect Microbiol* 10, p. 213. doi: 10.3389/fcimb.2020.00213

Weycker, D., Hansen, G. L. and Seifer, F. D. 2017. Prevalence and incidence of noncystic fibrosis bronchiectasis among US adults in 2013. *Chron Respir Dis* 14(4), pp. 377-384. doi: 10.1177/1479972317709649

White, J. R., Nagarajan, N. and Pop, M. 2009. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput Biol* 5(4), p. e1000352. doi: 10.1371/journal.pcbi.1000352

Wick, R. R., Judd, L. M., Gorrie, C. L. and Holt, K. E. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Computational Biology* 13(6), p. e1005595. doi: 10.1371/journal.pcbi.1005595

Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York:Springer-Verlag.

Wickham, H. et al. 2019. Welcome to the tidyverse. *J Open Source Softw* 4(43), p. 1686. doi: 10.21105/joss.01686

Wickham, H., François, R., Henry, L. and K, M. 2018. dplyr: A Grammar of Data Manipulation. Available at: https://CRAN.R-project.org/package=dplyr [Accessed: 8 February 2018].

Wiegmann, P. M., de Vries, Henk J. Blind, Knut. 2017. Multi-mode standardisation: A critical review and a research. *Res* 46(8), pp. 1370-1386. doi: 10.1016/j.respol.2017.06.002

Wiser, M. J. and Lenski, R. E. 2015. A Comparison of Methods to Measure Fitness in *Escherichia coli*. *PLoS One* 10(5), p. e0126210. doi: 10.1371/journal.pone.0126210

Wright, R. M. et al. 2001. Improved cultural detection of *Burkholderia cepacia* from sputum in patients with cystic fibrosis. *J Clin Pathol* 54(10), pp. 803-805. doi: 10.1136/jcp.54.10.803

Yabuuchi, E. and Oyama, A. 1971. *Achromobacter xylosoxidans* n. sp. from human ear discharge. *Jpn J Microbiol* 15(5), pp. 477-481. doi: 10.1111/j.1348-0421.1971.tb00607.x

Yamamoto, M. et al. 2012. Molecular characterization of IMP-type metallo-beta-lactamases among multidrug-resistant *Achromobacter xylosoxidans. J Antimicrob Chemother* 67(9), pp. 2110-2113. doi: 10.1093/jac/dks179

Yang, B., Wang, Y. and Qian, P. Y. 2016. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinform* 17, p. 135. doi: 10.1186/s12859-016-0992-y

Yuan, S., Cohen, D. B., Ravel, J., Abdo, Z. and Forney, L. J. 2012. Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One* 7(3), p. e33865. doi: 10.1371/journal.pone.0033865

Zemanick, E. T. and Hoffman, L. R. 2016. Cystic Fibrosis: Microbiology and Host Response. *Pediatr Clin North Am* 63(4), pp. 617-636. doi: 10.1016/j.pcl.2016.04.003

Zemanick, E. T. et al. 2017. Airway microbiota across age and disease spectrum in cystic fibrosis. *Eur Respir J* 50(5), doi: 10.1183/13993003.00832-2017

Zhang, B., Brock, M., Arana, C., Dende, C., Hooper, L. and Raj, P. 2020a. Impact of bead-beating intensity on microbiome recovery in mouse and human stool: Optimization of DNA extraction. *Front Cell Infect Microbiol* 11, p. 678522. doi: 10.1101/2020.06.15.151753

Zhang, D., Li, S., Wang, N., Tan, H. Y., Zhang, Z. and Feng, Y. 2020b. The Cross-Talk Between Gut Microbiota and Lungs in Common Lung Diseases. *Front Microbiol* 11, p. 301. doi: 10.3389/fmicb.2020.00301

Zlosnik, J. E., Mori, P. Y., To, D., Leung, J., Hird, T. J. and Speert, D. P. 2014. Swimming motility in a longitudinal collection of clinical isolates of *Burkholderia cepacia* complex bacteria from people with cystic fibrosis. *PLoS One* 9(9), p. e106428. doi: 10.1371/journal.pone.0106428

Zlosnik, J. E. A. et al. 2020. Epidemiology of *Burkholderia* Infections in People with Cystic Fibrosis in Canada between 2000 and 2017. *Ann Am Thorac Soc* 17(12), pp. 1549-1557. doi: 10.1513/AnnalsATS.201906-443OC

Zong, Z. Y. and Peng, C. H. 2011. *Ralstonia mannitolilytica* and COPD: a case report. *Eur Respir J* 38(6), pp. 1482-1483. doi: 10.1183/09031936.00046011

Appendix A

SI Table 1. Mock microbial standard B community composition

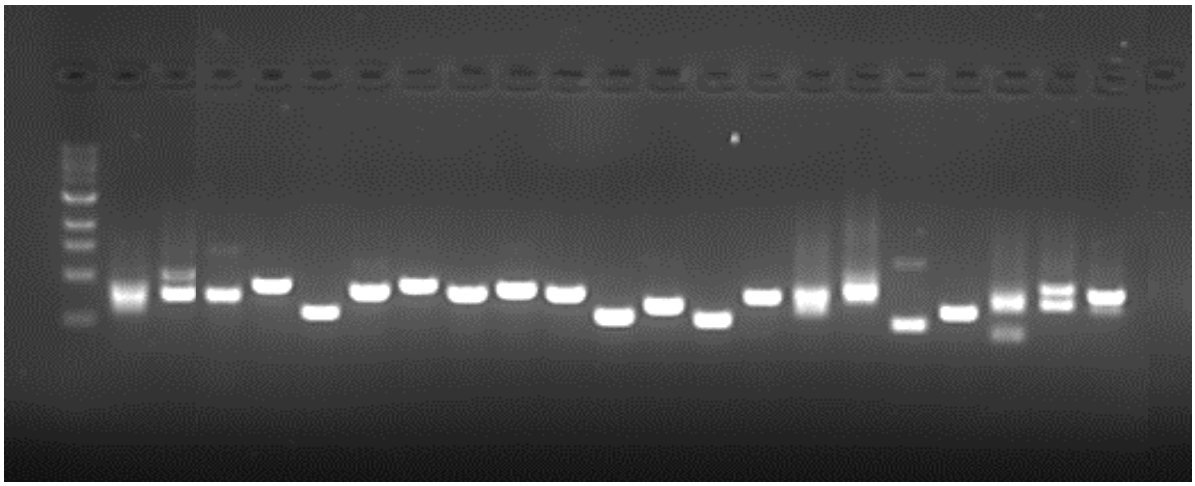
Species	Number of 16S rRNA gene operons	DNA concentration (pg/µl)
Acinetobacter baumannii	10000	8.2
Actinomyces odontolyticus	1000	1
Bacillus cereus	100000	45
Bacteroides vulgatus	1000	0.8
Clostridium beijerinckii	100000	44
Cutibacterium acnes	10000	8.8
Deinococcus radiodurans	1000	1
Enterococcus faecalis	1000	0.7
Escherichia coli	1000000	680
Helicobacter pylori	10000	8.6
Lactobacillus gasseri	10000	3.2
Listeria monocytogenes	10000	5
Neisseria meningitidis	10000	5.8
Pseudomonas aeruginosa	100000	160
Rhodobacter sphaeroides	1000000	1400
Staphylococcus aureus	100000	59
Staphylococcus epidermidis	1000000	510
Streptococcus agalactiae	100000	32
Streptococcus mutans	1000000	420
Streptococcus pneumoniae	1000	0.6

SI Table 2. Illumina indexed primers and sequencing primers

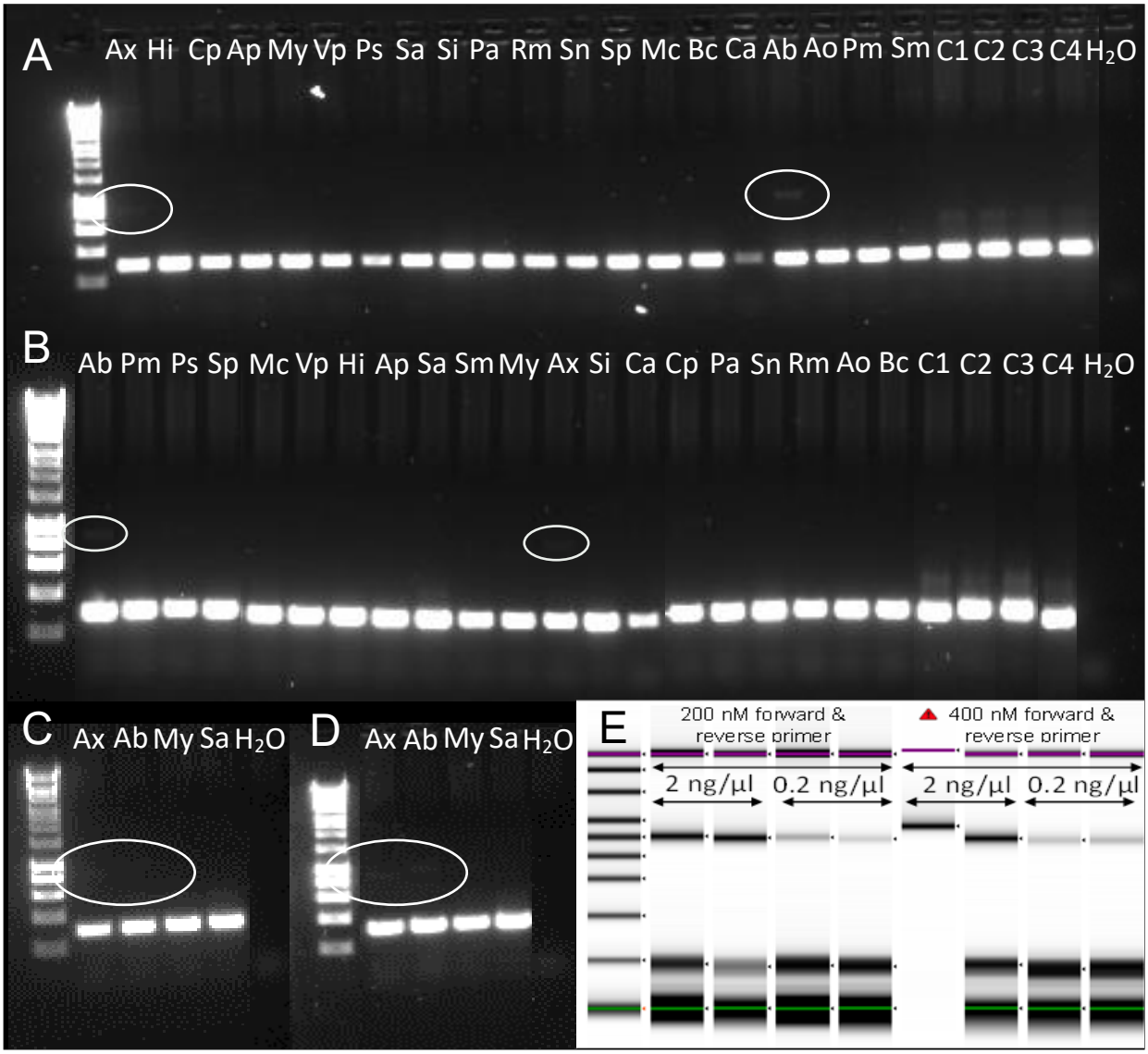
Index	Sequence
A501	AATGATACGGCGACCACCGAGATCTACACATCGTACGTATGGTAATTGTGTGCCAGCMGCCGCGGTAA
A502	AATGATACGGCGACCACCGAGATCTACACACTATCTGTATGGTAATTGTGTGCCAGCMGCCGCGGTAA
A503	AATGATACGGCGACCACCGAGATCTACACTAGCGAGTTATGGTAATTGTGTGCCAGCMGCCGCGGTAA
A504	AATGATACGGCGACCACCGAGATCTACACCTGCGTGTTATGGTAATTGTGTGCCAGCMGCCGCGGTAA
A505	AATGATACGGCGACCACCGAGATCTACACTCATCGAGTATGGTAATTGTGTGCCAGCMGCCGCGGTAA
A506	AATGATACGGCGACCACCGAGATCTACACCGTGAGTGTATGGTAATTGTGTGCCAGCMGCCGCGGTAA
A507	AATGATACGGCGACCACCGAGATCTACACGGATATCTTATGGTAATTGTGTGCCAGCMGCCGCGGTAA
A508	AATGATACGGCGACCACCGAGATCTACACGACACCGTTATGGTAATTGTGTGCCAGCMGCCGCGGTAA
B701	CAAGCAGAAGACGGCATACGAGATAAGTCGAGAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
B702	CAAGCAGAAGACGGCATACGAGATATACTTCGAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
B703	CAAGCAGAAGACGGCATACGAGATAGCTGCTAAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
B704	CAAGCAGAAGACGGCATACGAGATCATAGAGAAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
B705	CAAGCAGAAGACGGCATACGAGATCGTAGATCAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
B706	CAAGCAGAAGACGGCATACGAGATCTCGTTACAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT

Primer	Sequence
P1 V4	TATGGTAATTGTGTGCCAGCMGCCGCGGTAA
P2 V4	AGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
V4	ATTAGAWACCCBDGTAGTCCGGCTGACTGACT

Appendix B

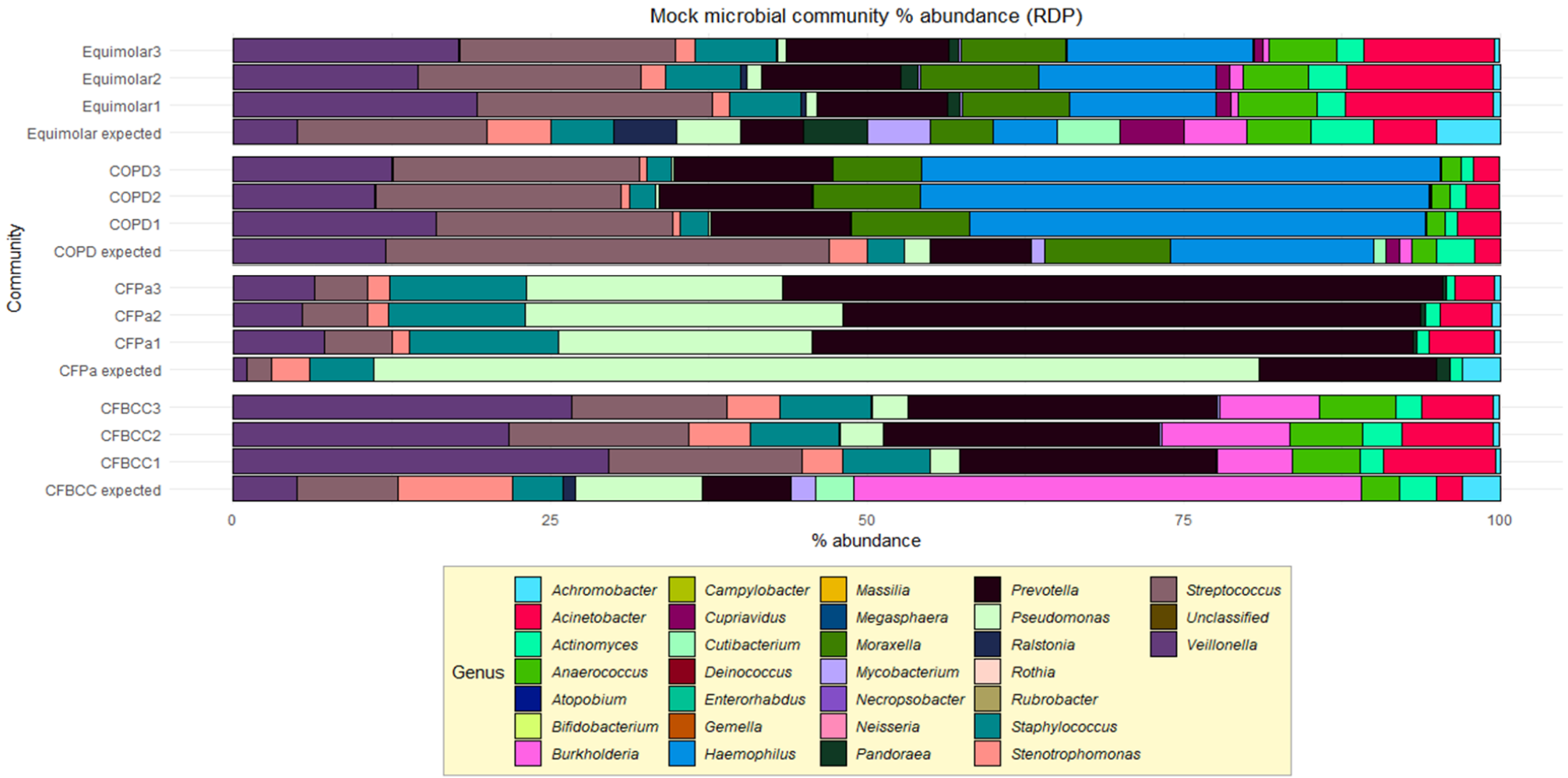


SI Figure 1 Single-species RISA PCR references visualised on a 1% agarose gel. A 1 kb ladder is included on the left followed by *H. influenzae, B. cepacia, A. xylosoxidans, M. abscessus, S. maltophilia, A. baumannii, R. mannitolilytica, P. sputorum, C. pauculus, S. intermedius, S. anginosus, S. pneumoniae, M. catarrhalis, S. aureus, V. parvula, C. acnes, A. odontolyticus, A. prevotii, P. melaninogenica, P. aeruginosa* and PCR nuclease-free water control. This image was edited to include the PCR control at the end of the gel.

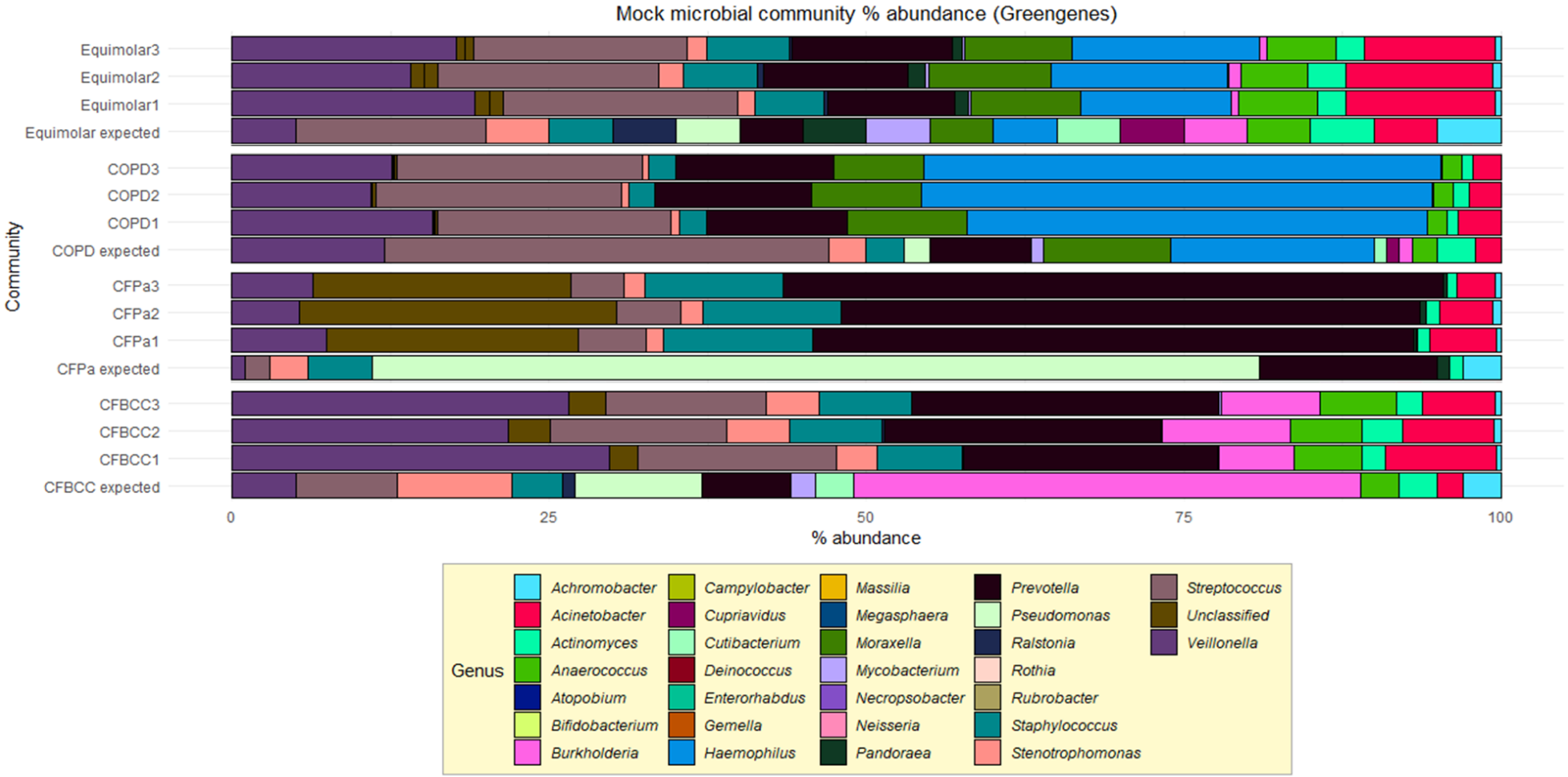


SI Figure 2 16S rRNA gene PCR optimisation. Each image portrays: A) a 30 second extension at 72°C using 400 nM of both forward and reverse primer. B) a 30 second extension at 68°C using 200 nM of both forward and reverse primer and a 5-minute final extension. C) Example gel using 400 nM of both forward and reverse primer using a 30 second extension at 68°C without a final extension. D) Example gel using 400 nM of both forward and reverse primer using a 30 second extension at 68°C with a 5-minute final extension. E) A comparison of *S. aureus* amplified DNA using a 200 nM and 400 nM of both forward and reverse primer using optimal PCR extension. The first lane shows the ScreenTape ladder, and the Tapestation upper and lower markers are highlighted in purple and green in each lane. Here, Diluted DNA was quantified using HS dsDNA Qubit reagents. Residual adapter primer is visible above the 25 bp lower marker (in green). Indexed primers were used in this ScreenTape, resulting in ~400 bp amplicons (including fifth lane which failed alignment). Abbreviations represent: Ab: *A. baumannii*, Ao: *A. odontolyticus*, Ax: *A. xylosoxidans*, Bc: *B. cepacia*, C1: CFBCC mock community, C2: CFPa mock community, C3: COPD mock community, C4: Equimolar mock community, Ca: *C. acnes*, Cp: *C. pauculus*, Hi: *H. influenzae*, Mc: *M. catarrhalis*, My: *M. abscessus*, Pa: *P. aeruginosa*, Pm: *P. melaninogenica*, Rm: *R. mannitolilytica*, Sn: *S. anginosus*, Sa: *S. aureus*, Si: *S. intermedius*, Sm: *S. maltophilia*, Sp: *S. pneumoniae*, Vp: *V. parvula*, H2O: PCR water negative control. Spurious amplification in single species samples have been circled and were ~800 bp and ≤20 ng/µl, based on the lowest concentration on a 1 kb ladder reference. PCR artifacts were also visible in mock community samples (~600 bp). Both brightness and contrast of each gel

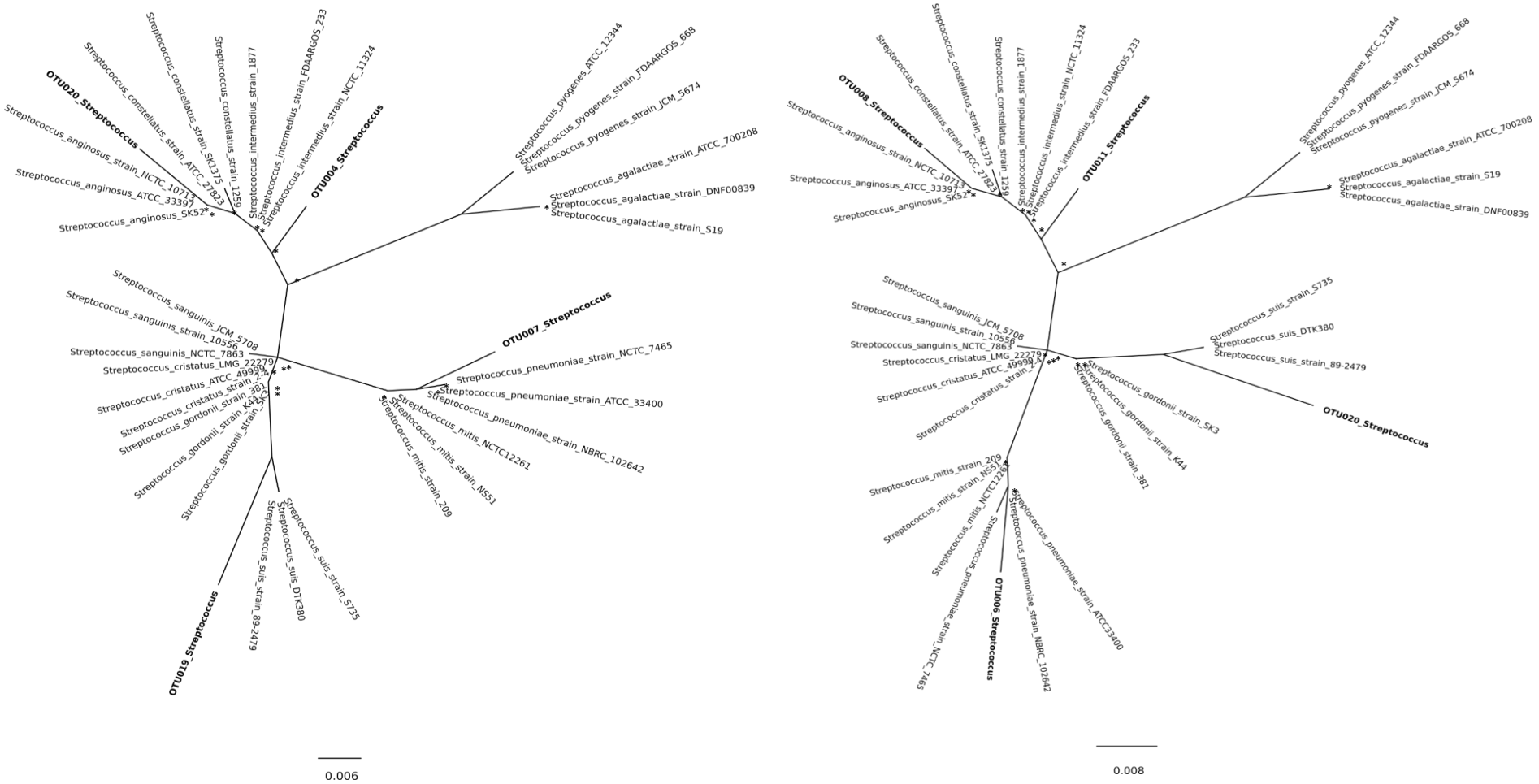
image was increased to visualise non-specific amplification. Images have been edited to rearrange samples in order and to remove samples that were discarded from this study.



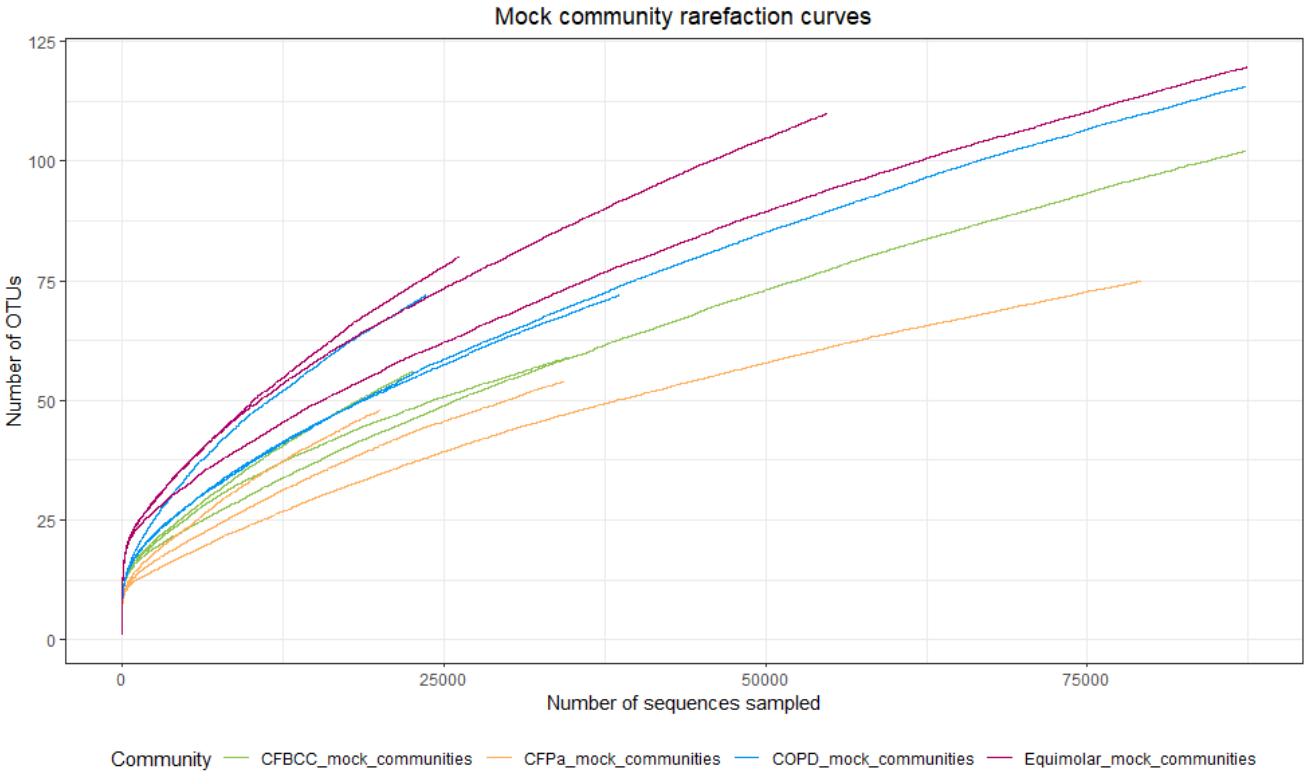
SI Figure 3 RDP mock community composition by % relative abundance. Due to the low number of genera detected, relative abundance was not capped prior to generating bar plots. Each sequenced run is shown to visualise batch variation in addition to expected community composition. All unclassified sequences were grouped together. Abbreviations include: Equimolar; equimolar community, COPD; COPD community, CFPa; cystic fibrosis *Pseudomonas* dominated community, CFBCC; cystic fibrosis *Burkholderia cepacia* complex dominated community. Bars were edited to make comparisons easier between groups.



SI Figure 4 Greengenes mock community composition by % relative abundance. Due to the low number of genera detected, relative abundance was not capped prior to generating bar plots. Each sequenced batch is shown to visualise batch variation in addition to expected community composition. All unclassified sequences were grouped together.Abbreviations include: Equimolar; equimolar community, COPD; COPD community, CFPa; cystic fibrosis *Pseudomonas* dominated community, CFBCC; cystic fibrosis *Burkholderia cepacia* complex dominated community. Bars were edited to make comparisons easier between groups.



SI Figure 5 Streptococcal OTU phylogenies. Streptococcal OTUs were taxonomically assigned through SILVA (left) and Greengenes (right) and a maximum likelihood phylogeny was constructed and edited in FigTree and Inkscape. Bootstrap values <80 are marked with an asterisk.



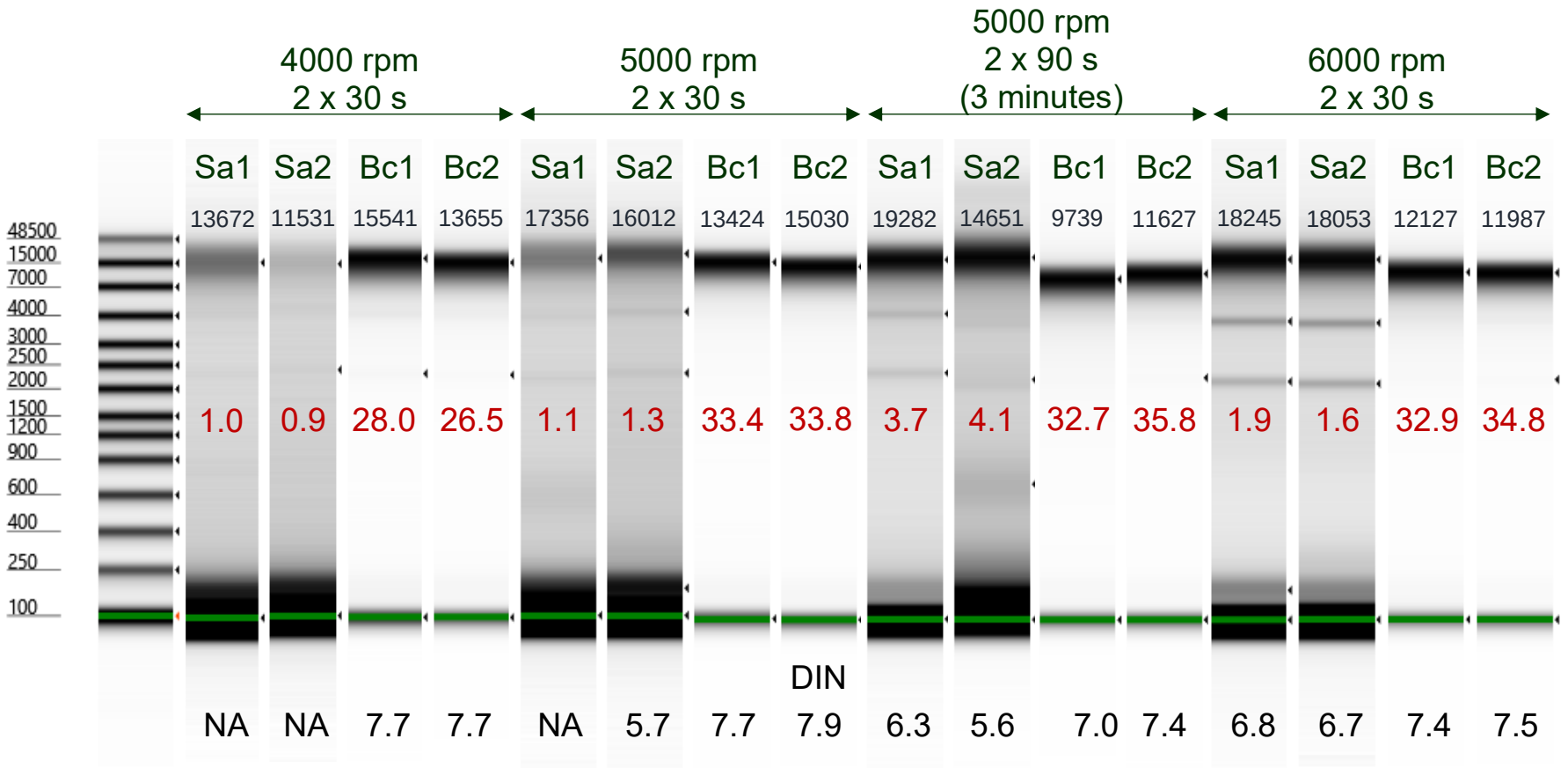
SI Figure 6 Mock community rarefaction curves. Each curve has been coloured according to mock community type.

SI Table 1 Mock community alpha diversity predictions[a,b]

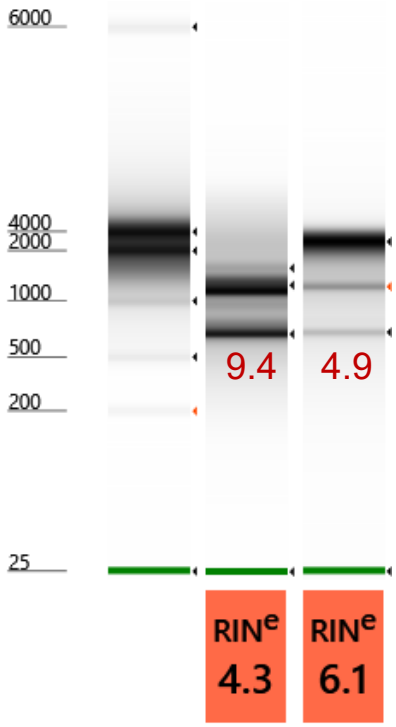
Community	Richness[c]			Diversity	
	ACE	Chao1	Sobs	InvSimpson	Shannon
CFBCC	14.4 259.4	14.3 179.3	14 ~76	5.3 6.3	2.1 2.1
CFPa	11.4 162.4	13.5 124	9 ~59	2 3.4	1.1 1.6
COPD	22.6 265.3	20.1 165.1	16 ~89	6 4.5	2.1 1.8
Eq	20 320.2	20 235.9	20 ~111	24.8 9.4	3 2.5

[a]Abbreviations include: CFBCC; CF *Burkholderia cepacia* complex-dominated mock community, CFPa; CF *Pseudomonas*-dominated mock community, COPD; COPD mock community, Eq; equimolar mock community.
[b]Predicted data has been included above mean experimental values.
[c]ACE and Chao1 calculated using SpadeR software https://chao.shinyapps.io/SpadeR/ (Chao et al. 2016).

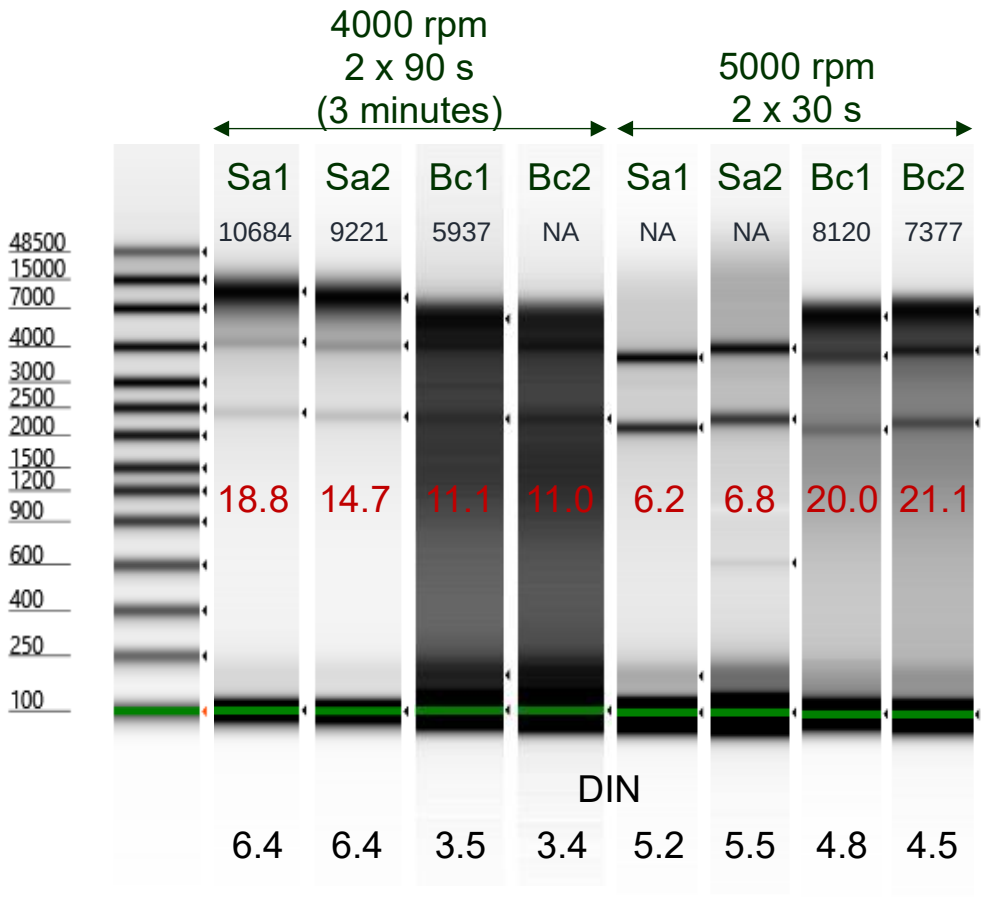
Appendix C



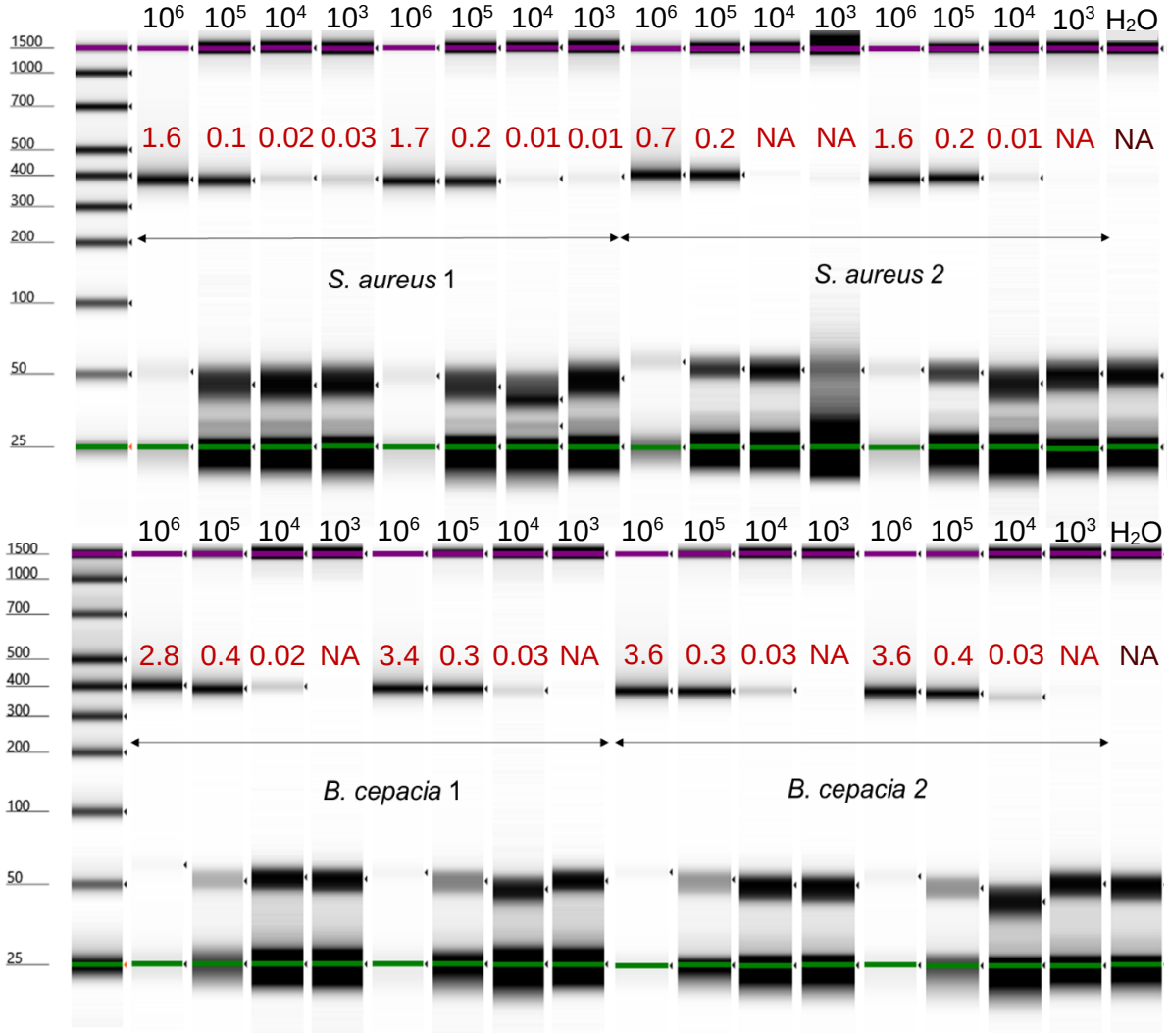
SI Figure 1 Supplemental EMAG ceramic bead-beating optimisation settings. This image has been edited so that each bead-beating speed setting is shown in order for each replicate of *S. aureus* (Sa) and *B. cepacia* (Bc). The first lane shows the genomic DNA ScreenTape ladder and the Tapestation lower marker is highlighted in green in each lane. Sample DNA molecular weight (in bp) has been included beneath each sample label (navy). Qubit scores (in ng/µl) have been annotated in each sample lane (red) and DNA integrity (DIN) scores are beneath each sample lane.



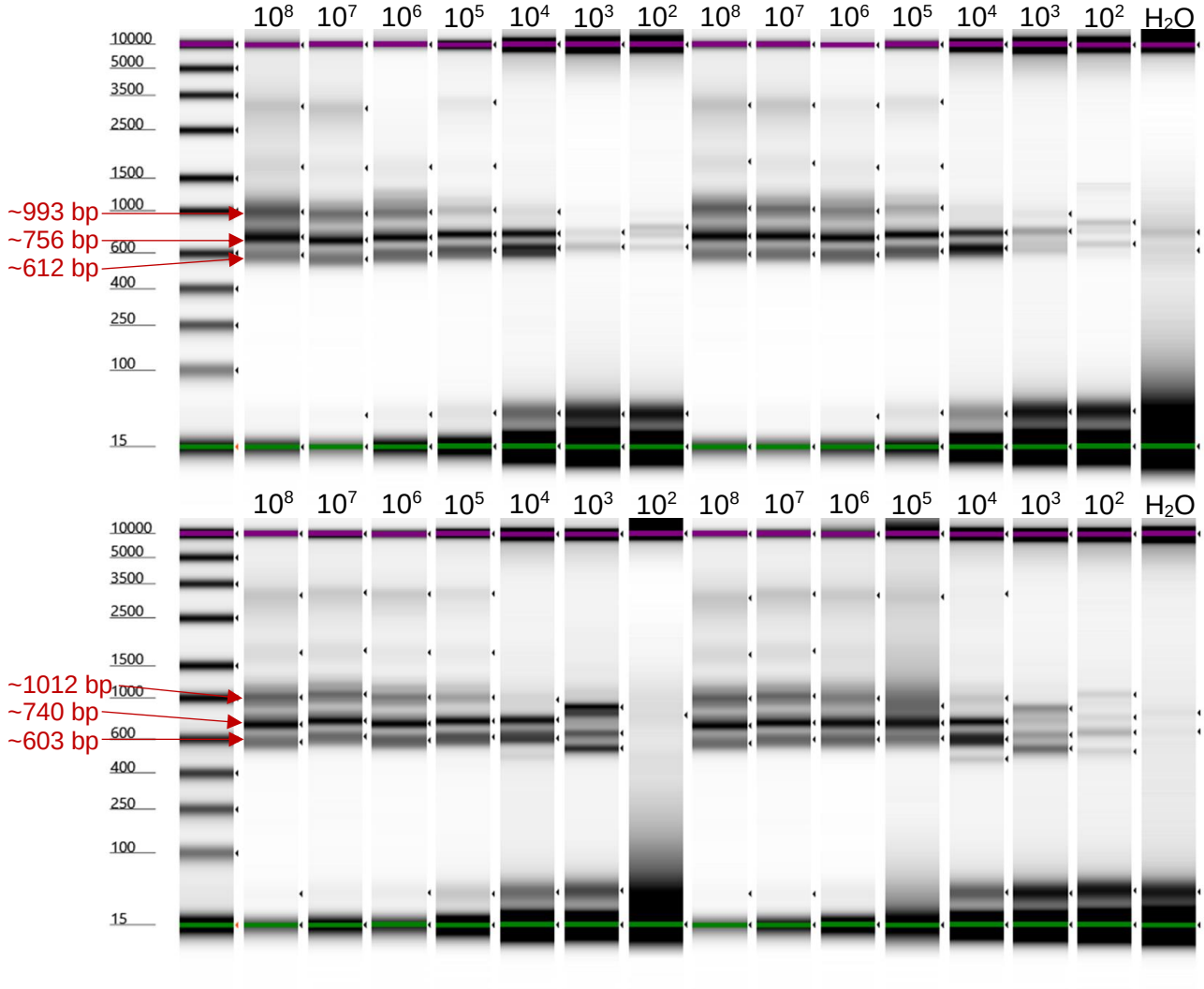
SI Figure 2 Example 2200 Agilent RNA ScreenTape visualising EMAG RNA co-extraction. The first lane shows the RNA ScreenTape ladder, and the Tapestation lower marker is highlighted in green in each lane RNA co-extraction was evident in EMAG gDNA extracts where distinct fragments could be observed at ~670 bp, >1200 bp (16S rRNA) and >2800 bp (23S rRNA). RNA samples were derived from *S. aureus* that was lysed using ZR BashingBead™ tubes at 5000 rpm (2 x 30 seconds; left), or with ceramic beads at 6000 rpm (2 x 30 seconds; right). Qubit scores (in ng/µl) has been annotated in each sample lane (red). RNA integrity number (RIN) is included where RIN can be interpreted as: 1=highest degradation and 10=highest quality.



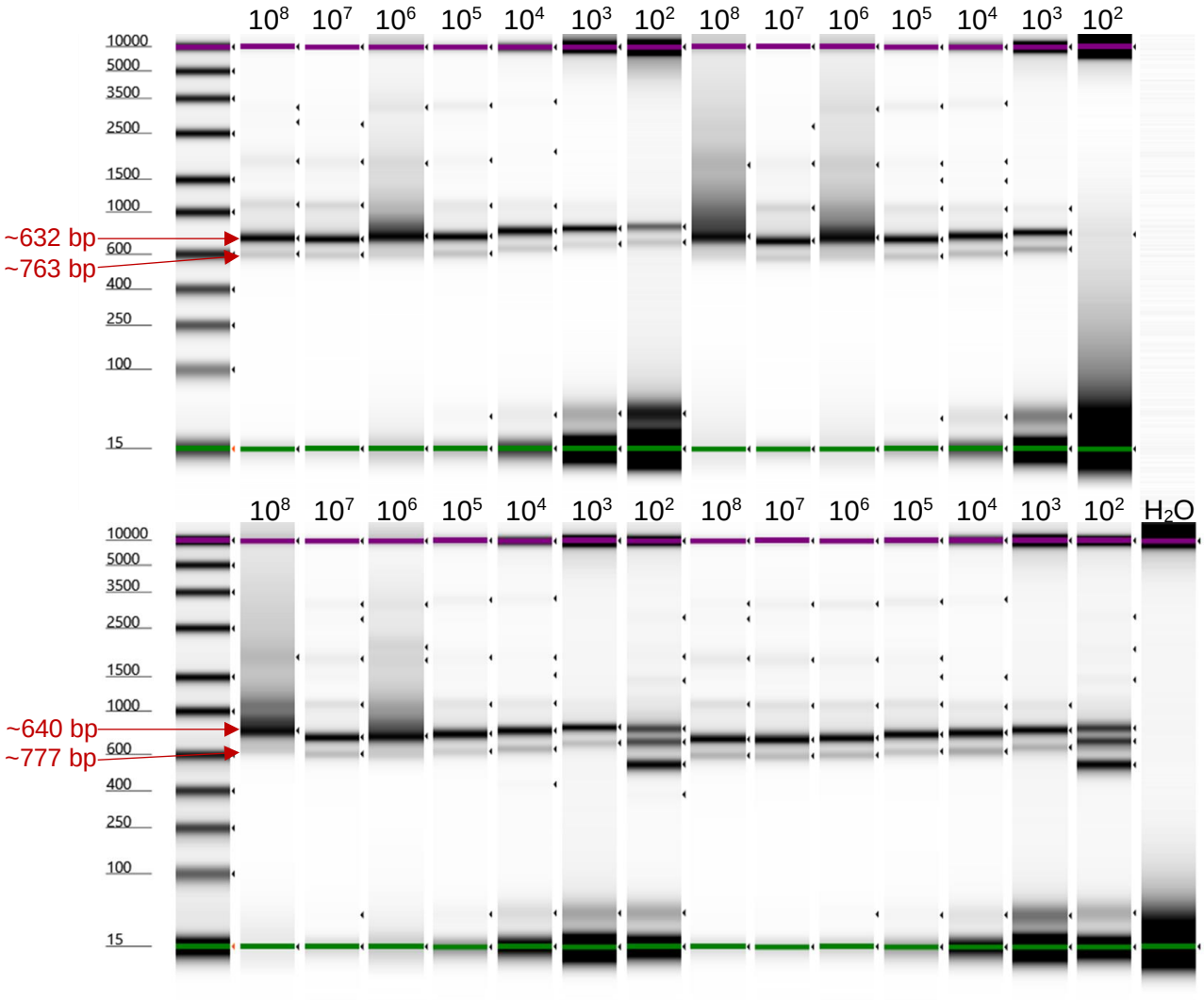
SI Figure 3 Supplementary EMAG ZR BashingBead™ lysis optimisation settings. Bead-beating speed and duration has been ordered on this composite image showing both *S. aureus* (Sa) and *B. cepacia* (Bc) replicates. The first lane shows the genomic DNA ScreenTape ladder and the Tapestation lower marker is highlighted in green in each lane. Sample DNA molecular weight (in bp) has been included beneath each sample label (navy). Qubit scores (in ng/µl) have been annotated in each sample lane (red) and DNA integrity (DIN) scores have been included last.



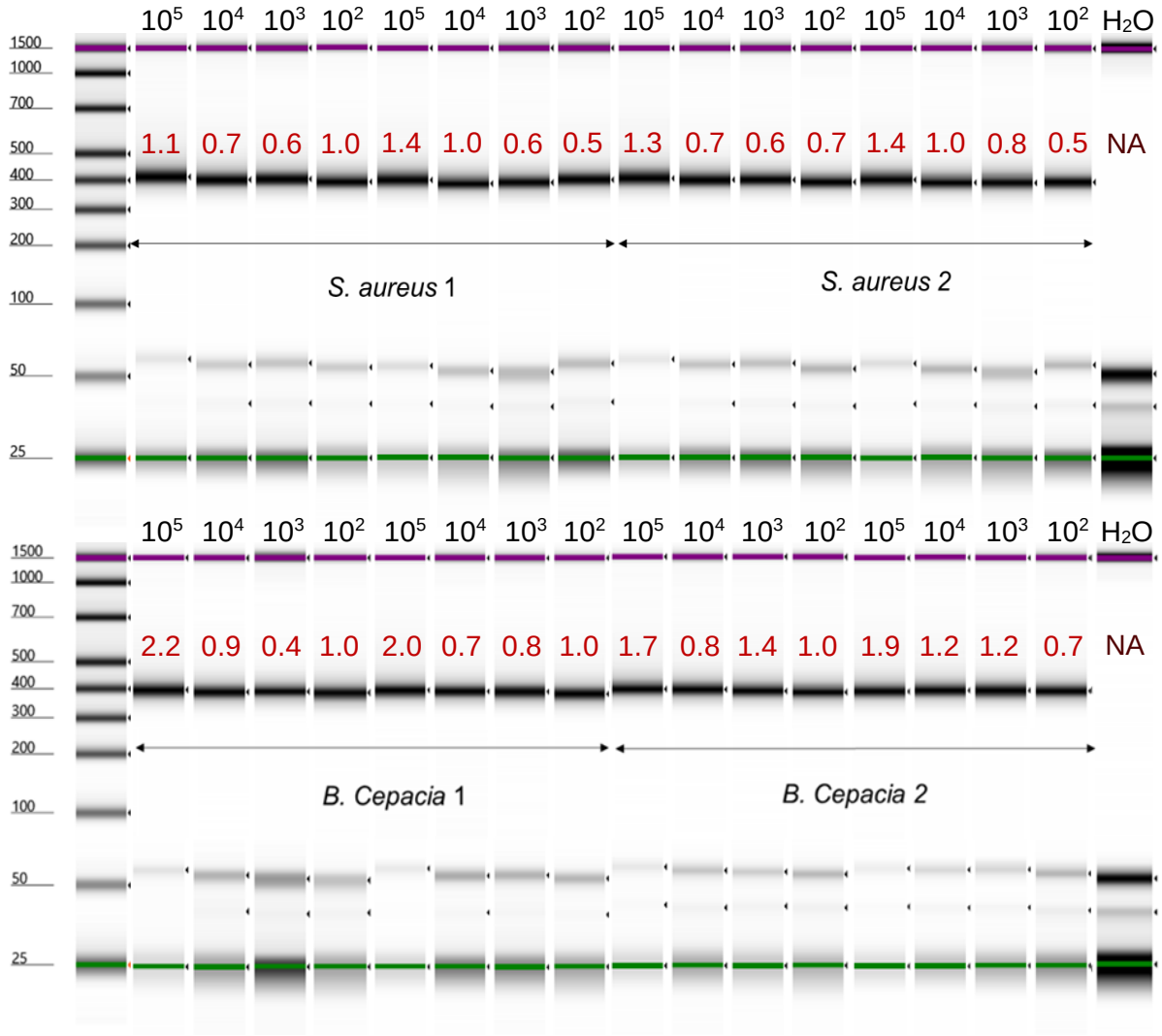
SI Figure 4 EMAG mechanical lysis 16S rRNA gene PCR limit of detection. Resulting amplified DNA was visualised with a 2200 TapeStation HS D1000 ScreenTape where this output has been edited to show 10^4 CFU/ml to 10^6 CFU/ml of each replicate in order. The first lane shows the ScreenTape ladder, and the Tapestation upper and lower markers are highlighted in purple and green in each lane. Qubit scores (in ng/µl) have been annotated in each sample lane. PCR nuclease-free water controls have been included last.



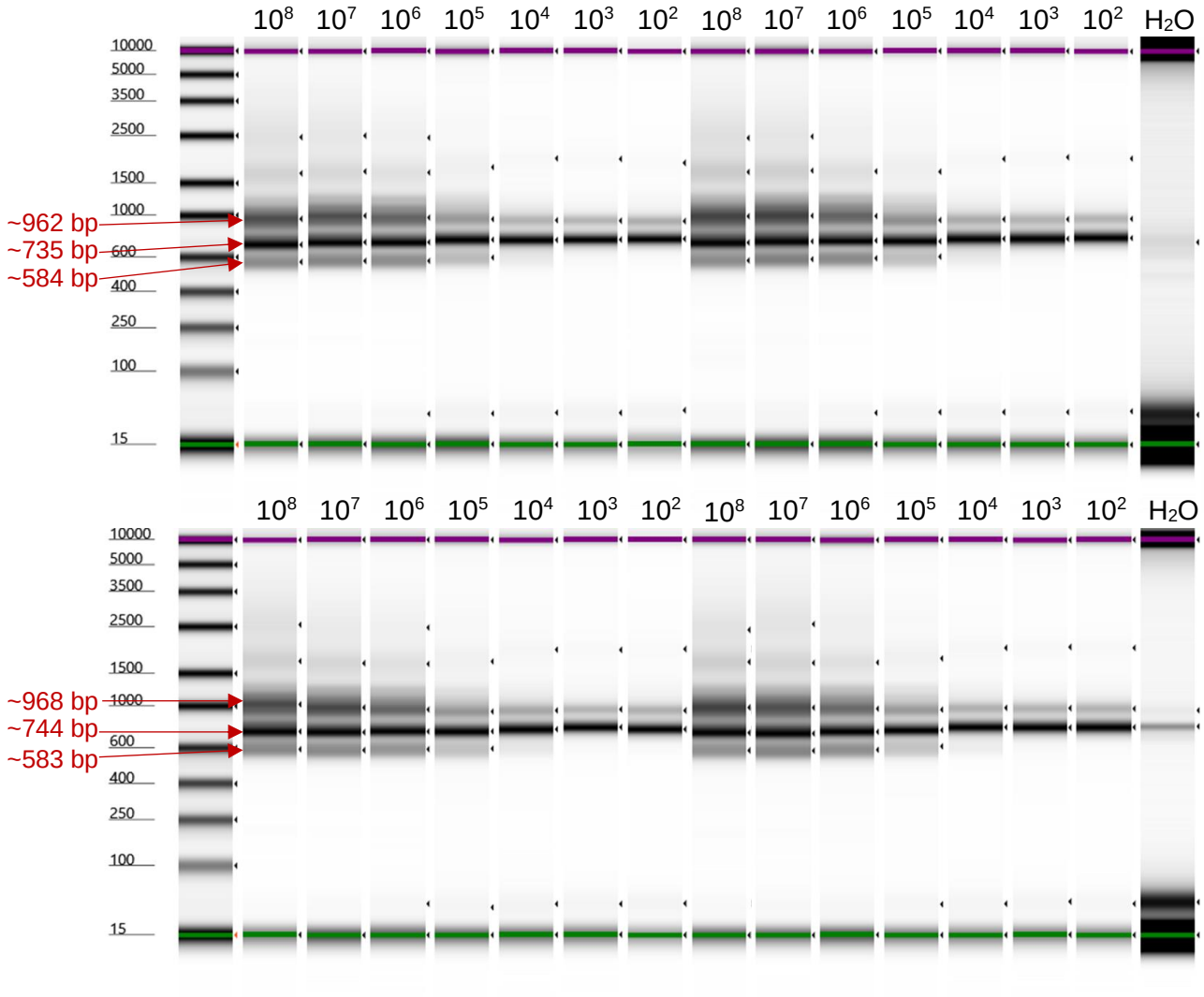
SI Figure 5 *S. aureus* EMAG mechanical lysis RISA PCR limit of detection. Amplified DNA was visualised with a 2200 TapeStation D5000 ScreenTape which has been edited to show 10^8 CFU/ml to 10^2 CFU/ml of each replicate in order. The first lane shows the ScreenTape ladder, and the Tapestation upper and lower markers are highlighted in purple and green in each lane. PCR nuclease-free water controls have been included last. In these controls, a 614 bp and 765 bp amplicon was measured (0.1 ng/µl and 0.3 ng/µl respectively; control 1) and a 661 bp and 833 bp amplicon (0.1 ng/µl each; control 2). Core *S. aureus* ITS amplicons have been annotated by calculating mean TapeStation ITS amplicon sizes for each ScreenTape.



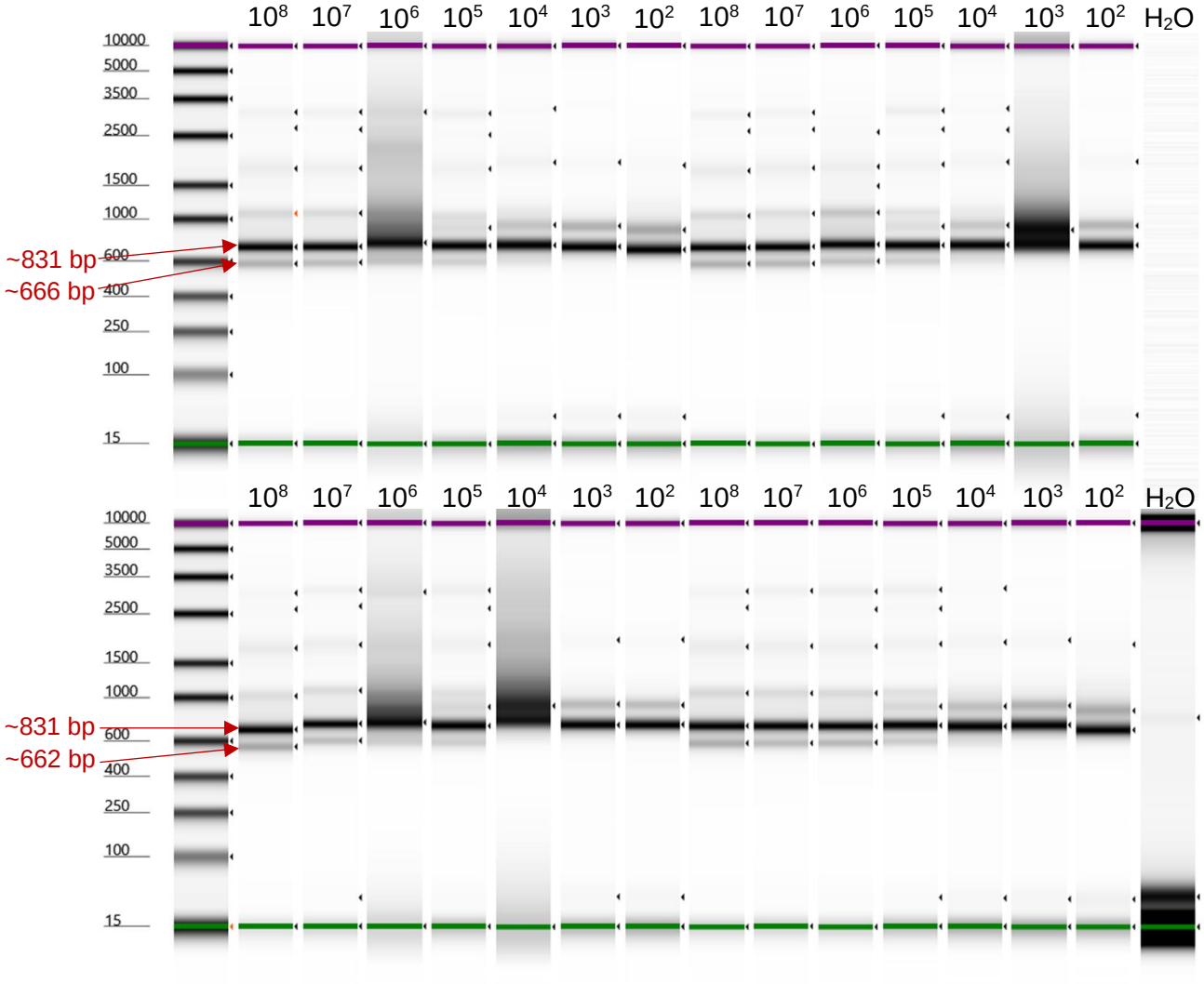
SI Figure 6 *B. cepacia* EMAG mechanical lysis RISA PCR limit of detection. Resulting amplified DNA was visualised with a 2200 TapeStation D5000 ScreenTape where its graphical output was edited to show 10^8 CFU/ml to 10^2 CFU/ml of each replicate in order. The first lane shows the ScreenTape ladder, and the Tapestation upper and lower markers are highlighted in purple and green in each lane. PCR nuclease-free water controls are included last. Note that one control lane failed detection although amplified sequences originating from PCR reagents were not significant in Figure 4.3**.** Core *B. cepacia* ITS amplicons have been annotated by calculating mean TapeStation ITS product sizes for each ScreenTape.



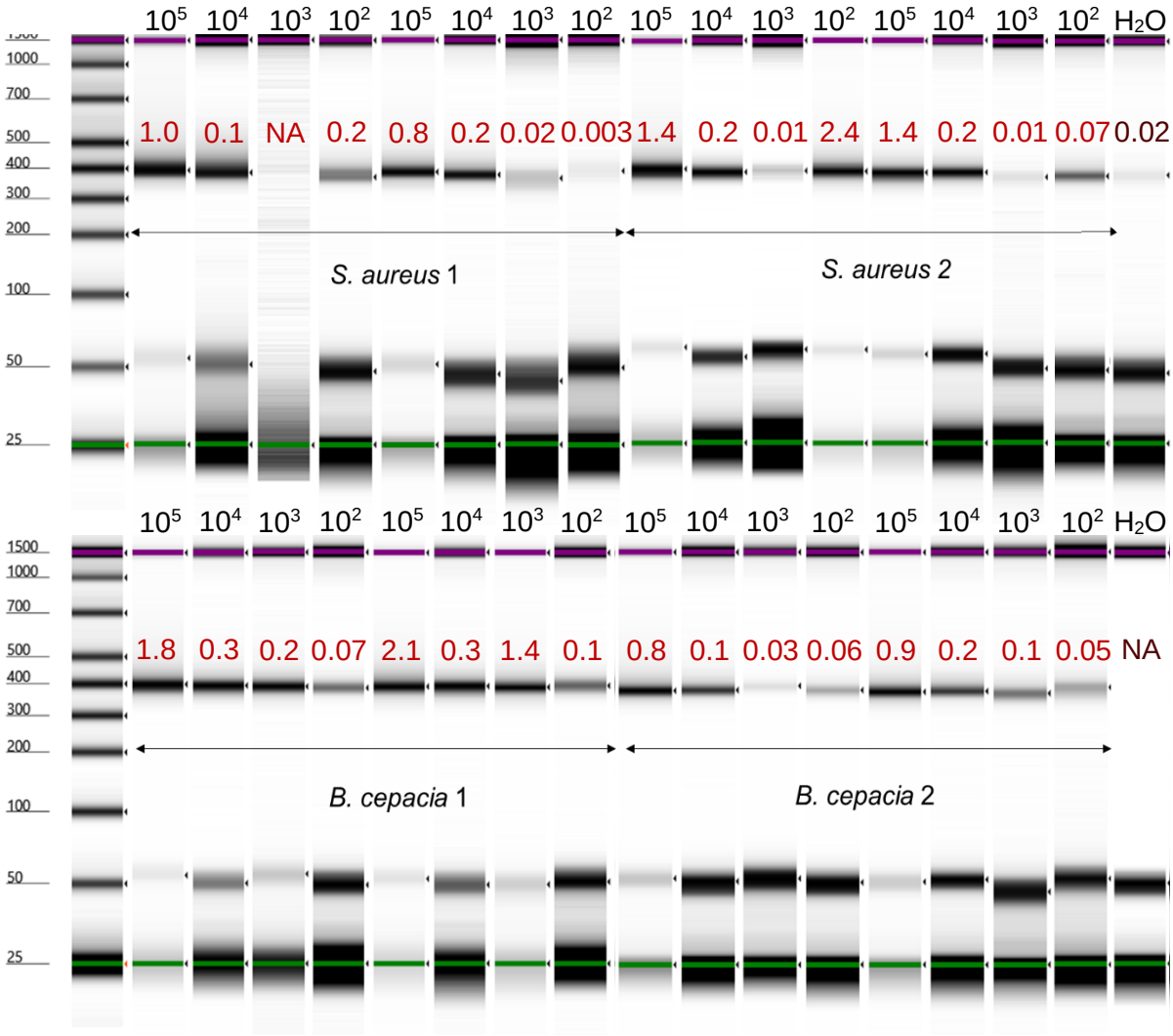
SI Figure 7 EMAG enzymatic lysis 16S rRNA gene PCR limit of detection. Resulting HS D1000 ScreenTape images were edited to show 10^2 CFU/ml to 10^5 CFU/ml of each replicate in order. The first lane shows the ScreenTape ladder, and the Tapestation upper and lower markers are highlighted in purple and green in each lane. Qubit scores (in ng/µl) have been annotated in each sample lane. PCR nuclease-free water controls have been included last.



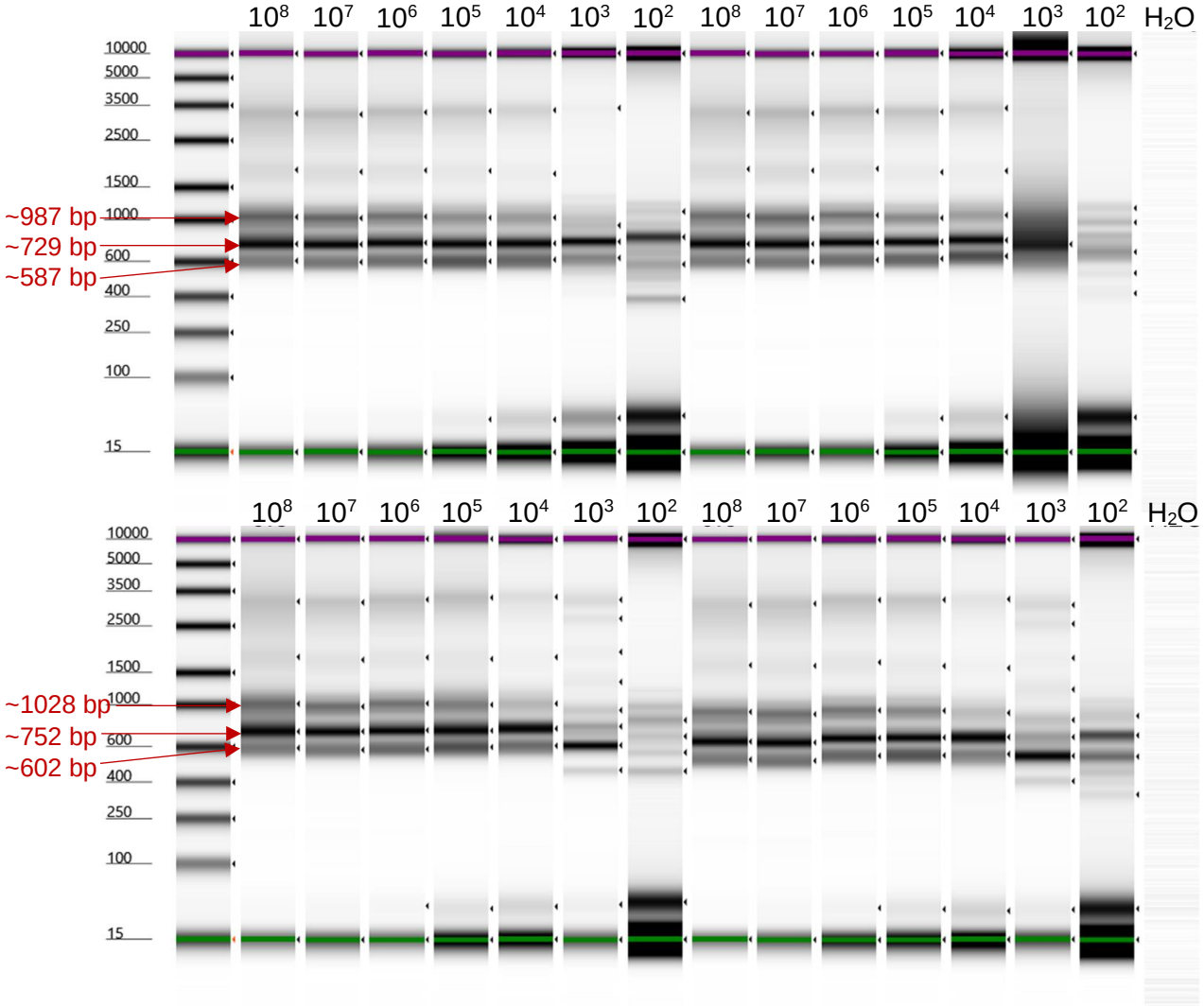
SI Figure 8 *S. aureus* EMAG enzymatic lysis RISA PCR limit of detection. Resulting TapeStation D5000 ScreenTape images were edited to include 10^8 CFU/ml to 10^2 CFU/ml of each replicate in order. The first lane shows the ScreenTape ladder, and the Tapestation upper and lower markers are highlighted in purple and green in each lane. PCR nuclease-free water controls are included. These controls consist of a 729 bp amplicon (0.1 ng/µl; control 1) and 780 bp and 949 p amplicons (0.3 ng/µl and 0.1 ng/µl respectively; control 2). Core *S. aureus* ITS amplicons have been annotated by calculating mean TapeStation ITS amplicon sizes for each ScreenTape.



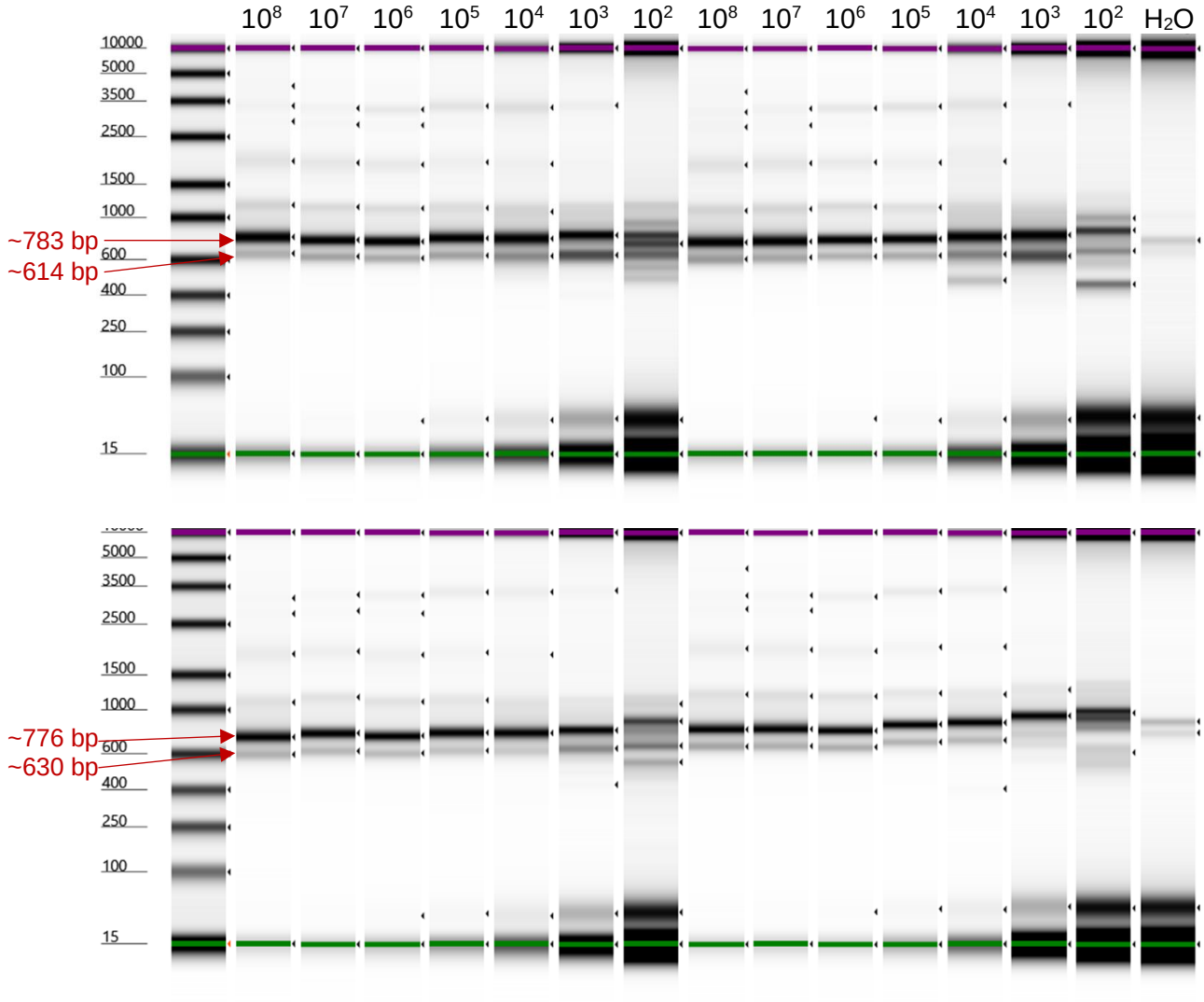
SI Figure 9 *B. cepacia* EMAG enzymatic lysis RISA PCR limit of detection. This composite image shows 10^8 CFU/ml to 10^2 CFU/ml of each replicate in order. The first lane shows the ScreenTape ladder, and the Tapestation upper and lower markers are highlighted in purple and green in each lane. PCR nuclease-free water controls are included. Note that one control lane failed detection, yet amplified sequences originating from PCR reagents were not detected on a 1% agarose gel (Figure 4.4). A second control identified a 790 bp (0.1 ng/µl) amplicon. Core *B. cepacia* ITS amplicons have been annotated by calculating mean TapeStation ITS product sizes for each ScreenTape.



SI Figure 10 ZymoBIOMICS™ 16S rRNA gene PCR limit of detection. This composite image shows 10^2 CFU/ml to 10^5 CFU/ml of all *S. aureus* and *B. cepacia* replicates in order. The first lane shows the ScreenTape ladder, and the Tapestation upper and lower markers are highlighted in purple and green in each lane. DNA concentration (in ng/µl) have been annotated in each sample lane. PCR nuclease-free water controls are included. Note one *S. aureus* sample that failed detection; refer to Figure 4.5.



SI Figure 11 *S. aureus* ZymoBIOMICS™ RISA PCR limit of detection. This ScreenTape was edited to show 10^2 CFU/ml to 10^8 CFU/ml of each replicate in order. The first lane shows the ScreenTape ladder, and the Tapestation upper and lower markers are highlighted in purple and green in each lane. PCR nuclease-free water controls failed detection, refer to Figure 4.5 where amplified sequences originating from PCR reagents were not absent. Core *S. aureus* ITS amplicons have been annotated by calculating mean TapeStation ITS product sizes for each ScreenTape.



SI Figure 12 *B. cepacia* ZymoBIOMICS™ RISA PCR limit of detection. 10^2 CFU/ml to 10^8 CFU/ml of each replicate is shown in order for each *S. aureus* and *B. cepacia* on this edited image. The first lane shows the ScreenTape ladder, and the Tapestation upper and lower markers are highlighted in purple and green in each lane. PCR nuclease-free water controls are included. Small amounts of amplified DNA were detected in these controls, equating to a 760 bp amplicon (0.1 ng/µl; control 1) and 701 bp and 803 bp amplicons (0.1 ng/µl and 0.2 ng/µl respectively; control 2). Core *B. cepacia* ITS amplicons have been annotated by calculating mean TapeStation ITS product sizes from each ScreenTape.

SI Table 1 Spike-in *S. aureus* and *B. cepacia in silico* ITS profiles and species producing similar profiles

Species	ITS predicted size (bp)
Staphylococcus aureus	**648, 743-757, 830**
Capnocytophaga sputigena	710, 717, 855, 1135
Veillonella parvula	883, 926, 1075
Burkholderia cepacia	**681, 830, 871**
Achromobacter deleyi	891
Achromobacter dolens	892
Achromobacter insuavis	876
Achromobacter pulmonis	884
Achromobacter ruhlandii	891
Acinetobacter baumannii	889
Acinetobacter modestus	889
Acinetobacter nosocomialis	889
Acinetobacter vivianii	935
Pandoraea aquatica	855
Pandoraea capi	844
Pandoraea fibrosis	831, 894-895
Pandoraea pneumonica	839
Pandoraea pnomenusa	835, 891
Pandoraea sputorum	848, 1911
Neisseria lactamica	910
Neisseria cineraea	885
Neisseria bacilliformis	861
Neisseria shayeganii	875
Neisseria mucosa	891
Neisseria flavescens	893
Neisseria subflava	893

SI Table 2 Comparison of *Staphylococcus* (Sa) and *Burkholderia* (Bc) 16S rRNA gene sequenced reads in sputum

Sample	Reads	% Sa/Bc	Sa/Bc reads
Sa1 10^6	24703	0.8	198
Sa1 10^5	21681	0.04	9
Sa1 10^4	29662	0.01	3
Sa1 10^3	23986	0	0
Sa1 10^2	20780	0.005	1
Sa2 10^6	22178	0.8	177
Sa2 10^5	24004	0.04	10
Sa2 10^4	23429	0.03	7
Sa2 10^3	24583	0.02	5
Sa2 10^2	23723	0	0
Sputum content Sa spike-in	22496	0.02	4
Bc1 10^6	17634	0.03	5
Bc1 10^5	17886	0.01	2
Bc1 10^4	18659	0.02	4
Bc1 10^3	16768	0	0
Bc1 10^2	18659	0	0
Bc2 10^6	17747	0.07	12
Bc2 10^5	14830	0	0
Bc2 10^4	19835	0	0
Bc2 10^3	17248	0	0
Bc2 10^2	18448	0	0
Sputum content Bc spike-in	16280	0.006	1
ZymoBIOMICS kit blank	31	9.6 (Sa); 3.2 (Bc)	3 (Sa); 1 (Bc)
Mucolyse	21	33.3 (Sa); 0 (Bc)	7 (Sa); 0 (Bc)
PCR nuclease-free water	22	4.5 (Sa); 0 (Bc)	1 (Sa); 0 (Bc)

Appendix D

SI Table 1 Patient clinical parameters and culture-based microbiological identification

ID[a]	Age	Sex	% FEV_1[b]	Disease severity	Antibiotics	Culture	Culture positive	MALDI-TOF
471	84	Male	37	Severe	No	*S. maltophilia*	1	*S. maltophilia*
483	62	Male	34	Severe	No	NRF, *P. aeruginosa, S. agalactiae*	3	*P. aeruginosa, S. agalactiae*
699	75	Female	44	Severe	No	NRF	1	Not Done
707	79	Female	63	Moderate	Ciprofloxacin	NRF, *H. influenzae*, Yeast	3	*H. influenzae*
718	75	Female	84	Mild	No	NRF, *P. aeruginosa*	2	*P. aeruginosa*
758	29	Female	122	Mild	No	NRF	1	Not Done
788	71	Female	52	Moderate	Amoxicillin	NRF, *H. influenzae*	2	*H. influenzae*
790	78	Female	94	Mild	Doxycycline, Amoxicillin	NRF	1	Not Done
913	75	Male	94	Mild	No	*P. aeruginosa, C. striatum*	2	*P. aeruginosa, C. striatum*
944	79	Male	87	Mild	No	NRF, Yeast	1	Not Done
945	88	Female	58	Moderate	No	*K. oxytoca, S. marcescens, P. aeruginosa*, Yeast	4	*K. oxytoca, S. marcescens, P. aeruginosa*
995	60	Female	44	Severe	No	NRF, *E. coli*, Yeast	3	*E. coli*
996	82	Female	77	Moderate	No	NRF, Yeast	2	Not Done
702	36	Male	31	Severe	No	NRF, *S. aureus, P. aeruginosa*, Yeast	4	*S. aureus, P. aeruginosa*
703	25	Male	92	Mild	No	NRF, *S. aureus, P. aeruginosa*, Fungus, Yeast	5	*S. aureus, P. aeruginosa*
704	26	Female	97	Mild	No	*S. aureus, P. aeruginosa, S. maltophilia*, Fungus	4	*S. aureus, P. aeruginosa, S. maltophilia*
757	20	Male	86	Mild	No	NRF, *H. influenzae*	2	*H. influenzae*
761	27	Female	122	Mild	Ceftaroline, Tobramycin	No Growth	0	Not Done
782	25	Female	34	Severe	No	NRF, *P. aeruginosa, S. aureus*	3	*P. aeruginosa, S. aureus*
942	28	Male	67	Moderate	No	NRF, *P. aeruginosa*, Yeast	3	*P. aeruginosa*
943	42	Male	65	Moderate	No	*P. aeruginosa*, Yeast	2	*P. aeruginosa*
955	42	Female	99	Mild	No	NRF, *P. aeruginosa*, Yeast, Fungus	4	*P. aeruginosa*
970	25	Male	20	Severe	No	NRF, *B. vietnamiensis, P. aeruginosa*, Yeast	4	*B. vietnamiensis, P. aeruginosa*
971	41	Female	17	Severe	No	*P. aeruginosa*, Yeast, Fungus	3	*P. aeruginosa*
972	35	Male	24	Severe	No	*S. maltophilia*, Fungus	2	*S. maltophilia*
440	58	Male	59	Moderate	Doxycycline	NRF, *H. influenzae*	2	*H. influenzae*
490	61	Female	14	Very severe	Amoxicillin, Clarithromycin	NRF	1	Not Done
708	63	Female	28	Very severe	No	*K. pneumoniae*	1	*K. pneumoniae*
717	73	Male	34	Severe	No	*S. pneumoniae*	1	Not Done
720	66	Male	25	Very severe	Doxycycline	*H. influenzae*	1	*H. influenzae*
721	59	Female	61	Moderate	Doxycycline	NRF	1	Not Done
732	54	Female	17	Very severe	Doxycycline	No Growth	0	Not Done
759	62	Male	59	Moderate	Doxycycline	NRF, *H. influenzae, S. pneumoniae, N. meningitis*, Yeast	5	*H. influenzae, N. meningitis*
775	32	Female	43	Severe	Colomycin	NRF	1	Not Done
977	70	Female	14	Very severe	Amoxicillin	NRF	1	Not Done
978	60	Female	58	Moderate	No	NRF, *H. influenzae*	2	*H. influenzae*
987	80	Female	81	Mild	Doxycycline	NRF, *K. pneumoniae*	2	*K. pneumoniae*
991	69	Male	24	Very severe	No	No Growth	0	Not Done

[a]Samples are coloured by disease category: BR; blue, CF; yellow, COPD; grey.
[b]Coloured % fev_1 predicted values are based on predictive mean matching multiple imputation.

SI Table 2 Highest proportional rare taxa compared to negative controls

Rare taxa	Rare in n samples	Sequenced reads range	Control reads range
Acinetobacter	3	10-18	1
Actinomyces	18	4-199	0.1
Anaerococcus	3	4-16	1
Atopobium	24	3-131	None
Burkholderia	1	5	1
Capnocytophaga	21	4-111	None
Fusobacterium	25	2-209	0.1-1
Haemophilus	14	4-180	0.1-5
Leptotrichia	17	4-112	None
Prevotella	7	2-91	1-3
Pseudomonas	19	2-28	2-3
Staphylococcus	17	2-53	3
Stenotrophomonas	3	2-5	None
Veillonella	7	15-145	1-3

SI Table 3 16S rRNA gene sequence % relative abundance of contaminating genera

Sample	N reads	% Relative abundance
Mucolyse 1	40	2.7% *Alloprevotella*, 2.7% Cy, 2.7% *Effusibacillus,* 2.7% *Escherichia-Shigella*, 2.7% *Fusobacterium*, 5.4% *Granulicatella*, 2.7% *Hydrogenophilus*, 5.4% *Micrococcus*, 5.4% *Prevotella*, 8.1% *Pseudomonas*, 8.1% *Staphylococcus*, 40.5% *Streptococcus*, 8.1% Unclassified *Bacteroidia*, 2.7% Unclassified *Xanthomonadaceae*
Mucolyse 2	21	0.6% *Actinomyces*, 2.6% *Anoxybacillus*, 28.2% *Burkholderia*, 0.6% *Fusobacterium*, 3.2% *Gemella*, 0.6% *Granulicatella*, 0.6% *Haemophilus*, 0.6% *Massilia*, 0.6% *Meiothermus*, 6.4% *Prevotella*, 11.5% *Pseudomonas*, 5.8% *Rothia*, 28.2% *Streptococcus*, 7.1% *Veillonella*, 1.9% Unclassified *Enterobacterales*, Unclassified *Pasteurellaceae*
PCR water control	4	33.3% *Escherichia-Shigella*, 33.3% *Streptococcus*, 33.3% *Veillonella*
Zymo kit negative	31	3.2% *Acinetobacter*, 3.2% *Anaerococcus*, 3.2% *Anoxybacillus*, 3.2% *Burkholderia*, 3.2% *Escherichia-Shigella*, 16.1% *Haemophilus*, 6.5% *Hydrogenophilus*, 6.5% *Meiothermus*, 9.7% *Prevotella*, 9.7% *Staphylococcus*, 3.2% *Streptococcus*, 9.7% *Veillonella*, 9.7% Unclassified bacteria, 9.7% Unclassified *Bacteroidia*, 3.2% Unclassified *Chitinophagaceae*

SI Table 4 Total bacterial phyla identified in sputum and its median relative abundance

Phylum	Number of reads	Median proportion of reads (%)	IQR (%)
Firmicutes	**379847**	**55.1**	**30.4-77.6**
Proteobacteria	**195336**	**15.8**	**3.8-58.2**
Bacteroidetes	**51216**	**4.6**	**0.7-11.5**
Actinobacteria	**42557**	**3.7**	**2.-8.1**
Fusobacteria	**9864**	**0.7**	**0.04-2**
Campylobacteria	1056	0.1	0.00-0.2
Spirochaetes	692	0	0.00-0.02
Synergistetes	206	0	0
Patescibacteria	45	0	0
Unclassified bacteria	34	0	0.00-0.02
Deinococcus-Thermus	7	0	0
Bdellovibrionota	5	0	0
Cyanobacteria	3	0	0
Planctomycetes	2	0	0
Verrucomicrobia	1	0	0

SI Table 5 *in silico* RISA ITS profiles for respiratory bacterial species identification[a]

Species	Predicted ITS size	Species	Predicted ITS size
Actinomyces bowdenii	780	*Neisseria meningitidis*	946-949
Actinomyces denticolens	657	*Neisseria mucosa*	891
Actinomyces gaoshouyii	616	*Neisseria shayeganii*	875
Actinomyces howellii	685	*Neisseria sicca*	972
Actinomyces johnsonii	720	*Neisseria subflava*	893
Actinomyces marseillensis	638	*Neisseria zalophi*	945
Actinomyces massiliensis	718	*Noviherbaspirillum autotrophicum*	481, 781, 789
Actinomyces naeslundii	711	*Noviherbaspirillum denitrificans*	707, 796, 803
Actinomyces odontolyticus	609-612	*Noviherbaspirillum humi*	817
Actinomyces oris	680	*Noviherbaspirillum soli*	804
Actinomyces polynesiensis	615	*Porphyromonas asaccharolytica*	1021
Actinomyces provencensis	756	*Porphyromonas gingivalis*	1026-1037
Actinomyces ruminicola	626	*Prevotella buccae*	548, 913, 2943
Actinomyces slackii	830, 838, 896	*Prevotella denticola*	1408, 1994
Actinomyces tangfeifanii	686	*Prevotella intermedia*	634, 787
Actinomyces urogenitalis	666	*Prevotella loescheii*	1092, 1335
Actinomyces viscosus	679	*Prevotella melaninogenica*	651-652, 844
Alloprevotella rava	765, 899	*Prevotella oris*	984
Alloprevotella tannerae	237, 628, 835, 1901	*Prevotella salivae*	1092
Bergeyella cardium	812	*Prevotella shahii*	1092, 1335
Bergeyella zoohelcum	872	*Pseudomonas aeruginosa*	753
Burkholderia ambifaria	663, 864-885	*Pseudomonas fluorescens*	790-840
Burkholderia cenocepacia	672-677, 811-824	*Pseudomonas fulva*	796
Burkholderia multivorans	829-830, 883, 937	*Pseudomonas mendocina*	795
Burkholderia vietnamiensis	665, 812, 821, 872	*Pseudomonas monteilii*	599
Capnocytophaga cynodegmi	928	*Pseudomonas stutzeri*	786
Capnocytophaga leadbetteri	742,744, 831, 849	*Rothia aeria*	601
Capnocytophaga ochracea	876	*Rothia dentocariosa*	624
Capnocytophaga sputigena	710, 717, 855, 1135	*Rothia halotolerans*	685
Capnocytophaga stomatis	843	*Rothia koreensis*	629
Corynebacterium imitans	789, 793-797, 1285	*Rothia kristinae*	706
Corynebacterium striatum	774, 1439, 1886	*Rothia mucilaginosa*	588
Escherichia coli	627-637, 713-733	*Rothia nasimurium*	636
Fusobacterium naviforme	907, 1093	*Scardovia inopinata*	663
Gemella haemolysans	490, 692	*Serratia ficaria*	601, 735, 748, 837, 1871
Gemella massiliensis	494, 691	*Serratia grimesii*	579, 736, 750
Gemella morbillorum	491, 699	*Serratia liquefaciens*	747-752, 764
Gemella sanguinis	486, 687	*Serratia marcescens*	726, 748-749, 824-826, 970
Granulicatella adaciens	503	*Serratia proteamaculans*	736, 750, 854
Haemophilus influenzae	757-759, 995-1015	*Serratia ureilytica*	587, 685, 753, 1710
Klebsiella aerogenes	488, 625-633, 782-790	*Staphylococcus aureus*	648, 743-757, 830
Klebsiella grimontii	634	*Stenotrophomonas maltophilia*	777-782, 826-829
Klebsiella pneumoniae	480, 621-633, 704-719,780-781	*Stomatobaculum longum*	698-699, 725
Klebsiella quasipneumoniae	488, 625-633, 717, 788, 951	*Streptococcus agalactiae*	563
Klebsiella oxytoca	488, 633-636, 790-792	*Streptococcus anginosus*	669
Klebsiella variicola	476, 621-633, 704, 707, 763-764, 845	*Streptococcus constellatus*	569-600
Lactobacillus amylovorus	484, 732-733	*Streptococcus dysgalactiae*	569
Lactobacillus brevis	497, 699	*Streptococcus gallolyticus*	556
Lactobacillus crispatus	486, 736	*Streptococcus gordonii*	530
Lactobacillus delbrueckii	500-502, 739-741	*Streptococcus intermedius*	672-673
Lactobacillus fermentum	479, 674-735	*Streptococcus mitis*	530
Lactobacillus gasseri	503, 729	*Streptococcus mutans*	670-671
Lactobacillus helveticus	484, 733-752	*Streptococcus oralis*	529
Lactobacillus johnsonii	501-503, 732-733	*Streptococcus pneumoniae*	529
Lactobacillus rhamnosus	504	*Streptococcus suis*	574-710
Lautropia dentalis	839	*Treponema brennaborense*	674, 798
Lautropia mirabilis	611, 837	*Treponema denticola*	670, 687
Leptotrichia buccalis	393, 562	*Treponema pallidum*	578, 588
Leptotrichia hofstadii	396, 562-563	*Treponema paraluiscuniculi*	578, 588
Leptotrichia hongkongensis	394, 558	*Treponema primitia*	685, 752
Leptotrichia shahii	391, 558	*Treponema succinifaciens*	511, 526
Leptotrichia trevisanii	399, 564	*Veillonella atypica*	563, 729, 875, 2657
Leptotrichia wadei	396, 565	*Veillonella dispar*	617, 725, 931, 1071
Neisseria bacilliformis	861	*Veillonella magna*	1173
Neisseria cineraea	885	*Veillonella nakazawae*	652, 723, 925, 950
Neisseria flavescens	893	*Veillonella parvula*	883, 926, 1075
Neisseria lactamica	910	*Veillonella ratti*	1658

[a]Where available, these ITS references cover all abundant genera identified in all sputum samples.

SI Table 6 RISA PCR species-level identification of bacterial species according to abundant 16S rRNA gene sequence data

ID	ITS size (bp)[a]	16S-based abundance data[b]	RISA and 16S-based sequence matching
471	477	41.2% Sn	*S. maltophilia* (partial, 726 bp)
	504	22.1% Ne	
	634	21.7% Str	*S. anginosus/S. intermedius/S. mutans/S. suis* (671 bp)
	671	3.5% Ge	*G. sanguinis* (477 bp, 690 bp)
	690	2.3% Pst	
	726	2.2% Ve	*V. dispar/V. nakazawae* (partial, 634 bp, 726 bp), *V. atypica* (partial, 726 bp)
	1148	1.7% Gr	*G. adaciens* (504 bp)
		1.5% Pv	*P. intermedia/P. melaninogenica* (partial, 634 bp), *P. denticola* (partial, 1448 bp)
483	**556**	96.6% Str	*S. agalactiae/S. dysgalactiae/S. gallolyticus* (556 bp)
		1.2% Ve	*V. atypica* (partial, 556 bp)
699	492	46.2% Ge	*G. haemolysans/G. massiliensis/G. morbillorum* (492 bp/499 bp, 686 bp)
	499	32.5% Str	*S. gordonii/S. mitis/S. oralis/S. pneumoniae* (533 bp), *S. anginosus/S. intermedius/S. mutans* (686 bp), *S. suis* (584 bp/686 bp/709 bp)
	533	5.1% Cp	*C. cynodegmi* (928 bp), *C. ochracea/C. stomatis* (849 bp)
	584	4.4% Pv	*P. oris* (988 bp), *P. buccae* (partial, 584 bp, 928 bp), *P. melaninogenica* (partial, 686 bp, 849 bp)
	686	4% Gr	*G. adaciens* (499 bp)
	709	2.6% Le	*L. buccalis/L. hofstadii/ L. trevisanii/L. wadei* (partial, 584 bp)
	849	1.5% Bg	*B. cardium/B. zoohelcum* (849 bp)
	928	1.2% Ve	*V. parvula* (partial, 849 bp, 928 bp), *V. nakazawae* (686 bp, 709 bp, 928 bp, 988 bp), *V. atypica* (partial, 584 bp, 709 bp, 849 bp), *V. dispar* (partial, 584 bp, 709 bp, 928 bp)
	988		
707	531	69.7% Ha	*H. influenzae* (843 bp, 1133 bp)
	843	13.8% Str	*S. gordonii/S. mitis/S. oralis/S. pneumoniae* (531 bp)
	1133	4.3% Pv	*P. melaninogenica* (partial, 843 bp), P. buccae (partial, 531 bp)
		4% Ne	*N. bacilliformis/N. cineraea/N. shayeganii* (843 bp)
		1.8% Ro	
		1.4% Le	*L. buccalis/L. hongkongensis/L. shahii* (partial, 531 bp)
718	537	28.8% Str	S. gordonii/S. mitis/S. oralis/S. pneumoniae (537 bp), *S. aniginosus/S. intermedius/S. mutans/S. suis* (676 bp)
	566	14.8% Pv	*P. intermedia* (603 bp, 731 bp), *P. buccae* (partial, 566 bp, 934 bp), *P. melaninogenica* (partial, 676 bp)
	603	14% Pa	*P. aeruginosa* (731 bp)
	676	10.9% Ve	*V. atypica* (partial, 566 bp, 934 bp), *V. dispar* (partial, 603 bp, 731 bp, 934 bp), *V. nakazawae* (partial, 676 bp, 731 bp, 934 bp), *V. parvula* (partial, 934 bp)
	731	8.2% Ro	*R. aeria* (603 bp), *R. halotolerans/R. nasimurium* (676 bp), *R. mucilaginosa* (566 bp)
	934	4.2% At	*A. howellii/A. tangfeifanii/A. viscosus* (676 bp), *A. odontolyticus* (603 bp), *A. johnsonii/A. naeslundii/A. provencensis* (731 bp)
	2177	4% Le	*L. buccalis/L. hofstadii/L. trevisanii/L. wadei* (partial, 566 bp)
		2.7% Ge	*G. sanguinis* (partial, 676 bp)
		2.7% Ne	*N. meningitidis/N. sicca/N. zalophi* (934 bp)
		2.3% Gr	
		2.1% St	*S. longum* (676 bp, 934 bp)
		1% Pst	
758	517	51.1% Str	*S. gallolyticus/S. suis* (570 bp), *S. gordonii/S. mitis/S. oralis/S. pneumoniae/S. gallolyticus* (544 bp)
	544	21.5% Ve	*V. atypica* (partial, 570 bp)
	570	11.6% La	*L. rhamnosus* (517 bp)

	2236	8.1% Ro	*R. mucilaginosa* (570 bp)
		2.1% At	*A. odontolyticus/A. polynesiensis* (570 bp)
		1.7% Gr	*G. adaciens* (517 bp)
		1.3% Le	*L. buccalis/L. hofstadii/L. trevisanii/L. wadei* (570 bp)
788	**532**	37.6% Ha	*H. influenzae* (851 bp, 1012 bp)
	564	33% Str	*S. gordonii/S. mitis/S. oralis/S. pneumoniae* (532 bp), *S. gallolyticus/S. suis* (564 bp)
	851	6.2% Pv	*P. salivae* (1012 bp), *P. buccae* (partial, 564 bp)
	1012	5.2% Ve	*V. magna* (1173 bp)
	1161	4.9% Gr	
		4.4% Ge	
		1.5% At	*A. slackii* (partial, 851 bp)
		1.2% Ro	*R. aeria* (564 bp)
790	**529**	86.2% Pa	*P. aeruginosa* (736 bp)
	603	6.8% Str	*S. gordonii/S. mitis/S. oralis/S. pneumoniae* (529 bp), *S. anginosus/S. intermedius/S. mutans/S. suis* (677 bp)
	677	3.4% Ve	*V. parvula* (partial, 924 bp), *V. nakazawae* (partial, 677 bp, 736 bp, 924 bp), *V. dispar* (partial, 603 bp, 736 bp, 924 bp)
	736	2.2% Ro	*R. aeria* (603 bp), *R. halotolerans/R. nasimurium* (677 bp), *R. kristinae* (736 bp)
	924		
	1725		
913	538	34.7% Str	*S. gallolyticus/S. suis* (570 bp), *S. gordonii/S. mitis/S. oralis/S. pneumoniae* (538 bp), *S. anginosus/S. intermedius/S. mutans/S. suis* (686 bp)
	570	30.8% Sa	*S. aureus* (partial, 646 bp)
	646	10% Ve	*V. parvula* (partial, 989 bp, 1056 bp), *V. dispar* (partial, 646 bp, 989 bp, 1056 bp), *V. nakazawae* (partial, 646 bp, 989 bp), *V. atypica* (partial, 570 bp)
	686	7.7% Pv	*P. oris* (989 bp), *P. buccae* (570 bp), *P. intermedia/P. melaninogenica* (partial, 646 bp)
	989	6.3% At	*A. howellii/A. oris/A. tangfeifanii/A. viscosus* (686 bp), *A. gaoshouyii/A. marseillensis/A. polynesiensis/A. ruminicola* (646 bp)
	1056	1.3% Ge	*G. sanguinis* (partial, 686 bp)
	2059	1.3% Ne	*N. meningitidis/N. sicca/N. zalophi* (989 bp)
		1.1% Ro	*R. dentocariosa/R. koreensis* (646 bp), *R. halotolerans* (685 bp), *R. mucilaginosa* (570 bp)
944	498	61.6% Str	*S. gallolyticus/S. suis* (573 bp), *S. gordonii/S. mitis/S. oralis/S. pneumoniae* (544 bp), *S. anginosus/S. intermedius/S. mutans/S. suis* (659 bp or 693 bp), *S. constellatus* (612 bp)
	508	8.6% Gr	*G. adaciens* (508 bp)
	544	7.5% Ge	*G. haemolysans/G. massiliensis/G. morbillorum* (498 bp, 693 bp)
	573	5.9% Ro	*R. dentocariosa/R. koreensis/R. nasimurium* (659 bp), *R. halotolerans/R. kristinae* (693 bp), *R. aeria* (612 bp), *R. mucilaginosa* (573 bp),
	612	5.1% Ve	*V. parvula* (partial, 950 bp, 1056 bp), *V. dispar* (partial, 612 bp, 950 bp, 1056 bp), *V. nakazawae* (partial, 659 bp, 950 bp), *V. atypica* (partial, 573 bp)
	659	4% Pv	*P. oris* (950 bp), *P. intermedia/P. melaninogenica* (partial, 659 bp)
	693	2.7% Pst	
	950	2% At	*A. gaoshouyii/A. polynesiensis/A. odontolyticus* (612 bp), *A. denticolens/A. marseillensis/A. ruminicola/A. urogenitalis* (659 bp), *A. howellii/A. oris/A. tangfeifanii/A. viscosus* (693 bp)
	1056	1.3% Ne	*N. lactamica/N. meningitidis/N. sicca/N. zalophi* (950 bp)
	1923		
	2256		
945	571	77.9% Ent	*S. liquefaciens/S. marcescens/S. proteamaculans* (partial, 715 bp), *S. ficaria/S. grimesii/S. ureilytica* (partial, 571 bp, 715 bp)
	715	12.6% Pa	*P. aeruginosa* (715 bp)
		6.9% Eb	*K. pneumoniae/K. quasipneumoniae/K. variicola* (partial, 571 bp, 715 bp), *K. aerogenes/K. grimontii/K. oxytoca* (partial, 571 bp)
		1.9% Str	*S. constellatus/S. gallolyticus/S. suis* (571 bp)
995	**542**	35.5% Str	*S. gordonii/S. mitis/S. oralis/S. pneumoniae (*542 bp), *S. anginosus/S. intermedius/S. mutans/S. suis* (684 bp), *S. gallolyticus/S. suis* (568 bp), *S. constellatus* (609 bp)
	568	14.7% Ve	*V. parvula* (partial, 911 bp, 1022 bp), *V. dispar* (partial, 609 bp, 911 bp, 1022 bp), *V. nakazawae* (partial, 654 bp, 911 bp), *V. atypica* (partial, 586 bp)
	609	10.6% Al	*A. rava* (partial, 911 bp), *A. tannerae* (partial, 654 bp, 2094 bp)

	654	9.9% Pv	*P. buccae* (542 bp, 911 bp), *P. salivae* (1022 bp), *P. denticola* (1592 bp, 2094 bp), *P. intermedia/P. melaninogenica* (partial, 654 bp)
	684	6.7% Fu	*F. naviforme* (partial, 907 bp, 2094 bp)
	911	3.7% Ro	*R. halotolerans/R. kristinae* (684 bp), *R. nasimuium* (654 bp), *R. aeria/R. dentocariosa/R. koreensis* (609 bp)
	1022	3.2% At	*A. gaoshouyii/A. odontolyticus/A. polynesiensis/A. ruminicola* (609 bp), *A. howellii/A. oris/A. tangfeifanii/A. viscosus* (684 bp), *A. denticolens/A. marseillensis/A. urogenitalis* (654 bp)
	1592	2.1% Ge	*G. sanguinis* (partial, 684 bp)
	2094	1.9% Tr	*T. brennaborense/T. denticola/T. primitia* (partial, 684 bp), *T. palladum/T. paraluiscuniculi/T. succinifaciens* (partial, 568 bp)
	2237	1.8% Gr	
996	500	48.8% Str	*S. gordonii/S. mitis/S. oralis/S. pneumoniae* (538 bp)
	538	15.5% Gr	*G. adaciens* (500 bp)
		8.5% Ve	*V. atypica* (partial, 538 bp)
		6.9% Ro	
		5.5% At	
		3.3% Pv	P. buccae (partial, 538 bp)
		2.2% Ge	
		1.8% Pst	
		1.6% St	
		1.2% Ne	
702	538	56.7% Sa	*S. aureus* (partial, 690 bp, 730 bp, 917 bp)
	690	21.2% Pa	*P. aeruginosa* (730 bp)
	730	8.3% Str	*S. gordonii/S. mitis/S. oralis/S. pneumoniae* (538 bp)
	917	2.1% Pv	*P. buccae* (538 bp, 917 bp), *P. loescheii/P. salivae/P. shahii* (partial, 1048), *P. melaninogenica* (partial, 690 bp)
	1048	1.7% Fu	*F. naviforme* (917 bp, 1048 bp)
		1.7% Al	*A. rava* (917 bp)
		1.5% Ve	*V. parvula* (partial, 917 bp, 1048 bp), *V. nakazawae* (partial, 690 bp, 730 bp, 917 bp), *V. dispar* (730 bp, 917 bp, 1048 bp), *V. atypica* (partial, 538 bp, 730 bp, 917 bp)
		1.3% Ro	*R. halotolerans/R. kristinae* (690 bp)
		1.2% Pst	
		1.1% Gr	
703	**539**	46% Sa	*S. aureus* (partial, 684 bp, 748 bp, 937 bp)
	569	16.2% Str	*S. gordonii/S. mitis/S. oralis/S. pneumoniae* (539 bp), *S. anginosus/S. intermedius/S. mutans/S. suis* (684 bp)
	601	8.9% Pa	*P. aeruginosa* (748 bp)
	684	5.5% Ro	*R. halotolerans/R. kristinae* (684 bp)
	748	4% Pv	*P. loescheii/P. salivae/P. shahii* (partial 1057 bp), *P. buccae* (partial, 539 bp, 937 bp), *P. melaninogenica* (partial, 684 bp)
	937	2.9% Pst	
	1057	2.5% Ve	*V. parvula* (partial, 937 bp, 1057 bp), *V. dispar* (partial, 748 bp, 937 bp, 1057 bp), *V. atypica* (partial, 569 bp, 748 bp, 937 bp), *V. nakazawae* (partial, 684 bp, 748 bp, 937 bp)
		2% Fu	*F. naviforme* (937 bp, 1057 bp)
		1.9% Po	*P. asaccharolytica/P. gingivalis* (1057 bp)
		1.7% At	*A. howellii/A. oris/A. tangfeifanii/A. viscosus* (684 bp), *A. johnsonii/A. naeslundii* (748 bp)
		1.7% Ge	*G. haemolysans/G. massiliensis/G. morbillorum* (partial, 684 bp)
		1.7% Pva	
		1% Ne	*N. lactamica/N. meningitidis/N. zalophi* (937 bp)
704	536	40.9% Sa	*S. aureus* (partial, 642 bp, 708 bp)
	566	37.9% Pa	*P. aeruginosa* (708 bp)
	624	8.5% Str	*S. gordonii/S. mitis/S. oralis/S. pneumoniae* (536 bp), *S.anginosus/S. intermedius/S. mutans/S. suis* (656 bp)
	642	2.3% Sn	*S. maltophilia* (partial, 708 bp)

	656	2% Pst	
	708	1.6% Sne	
	1066	1.5% Ge	
		1.1% Pv	*P. intermedia/P. melaninogenica* (partial, 656 bp), *P. loescheii/P. salivae/P. shahii* (partial, 1066 bp)
		1% Gr	
757	494	30.6% Ha	*H. influenzae* (888 bp, 1190 bp)
	536	20.5% Str	*S. gordonii/S. mitis/S. oralis/S. pneumoniae* (536 bp), *S. anginosus/S. intermedius/S. mutans/S. suis* (684 bp)
	553	7.3% Pst	
	684	6.8% Ve	*V. magna* (1190 bp), *V. parvula* (partial, 888 bp, 1190 bp), *V. nakazawae* (partial, 684 bp, 888 bp), *V. atypica* (553 bp, 684 bp)
	888	6.2% Pv	*P. melaninogenica* (684 bp, 888 bp), *P. loescheii/P. salivae/P. shahii* (partial, 1190 bp), *P. buccae* (partial, 536 bp, 888 bp)
	1190	6.1% Ro	*R. halotolerans/R. kristinae* (684 bp)
		4.2% Ge	*G. haemolysans/G. massiliensis/G. morbillorum* (494 bp, 684 bp)
		3.4% At	*A. howellii/A. oris/A. tangfeifanii/A. viscosus* (684 bp)
		3.3% Ne	*N. cineraea/N. flavescens/N. mucosa/N. subflava* (888 bp)
		2.9% Gr	
		2.4% Le	*L. buccalis/L. hongkongensis/L. shahii* (553 bp)
		1.9% Fu	
761	498	48.7% Str	*S. gordonii/S. mitis/S. oralis/S. pneumoniae* (532 bp)
	532	17.4% Ro	*R. aeria* (607 bp)
	607	14.7% Gr	*G. adiacens* (498 bp)
	844	6.2% Lau	*L. dentalis* (844 bp)
		3.8% Sc	
		2.4% Ve	*V. parvula* (partial, 883 bp), *V. atypica* (partial, 844 bp), *V. dispar* (partial, 607 bp)
		1.2% Nv	N. humi/N. soli (844 bp), N. autotrophicum/N. denitrificans (partial, 844 bp)
782	498	54.3% Pa	*P. aeruginosa* (719 bp)
	536	23.1% Str	*S. gordonii/S. mitis/S. oralis/S. pneumoniae* (536 bp), *S. gallolyticus/S. suis* (564 bp)
	564	4.9% Ge	*G. haemolysans/G. morbillorum/G. massiliensis* (partial, 498 bp)
	719	3.7% Ve	*V. dispar/V. nakazawae* (partial, 719 bp), *V. atypica* (partial, 564 bp)
		3.3% Ha	
		2.5% Pv	*P. buccae* (partial, 564 bp)
		1.5% Ro	*R. kristinae* (719 bp), *R. mucilaginosa* (564 bp)
		1% Pst	
942	**558**	62.7% Ro	*R. aeria* (601 bp)
	601	29.3% Str	*S. gallolyticus* (558 bp), *S. constellatus* (601 bp)
	1265	3.5% Pa	
		2.4% Ve	*V. magna* (1173 bp)
943	544	50% Pa	*P. aeruginosa* (719 bp)
	575	28.7% Pv	*P. intermedia/P. melaninogenica* (663 bp, 819 p)
	663	18.2% Str	*S. gordonii/S. mitis/S. oralis/S. pneumoniae* (544 bp), *S. gallolyticus* (575 bp),
	719	1.6% Ve	*V. nakazawae* (partial, 663 bp, 719 bp, *V. atypica* (partial, 575 bp, 719 bp), *V. dispar* (partial, 719 bp)
	819		
955	496	46.7% Pa	*P. aeruginosa* (712 bp)
	541	13.2% Pv	*P. oris* (984 bp), *P. intermedia/P. melaninogenica* (partial, 680 bp), *P. buccae* (partial, 541 bp, 945 bp)
	680	7.7% Str	*S. gordonii/S. mitis/S. oralis/S. pneumoniae* (541 bp), *S.anginosus/S. intermedius/S. mutans/S. suis* (680 bp)
	712	5.6% Ve	*V. parvula*(partial, 945 bp), *V. nakazawae* (680 bp, 712 bp, 945 bp, 984 bp), *V. dispar* (partial, 712 bp, 945 bp), *V. atypica* (partial, 541 bp, 712 bp)
	945	5% Pst	
	984	3.8% Ge	*G. haemolysans/G. massiliensis/G. morbillorum* (496 bp, 680 bp)

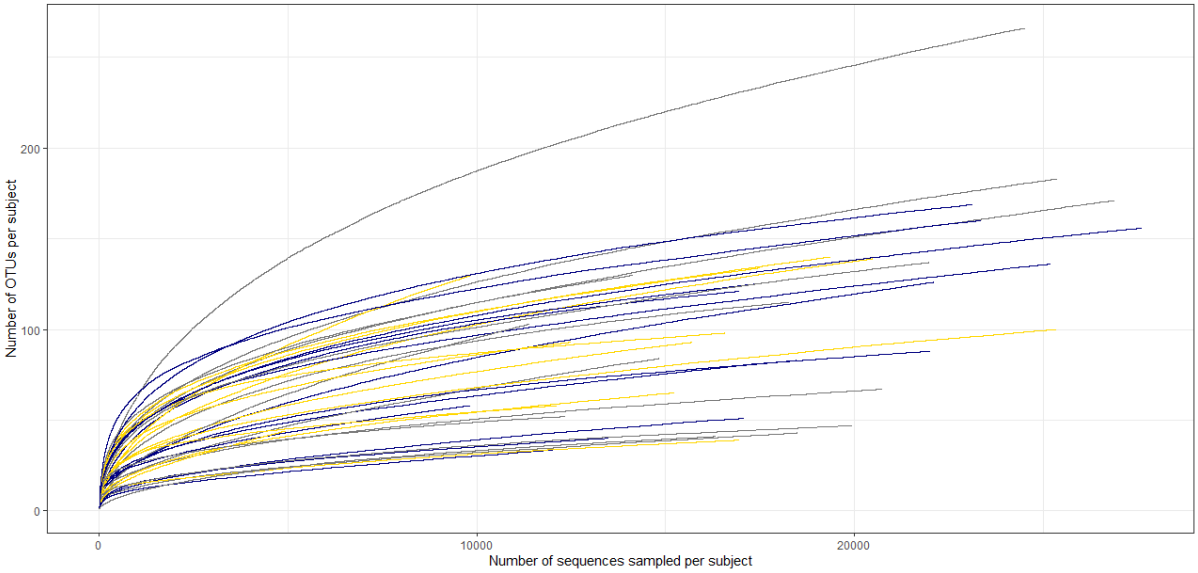
		3.2% Fu	*F. naviforme* (partial, 945 bp)
		2.7% Gr	*G. adaciens* (496 bp)
		1.9% Ne	*N. lactamica/N. meningitidis/N. sicca/N. zalophi* (945 bp)
		1.8% Le	*L.buccalis/L.hofstadii/L. hongkongensis/L. shahii* (partial, 541 bp)
		1.4% Ro	*R. halotolerans/R. nasimurium* (685 bp), *R. mucilaginosa* (541 bp), *R. kristinae* (712 bp)
		1.3% Po	
970	497	41.8% Bu	*B. vietnamiensis* (partial, 653 bp, 767 bp), *B. cenocepacia* (653 bp, 767 bp), *B. multivorans* (partial 653 bp, 767 bp
	532	29.2% Str	*S. gordonii/S. mitis/S. oralis/S. pneumoniae* (532 bp), *S. anginosus/S. intermedius/S. mutans/S. suis* (653 bp)
	552	9.7% Ve	*V. dispar* (653 bp, 767 bp, 985 bp, 1097 bp), *V. parvula* (partial, 985 bp, 1097 bp), *V. nakazawae* (partial, 653 bp, 767 bp, 985 bp), *V. atypica* (552 bp, 767 bp)
	653	6.5% Ge	*G. haemolysans/G. massiliensis/G. morbillorum* (497 bp, 653 bp)
	767	5.7% Pv	*P. oris* (985 bp), *P. intermedia* (653 bp, 767 bp), *P. buccae* (552 bp), *P. loescheii/P. salivae/P. shahii* (partial, 1097 bp), *P. melaninogenica* (partial, 653 bp)
	985	4.6% Ro	*R. dentocariosa/R. koreensis/R. nasimurium* (653 bp), *R mucilaginosa* (552 bp)
	1097	1.1% Gr	*G. adaciens* (497 bp)
		1.1% At	*A. denticolens/A. marseillensis/A. urogenitalis* (653 bp), *A. bowdenii/A. johnsonii/A. provencensis* (767 bp)
971	538	48.5% Str	*S. gordonii/S. mitis/S. oralis/S. pneumoniae* (538 bp), *S. constellatus* (600 bp)
	597	17% Pv	*P. buccae* (partial, 538 bp)
	717	6.8% Ro	*R. mucilaginosa* (597 bp)
	1698	5.6% Gr	
		5.2% Ve	*V. ratti* (1698 bp), *V. atypica/V. dispar* (partial, 597 bp, 717 bp)
		4.5% Ne	
		4% Ha	
		3.4% Pst	
		1.9% Fu	
972	498	64.1% Sn	*S. maltophilia* (partial, 740 bp),
	740	8.7% Ge	*G. haemolysans/G. massiliensis/G. morbillorum* (partial, 498 bp)
		6.5% Str	
		6.5% Pv	
		4.4% At	*A. bowdenii/A. johnsonii/A. massiliensis/A. provencensis* (740 bp)
		3.4% Cp	*C. leadbetteri/C. sputigena* (740 bp)
		1.6% Pva	
		1.1% Ro	*R. kristinae* (740 bp)
440	500	62.1% Str	*S. gordonii/S. mitis/S. oralis/S. pneumoniae* (532 bp), *S. anginosus/S. intermedius/S. mutans/S .suis* (683bp)
	532	12.4% Pv	*P. melaninogenica* (683 bp, 888 bp), *P. buccae* (partial, 532 bp, 888 bp), *P. intermedia* (partial, 683 bp)
	683	7.4% Ve	*V. parvula* (partial, 888 bp), *V. atypica* (partial, 532 bp, 683 bp, 888 bp), *V. dispar* (partial, 888 bp)
	888	4.4% Ha	*H. influenzae* (888 bp)
	1760	4% Gr	*G. adaciens* (500 bp)
		2.5% At	*A. howellii/A. oris/A. tangfeifanii/A. viscosus* (888 bp)
		2.3% Ge	*G. sanguinis* (partial, 683 bp)
		1.3% Ro	*R. halotolerans/R. kristinae* (683 bp)
490	507	69.6% Pa	*P. monteilii* (623 bp), *P. fluorescens/P. fulva/P. mendocina/P. stutzeri* (840 bp)
	538	17.6% Str	*S. gordonii/S. mitis/S. oralis/S. pneumoniae* (538 bp)
	574	3.8% Pv	*P. melaninogenica* (623 bp, 840 bp), *P. denticola* (partial, 1455 bp), *P. intermedia* (partial, 623 bp)
	623	2.6% Ve	*V. parvula* (partial, 840 bp), *V. dispar/V. nakazawae* (partial, 623 bp), *V. atypica* (partial, 774 bp)
	840	1.9% La	*L. rhamnosus* (507 bp), *L. brevis/L. delbrueckii/L. gasseri/L. johnsonii/L. rhamnosus* (partial, 507 bp)
	1455	1.7% Ro	*R. mucilaginosa* (574 bp), *R. dentocariosa/R. koreennsis/R. nasimurium* (623 bp)

	1784	1.4% Gr	*G. adaciens* (507 bp)
708	270 495 539 583 605 634 **762**	62.8% Eb 18.7% Ge 8.2% Str 6.3% Ro 3.6% La	*K. pneumoniae* (495 bp, 634 bp, 762 bp) *G. haemolysans/G. massiliensis/G. morbillorum* (partial, 495 bp) *S. gordonii/S. mitis/S. oralis/S. pneumoniae* (539 bp), *S. constellatus* (605 bp) *R. mucilaginosa* (583 bp), *R. dentocariosa/R. koreensis/R. nasimurium* (634 bp), *R. aeria* (605 bp), *L. amylovorans/L. crispatus/L. delbrueckii/L. helveticus* (495, 762 bp)
717	**524**	91.4% Str 1.8% Gr 1.5% Ro 1% Ve	*S. pneumoniae* (524 bp) *G. adaciens* (524 bp) *V. atypica* (partial, 524 bp)
720	**538** 835 1155	89.2% Ha 4.6% Str 1.6% Pv 1% Ge	*H. influenzae* (835 bp, 1155 bp) *S. gordonii/S. mitis/S. oralis/S. pneumoniae* (538 bp) *P. melaninogenica* (partial, 835 bp), *P. buccae* (partial, 538 bp)
721	**541** 574 649	63.7% Str 10.4% Pv 9% Ve 6.3% At 4.1% Ro 1.6% Gr 1.2% Ge	*S. gordonii/S. mitis/S. oralis/S. pneumoniae* (541 bp), *S. gallolyticus/S. suis* (571 bp), *S. anginosus/S. intermedius/S. mutans* (649 bp) *P. melaninogenica/P. intermedia* (parial, 649 bp), *P. buccae* (partial, 548 bp) *V. nakazawae* (partial, 649 bp), *V. atypica/V. dispar* (574 bp) *A. denticolens/A. johnsonii/A. ruminicola/A. marseillensis* (649 bp)
732	**536** 573 754 1490 1952	93.8% Str 1.6% Cy 1% Ve 1% La	*S. gordonii/S. mitis/S. oralis/S. pneumoniae* (536 bp), *S. gallolyticus/S. suis* (573 bp) *C. striatum* (754 bp, 1490 bp, 1952 bp) *V. atypica/V. dispar* (patial, 573 bp, 754 bp), *V. nakazawae* (partial, 754 bp), *L. crispatus/L. delbrueckii/L. helveticus/L. johnsonii* (partial, 754 bp)
759	284 501 **538** 570 930 1033	65.5% Str 9.5% Ge 6% Ne 4.9% Ha 2.4% Pv 1.5% Ro 1.5% Ve 1.3% Gr 1.1% Lea	*S. gordonii/S. mitis/S. oralis/S. pneumoniae* (538 bp), *S. suis/S. gallolyticus* (570 bp) *G. haemolysans/G. massiliensis/G. morbillorum* (partial, 501 bp) *N. lactamica/N. meningitidis/N. zalophi* (930bp) *H. influenzae* (partial, 1033 bp), *P. buccae* (538 bp, 930 bp), *P. loescheii/P. salivae/P. shahii* (partial, 1033 bp) *R. aeria/R. mucilaginosa* (570 bp) *V. parvula* (partial, 930 bp, 1033 bp), V. *dispar* (partial, 930 bp) *G. adaciens* (501 bp)
775	**538** **567** 601 674 2190	39.4% Str 19.4% Ve 13.4% Pv 6.3% Ro 5.7% At 3.2% Gr 1.9% Fu 1.9% La 1.4% Sc 1% Pva	*S. anginosus/S. intermedius/S. mutans/S. suis* (674 bp), *S. gallolyticus/S. suis* (567 bp), *S. gordonii/S. mitis/S. oralis/S. pneumoniae* (538bp), *S. constellatus* (600 bp) *V. atypica* (partial, 567 bp), *V. dispar* (partial, 601 bp), *V. nakazawae* (partial, 674 bp) *P. intermedia/P. melaninogenica* (partial, 674 bp), *P. buccae* (partial, 567 bp) *R. aeria* (603 bp), *R. halotolerans/R. kristinae/R. nasimurium* (674 bp), *A. gaoshouyii/A. massiliensis/A. odontolyticus/A. polynesiensis* (601 bp), *A. denticolens/A. oris/A. tangfeifanii/A. viscosus* (674 bp) *L. brevis/L. fermentum* (partial, 674 bp) *S. inopinata* (674 bp)

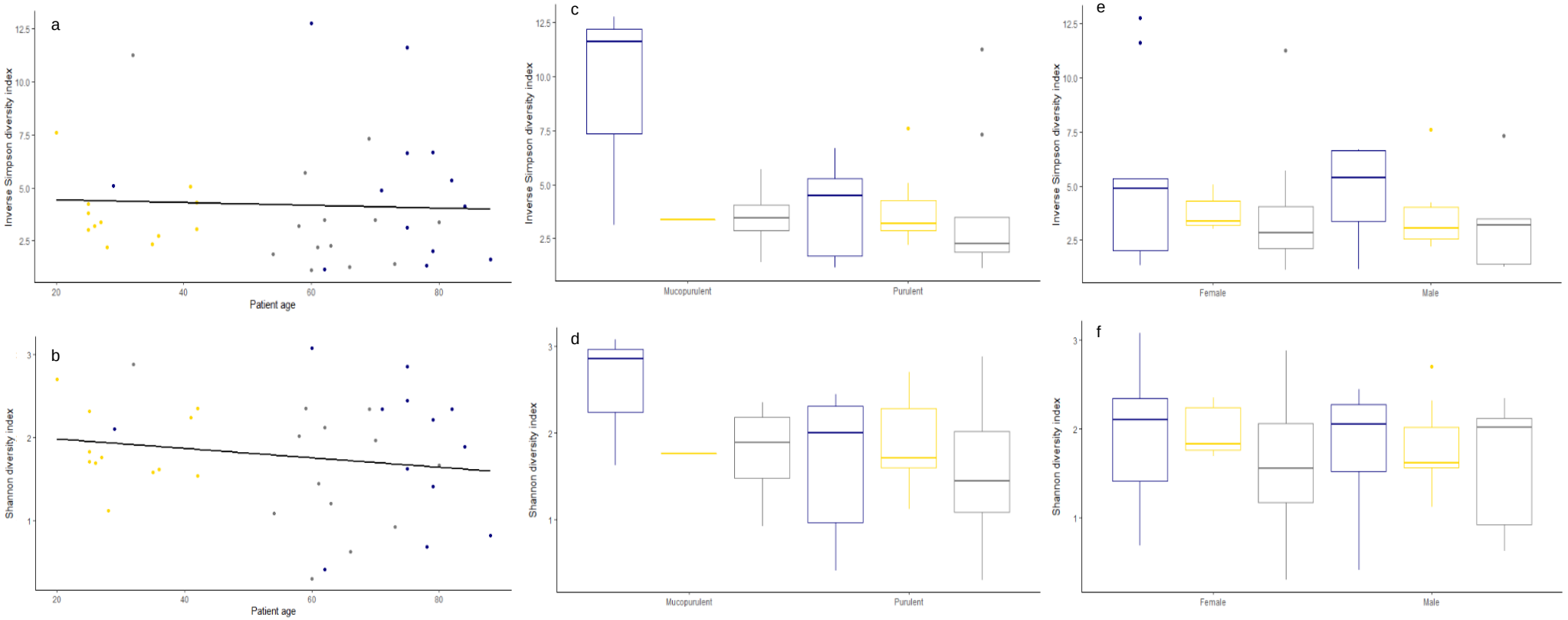
977	505 **545** 683 725 843 879	50.3% Gr 17.6% Str 14.3% Ve 5.7% Pv 2.2% Le 2.1% Pst 1.6% Ha 1.5% At	*G. adaciens* (505 bp) *S. anginosus/S. intermedius/S. mutans/S. suis* (683 bp), *S. gordonii/S. mitis/S. oralis/S. pneumoniae* (545 bp) *V. parvula* (partial, 879 bp), *V. nakazawae* (partial, 683 bp, 725 bp), *V. atypica* (partial, 545 bp, 725 bp, 879 bp), *V. dispar* (partial, 725 bp) *P. melaninogenica/P. intermedia* (683 bp, 843 bp), *P. buccae* (partial, 545 bp, 879 bp) *L. buccalis/L. hongkongensis/L. shahii* (partial, 545 bp) *H. influenzae* (partial, 843 bp) *A. howellii/A. oris/A. tangfeifanii/A. viscosus* (683 bp), *A. johnsonii/A. naeslundii/A. provencensis* (725 bp)
978	533 551 **831** 1161	93.9% Ha 5.3% Str	*S. gordonii/S. mitis/S. oralis/S. pneumoniae* (533 bp), *S. gallolyticus* (551 bp) *H. influenzae* (831 bp, 1161 bp)
987	493 **531** 679 764 914	66.7% Str 17.4% Ro 8.5% Ge 3.4% Pst 1.4% Ent 1.1% Ve	*S. gordonii/S. mitis/S. oralis/S. pneumoniae* (531 bp), *S. anginosus/S.intermedius/S. mutans/S. suis* (679 bp) *R. halotolerans/R. kristinae/R. nasimurium* (679 bp) *G. haemolysans/G. massiliensis/G. morbillorum* (partial, 493 bp) *K. aerogenes/K. penumoniae* (partial, 493 bp, 679 bp, 764 bp, 914 bp) *V. nakazawae* (partial, 679 bp, 764 bp, 914 bp), *V. parvula* (partial, 914 bp), *V. atypica* (partial, 531 bp, 764 bp, 914 bp), *V. dispar* (partial, 764 bp, 914 bp)
991	500 **538** 574 613 681	29.3% Str 17.9% Ne 14.1% Pv 13.4% Ve 5.8% Sn 4.6% At 4.3% Pst 3% Ge 1.5% Lch 1.4% St	*S. gordonii/S. mitis/S. oralis/S. pneumoniae* (538 bp), *S. anginosus/S. intermedius/S. mutans/S. suis* (681 bp), *S. gallolyticus* (574 bp), *S. constellatus* (613 bp) *P. mlaninogenica* (partial, 681 bp), *P. intermedia* (partial, 613 bp), *P. buccae* (partial, 538 bp) *V. dispar* (partial, 613 bp), *V. nakazawae* (partial, 681 bp), *V. atypica* (partial, 538 bp) *A. gaoshouyii/A. polynesiensis/A. odontolyticus* (613 bp), *A. howelii/A. oris/A. tangfeifanii/A. viscosus* (681 bp) *G. haemolysans/G. morbillorum/G. massiliensis* (500 bp, 681 bp) *S. longum* (partial, 681 bp)

[a]ITS amplicons in bold were most concentrated within a sample.

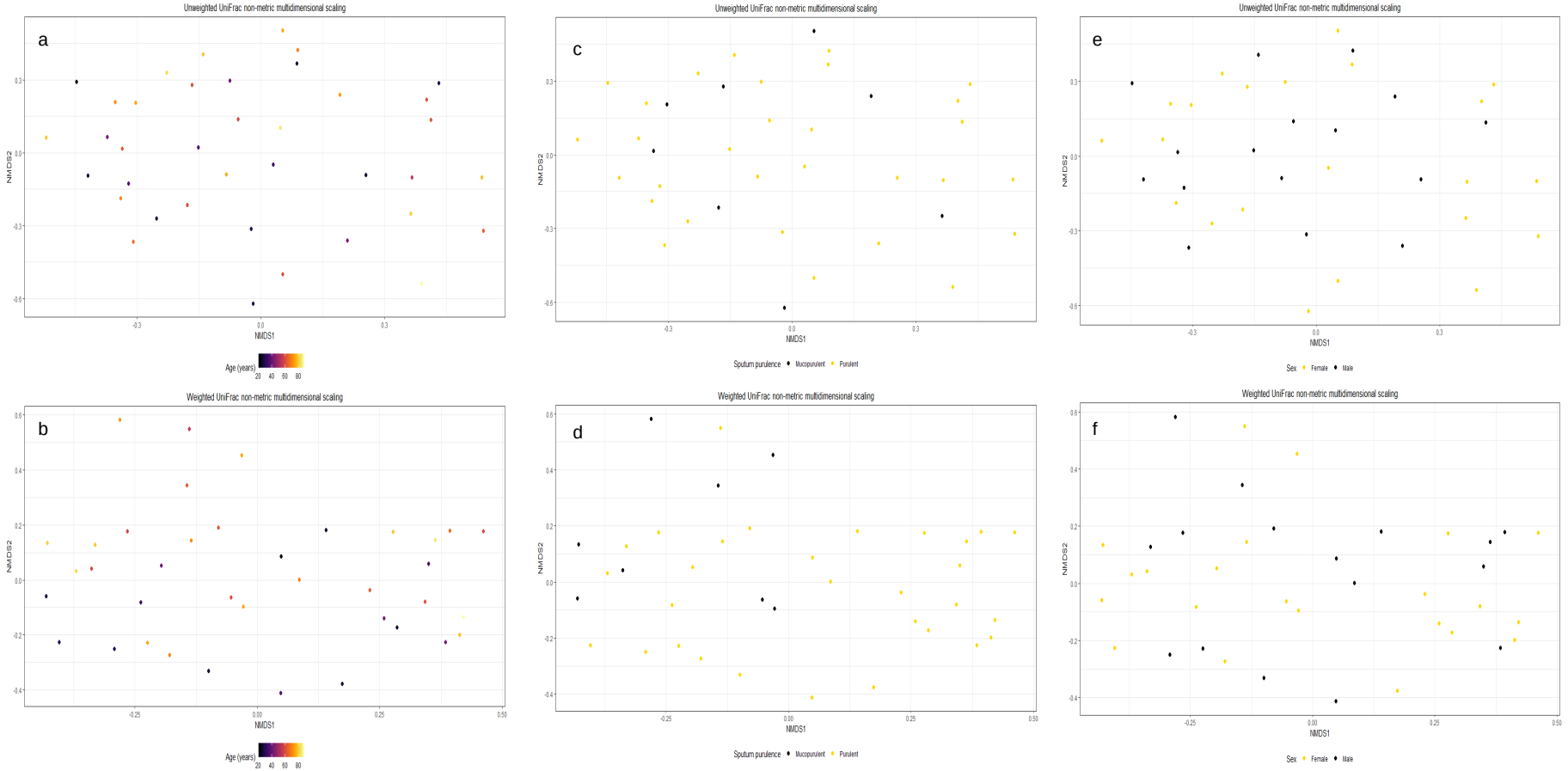
[b]Abbreviations include: Al; *Alloprevotella*, At; *Actinomyces*, Bg; *Bergeyella*, Bu; *Burkholderia*, Cp; *Capnocytophaga*, Cy; *Corynebacterium*, Eb; *Enterobacterales*, Ent; *Enterobacteriaceae*, Fu; *Fusobacterium*, Ge; *Gemella*, Gr; *Granulicatella*, Ha; *Haemophilus*, La; *Lactobacillus*, Lau; *Lautropia*, Lch; *Lachnoanaerobaculum*, Le; *Leptotrichia*, Lea; *Leptotrichiaceae*, Ne; *Neisseria*, Nv; *Noviherbaspirillum*, Pa; *Pseudomonas*, Po; *Porphyromonas*, Pst; *Pasteurellaceae*, Pv; *Prevotella*, Pva; *Prevotellaceae*, Ro; *Rothia*, Sa; *Staphylococcus*, Sc; *Scardovia*, Sn; *Stenotrophomonas*, Sne; *Sneathia*, St; *Stomatobaculum*, Str; *Streptococcus*, Tr; *Treponema*, Ve; *Veillonella*.



SI Figure 1 Rarefaction curves of sequenced sputa. Each sample has been coloured according to disease diagnosis: BR; navy, CF; gold, COPD; grey.



SI Figure 2 Alpha diversity correlated to patient age, sputum purulence and sex. Both Inverse Simpson (a) and Shannon diversity (b) was plotted against patient age and was statistically compared through multiple linear regression. A regression line has been included where each point has been coloured according to disease category (BR; navy, CF; gold, COPD; grey). InvSimpson (c) and Shannon (d) diversity was plotted against mucopurulent and purulent sputa. InvSimpson (e) and Shannon (f) diversity was also plotted against patient sex which was divided into separate disease categories. Boxplots were coloured according to disease diagnoses.

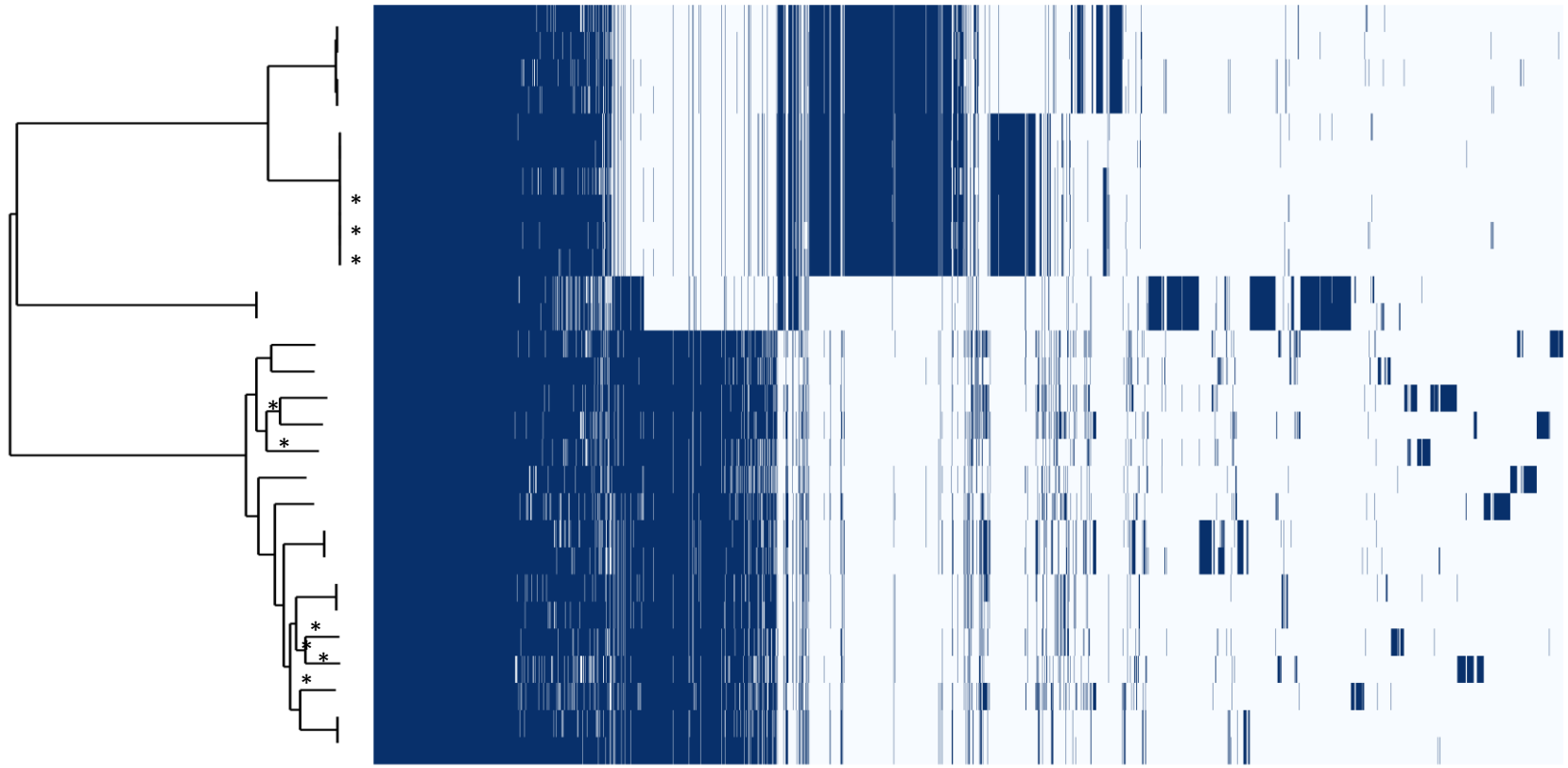


SI Figure 3 Bacterial community membership and structure compared to patient age, sputum purulence, and sex. Differences in community membership (unweighted UniFrac) was plotted against patient age (a), sputum purulence (b), and sex (c). Differences in community structure (weighted UniFrac) was plotted against patient age (d), sputum purulence (e), and patient sex (f). Patient age was coloured as a gradient using the R Viridis package.

Appendix E

SI Table 1 *Achromobacter* spp. Genomic features and statistics

Clinical ID	Genomic ID	Size (bp)	%GC	Contigs	N50	L50	CDS	RNAs	rRNA	tmRNA	tRNA
18179	*A. insuavis*	6545660	68.4	112	96456	22	5816	75	3	1	71
19065	*A. insuavis*	6676555	68.3	133	86035	23	5943	79	4	1	74
19114	*A. insuavis*	6956129	68.1	109	120513	15	6233	76	4	1	71
19115	*A. insuavis*	7034037	68.1	99	147245	15	6305	76	4	1	71
19116	*A. insuavis*	6784170	68.1	99	132544	17	6081	70	4	1	65
19117	*A. insuavis*	6980565	68.1	176	73686	33	6250	76	4	1	71
19118	*A. insuavis*	6953259	68.2	89	152439	17	6214	75	4	1	70
20885	*A. insuavis*	6635970	68.3	99	97897	20	5902	74	3	1	70
20886	*A. insuavis*	6622380	68.3	113	114308	23	5906	72	3	1	68
23032	*A. insuavis*	6954773	68.1	106	110479	20	6215	76	4	1	71
23681	*A. ruhlandii*	6106634	67.8	96	111202	18	5483	68	4	1	63
23696	*A. ruhlandii*	6226266	67.8	100	104600	18	5578	72	4	1	67
7901	*A. xylosoxidans*	6305324	67.5	79	139750	16	5692	57	4	1	52
11529	*A. xylosoxidans*	6213766	67.6	91	168070	12	5652	66	4	1	61
13945	*A. xylosoxidans*	6561758	67.6	68	172711	15	6000	61	4	1	56
16967	*A. xylosoxidans*	6178511	67.8	73	161659	12	5595	67	4	1	62
18454	*A. xylosoxidans*	6760908	67.4	72	196380	10	6216	67	4	1	62
19179	*A. xylosoxidans*	6164203	67.8	85	112923	16	5574	66	4	1	61
19519	*A. xylosoxidans*	6261297	67.7	83	164996	12	5701	64	4	1	59
20408	*A. xylosoxidans*	6327736	67.8	70	204591	11	5707	69	4	1	64
20777	*A. xylosoxidans*	6510956	67.6	56	182297	11	5945	66	4	1	61
22087	*A. xylosoxidans*	6735159	67.4	64	201296	11	6158	63	4	1	58
22114	*A. xylosoxidans*	6633523	67.5	82	145574	15	6077	62	4	1	57
22569	*A. xylosoxidans*	6238863	67.9	79	164073	12	5622	69	4	1	64
23826	*A. xylosoxidans*	6287573	67.6	99	105875	16	5644	66	4	1	61
24739	*A. xylosoxidans*	6357181	67.7	74	181199	10	5757	61	4	1	56
25103	*A. xylosoxidans*	6477026	67.6	52	209853	11	5980	66	4	1	61
26400	*A. xylosoxidans*	6383850	67.6	65	177709	11	5756	59	4	1	54



SI Figure 1 *Achromobacter* spp. roary pangenome matrix and core genome SNP-sites maximum likelihood phylogenetic tree. Shared genes are highlighted in blue while unique genes are blank, demarcating clade specific presence and absence of genes. A graphical output was generated using the roary_plots.py Python script (Galardini 2020). Here, A ML phylogenomic tree of the 28 CF *Achromobacter* isolates was constructed from a Roary core genome alignment with SNP-sites using RAxML and the GTR gamma model. This phylogenetic tree was manipulated in FigTree and was joined with the Roary gene presence and absence dataset to generate this figure. Node confidence (bootstrap) is shown as an asterisk if below 100.

SI Table 2 MALDI-TOF identification of CF *Achromobacter* spp.

Clinical ID	Organism (best match) [a]	Score [b]	Genomic ID
18179	*A. xylosoxidans*	1.98	*A. insuavis*
19065	*A. xylosoxidans*	2.02	*A. insuavis*
19114	*A. xylosoxidans*	2.13	*A. insuavis*
19115	*A. xylosoxidans*	2.04	*A. insuavis*
19116	*A. xylosoxidans*	1.89	*A. insuavis*
19117	*A. xylosoxidans*	2.10	*A. insuavis*
19118	*A. xylosoxidans*	2.02	*A. insuavis*
20885	*A. xylosoxidans*	2.08	*A. insuavis*
20886	*A. xylosoxidans*	2.06	*A. insuavis*
23032	*A. xylosoxidans*	2.05	*A. insuavis*
23681	*A. xylosoxidans*	2.06	*A. ruhlandii*
23696	*A. xylosoxidans*	2.02	*A. ruhlandii*
7901	*A. xylosoxidans*	2.30	*A. xylosoxidans*
11529	*A. xylosoxidans*	2.32	*A. xylosoxidans*
13945	*A. xylosoxidans*	2.23	*A. xylosoxidans*
16967	*A. xylosoxidans*	2.45	*A. xylosoxidans*
18454	*A. xylosoxidans*	2.39	*A. xylosoxidans*
19179	*A. xylosoxidans*	2.35	*A. xylosoxidans*
19519	*A. xylosoxidans*	2.11	*A. xylosoxidans*
20408	*A. xylosoxidans*	2.06	*A. xylosoxidans*
20777	*A. xylosoxidans*	2.40	*A. xylosoxidans*
22087	*A. xylosoxidans*	2.18	*A. xylosoxidans*
22114	*A. xylosoxidans*	2.28	*A. xylosoxidans*
22569	*A. xylosoxidans*	2.23	*A. xylosoxidans*
23826	*A. xylosoxidans*	2.37	*A. xylosoxidans*
24739	*A. xylosoxidans*	2.30	*A. xylosoxidans*
25103	*A. xylosoxidans*	2.34	*A. xylosoxidans*
26400	*A. xylosoxidans*	2.03	*A. xylosoxidans*

[a] Includes Bruker Biotyper closest reference match.

[b] Scores relate to the Bruker Biotyper confidence scores for the correct identification of a target organism. Scores >2 were considered highly confident and are highlighted in green. Scores <2 denote a lower confidence. These scores are highlighted in yellow.

SI Table 3 *CF Achromobacter* spp. MLSA sequence types

Clinical ID	Genomic ID	ST [a]	New ST	*nusA*	*rpoB*	*eno*	*gltB*	*lepA*	*nuoL*	*nrdA*
18179	*A. insuavis*	215~	Unknown	97	51	18	18	57~	23	19
19065	*A. insuavis*	215~	Unknown	97	51	18	18	57~	23	19
19114	*A. insuavis*	304~	Unknown	92~	112	20	19	112	24	19
19115	*A. insuavis*	304~	Unknown	92~	112	20	19	112	24	19
19116	*A. insuavis*	304~	Unknown	92~	112	20	19	112	24	U
19117	*A. insuavis*	304~	Unknown	92~	112	20	19	112	24	19
19118	*A. insuavis*	304~	Unknown	92~	112	20	19	112	24	19
20885	*A. insuavis*	215~	Unknown	97	51	18	18	57~	23	19
20886	*A. insuavis*	215~	Unknown	97	51	18	18	57~	23	19
23032	*A. insuavis*	304~	Unknown	92~	112	20	19	112	24	19
23681	*A. ruhlandii*	265~	Unknown	96	93~	26	16	46~	U	84~
23696	*A. ruhlandii*	265~	Unknown	96	93~	26	16	46~	19~	84~
7901	*A. xylosoxidans*	180	Match	1	26	2	2	59	8	2
11529	*A. xylosoxidans*	24~	Unknown	105	39	8	1	59	U	3
13945	*A. xylosoxidans*	28~	Unknown	U	40	2	2	59	8	2
16967	*A. xylosoxidans*	184~	Unknown	81	21	2	4	59	8	U
18454	*A. xylosoxidans*	27	Match	14	26	2	2	59	13	3
19179	*A. xylosoxidans*	184~	Novel ST	81	21	2	4	59	8	2
19519	*A. xylosoxidans*	176~	Unknown	14	40	2	1	59~	4	4
20408	*A. xylosoxidans*	237~	Unknown	7	100~	2	4	62	8	1
20777	*A. xylosoxidans*	184~	Unknown	81	84	2	2	59	8	U
22087	*A. xylosoxidans*	182	Match	6	26	2	4	62	4	2
22114	*A. xylosoxidans*	182	Match	6	26	2	4	62	4	2
22569	*A. xylosoxidans*	237~	Unknown	7	100~	2	4	62	8	1
23826	*A. xylosoxidans*	27~	Unknown	14	26~	2	2	82~	8	3
24739	*A. xylosoxidans*	290	Match	4	40	2	4	62	8	2
25103	*A. xylosoxidans*	311~	Unknown	14	39	8	4	59	U	4
26400	*A. xylosoxidans*	12~	Novel ST	1	21	2	2	59	8	2

[a] STs are reported using the set default values to report known sequence types (STs), novel STs and genuine partial ST partial matches within the GitHub MLST package (Seemann 2018a). When interpreting the type of match provided, the ~ represents a novel, full length allele similar to that corresponding allele was found. If one of the alleles was not found in the existing *Achromobacter* MLST database, it was flagged as “U” meaning unknown.

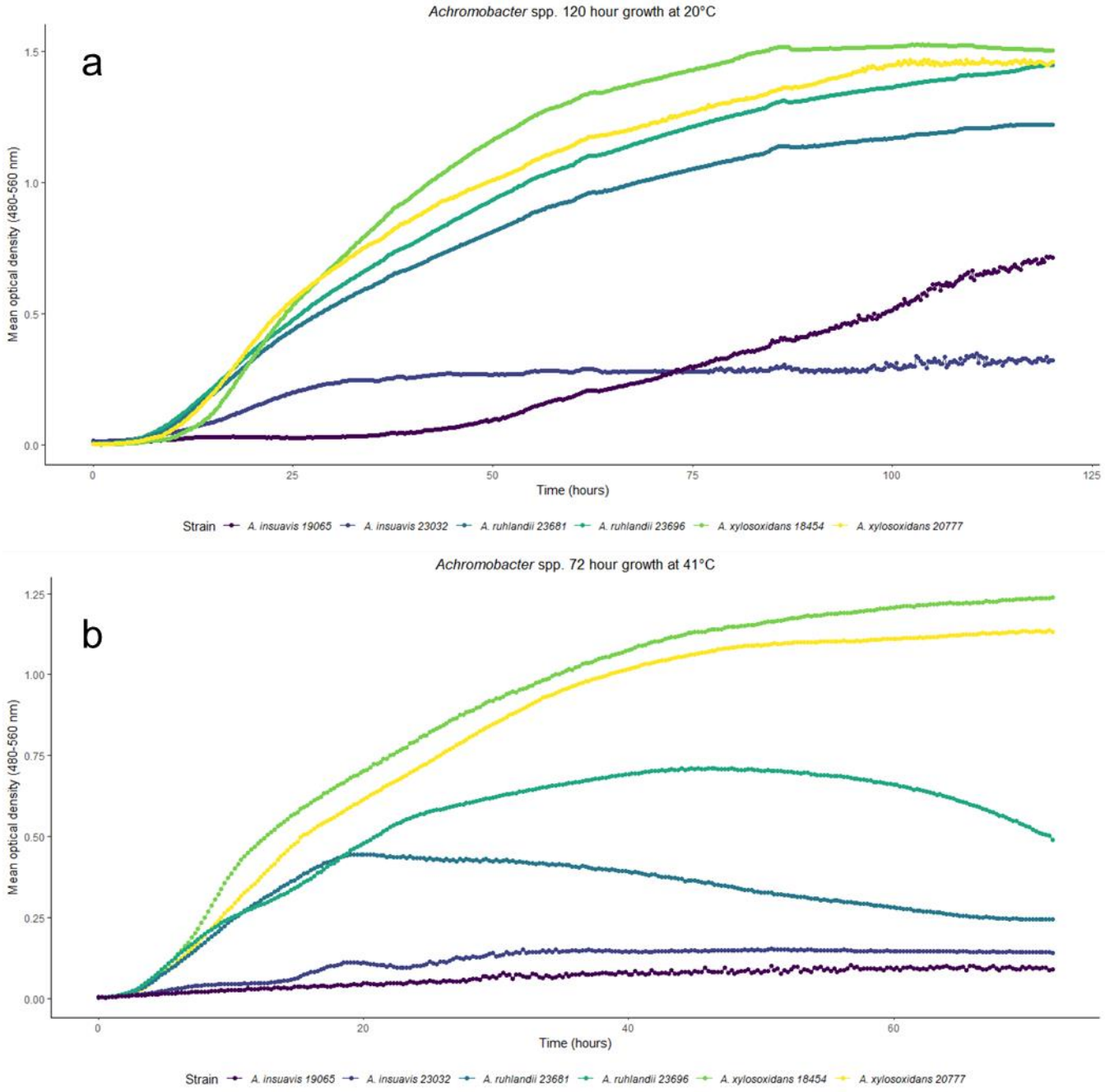
SI Table 4 *in silico* CF *Achromobacter* spp. Roary and ARIBA-CARD AMR profiles

***In silico* AMR genes**		***A. insuavis***	***A. ruhlandii***	***A. xylosoxidans***
ARIBA-CARD	***axyXY-oprZ***	10	2	16
	$bla_{OXA-114a}$	0	0	16
	$bla_{OXA-258}$	0	2	0
	$bla_{OXA-243}$	5	0	0
	$bla_{OXA-457}$	4	0	0
	sul2	0	0	1
Roary[a]	bla_{OXA-18}	10	2	16
	mexAB-oprM	10	2	16
	ampC	10	2	16
	blaP	0	1	7
	blh	10	2	18
	blm	10	2	16
	ccrA	0	0	12
	tetA	12	4	32
	bcr	30	4	32
	ble	10	0	0
	phoQ	9	0	16
	phoP	9	4	16
	macAB	10	2	16

[a] Interpreted from the Roary pangenome gene presence/absence matrix which includes gene variants, hence some values may be over 28.

SI Table 5 Pure culture CF *Achromobacter* and CF pathogen panel 20°C and 41°C growth parameters

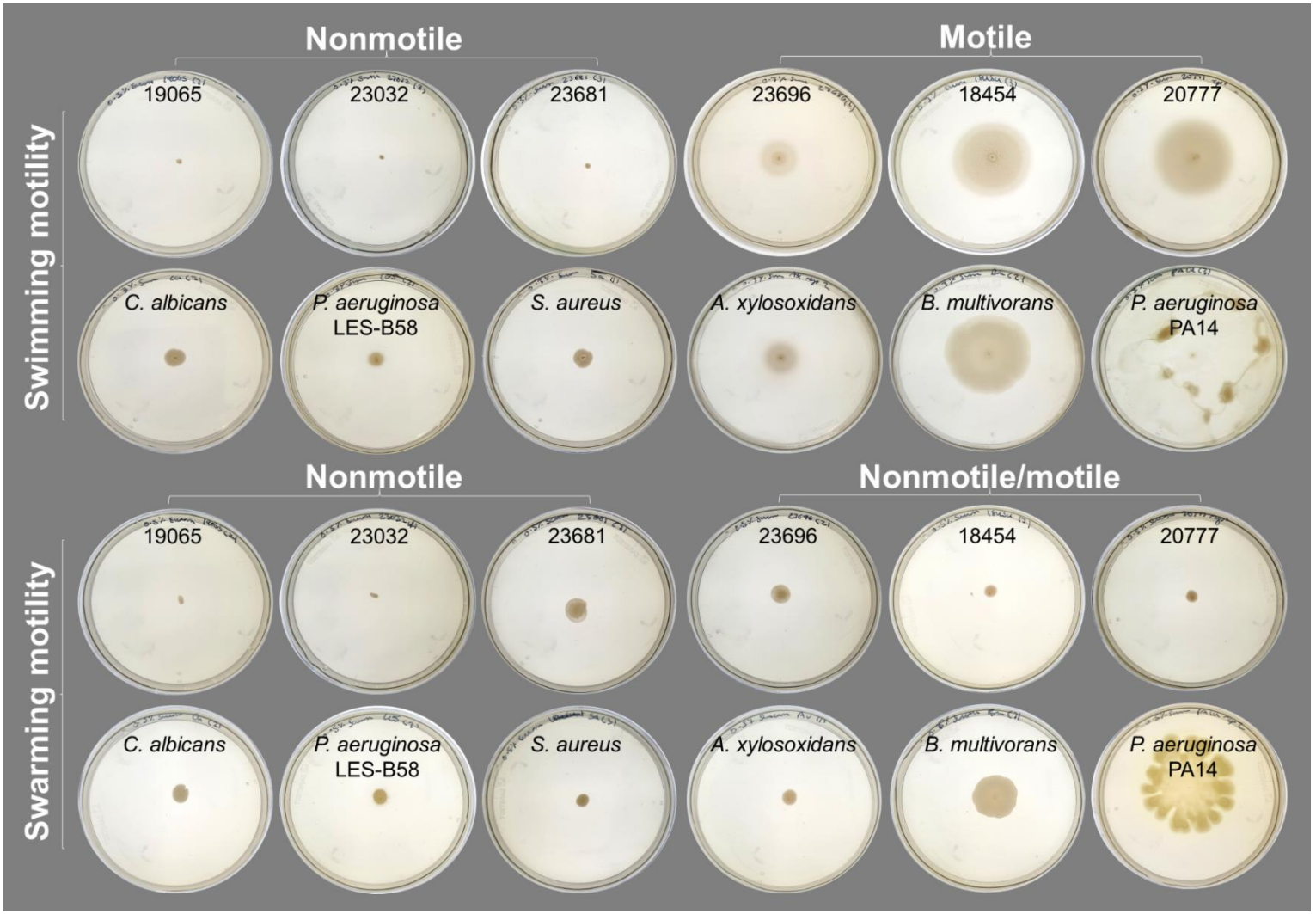
Species	Strain	Temperature	Lag phase (λ) in hours	Carrying capacity (k)	Growth rate (μ) in hours	Doubling time (g) in hours	Maximum predicted OD	Observed maximum OD	RMSE	R^2
A. insuavis	19065	20°C	44.4	0.9	0.1	0.6	0.7	0.7	0.02	1.00
A. insuavis	19065	41°C	12.5	0.1	0.1	0.3	0.1	0.1	0.004	0.98
A. insuavis	23032	20°C	7.9	0.3	0.2	0.2	0.3	0.3	0.02	0.97
A. insuavis	23032	41°C	4.5	0.04	0.6	0.05	0.04	0.04	0.007	0.98
A. ruhlandii	23681	20°C	6.7	1.2	0.1	0.4	1.2	1.2	0.05	0.99
A. ruhlandii	23681	41°C	4	0.4	0.5	0.1	0.4	0.2	0.06	0.7
A. ruhlandii	23696	20°C	6.9	1.4	0.1	0.5	1.4	1.4	0.06	0.99
A. ruhlandii	23696	41°C	3.1	0.7	0.2	0.2	0.7	0.5	0.05	0.95
A. xylosoxidans	18454	20°C	11.6	1.5	0.1	0.3	1.5	1.5	0.05	0.99
A. xylosoxidans	18454	41°C	4.5	1.2	0.1	0.2	1.2	1.2	0.06	0.98
A. xylosoxidans	20777	20°C	9.3	1.4	0.1	0.4	1.4	1.5	0.07	0.98
A. xylosoxidans	20777	41°C	3.8	1.1	0.1	0.2	1.1	1.1	0.04	0.99



SI Figure 2 Single *Achromobacter* spp. growth curves. Sample OD (480 nm-560 nm) was plotted at every 15-minute interval over 120-h at 20°C (a) and for 72-h at 41°C (b).

SI Table 6 Aligned rank transform ANOVA *post hoc* statistical comparisons

Comparison	Estimate	SE	t.ratio	Adjusted P value
A. insuavis 19065 & *A. xylosoxidans* 18454	14.3	3.4	4.3	0.0002
A. insuavis 23032 & *A. xylosoxidans* 18454	-23.6	3.4	-7.1	P<0.0001
A. ruhlandii 23681 & *A. xylosoxidans* 18454	-18.4	3.4	-5.5	P<0.0001
A. xylosoxidans 20777 & *A. xylosoxidans* 18454	9.7	3.4	2.9	0.008
A. insuavis 23032 & *A. insuavis* 19065	-37.9	3.4	-11.3	P<0.0001
A. ruhlandii 23681 & *A. insuavis* 19065	-32.7	3.4	-9.8	P<0.0001
A. ruhlandii 23696 & *A. insuavis* 19065	-14.3	3.4	-4.3	0.0002
A. insuavis 23032 & *A. xylosoxidans* 20777	-33.3	3.4	-10.0	P<0.0001
A. ruhlandii 23681 & *A. xylosoxidans* 20777	-28.1	3.4	-8.4	P<0.0001
A. ruhlandii 23696 & *A. xylosoxidans* 20777	-9.7	3.4	-2.9	0.008
A. ruhlandii 23696 & *A. insuavis* 23032	23.6	3.4	7.1	P<0.0001
A. ruhlandii 23696 & *A. ruhlandii* 23681	18.4	3.4	5.5	P<0.0001
37°C & 20°C	-12.0	2.9	-4.2	0.0002
41°C & 20°C	-32.8	2.9	-11.5	P<0.0001
41°C & 37°C	-20.8	2.9	-7.3	P<0.0001
A. xylosoxidans 18454:20°C & *A. xylosoxidans* 18454:37°C	-23.5	6.1	-3.8	0.002
A. xylosoxidans 20777:20°C & *A. xylosoxidans* 20777:37°C	-15.7	6.1	-2.6	0.03
A. ruhlandii 23681:20°C & *A. ruhlandii* 23681:37°C	-15.7	6.1	-2.6	0.03
A. xylosoxidans 18454:37°C & A. insuavis 19065:37°C	15.7	6.1	2.6	0.03
A. xylosoxidans 18454:37 & *A. insuavis* 23696:37°C	23.5	6.1	3.8	0.002
A. xylosoxidans 20777:37°C & *A. ruhlandii* 23696:37°C	15.7	6.1	2.6	0.03
A. insuavis 23032:37°C & *A. ruhlandii* 23696:37°C	15.7	6.1	2.6	0.03
A. ruhlandii 23681:37°C & *A. ruhlandii* 23696:37°C	15.7	6.1	2.6	0.03
A. ruhlandii 23681:20°C & *A. ruhlandii* 23681:41°C	-32.0	6.1	-5.2	0.0001
A. ruhlandii 23696:20°C & *A. ruhlandii* 23696:41°C	-23.5	6.1	-3.8	0.002
A. xylosoxidans 18454:37°C & *A. xylosoxidans* 18454:41°C	23.5	6.1	3.8	0.002
A. xylosoxidans 20777:37°C & *A. xylosoxidans* 20777:41°C	15.7	6.1	2.6	0.03
A. insuavis 23032:37°C & *A. insuavis* 23032:41°C	-19.3	6.1	-3.1	0.011
A. ruhlandii 23681:37°C & A. ruhlandii 23681:41°C	-16.3	6.1	-2.7	0.03
A. ruhlandii 23696:37°C & A. ruhlandii 23696:41°C	-23.5	6.1	-3.8	0.002
A. xylosoxidans 18454:41°C & *A. insuavis* 23032:41°C	-35.0	6.1	-5.7	P<0.0001
A. xylosoxidans 18454:41°C & *A. ruhlandii* 23681:41°C	-32.0	6.1	-5.2	0.0001
A. xylosoxidans 18454:41°C & *A. ruhlandii* 23696:41°C	-23.5	6.1	-3.8	0.002
A. insuavis 19065:41°C & *A. insuavis* 23032:41°C	-35.0	6.1	-5.7	P<0.0001
A. insuavis 19065:41°C & *A. ruhlandii* 23681:41°C	-32.0	6.1	-5.2	0.0001
A. insuavis 19065:41°C & *A. ruhlandii* 23696:41°C	-23.5	6.1	-3.8	0.002
A. xylosoxidans 20777:41°C & *A. insuavis* 23032:41°C	-35.0	6.1	-5.7	P<0.0001
A. xylosoxidans 20777:41°C & *A. ruhlandii* 23681:41°C	-32.0	6.1	-5.2	0.0001
A. xylosoxidans 20777:41°C & *A. ruhlandii* 23696:41°C	-23.5	6.1	-3.8	0.002



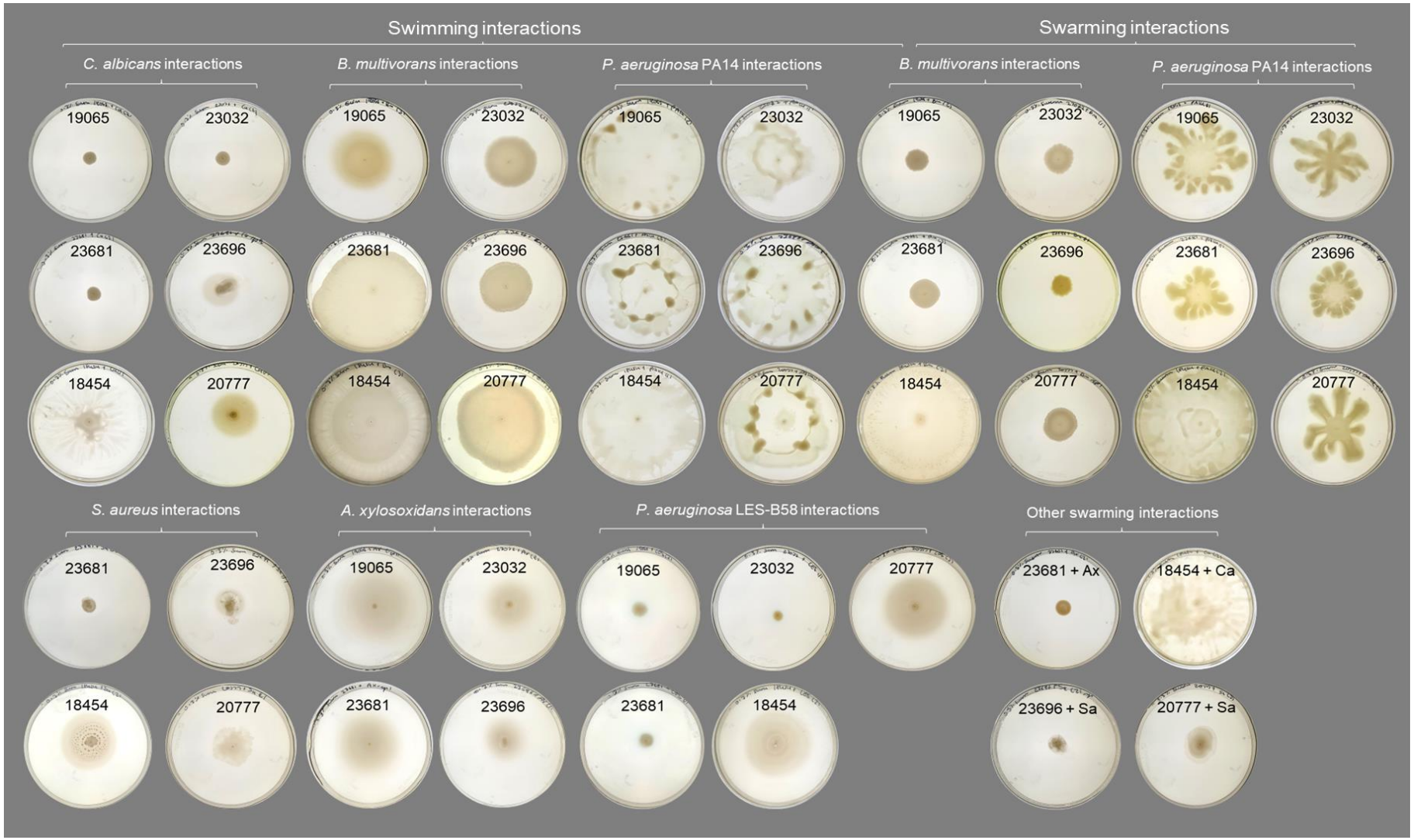
SI Figure 3 Complete clinical *Achromobacter* spp. and CF pathogen pure culture swimming and swarming motility. Images have been cropped and corrections have been made based on image brightness and saturation to facilitate viewing. The *A. xylosoxidans* reference refers to *A. xylosoxidans* LMG 1863.

SI Table 7 Swimming and swarming motility of clinical *Achromobacter* spp. and CF pathogens in pure culture and coculture[a]

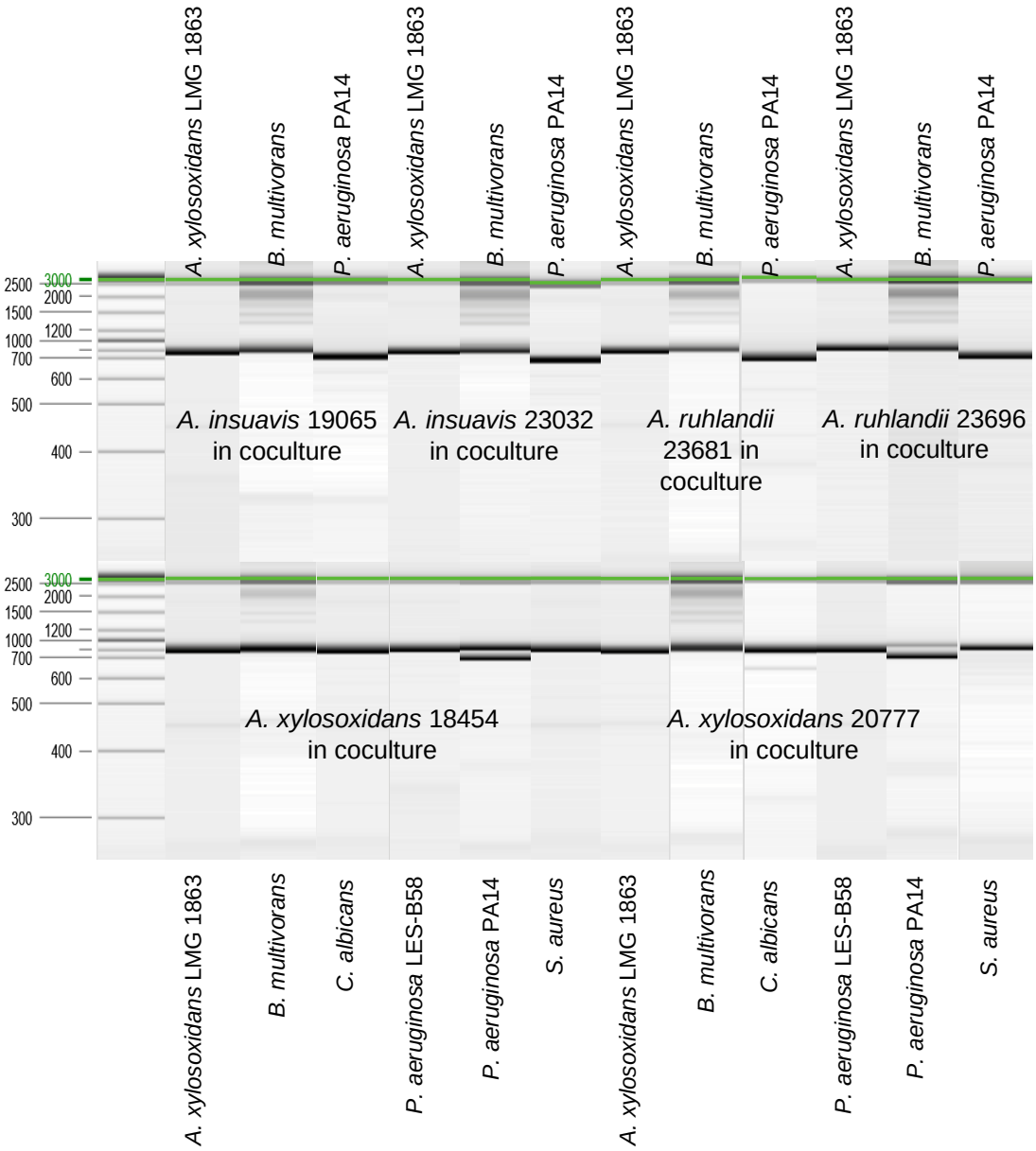
	swimming				swarming			
	24-h motility		48-h motility		24-h motility		48-h motility	
A. xylosoxidans LMG 1863	5.3	M	13.2	M	1.6	N	3.9	N
B. multivorans	17.2	M	22.5	M	6.5	M	10.1	M
C. albicans	3.3	N	4.8	N	3.3	N	4.9	N
P. aeruginosa LES-B58	2.1	N	4.4	N	0.9	N	3.6	N
P. aeruginosa PA14	42.7	M*	45.0	M*	14.8	M	21.0	M
S. aureus	2.9	N	4.9	N	2.0	N	3.4	N
A. insuavis 23032	0.0	N	1.2	N	0.0	N	0.9	N
A. insuavis 19065	0.0	N	1.0	N	0.0	N	1.3	N
A. ruhlandii 23681	0.0	N	0.8	N	3.7	N	5.2	M[b]
A. ruhlandii 23696	1.1	N	8.6	M	4.1	N	5.3	M[b]
A. xylosoxidans 18454	10.6	M	21.6	M	1.0	N	3.5	N
A. xylosoxidans 20777	8.4	M	21.4	M	1.1	N	3.5	N
A. insuavis 23032 & *S. aureus*	2.5	N	4.4	N	1.7	N	3.4	N
A. insuavis 23032 & *A. xylosoxidans* LMG 1863	5.7	M	24.0	M	0.2	N	3.1	N
A. insuavis 23032 & *B. multivorans*	14.3	M	21.0	M	6.3	M	8.5	M
A. insuavis 23032 & *C. albicans*	3.8	N	5.9	M	2.6	N	4.3	N
A. insuavis 23032 & *P. aeruginosa* LES-B58	2.9	N	7.4	M	0.0	N	1.8	N
A. insuavis 23032 & *P. aeruginosa* PA14	32.2	M*	45.0	M*	16.3	M	26.5	M
A. insuavis 19065 & *S. aureus*	2.6	N	4.7	N	1.5	N	3.4	N
A. insuavis 19065 & *A. xylosoxidans* LMG 1863	9.3	M	31.7	M*	0.6	N	3.1	N
A. insuavis 19065 & *B. multivorans*	15.3	M	29.0	M	5.8	M	8.0	M
A. insuavis 19065 & *C. albicans*	3.3	N	5.4	M	2.8	N	4.6	N
A. insuavis 19065 & *P. aeruginosa* LES-B58	2.3	N	6.3	M	0.0	N	2.3	N
A. insuavis 19065 & *P. aeruginosa* PA14	34.0	M*	45.0	M*	18.9	M	26.9	M
A. ruhlandii 23681 & *A. xylosoxidans* LMG 1863	8.7	M	28.1	M	3.2	N	5.2	M[b]
A. ruhlandii 23681 & *B. multivorans*	30.8	M*	41.1	M*	7.3	M	11.3	M
A. ruhlandii 23681 & *C. albicans*	4.4	N	6.0	M	3.8	N	4.8	N
A. ruhlandii 23681 & *P. aeruginosa* LES-B58	3.2	N	7.0	M	0.6	N	2.6	N
A. ruhlandii 23681 & *P. aeruginosa* PA14	40.1	M*	45.0	M*	15.9	M	24.3	M
A. ruhlandii 23681 & *S. aureus*	3.5	N	6.7	M	2.0	N	4.7	N
A. ruhlandii 23696 & *A. xylosoxidans* LMG 1863	6.1	M	16.2	M	2.5	N	4.3	N
A. ruhlandii 23696 & *B. multivorans*	11.2	M	17.4	M	4.5	N	6.7	M
A. ruhlandii 23696 & *C. albicans*	3.7	N	7.9	M	2.8	N	4.7	N
A. ruhlandii 23696 & *P. aeruginosa* LES-B58	2.3	N	3.8	N	0.8	N	2.2	N
A. ruhlandii 23696 & *P. aeruginosa* PA14	37.8	M*	45.0	M*	16.6	M	21.5	M
A. ruhlandii 23696 & *S. aureus*	3.5	N	9.2	M	2.4	N	5.9	M[b]
A. xylosoxidans 18454 & *B. multivorans*	16.3	M	43.3	M*	7.3	M	44.3	M*
A. xylosoxidans 18454 & *C. albicans*	40.1	M*	45.0	M*	23.0	M	45.0	M*
A. xylosoxidans 18454 & *P. aeruginosa* LES-B58	7.6	M	18.3	M	1.3	N	4.0	N
A. xylosoxidans 18454 & *P. aeruginosa* PA14	42.0	M*	45.0	M*	14.3	M	45.0	M*
A. xylosoxidans 18454 & *S. aureus*	9.8	M	24.2	M	2.8	N	4.5	N
A. xylosoxidans 20777 & *B. multivorans*	14.5	M	22.6	M	6.9	M	9.8	M
A. xylosoxidans 20777 & *C. albicans*	7.3	M	16.8	M	2.9	N	4.9	N
A. xylosoxidans 20777 & *P. aeruginosa* LES-B58	7.3	M	18.4	M	0.9	N	2.9	N
A. xylosoxidans 20777 & *P. aeruginosa* PA14	39.8	M*	45.0	M*	33.3	M*	34.8	M*
A. xylosoxidans 20777 & *S. aureus*	6.9	M	16.7	M	2.5	N	7.4	M

[a] Highly motile plates are denoted with an asterisk.

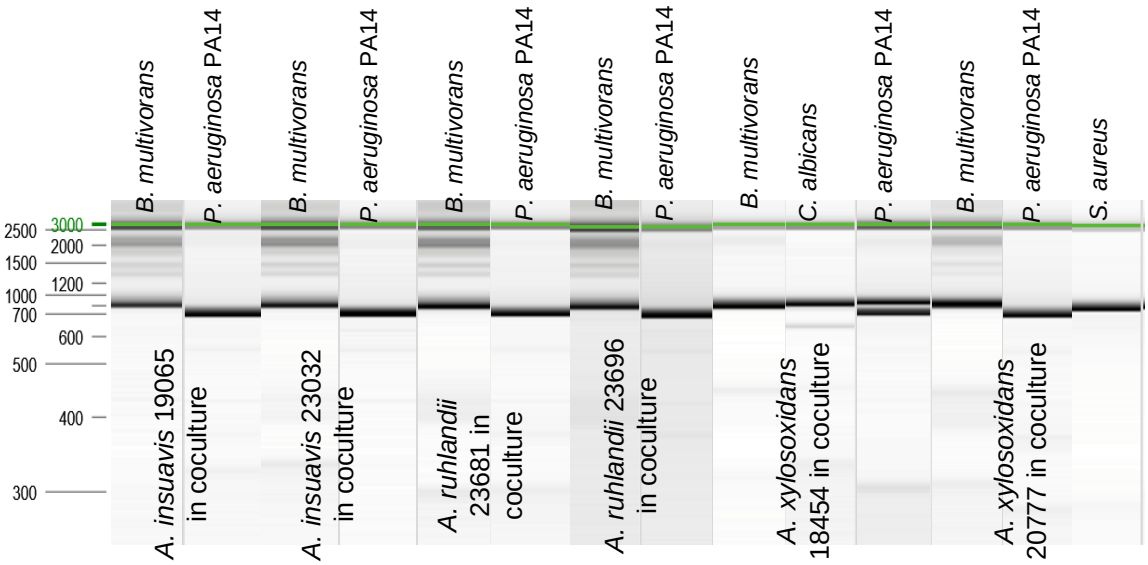
[b] Swarming motility was not considered but as a result of pooling mucoid colonies.



SI Figure 4 Mixed pathogen positive swimming and swarming motility. Images have been cropped and corrections have been made based on image brightness and saturation to facilitate viewing. The *A. xylosoxidans* reference refers to *A. xylosoxidans* LMG 1863.



SI Figure 5 RISA PCR amplified ITS regions of mixed pathogen cultures from swimming plates. Each gel image processed through the QIAxcel has been edited in order with a DNA size reference ladder to the left, and the QIAxcel upper marker is highlighted in green in each lane. Each interaction has been labelled where secondary pathogens have been labelled on the outside.



SI Figure 6 RISA PCR amplified ITS regions of mixed pathogen cultures from swarming plates. Each gel image processed through the QIAxcel has been edited in order with a DNA size reference ladder to the left, and the QIAxcel upper marker is highlighted in green in each lane. Each interaction has been labelled where secondary pathogens have been labelled on the outside.

SI Table 8 Regression model statistics for pure and mixed culture bacteria motility assays

Pure culture comparison	Interaction	Estimate	SE	t ratio	P value
Swimming assays					
A. insuavis 19065	*A. xylosoxidans* LMG 1863 & *A. insuavis* 19065	2.2	0.06	34.4	P<0.0001
A. insuavis 19065	*B. multivorans* & *A. insuavis* 19065	2.3	0.06	35.6	P<0.0001
A. insuavis 19065	*P. aeruginosa* PA14 & *A. insuavis* 19065	2.6	0.06	39.9	P<0.0001
A. insuavis 23032	*A. xylosoxidans* LMG 1863 & *A. insuavis* 23032	2.0	0.08	25.6	P<0.0001
A. insuavis 23032	*B. multivorans* & *A. insuavis* 23032	2.2	0.08	27.9	P<0.0001
A. insuavis 23032	*P. aeruginosa* PA14 & *A. insuavis* 23032	2.5	0.08	32.2	P<0.0001
A. ruhlandii 23681	*A. xylosoxidans* LMG 1863 & *A. ruhlandii* 23681	2.6	0.1	24.0	P<0.0001
A. ruhlandii 23681	*B. multivorans* & *A. ruhlandii* 23681	3.0	0.1	27.2	P<0.0001
A. ruhlandii 23681	*P. aeruginosa* PA14 & *A. ruhlandii* 23681	3.1	0.1	27.9	P<0.0001
A. ruhlandii 23696	*A. xylosoxidans* LMG 1863 & *A. ruhlandii* 23696	0.9	0.1	7.8	0.0006
A. ruhlandii 23696	*B. multivorans* & *A. ruhlandii* 23696	1.1	0.1	9.4	0.0003
A. ruhlandii 23696	*P. aeruginosa* PA14 & *A. ruhlandii* 23696	1.6	0.09	18.1	P<0.0001
A. xylosoxidans 18454	*B. multivorans* & *A. xylosoxidans* 18454	0.2	0.06	3.9	0.0109
A. xylosoxidans 18454	*C. albicans* & *A. xylosoxidans* 18454	0.5	0.06	7.5	0.0004
A. xylosoxidans 18454	*P. aeruginosa* PA14 & *A. xylosoxidans* 18454	0.5	0.06	7.7	0.0003
A. xylosoxidans 20777	*P. aeruginosa* PA14 & *A. xylosoxidans* 20777	0.5	0.07	7.7	0.0003
Swarming assays					
A. insuavis 19065	*B. multivorans* & *A. insuavis* 19065	0.7	0.1	13.3	P<0.0001
A. insuavis 19065	*P. aeruginosa* PA14 & *A. insuavis* 19065	1.3	0.1	26.9	P<0.0001
A. insuavis 23032	*B. multivorans* & *A. insuavis* 23032	0.8	0.04	21.2	P<0.0001
A. insuavis 23032	*P. aeruginosa* PA14 & *A. insuavis* 23032	1.3	0.04	37.7	P<0.0001
A. ruhlandii 23681	*B. multivorans* & *A. ruhlandii* 23681	0.3	0.04	8.5	0.0003
A. ruhlandii 23681	*P. aeruginosa* PA14 & *A. ruhlandii* 23681	0.8	0.04	20.2	P<0.0001
A. ruhlandii 23696	*P. aeruginosa* PA14 & *A. ruhlandii* 23696	0.7	0.02	30.3	P<0.0001
A. xylosoxidans 18454	*B. multivorans* & *A. xylosoxidans* 18454	1	0.04	23.7	P<0.0001
A. xylosoxidans 18454	*C. albicans* & *A. xylosoxidans* 18454	1.3	0.04	31.4	P<0.0001
A. xylosoxidans 18454	*P. aeruginosa* PA14 & *A. xylosoxidans* 18454	1.2	0.04	27.8	P<0.0001
A. xylosoxidans 20777	*B. multivorans* & *A. xylosoxidans* 20777	0.5	0.07	6.4	0.001
A. xylosoxidans 20777	*P. aeruginosa* PA14 & *A. xylosoxidans* 20777	1.4	0.07	19.3	P<0.0001
A. xylosoxidans 20777	*S. aureus* & *A. xylosoxidans* 20777	0.2	0.07	3.1	0.04

SI Table 9 QIAxcel ITS amplicon size, concentration, and species identification from swimming plates

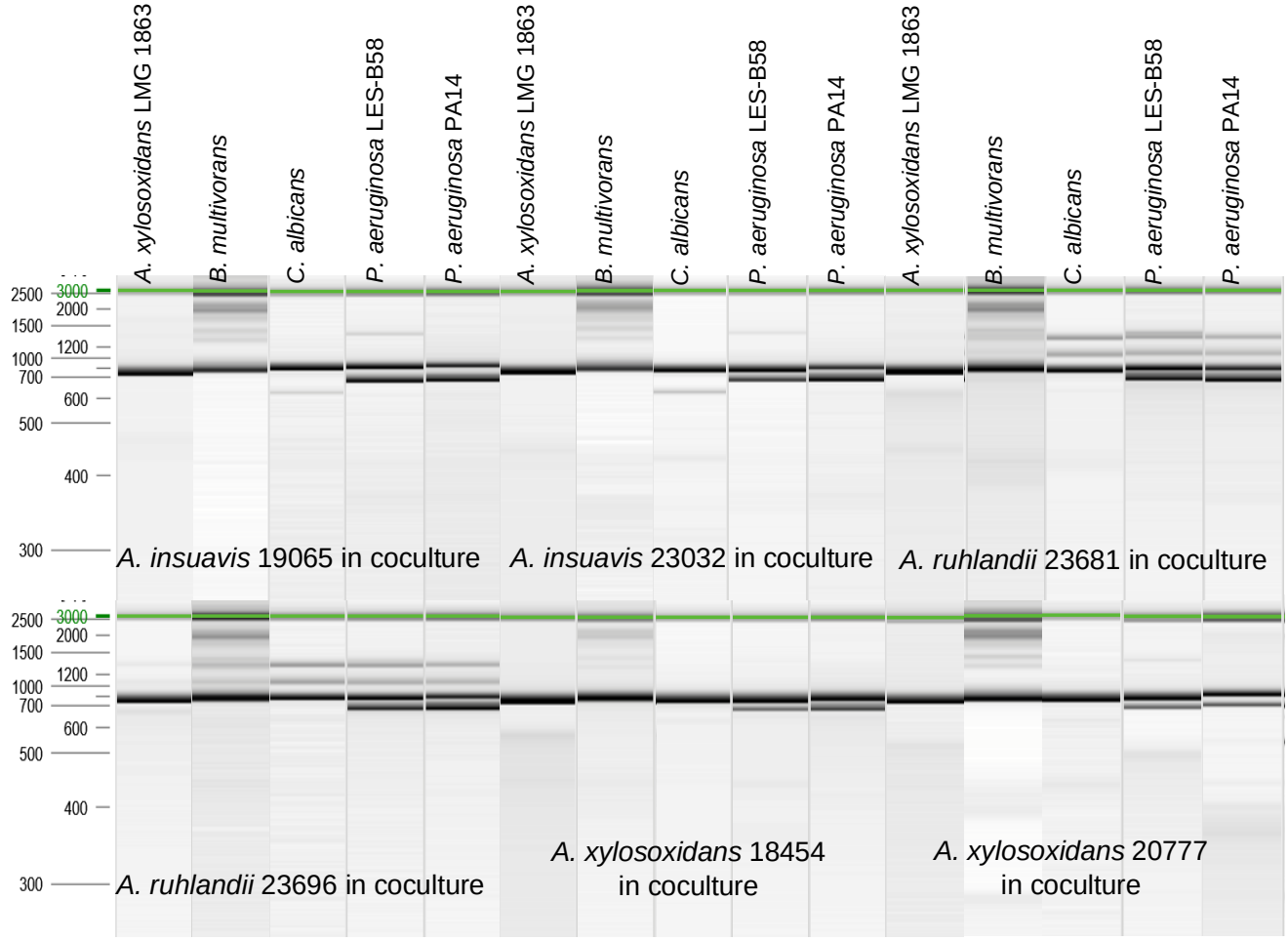
Interaction swimming	clinical *Achromobacter* isolate		CF reference		% ratio
	ITS (bp)	(ng/µl)	ITS (bp)	(ng/µl)	
A. insuavis 19065 & *A. xylosoxidans* LMG 1863			778.3	35.1	
A. insuavis 19065 & *B. multivorans*			808	12.9	
A. insuavis 19065 & *P. aeruginosa* PA14			739.7	23.1	
A. insuavis 23032 & *A. xylosoxidans* LMG 1863			802.3	20.4	
A. insuavis 23032 & *B. multivorans*			789.7	12.5	
A. insuavis 23032 & *P. aeruginosa* PA14			699	20	
A. ruhlandii 23681 & *A. xylosoxidans* LMG 1863			769.7	34.5	
A. ruhlandii 23681 & *B. multivorans*			811.3	9.9	
A. ruhlandii 23681 & *P. aeruginosa* PA14			689	27.3	
A. ruhlandii 23696 & *A. xylosoxidans* LMG 1863			818.5	20.1	
A. ruhlandii 23696 & *B. multivorans*			837.3	8.8	
A. ruhlandii 23696 & *P. aeruginosa* PA14			705.3	15.3	
A. xylosoxidans 18454 & *B. multivorans*			810.3	21.4	
A. xylosoxidans 18454 & *C. albicans*	773.7	32.5			100
A. xylosoxidans 18454 & *P. aeruginosa* LES-B58	785.3	35.2			100
A. xylosoxidans 18454 & *P. aeruginosa* PA14	818.7	19.4	691.0	17.7	52.2
A. xylosoxidans 18454 & *S. aureus*	778.7	32.3			100
A. xylosoxidans 20777 & *B. multivorans*			806	15.9	
A. xylosoxidans 20777 & *C. albicans*	781.7	30.3			100
A. xylosoxidans 20777 & *P. aeruginosa* LES-B58	798.3	26			100
A. xylosoxidans 20777 & *P. aeruginosa* PA14	706.7	11	866.7	8.1	57.7
A. xylosoxidans 20777 & *S. aureus*	857.7	5.7			100

SI Table 10 QIAxcel ITS amplicon size, concentration, and species identification from swarming plates

Interaction swarming	clinical *Achromobacter* isolate		CF reference		% ratio
	ITS (bp)	(ng/µl)	ITS (bp)	(ng/µl)	
A. insuavis 19065 & *B. multivorans*			811.3	16.4	
A. insuavis 19065 & *P. aeruginosa* PA14			706	25.5	
A. insuavis 23032 & *B. multivorans*			809	16.8	
A. insuavis 23032 & *P. aeruginosa* PA14			738.7	20.2	
A. ruhlandii 23681 & *B. multivorans*			813	18.1	
A. ruhlandii 23681 & *P. aeruginosa* PA14			696.3	32	
A. ruhlandii 23696 & *B. multivorans*			852.3	7.5	
A. ruhlandii 23696 & *P. aeruginosa* PA14			751.3	12.2	
A. xylosoxidans 18454 & *B. multivorans*			814	33.6	
A. xylosoxidans 18454 & *C. albicans*	810	34.9			100
A. xylosoxidans 18454 & *P. aeruginosa* PA14	869.7	19.1	720.7	17.8	51.7
A. xylosoxidans 20777 & *B. multivorans*			815.3	18.7	
A. xylosoxidans 20777 & *P. aeruginosa* PA14			693.3	32.7	
A. xylosoxidans 20777 & *S. aureus*	794.7	34.4			100

SI Table 11 Biofilm regression model for clinical *Achromobacter* and CF pathogen pure culture and coculture

Strain	Mean difference	Estimate	SE	t-ratio	P value
Pure culture lower biofilm formers 24h					
A. insuavis 19065 & *C. albicans*	-0.05	-0.2	0.03	-4.8	0.004
A. insuavis 19065 & *P. aeruginosa* LES-B58	-0.03	-0.1	0.03	-4.5	0.01
A. insuavis 23032 & *C. albicans*	-0.05	-0.2	0.03	-5.8	0.0001
A. insuavis 23032 & *P. aeruginosa* LES-B58	-0.04	-0.2	0.03	-5.5	0.0002
A. ruhlandii 23681 & *C. albicans*	-0.05	-0.2	0.03	-5.	0.002
A. ruhlandii 23681 & *P. aeruginosa* LES-B58	-0.03	-0.2	0.03	-4.7	0.005
A. ruhlandii 23696 & *C. albicans*	-0.05	-0.2	0.03	-5	0.002
A. ruhlandii 23696 & *P. aeruginosa* LES-B58	-0.03	-0.2	0.03	-4.6	0.008
Pure culture high biofilm formers 24h					
A. xylosoxidans 18454 & *P. aeruginosa* PA14	-0.3	-0.3	0.03	-9.3	P<0.0001
A. xylosoxidans 20777 & *P. aeruginosa* PA14	-0.3	-0.3	0.03	-8	P<0.0001
A. xylosoxidans LMG 1863 & *P. aeruginosa* PA14	-0.3	-0.3	0.03	-10	P<0.0001
B. multivorans & *P. aeruginosa* PA14	-0.3	-0.3	0.03	-8.6	P<0.0001
24-h to 48-h significant differences					
A. xylosoxidans 18454 & *A. xylosoxidans* LMG 1863	-0.1	-0.2	0.03	-5.6	0.0002
A. xylosoxidans 18454 & *B. multivorans*	-0.2	-0.2	0.03	-7	P<0.0001

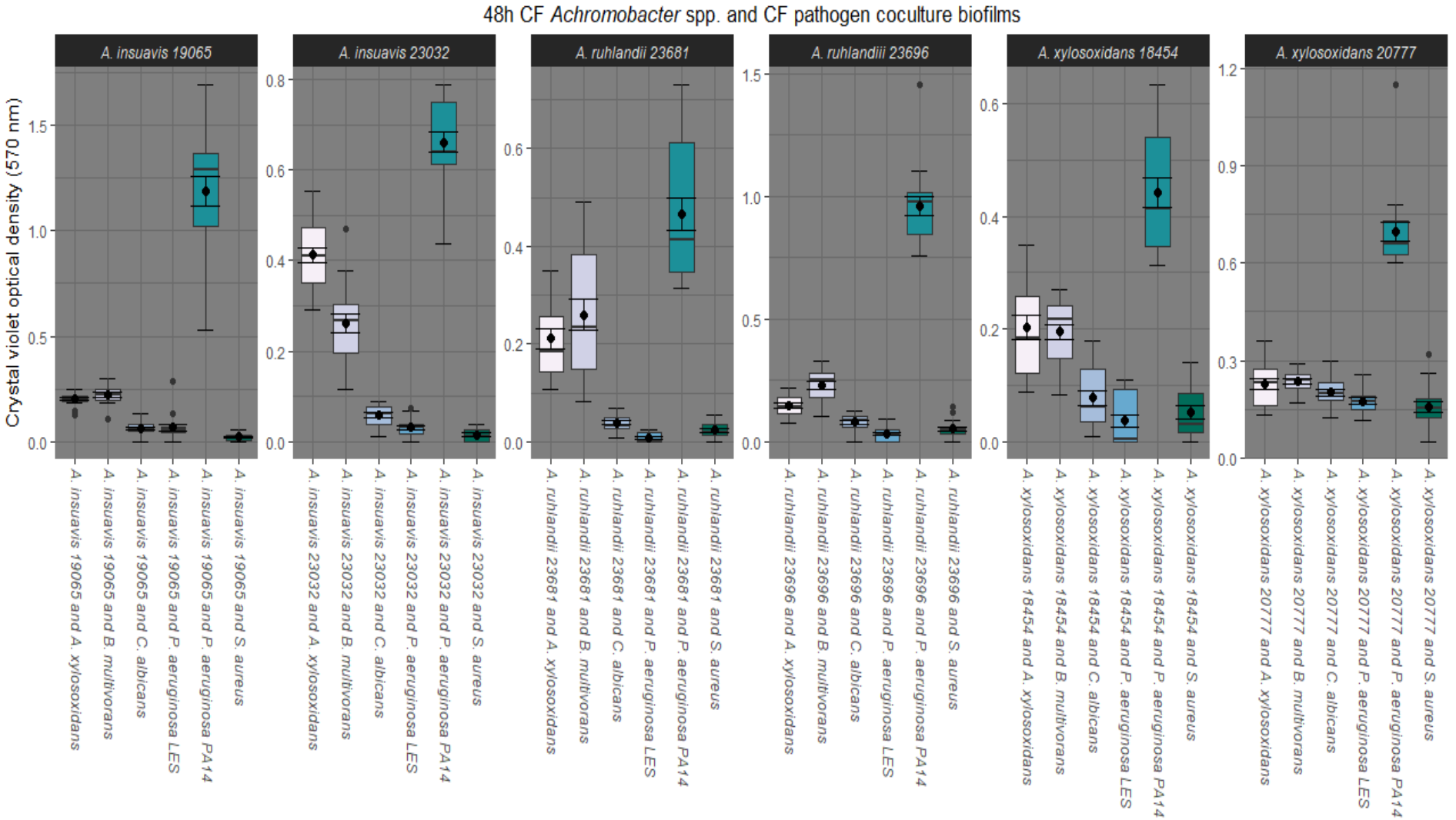


SI Figure 7 RISA PCR amplified ITS regions of mixed pathogen cultures on an abiotic surface. Each gel image was edited to visualise each interaction in order where a DNA size ladder has been included on the left, and the QIAxcel upper marker is highlighted in green in each lane. Each interaction has been labelled where secondary pathogens have been labelled on the outside.

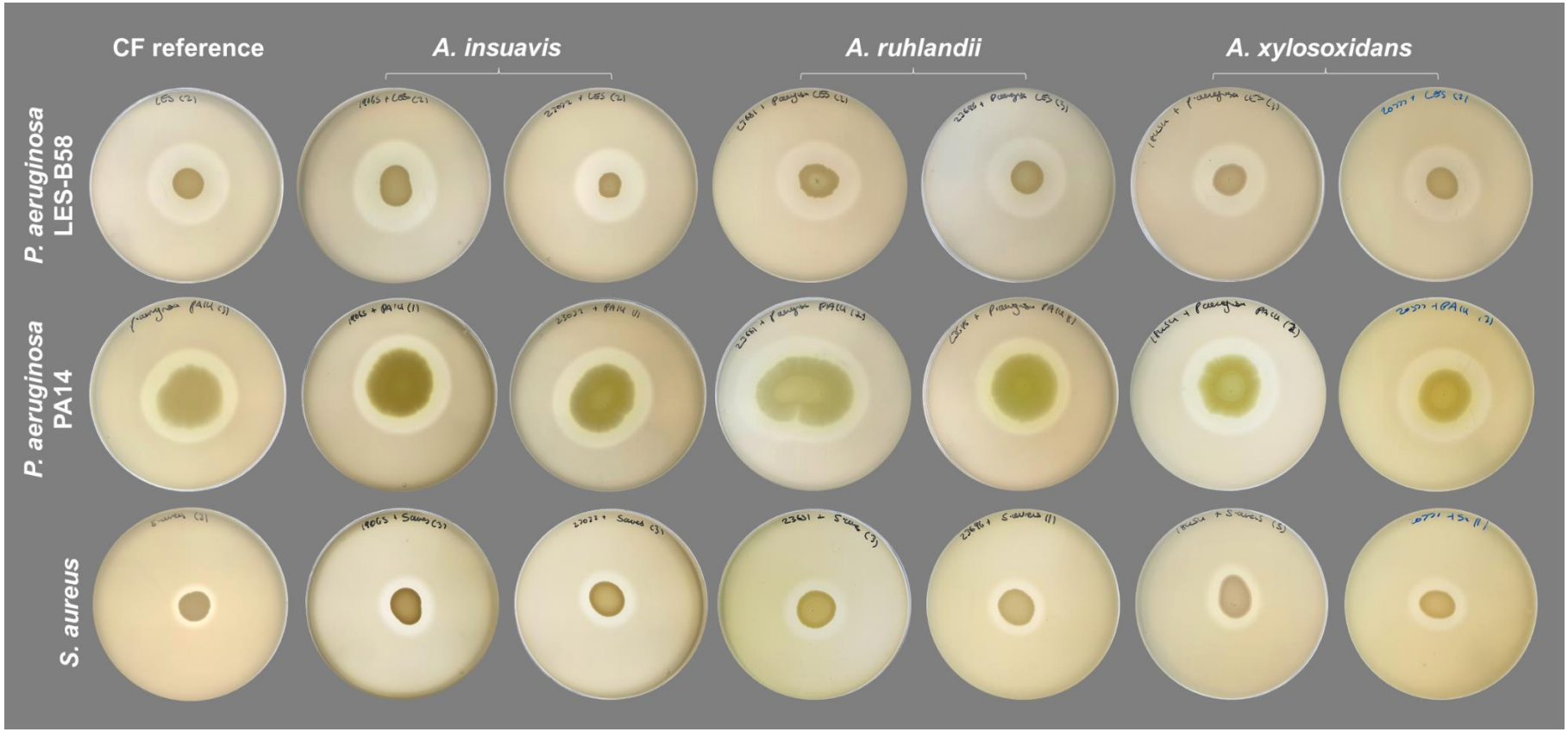
SI Table 12 QIAxcel ITS amplicon size, concentration, and species identification from 1:1 mixed pathogen cultures on abiotic surfaces[a,b]

Interaction	clinical *Achromobacter* isolate		CF reference		% *Achromobacter*
	ITS (bp)	(ng/µl)	ITS (bp)	(ng/µl)	
A. insuavis 19065 & *A. xylosoxidans* LMG 1863			782	27.5	
A. insuavis 19065 & *B. multivorans*			820	9.5	
A. insuavis 19065 & *C. albicans*	885	20			100
A. insuavis 19065 & *P. aeruginosa* LES-B58	884	18.6	695	17.1	52.1
A. insuavis 19065 & *P. aeruginosa* PA14	927	9.4	699	15.3	38.2
A. insuavis 19065 & *P. aeruginosa* PA14 (48-h)	964	1.9	709	12.2	13.4
A. insuavis 23032 & *A. xylosoxidans* LMG 1863			778	34.7	
A. insuavis 23032 & *B. multivorans*			824	9.2	
A. insuavis 23032 & *C. albicans*	862	14.2			100
A. insuavis 23032 & *P. aeruginosa* LES-B58	842	15.2	698	11.9	56.1
A. insuavis 23032 & *P. aeruginosa* PA14	895	8	708	11.7	40.8
A. insuavis 23032 & *P. aeruginosa* PA14 (48-h)	890	6.3	700	12.3	33.8
A. ruhlandii 23681 & *A. xylosoxidans* LMG 1863			778	35.9	
A. ruhlandii 23681 & *B. multivorans*			844	7.7	
A. ruhlandii 23681 & *C. albicans*	807, 1114, 1354	22.7, 5.8, 5.5			100
A. ruhlandii 23681 & *P. aeruginosa* LES-B58	826, 1121, 1384	15.3, 3.7, 3.3	696	15.3	59.3
A. ruhlandii 23681 & *P. aeruginosa* PA14	867, 1143, 1379	10.1, 2.4, 2.1	707	15	49.5
A. ruhlandii 23681 & *P. aeruginosa* PA14 (48-h)	881, 1143, 1385	5.5, 1.4, 1	709	14.4	35.4
A. ruhlandii 23696 & *A. xylosoxidans* LMG 1863			775	33.4	
A. ruhlandii 23696 & *B. multivorans*	*1178, 1391	*0.8, 2.3	834	10.1	
A. ruhlandii 23696 & *C. albicans*	811, 1116, 1354	21.1, 5.8, 5.5			100
A. ruhlandii 23696 & *P. aeruginosa* LES-B58	837, 1126, 1365	14.8, 3.9, 4.2	706	12.7	64.2
A. ruhlandii 23696 & *P. aeruginosa* PA14	843, 1124, 1370	11.7, 3.1, 2.6	697	15.8	52.4
A. ruhlandii 23696 & *P. aeruginosa* PA14 (48-h)	876, 1118, 1371	5.5, 1.8, 1.2	696	17.8	32.3
A. xylosoxidans 18454 & *B. multivorans*			830	22.4	
A. xylosoxidans 18454 & *C. albicans*	780	37.7			100
A. xylosoxidans 18454 & *P. aeruginosa* LES-B58	796	30.5	698	9	77.2
A. xylosoxidans 18454 & *P. aeruginosa* PA14	819	22.3	698	12.2	64.6
A. xylosoxidans 18454 & *P. aeruginosa* PA14 (48-h)	820	20.2	697	11.9	62.9
A. xylosoxidans 20777 & *B. multivorans*			853	10.9	
A. xylosoxidans 20777 & *C. albicans*	780	36.7			100
A. xylosoxidans 20777 & *P. aeruginosa* LES-B58	807	21.2	701	7.4	74.1
A. xylosoxidans 20777 & *P. aeruginosa* PA14	841	14.7	702	9.2	61.4
A. xylosoxidans 20777 & *P. aeruginosa* PA14 (48-h)	820	17.6	697	10.3	63.2

[a] An asterisk denotes a value not being present in every replicate, therefore ratios could not be calculated.
[b] *S. aureus* did not extract with *A. ruhlandii* or *A. insuavis* and were dropped from analysis.



SI Figure 8 48-h mixed pathogen biofilm OD measurements. Standard error bars have been included around the centre of each plot in addition to standard deviation. Each interaction was coloured according to secondary pathogens which was accomplished in R Color Brewer (Neuwirth 2015). The *A. xylosoxidans* reference refers to *A. xylosoxidans* LMG 1863.



SI Figure 9 Pure and mixed culture bacteria protease production. Images have been edited in order and corrections have been made based on picture brightness and contrast. The *A. xylosoxidans* reference refers to *A. xylosoxidans* LMG 1863.

SI Table 13 Regression model statistics for pure and mixed culture bacteria protease production

Interaction	Estimate	SE	DF	t ratio	P value
23681 & *P. aeruginosa* PA14 vs. *S. aureus* (24-h)	37188.7	7253	21	5.1	0.01
19065 & *P. aeruginosa* LES-B58 (24-h-48-h)	-32782.5	7212	189	-4.5	0.006
19065 & *P. aeruginosa* PA14 (24-h-48-h)	-55584.7	7212	189	-7.7	P<0.0001
23032 & *P. aeruginosa* PA14 (24-h-48-h)	-72391.1	7212	189	-10	P<0.0001
23681 & *P. aeruginosa* LES-B58 (24-h-48-h)	-31400.1	7212	189	-4.4	0.01
23681 & *P. aeruginosa* PA14 (24-h-48-h)	-80441.9	7212	189	-11.2	P<0.0001
23696 & *P. aeruginosa* LES-B58 (24-h-48-h)	-30760.8	7212	189	-4.3	0.02
23696 & *P. aeruginosa* PA14 (24-h-48-h)	-72677.7	7212	189	-10.1	P<0.0001
18454 & *P. aeruginosa* LES-B58 (24-h-48-h)	-33870.7	7212	189	-4.7	0.004
18454 & *P. aeruginosa* PA14 (24-h-48-h)	-50584.4	7212	189	-7	P<0.0001
20777 *& P. aeruginosa* LES-B58 (24-h-48-h)	-30376	7212	189	-4.2	0.02
20777 *& P.* aeruginosa PA14 (24-h-48-h)	-54075.8	7212	189	-7.5	P<0.0001
P. aeruginosa LES-B58 (24-h-48-h)	-29018	7212	189	-4	0.04
P. aeruginosa PA14 (24-h-48-h)	-36064.6	7212	189	-5	0.001
18454 & *P. aeruginosa* PA14 vs. 18454 *& S. aureus* (48-h)	61070	12137	21	5	0.02
19065 & *P. aeruginosa* PA14 vs. 19065 *& S. aureus* (48-h)	61589.1	12137	21	5.1	0.02
23032 & *P. aeruginosa* LES-B58 vs. 23032 *& P. aeruginosa* PA14 (48-h)	-64129.3	12137	21	-5.3	0.01
23032 & *P. aeruginosa* PA14 vs. 23032 *& S. aureus* (48-h)	81639.4	12137	21	6.7	0.0005
23681 & *P. aeruginosa* LES-B58 vs. 23681 *& P. aeruginosa* PA14 (48-h)	-76051.9	12137	21	-6.3	0.001
23681 & *P. aeruginosa* PA14 vs. 23681 *& S. aureus* (48-h)	101355.3	12137	21	8.4	P<0.0001
23681 & *P. aeruginosa* PA14 vs. *P. aeruginosa* PA14 (48-h)	57552.7	12137	21	4.7	0.03
23696 & *P. aeruginosa* PA14 vs. 23696 *& S. aureus* (48-h)	57552.7	12137	21	4.7	0.03
20777 & *P. aeruginosa* PA14 vs. 20777 *& S. aureus* (48-h)	56545.6	12137	21	4.7	0.04