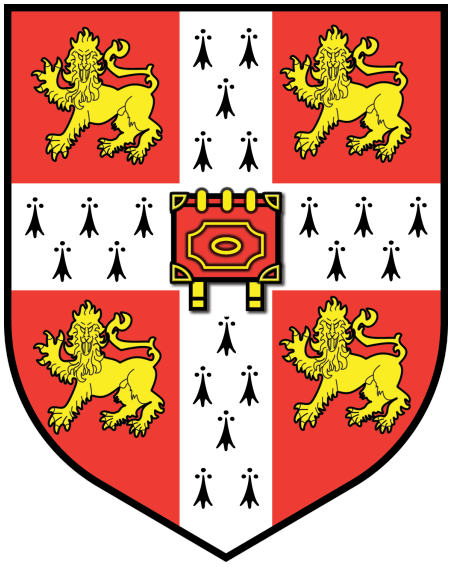
Pseudomonas aeruginosa genetics and virulence in cystic fibrosis and bacteraemia

Samuel E. Kidman

Wellcome Trust Sanger Institute
Clare College, University of Cambridge

August 2019



This dissertation is submitted for the degree of
Doctor of Philosophy

Declaration

This dissertation is the result of my own work, carried out between June 2016 and August 2019 under the supervision of Professor Julian Parkhill, and includes nothing which is the outcome of work done in collaboration except as declared in the Preface.

It is not substantially the same as any that I have submitted, or is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text

It does not exceed the 60,000 word limit set by the Biological Sciences Degree Committee.

Samuel E. Kidman

August 2019

Pseudomonas aeruginosa genetics and virulence in cystic fibrosis and bacteraemia

Samuel E. Kidman

Summary

Pseudomonas aeruginosa is an opportunistic pathogen that can invade and colonise the lungs of people with cystic fibrosis (CF), cause septic shock through bacteraemia infections, and lead to serious infection of burn injuries. It is one of the most critical multi-drug resistant bacteria, and is associated with high morbidity and mortality.

A total of 4,094 *P. aeruginosa* isolates were sampled from nine patients with CF over a six-month time period. These isolates were collected from sputum samples during stable, acute, and recovery timepoints from periods of sudden and rapid lung function decline, called acute pulmonary exacerbations (APEs). These isolates were previously analysed for the presence and absence of ten virulence-related phenotypes.

The *P. aeruginosa* isolates were whole-genome sequenced to investigate the inter- and intra-patient genotypic diversity, associations with phenotypic diversity, and adaptation within the CF lung. Each of the nine patients with CF were colonised with a distinct clone of *P. aeruginosa*. Six patients were infected with well-characterised, highly-transmissible strains of either the Liverpool Epidemic Strain (LES) or the Manchester Epidemic Strain (MES). The remaining three patients were infected with novel sequence types (STs); ST3307 or ST3308. Putative transmission was identified between the two patients infected with ST3307. Two large deletions in genetic regions commonly associated with progression from acute to chronic infection were identified in ST3307.

The acquisition of the LES by one of the patients was very recent, estimated to have occurred within the two years prior to the study. This recent acquisition provides an insight into the immediate adaptation of *P. aeruginosa* to the CF lung, with adaptation observed in genetic regions associated with progression from acute to chronic *P. aeruginosa* infection.

The timepoints for each APE within the individual patients were not associated with variation in the diversity of the populations of isolates. This was confirmed by random distribution of phylogenetic clusters with respect to each APE timepoint for most patients, suggesting that APEs, and the treatment of APEs, do not substantially affect the diversity of the *P. aeruginosa* population within the patient lung.

Genome-wide association studies (GWAS) were carried out on the CF isolates, to investigate any associations with the ten previously-tested virulence-related phenotypes. Population structure could be effectively controlled for in this highly structured dataset, using linear mixed models. Multiple GWAS approaches were required to capture the different classes of genetic variation, resulting in the identification of biologically relevant associations for complex phenotypes, most notably a premature stop-codon in the global transcriptional regulator *rhlR*, as well as several novel, potentially significant associations.

An additional 352 *P. aeruginosa* isolates from patients with bacteraemia were also whole-genome sequenced. These isolates were sourced from both a local collection and from a UK-wide surveillance collection, and broadly match the defined population structure of *P. aeruginosa*. Three STs were overrepresented in this dataset, which are associated with virulence and multi-drug resistance; ST175, ST253 and ST395. One of these overrepresented STs, ST175, was distributed across the UK, shows significant geographical clustering and temporal signal, and is predicted to have been introduced into the UK between the late 1980s and the early 1990s. Antimicrobial resistance profiles showed that current therapeutic options are still viable for most *P. aeruginosa* bacteraemia infections, and that colistin is still effective against the most multi-drug resistant isolates.

Acknowledgements

First and foremost I would like to acknowledge my supervisor, Professor Julian Parkhill, for initially agreeing to take me on as a PhD student, even though I had absolutely no understanding or experience of bioinformatics. Julian has been extremely patient and accommodating throughout the entire process, and I am grateful for the space he has provided to let me grow and develop my skills over the last 3 years. I wouldn't be in the position I am today if it wasn't for my co-supervisor, Dr Martin Welch. It's through a passing conversation with Martin when I first joined the University that I was first introduced to the project I would eventually undertake for my PhD. Throughout the entire process Martin has been incredibly supportive, kept my project grounded, and has shown enthusiasm to take the results in this dissertation further. Upon starting the project, I was introduced to Dr Josephine Bryant, who had the unenviable task of teaching me bioinformatics in the first 18 months of the project. It's through Josie's support and guidance that I gained the basic skills that underpin the knowledge I have today.

This project would not have been possible without the monumental efforts of Emem Ukor. Emem collected and processed the sputum samples, isolated all *P. aeruginosa* used in the cystic fibrosis chapters of this dissertation, and performed all phenotyping assays by hand. As well as a colleague, Emem has been a good friend, and is the kindest person I have met throughout this PhD.

My next acknowledgment is for my wonderful fiancée, Jacqueline Gill. The days and weeks of proofreading means that Jacqueline is the only person that will ever read this dissertation more than once. Without the unwavering support of Jacqueline by my side, the entire process would have been infinitely more difficult, stressful, and way less fun. The countless hours we spent discussing methodologies, results, and improvements have shaped this dissertation into what it is, and this journey we have taken together is one that I will remember forever.

My final acknowledgment goes to my family, especially my Mum and Dad, James, Henry, Casey, and little Mia, who continue to support me in any way they can, no matter what ventures I decide to undertake, and for putting up with me as I tried to explain the contents of this document over the last four years.

Submitting this dissertation draws to a close a challenging, but worthwhile chapter of my life. The technical skills developed throughout this project have equipped me to take the next step in my career with confidence.

Table of Contents

1 Introduction **13**

1.1 Cystic fibrosis **13**

1.1.1 Cystic fibrosis prevalence 13

1.1.2 Cystic fibrosis is caused by mutations in the CFTR gene 14

1.1.3 Classification of CFTR mutations 15

1.1.4 Treatment of CFTR mutations 17

1.1.5 Acute pulmonary exacerbations 19

1.1.6 Bacterial infections and cystic fibrosis 20

1.2 *Pseudomonas aeruginosa* **22**

1.2.1 *P. aeruginosa* and cystic fibrosis 23

1.2.2 Switch from acute to chronic *P. aeruginosa* infection 24

1.3 Phenotypic changes associated with chronic *P. aeruginosa* CF infection **25**

1.3.1 Biofilm formation 25

1.3.2 Secretion systems 26

1.3.3 Swimming and twitching motility 27

1.3.4 Siderophores 28

1.3.5 Rhamnolipids 29

1.3.6 Quorum sensing and virulence 30

1.3.7 Proteases 31

1.3.8 Type VI secretion systems 32

1.3.9 Hypermutation 33

1.4 *P. aeruginosa* and bacteraemia **34**

1.4.1 Virulence of bacteraemia-associated *P. aeruginosa* 34

1.5 Antibiotic-resistant *P. aeruginosa* **35**

1.5.1 Active efflux causes intrinsic resistance to many antibiotics 35

1.5.2 Resistance to β-lactams 37

1.5.3 Resistance to fluoroquinolones 38

1.5.4 Resistance to aminoglycosides 38

1.5.5 Resistance to other classes of antibiotics 39

1.5.6 Plasmid-mediated resistance 40

1.6 *P. aeruginosa* population structure **41**
1.7 Associating phenotype with genotype through genome-wide association studies **42**
1.7.1 Challenges associated with bacterial GWAS 44
1.7.2 Bacterial GWAS methodologies 45
1.8 Aims and objectives **47**
1.8.1 Importance and novelty of this study 50

2 Genomic surveillance of bacteraemia-associated *Pseudomonas aeruginosa* in the UK and Ireland **51**
2.1 Introduction **52**
2.1.1 Aims 53
2.2 Methods **55**
2.2.1 Isolate selection and DNA sequencing 55
2.2.2 MLST assignment 56
2.2.3 Phylogeny and pangenome 56
2.2.4 Investigating overrepresented clades 57
2.2.5 AMR profiling 58
2.3 Results **59**
2.3.1 Isolate selection and DNA sequencing 59
2.3.2 MLST prevalence 60
2.3.3 Population phylogeny 63
2.3.4 Investigating overrepresented clades 67
2.3.4.1 Sequencing statistics 68
2.3.4.2 Genomic islands 70
2.3.4.3 Phage identification 72
2.3.4.4 Virulence factors 74
2.3.4.5 Pairwise SNP distances and transmission 77
2.3.4.6 Temporal analysis 80
2.3.4.7 European context of ST253 and ST175 82
2.3.5 AMR surveillance 85
2.3.5.1 Local collection phenotypic AMR profiles 85
2.3.5.2 BSAC phenotypic AMR profiles 88
2.3.5.3 Phylogenetic correlation with phenotypic AMR 90

2.3.5.4 Genetic mechanisms of AMR 93
2.4 Discussion **96**

3 Genomic diversity of exacerbation-related *Pseudomonas aeruginosa* from nine patients with cystic fibrosis **99**
3.1 Introduction **100**
3.1.1 Aims 102
3.2 Methods **103**
3.2.1 Sample selection 103
3.2.1.1 Patient recruitment 103
3.2.1.2 Sample growth 104
3.2.2 DNA sequencing and quality checks 104
3.2.2.1 Mapping to internal references and specific genes 106
3.2.2.2 Sequence manipulation 106
3.2.3 Phylogeny 106
3.2.4 Genomic clustering 107
3.2.5 Recombination analysis 107
3.2.6 Pangenome analysis 108
3.2.7 Temporal analysis 108
3.2.8 Transmission analysis 108
3.2.9 Phylogenetic clustering 109
3.2.10 *P. aeruginosa* population context 109
3.2.11 Genetic AMR profiling 109
3.3 Results **110**
3.3.1 DNA sequencing and quality checks 110
3.3.1.1 Isolate metadata 110
3.3.2 Genomic clustering 111
3.3.3 Individual patient phylogenies 115
3.3.4 Intrapatient diversity 117
3.3.4.1 Early adaptation to the CF lung 120
3.3.5 Recombination 122
3.3.6 Gene presence-absence within the pangenome 125
3.3.6.1 Accessory genome 127

3.3.7 Context of CF isolates within the *P. aeruginosa* population 132
3.3.8 Patient-to-patient transmission 135
3.3.8.1 Estimating date of infection 138
3.3.9 Phylogenetic clustering with exacerbation state 141
3.3.10 AMR prediction 146
3.3.10.1 Mex efflux systems 146
3.3.10.2 Non-Mex efflux systems 149
3.3.10.3 Additional mechanisms of resistance 149
3.4 Discussion **153**

4 Application of genome-wide association studies to identify associations between virulence-related phenotypes and exacerbation-related *Pseudomonas aeruginosa* genotypes from nine patients with cystic fibrosis **157**
4.1 Introduction **158**
4.1.1 Aims 159
4.2 Methods **160**
4.2.1 Phenotypic assays 160
4.2.1.1 Protease production 160
4.2.1.2 Siderophore production 160
4.2.1.3 Rhamnolipid production 161
4.2.1.4 Motility 161
4.2.1.5 Biofilm 162
4.2.1.6 Quorum sensing molecules 162
4.2.2 Preparing genomes for GWAS 162
4.2.3 Running GWAS on SNPs 163
4.2.4 Running GWAS on gene presence-absence 163
4.2.5 Unitigs 164
4.2.6 Analysing GWAS outputs 164
4.2.6.1 Analysing YfiR variants 164
4.2.7 Estimating phenotypic co-occurrence 165
4.3 Results **166**
4.3.1 Phenotypic clustering within patient groups 166
4.3.1.1 Phenotypic co-occurrence 171

4.3.2 Linear models are unsuitable for GWAS in highly structured datasets 173
4.3.3 Using subsampled groups to control for population structure 175
4.3.4 Aggregating non-synonymous SNPs in subsampled groups 179
4.3.5 Linear mixed models are effective at controlling for population structure in structured datasets 182
4.3.6 Aggregating non-synonymous SNPs in the entire dataset 185
4.3.7 Gene presence-absence in subsampled groups can identify major gene truncations 190
4.3.7.1 Gene presence-absence for the entire dataset 192
4.3.8 Using kmers to identify complex variation 193
4.3.9 Accounting for missing data 194
4.3.10 Investigating global transcriptional regulators 197
4.4 Summary of results and discussion **201**

5 Conclusions **207**
5.1 Restatement of the research aims 207
5.2 Key findings and future directions 209
5.2.1 Bacteraemia-associated *P. aeruginosa* surveillance 209
5.2.2 Diversity of exacerbation-related *P. aeruginosa* 210
5.2.3 GWAS of complex *P. aeruginosa* phenotypes 212
5.3 Concluding remarks **214**

6 References **215**

7 Appendices **263**
7.1 Chapter 3 metadata **263**
7.2 Chapter 4 metadata **264**
7.3 Chapter 5 metadata **265**

Chapter 1

Introduction

1.1 Cystic fibrosis

1.1.1 Cystic fibrosis prevalence

Cystic fibrosis (CF) is an autosomal recessive disease that affects roughly 70,000 people worldwide (Cutting, 2014). It is the most prevalent life-limiting recessive genetic disorder affecting the caucasian population, with an incidence rate of 1 in every 2,500 births in populations of European or American descent (Cohen-Cymberknoh *et al.*, 2011). CF is less common in other demographics, with studies estimating the prevalence of CF in populations of African descent as roughly 1 in 14,000 births (Padoa et al., 1999), and in populations of Asian descent between 1 in 100,000 and 1 in 350,000 (Mirtajani *et al*., 2017).

CF affects roughly 10,000 people in the UK, with life expectancy of males around 48 years of age, and for females around 43 years of age (Keogh *et al.*, 2018). However, UK mortality rates are declining 2% year-on-year, suggesting that those born with CF in 2019 will survive to their mid-60s (Keogh *et al.*, 2018). In the UK, all newborns have been screened for markers of cystic fibrosis through a heel-prick test since 2007, and since 2002 in Scotland (McCormick *et al.*, 2002; Lim *et al.*, 2014). The levels of immunoreactive trypsinogen are measured, and a DNA test is carried out to confirm the presence of CF alleles if the levels of immunoreactive trypsinogen are raised (McCormick *et al.*, 2002).

1.1.2 Cystic fibrosis is caused by mutations in the CFTR gene

CF is caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene found on chromosome 7, which was first identified as the cause of CF in 1989 (Riordan *et al.*, 1989). The CFTR protein is an essential protein for the maintenance of ion homeostasis, and its normal function is to transport chloride across the epithelial cell membrane (Riordan, 2014). If the CFTR gene is ineffective, the chloride ions are not transported across the membrane (Verkman *et al.*, 2003). This draws water away from the extracellular space by osmosis, leading to the dehydration of the extracellular space (Figure 1.1). The CFTR protein is found on the membranes of epithelial cells in several organs within the body, including the lungs, intestines and pancreas (Collawn & Matalon, 2014). Abnormal function of the CFTR

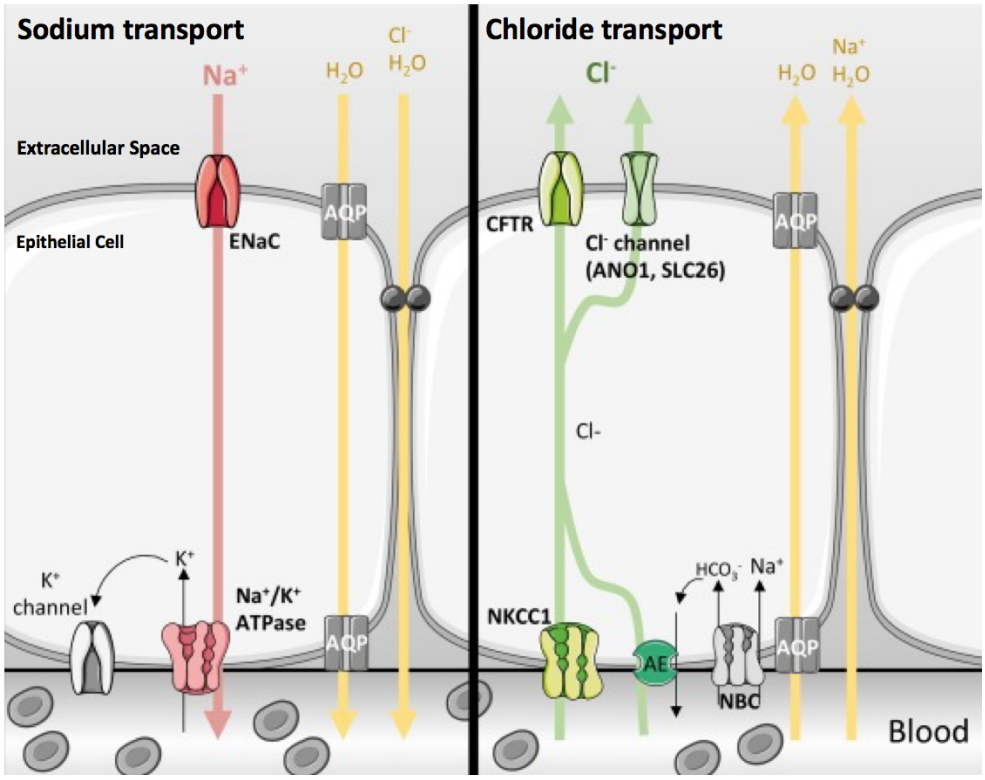


Figure 1.1 In the healthy lung, the CFTR protein transports chloride ions out of the epithelial cells and into the extracellular space, maintaining ion homeostasis. When the CFTR protein does not work effectively, it leads to a build up of chloride ions inside the epithelial cells and draws water out of the lung by osmosis. This dehydration is the cause of the thick and sticky mucus which is characteristic of CF. Image used under Creative Commons v4.0 license from Saint-Criq & Gray, 2016.

protein therefore draws water out of the mucus that lines these organs and into the epithelial cells, leading to the mucous becoming abnormally thick and sticky (Quinton, 1999).

In the gut, an increase in viscosity and a build up of mucus leads to the inability to properly absorb nutrients, which leads to malnutrition and weight-loss (Culhane *et al.*, 2013). The thick and sticky mucus covering the lining of the lungs causes several problems for patients with CF. Lung mucus is a mixture of water and heavily glycosylated proteins, called mucins, which are secreted into the lung in order to trap foreign particles, chemicals, and bacteria (Fahy & Dickey, 2010). In a healthy lung, this mucus is removed through the coordinated movement of ciliated cells, and by coughing, both of which are very effective forms of airway mucosal clearance (Stannard & O'Callaghan, 2006). In the CF lung, the dehydrated mucus cannot be removed through the action of the cilia due to its increased viscosity (Wilson *et al.*, 2019). This leads to a build-up of mucus. Difficulty breathing is a primary symptom experienced by patients with CF, as the mucus build-up constricts the airway (Filkins & O'Toole, 2015). As well as difficulty breathing, the inability to effectively clear the airway leads to an increased risk of infection from the bacteria trapped within the mucus (Filkins & O'Toole, 2015).

Because patients with CF have impaired airway clearance via ciliary action, they have to take a more active role in clearing their own airways. Most CF patients perform airway clearance techniques daily, through a variety of techniques that can be tailored to the patient's physical ability (Sawicki *et al.*, 2009). These airway clearance techniques are designed to help loosen the thickened mucus, to enable easier clearance through coughing. These techniques include breathing exercises and physiotherapy, and can be combined with mucus thinners and airway-clearing apparatus to ensure that as much of the mucus can be cleared as possible (Wilson *et al.*, 2019). Currently, no airway clearance technique stands out as more beneficial than any of the others, and therefore patients with CF have the freedom to choose the most convenient and practical clearance techniques that best suit them and their needs (Wilson *et al.*, 2019).

1.1.3 Classification of CFTR mutations

Over 1,900 variants of the CFTR gene have been identified (Vallieres & Elborn, 2014). To date, these variants have been assigned to seven different classes of CFTR mutation (De Boeck & Amarel, 2016), which vary from more-severe disease (Class I, II, III, and VII), to less-severe disease (Class IV, V, VI) (Table 1.1).

De Boeck and Amaral's classification	Class VII	Class I	Class II	Class III	Class IV	Class V	Class VI
CFTR defect	No mRNA	No protein	No traffic	Impaired gating	Decreased conductance	Less protein	Less stable
Mutation examples	Dele2,3(21 kb), 1717-1G→A	Gly542X, Trp1282X	Phe508del, Asn1303Lys, Ala561Glu	Gly551Asp, Ser549Arg, Gly1349Asp	Arg117His, Arg334Trp, Ala455Glu	3272-26A→G, 3849+10 kg C→T	c. 120del123, rPhe580del
Clinical features (global aspect)	More-severe disease				Less-severe disease		

Table 1.1 The classes of CFTR mutations. Table adapted with permission from Marson *et al.*, 2016.

Class I mutations are classified by the presence of premature stop codons and nonsense mutations within the CFTR gene, which leads to an absence of effective protein in the cell (Brodlie *et al.*, 2015). Class VII mutations also lead to a complete absence of the CFTR protein, instead through an absence of mRNA transcription induced by mutations in promoter regions (Marson *et al.*, 2016).

The most common disease-causing variant is a three base-pair deletion, which corresponds to phenylalanine at position 508 (ΔF508) in both copies of the CFTR gene (Lukacs & Verkman, 2012). The ΔF508 mutation is the most prevalent mutation in the European caucasian CF population, and is a Class II mutation. Class II mutations are classified by the correct synthesis of a protein, but lack of expression at the cell surface (Brodlie *et al.*, 2015). In the case of ΔF508, the mutant CFTR protein cannot maintain a stable 3D shape, leading to the intracellular degradation of the protein before it reaches the cell membrane (Mauiri *et al.*, 2017).

Class III mutations lead to expression of a protein at the cellular membrane that is ineffective at transporting ions. For example, the Class III G551D mutation prevents ATP from opening the channel and allowing effective chloride transport. G551D is the second-most prevalent mutation in the European caucasian CF population (Brodlie *et al.*, 2015).

Class IV to VI mutations lead to less severe CF disease (Marson *et al.*, 2016). Class IV mutations lead to functional, but partial channel conductance. These are caused by mutations that partially block the transmembrane pore, so that traffic through the CFTR protein is considerably reduced (Veit *et al.*, 2016). Class V mutations lead to low levels of protein expression due to altered mRNA splicing, estimated at around 4% of healthy CFTR expression, leading to a more mild form of the disease (Highsmith *et al.*, 1997). Finally, Class VI mutations lead to a high turnover of protein due to CFTR instability (Marson *et*

al., 2016). This is typically caused by mutations towards the C-terminus, which lead to correct protein synthesis, expression, and function, but leads to degradation rates that are 5-6x greater than normal (Haardt *et al*., 1999). These Class IV-VI mutations are less severe, and lead to a milder form of CF disease that may remain undiagnosed until adulthood (Schram, 2012).

1.1.4 Treatment of CFTR mutations

In 2011/12, a new class of drug molecules, known as "potentiators", were approved by the FDA and EMA, to target and improve defective CFTR function (Chaudary, 2018). The aim of the new drug class is to treat the underlying cause of CF, rather than the symptoms (Fohner *et al.*, 2017). Ivacaftor was the first of these molecules approved to treat eleven different Class III CFTR mutations, including G551D (Ramsey *et al.*, 2011; Fohner *et al.*, 2017). Ivacaftor aims to open the defective channel and allow the free movement of chloride ions, restoring effective ion transport (Chaudary, 2018) (Figure 1.2). The exact mechanism is unknown, but Ivacaftor is believed to stabilise the open state of the chloride channel and decrease the corresponding rate of channel closure (Fohner *et al.*, 2017). Ivacaftor clinical trials indicated an increase in lung function of 10.2%, compared to a decline of 0.2% of the placebo group, and a 55% decrease in the frequency of hospitalisation across the study period (Ramsey *et al.*, 2011). Follow up studies have also suggested that the rate of lung function decline is decreased by 50% (Sala & Jain, 2018).

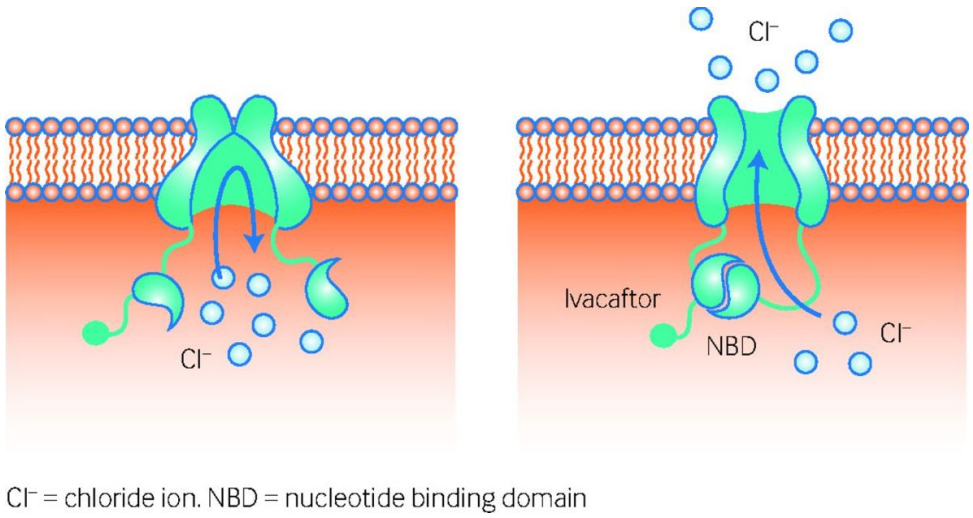


Figure 1.2 Ivacaftor stabilises the open state of the chloride channel, reversing the effects of the G551D mutation. Image adapted with permission from Barry *et al.*, 2018.

Drugs that treat Class II mutations, such as ΔF508, are more profitable, as Class II mutations are the most prevalent mutations within the CF population. However, correcting the error in protein-synthesis requires an initial step to aid the folding process of the CFTR protein within the cell before the protein reaches the cell surface membrane. Vertex Pharmaceuticals, the company behind Ivacaftor, developed another molecule, Lumacaftor, which can be used in conjunction with Ivacaftor to treat the ΔF508 mutation (Mayer, 2016). Lumacaftor is a member of the corrector drug class, which stabilises the interactions between two protein domains that are not formed when the ΔF508 mutation is present. This prevents the degradation of the CFTR protein, allowing its expression at the cell membrane (Deeks, 2016). Ivacaftor then maintains the CFTR ion channel in the open state (Deeks, 2016). Two clinical trials showed modest gains of 2.8-3.3% increase in lung function compared to the placebo group, and a 10% decrease in the frequency of hospitalisations across the study period (Mayer, 2016). Follow up studies suggest a decrease in the rate of decline in lung function over time of 40% (Sala & Jain, 2018).

Vertex Pharmaceuticals have recently developed a replacement drug for Lumacaftor, called Tezacaftor, which has been shown to improve lung function by 4.6-6.8%, and decrease the frequency of hospitalisation by 35%, which is over twice the improvement seen in clinical trials for Lumacaftor (Sala & Jain, 2018).

Two triple combination therapies to treat ΔF508 patients have completed Phase 2 clinical trials. One of these triple combination therapies involves both Tezacaftor as a corrector and Ivacaftor as a potentiator, plus VX-445, a new compound that works as an additional corrector. This therapy indicated an improvement of lung function of 11.0-13.3% (Sala & Jain, 2018). As well as a greater increase in lung function, this new triple combination therapy is associated with significantly reduced adverse effects compared to both Lumacaftor/Ivacaftor and Tezacaftor/Ivacaftor (Sala & Jain, 2018). A press release in March 2019 by Vertex Pharmaceuticals indicated the completion of Phase 3 clinical trials, showing an increase in lung function of 10.0-13.8% compared to a placebo group. Vertex Pharmaceuticals plans to register this combination therapy at the end of 2019 (Vertex Pharmaceuticals, 2019).

In March 2019, Proteostasis Therapeutics released a press statement citing successful Phase 1 clinical trials of a new triple combination therapy (Proteostasis Therapeutics, 2019), which is made up of three novel compounds; PTI-801, PTI-808 and PTI-428. PTI-801 is the corrector molecule, PTI-808 is a potentiator molecule, and PTI-428 is from a novel class of therapeutics called amplifier molecules, which increase the amount of CFTR protein in cells and tissues (Molinski *et al.*, 2017). The results of the 14-day

Phase 1 clinical trial were extremely positive, indicating no plateau for lung function improvement (Proteostasis Therapeutics, 2019). The results of the Phase 2 clinical trials have not yet been made publically available.

1.1.5 Acute pulmonary exacerbations

CF is a chronic disease that progresses over time towards greater lung mucosal blockage, lung damage, and bacterial infection (Pittman & Ferkol, 2015). People with CF can undergo periods of sudden and rapid worsening symptoms, termed "acute pulmonary exacerbations" (APEs), which account for much of the morbidity and mortality associated with CF (Bhatt, 2013). These APEs are associated with increased lung inflammation and damage, and requires aggressive antibiotic treatment. After each APE, symptoms improve, but lung function never quite returns to original baseline values (Hoffman, 2013) (Figure 1.3). There is currently no consensus on what clinical markers should be used to define an exacerbation, but the symptoms typically include a worsening cough, chest pain, shortness of breath, weight loss, and increased lung function decline (Cogen *et al.*, 2017).

A study of 241 patients in the USA observed that 88% of patients with CF experience at least one APE every year (Rubin *et al.*, 2017), and that they are one of the main causes of respiratory failure that eventually leads to patient death (Martin *et al.*, 2016). However, there is currently no recognised optimal treatment for APEs (Schechter *et al*., 2017). APEs are typically treated with antibiotics, usually by

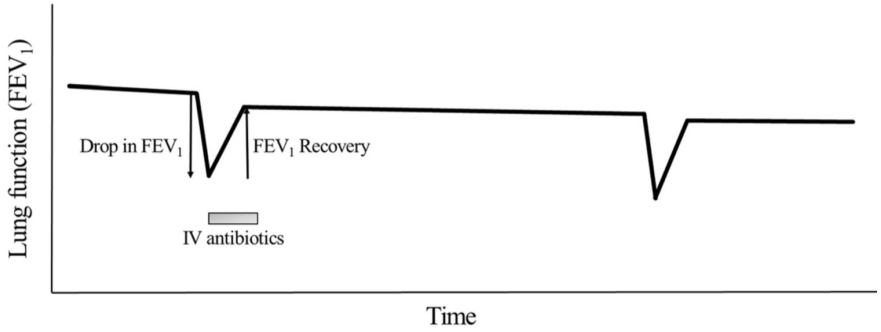


Figure 1.3 Periods of APE are associated with a drop in lung function. Lung function never quite returns to baseline values after an APE. Figure used under Creative Commons v4.0 license, from Espel *et al.*, 2017.

intravenous single or combination therapy for between 7 and 10 days (Bhatt, 2013). This is a significant burden to the patient as this typically requires a hospital stay of around 2 weeks (Cogen & Rosenfeld, 2017). A 2017 study of 2,700 CF patients undergoing APEs in the USA found a positive correlation between inpatient hospital stay, and recovery to 90% of baseline lung function, whereas a negative correlation was observed with home treatment (Schechter *et al*., 2017). This suggests that aggressive hospital treatment with antibiotics is the best treatment course for a patient.

The cause of APEs is not currently understood (Bhatt, 2013). It has been suggested that viruses may play a part in triggering APEs, however, viruses are not commonly isolated from sputum samples collected during an APE (Whelan & Surette, 2015). APEs usually respond to the administration of antibiotics, suggesting that their cause may be related to the bacterial community within the CF lung. However, a change in bacterial load was not found to be associated with the onset of an exacerbation (Stressmann *et al.*, 2011), and the diversity of bacteria within the CF lung remains relatively constant over the course of an exacerbation (Hoffman, 2013; Cuthbertson *et al.*, 2016). Antibiotic therapies have been found to be effective even in patients with bacteria-negative APE samples, suggesting that if the cause of an exacerbation is bacterial, the mechanisms are complex (Hoffman, 2013).

1.1.6 Bacterial infections and cystic fibrosis

In healthy lungs, the mucus lining provides the first line of the host innate defence against bacterial infection (Bals & Hiemstra, 2004). The components of mucus, called mucins, provide a plentiful supply of carbon and nitrogen that can be broken down into free amino acids, sugars, and short-chain fatty acids for bacterial growth (Flynn *et al.*, 2016). The impaired clearance of lung mucus in CF is advantageous for bacteria, which consequently colonise the lungs of patients with CF. Recent developments using culture-independent techniques to characterise the bacterial communities of the lung, such as deep sequencing, have shown that the microbial community within the CF lung is complex and polymicrobial (Filkins & O'Toole, 2015). Between 50 and 200 individual taxa have been identified in a single sputum sample from patients with CF (Filkins & O'Toole, 2015), however this is less diverse than the airway microbiome of healthy individuals (Hery-Arnaud *et al.*, 2019). Fungi and viruses may also coexist with bacteria within the CF lung, and it has been speculated that these may cause more severe symptoms during periods of APE (Etherington *et al.*, 2014; Willger *et al.*, 2014).

The reduction in microbial diversity and microbial burden caused by antibiotic treatment for bacterial lung infections is short-lived. Within 8-10 days after receiving antibiotic treatment, the bacterial load and composition return to pre-treatment levels (Smith *et al.*, 2014). Microbial diversity within the CF lung decreases with progression of the disease, possibly as a consequence of more aggressive antibiotic therapy, or with the introduction of environmental and highly-virulent bacteria, such as *P. aeruginosa*, as these organisms adapt and out-compete other species within the lung (Zhao *et al.*, 2012).

Studies on the lung microbiota have typically focussed on major bacterial species that are thought to contribute the most to morbidity and mortality within the CF lung (Harrison, 2007). Over time, the prevalence of the major CF-associated bacteria within the CF lung changes (Figure 1.4). Many bacteria that colonise the lung early in the progression of CF are those that typically colonise the nose or upper airway in healthy people, such as *Haemophilus influenzae* or *Staphylococcus aureus*. Over time, patients with CF are exposed to other bacteria from the environment, or through contact with other CF patients in treatment centres or support groups (Lyczak *et al.*, 2002). This causes the composition of bacteria within the lung to change. Probability of colonisation with *P. aeruginosa* increases as a patient ages, and is associated with a worse prognosis, more frequent APEs, and general lung function decline (Crull *et al.*, 2016; Harrison, 2007).

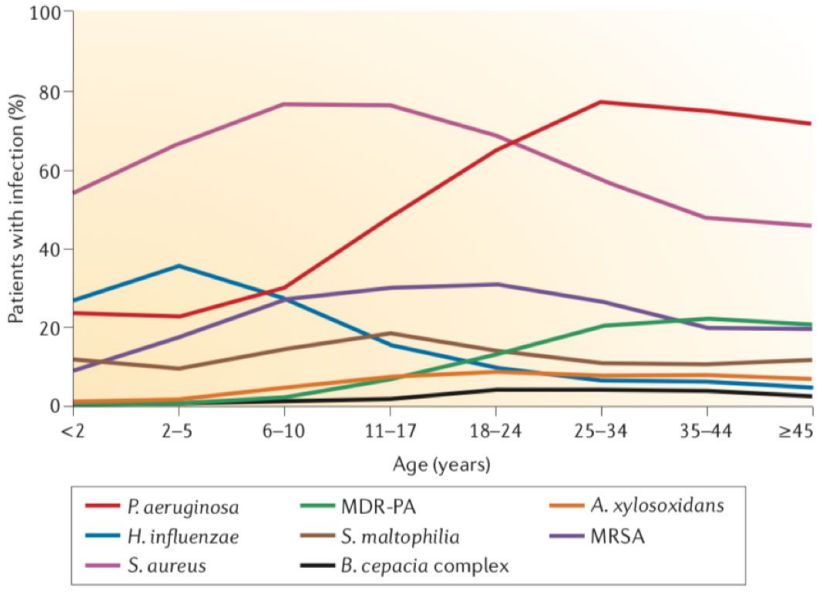


Figure 1.4 The prevalence of the major CF-associated pathogens within the lung, plotted against the age of the patient. MDR-PA = multidrug resistant *P. aeruginosa*, MRSA = methicillin resistance *S. aureus* Figure adapted with permission from Folkesson *et al.*, 2012.

Bacterial infection within the CF lung is the main driver of lung inflammation and damage (Downey *et al.*, 2008). Specifically, the severe and sustained immune response to lung infection is responsible for the majority of the destruction to the lung tissue (Downey *et al.*, 2008). The initial bacterial infection causes the migration of large numbers of neutrophils into the alveolar space, where the release of proteases, elastases and inflammatory cytokines contributes to the damage of the lung tissue (Laval *et al.*, 2016). This immune response recruits monocytes and macrophages, which contribute further to lung inflammation through the release of further inflammatory cytokines, exacerbating the tissue damage (Bruscia & Bonfield, 2016). The lungs of patients with CF are constantly colonised with bacteria. Therefore the inflammatory response, and associated lung damage, is occurring throughout the patient's life, contributing to lung function decline and the progression of CF disease (Lin & Kazmierczak, 2017).

1.2 *Pseudomonas aeruginosa*

Pseudomonas aeruginosa are Gram-negative, aerobic bacteria, belonging to the family *Pseudomonadaceae*. Members of the *Pseudomonas* genus demonstrate a great deal of metabolic diversity, and are able to thrive in a wide range of niches, including soil, marshland, coastal marine habitats, and plant and animal tissues (Stover *et al*., 2000). *P. aeruginosa* bacteria are ubiquitous in the environment and are usually harmless, but they can act as opportunistic pathogens, causing a wide array of acute or chronic infections in immunocompromised or other at-risk patient groups (Moradali *et al*., 2017).

P. aeruginosa genomes are large and vary in size (5.5-7 Mbp) and gene content (Mathee *et al.*, 2008). Their large and varied genome reflects the ability of *P. aeruginosa* to colonise many environmental settings, encoding genes relevant for survival in many niches (Klockgether *et al.*, 2011). Accessory genes have been historically acquired through mobile elements, such as plasmids, phage, and integrating conjugative elements, and aid in the adaptation of *P. aeruginosa* to their specific niche (Kidd *et al.*, 2012). The *P. aeruginosa* genome also shows genomic plasticity within each niche, with different isolates within the same niche able to acquire novel genetic content, or undergo deletion events, to differentiate themselves from other isolates (Klockgether *et al.*, 2011). The *P. aeruginosa* genome contains 690 regulatory genes (~12% of the genome), which is the third largest percentage of regulatory gene content of all bacterial species, behind *Escherichia coli* and *Bacillus subtilis* (Galan-Vasquez *et al.*, 2011).

1.2.1 *P. aeruginosa* and cystic fibrosis

P. aeruginosa is a very important clinical bacteria, and is one of the most important pathogens to consider in CF lung infection, as the inflammatory response to *P. aeruginosa* infection is the predominant cause of morbidity and mortality in CF patients (Bhagirath *et al.*, 2016). The prevalence of chronic *P. aeruginosa* in the CF population has been estimated between 50 and 80% (Crull *et al.*, 2016), and is strongly associated with more severe symptoms (Vongthilath *et al.*, 2019). Once a patient has acquired a *P. aeruginosa* strain, that strain will adapt to the CF lung and typically remain until the end of that patient's life (Stover *et al.*, 2000).

How *P. aeruginosa* infection is acquired is still relatively unknown. There have been many documented cases of direct transmission between patients as a result of prolonged socialisation, such as at CF rehabilitation centres, holiday parks, or close friends and family (Schelstraete *et al.*, 2008). However, this forms the minority of infection cases, with the majority of newly-acquired *P. aeruginosa* infection thought to have been picked up from the environment (Psoter *et al.*, 2014; Schelstraete *et al.*, 2008).

Once *P. aeruginosa* gets into the CF lung, it is faced with a hostile and stressful environment that it must overcome. Stressors include osmotic stress from the thick mucus, oxidative stress, host immune response, competition from other bacteria, and antibiotic treatment (Hector *et al.*, 2014; Andersson & Hughes, 2014; Winstanley *et al.*, 2016).

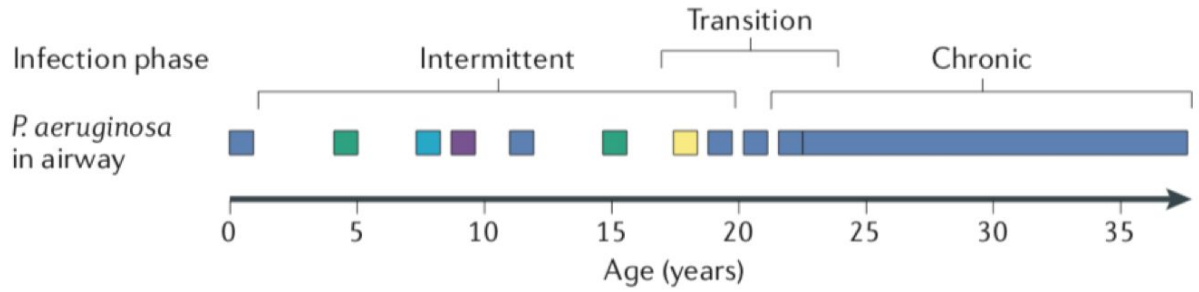


Figure 1.5 *P. aeruginosa* infection can occur from birth. Intermittent infections can be eradicated through aggressive antibiotic treatment, but eventually, the *P. aeruginosa* infection will transition from acute to chronic. Figure adapted with permission from Folkesson *et al.*, 2012.

During childhood, *P. aeruginosa* can intermittently infect the patient's lung (Folkesson *et al.*, 2012). During this period of intermittent infection, treatment with antibiotics can effectively clear the infection and prevent the development of chronic infection (Hoiby *et al.*, 2005). However, the patient often becomes reinfected at a later date, with the same *P. aeruginosa* strain in a quarter of cases (Munck *et al.*, 2001). Eventually, the intermittent colonisation of *P. aeruginosa* of the CF lung transitions into chronic infection (Figure 1.5), which is diagnosed by the continuous sampling of *P. aeruginosa* isolates from sputum, and the development of *P. aeruginosa*-specific immune response markers (Folkesson *et al.*, 2012).

1.2.2 Switch from acute to chronic *P. aeruginosa* infection

The reason why *P. aeruginosa* infection switches from intermittent to chronic is not well understood, but the mechanism by which this switch takes place has been extensively studied. A change in the gene expression profile, controlled by global two-component sensor kinase transcriptional factors, leads to chronic infection (Balasubramanian *et al.*, 2013), through the repression of the transcription factor RsmA. *P. aeruginosa* genomes encode roughly 127 of these two component systems, over twice that encoded within *E. coli*, which reflects the ability of *P. aeruginosa* to adapt to a variety of niches by changing gene expression (Balasubramanian *et al.*, 2013). During acute infection, the global regulator RetS negatively regulates the global regulator GacS, which upregulates RsmA through interactions with GacS, RsmY and RsmZ (Burrowes *et al.*, 2006) (Figure 1.6). The chronic phenotype occurs when the global transcriptional regulator LadS senses calcium in the environment, and acts with the opposite effect to RetS (Broder *et al.*, 2016), activating GacS, which leads to a repression of the RsmA regulator, promoting the downregulation of virulence (Francis *et al.*, 2018; Robledo-Avila *et al.*, 2018) (Figure 1.6).

This switch frequently occurs with a knockout of the *mucA* regulatory gene, enabling the transcription of several genes via an *algU*-encoded sigma-factor (Martin *et al*., 1994). These genes further downregulate expression of genes related to motility and virulence factors, lead to overexpression of stress response pathways, and enhance persistence within the CF lungs (Folkesson *et al.*, 2012). This knockout induces the overproduction of alginate, a secreted exopolysaccharide which forms a protective mucoid matrix around the colonies that provides additional protection from the host immune system and antibiotic treatments (Pritt *et al.*, 2007).

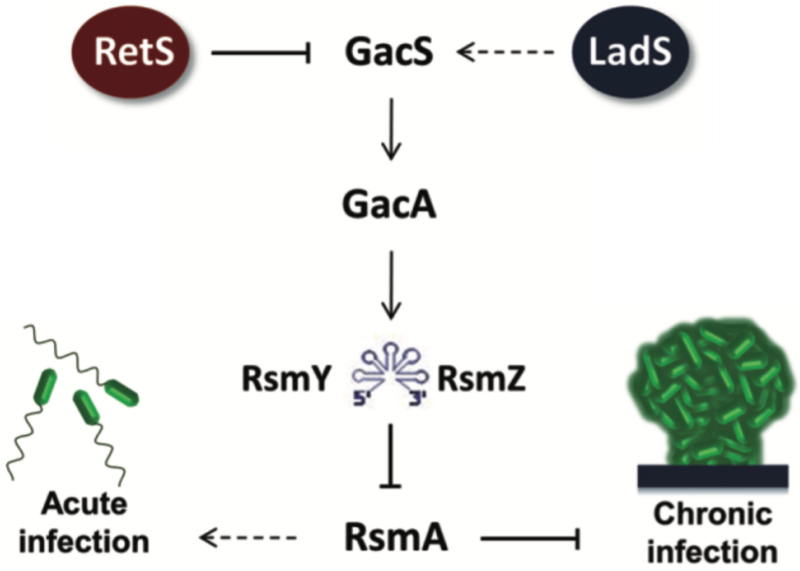


Figure 1.6 The regulatory cascade that controls the switch from acute infection to chronic infection. During acute infection, RetS represses GacS, which enables the upregulation of acute infection by RsmA. During chronic infection, LadS activates GacS, which leads to repression of RsmA, allowing the development of chronic infection. Figure adapted with permission from Moscoso *et al*., 2011.

Once a *P. aeruginosa* infection has become chronic, eradication from the lungs becomes impossible, even when treated with an aggressive regimen of antibiotics (Hurley *et al.*, 2012). Resistance arises not only through the protective mucoid layer, but also through intrinsic resistance to many antibiotics through a variety of mechanisms summarised in section 1.5 (Mesaros *et al.*, 2007). Because of its intrinsic resistance and ability to develop and acquire novel resistance variants, it is recommended that combination therapies are used to manage *P. aeruginosa* infection (Kapoor & Murphy, 2018).

1.3 Phenotypic changes associated with chronic *P. aeruginosa* CF infection

1.3.1 Biofilm formation

During the switch from acute infection to chronic infection, several important phenotypic changes occur. Initially, production of alginate causes the bacteria to develop the mucoid phenotype, forming a protective exopolysaccharide layer around the colony (Folkesson *et al.*, 2012).

The alginate forms the basis of a bacterial biofilm, which in the case of CF, is composed mostly of alginate, other polysaccharides, exoproteins, and extracellular DNA (Hoiby *et al.*, 2010). The molecules abundant in the formation of biofilm include alginate for biofilm stability, pel and psl proteins that form a scaffold for correct biofilm structure, cellulose that creates a strong biofilm, and rhamnolipids for host immune cell toxicity and biofilm dispersion (Mann & Wozniak, 2012; da Silva *et al.*, 2019).

The persistence of biofilm in the CF lung is one of the major causes of inflammation, immune response, and lung damage, and hence is a major cause of CF morbidity and mortality (Hoiby *et al.*, 2010). Lung inflammation can be caused by the interaction of host immune cells with each component of the biofilm without coming into direct contact with the bacteria themselves (Ciszek-Lenda *et al.*, 2019). This causes the immune cells to release pro-inflammatory factors such as TNF-α and Interleukin-6 (Ciszek-Lenda *et al.*, 2019).

1.3.2 Secretion systems

The Type III Secretion System (T3SS) is critical in acute infections, as it facilitates the delivery of toxins into foreign cells, such as the host immune system or other bacteria (Hauser, 2009). CF patients produce antibodies against T3SS effector proteins, suggesting that the T3SS is active during acute infection (Hauser, 2009). However, CF *P. aeruginosa* lose their ability to use the T3SS as chronic infection develops (Hauser, 2009). Downregulation of T3SS is facilitated through three different mechanisms. Firstly, the upregulation of alginate and biofilm production through mutations in *mucA* leads to a downregulation of genes responsible for T3SS (Wu *et al.*, 2004). Secondly, the GacS repression of RsmA, which occurs during the transcriptomic switch from acute to chronic infection, is also known to repress T3SS (Hauser, 2009). The third mechanism is through mutations in the T3SS machinery itself, which inactivate the T3SS or reduce its effectiveness (Smith *et al.*, 2006a).

The effector toxins of the T3SS are ExoS, ExoU, ExoT and ExoY, however, not all *P. aeruginosa* strains encode all four of these effector proteins (Feltman *et al.*, 2001). Whilst the presence of the four effector proteins are variable, different *P. aeruginosa* tend to either encode ExoU or ExoS (Hogardt & Heesemann, 2011). ExoU is a more virulent exotoxin, which initiates rapid host cell lysis that is very effective against host cell immune cells. In contrast, the ExoS exotoxin encourages delayed cell death through apoptosis (Schulert *et al.*, 2003). Maintaining production of T3SS is associated with more

aggressive disease and poor prognosis, and so downregulation of the T3SS during chronic infection is beneficial for maintaining a more symbiotic host-pathogen relationship (Galle *et al.*, 2012).

Other *P. aeruginosa* secretion systems are found to be downregulated during the switch from acute to chronic infection. For example, the T2SS are responsible for the secretion of several proteases, lipases, phosphatases, phospholipases, elastases, siderophores and toxins. Expression of the T2SS machinery is repressed due to RsmA repression during the switch from acute to chronic infection (Filloux, 2011).

The T1SS machinery is responsible for the secretion of alkaline protease, which blocks phagocytosis of host neutrophils (Laarman *et al.*, 2012). The T1SS is downregulated through mutations in the global regulatory gene *lasR* (Balasubramanian *et al.*, 2013), which also regulates the production of other quorum sensing molecules that regulate the expression of various other virulence factors in turn (Pena *et al.*, 2019). Knockout mutations in *lasR* are frequently found in isolates which have switched from acute to chronic infection (LaFayette *et al.*, 2015), and can explain the downregulation of virulence factors during this switch (Hogardt & Heesemann, 2011).

1.3.3 Swimming and twitching motility

Another phenotypic change related to chronic *P. aeruginosa* infection is the loss of certain motility mechanisms. Expression of both the Type IV pili and the bacterial flagella, responsible for twitching motility and swimming motility respectively, are downregulated during the switch from acute to chronic infection (Hogardt & Heesemann, 2011).

Twitching motility is characterised by the extension of long fibres called the Type IV pili, adhesion of the pili to a surface, and retraction of the pili, which pulls the bacteria along the surface (Wall & Kaiser, 1999). Their function is to facilitate the initial attachment of the bacterium to a surface, followed by the exploration of that surface. The Type IV pilus has been shown to be an important virulence factor in *P. aeruginosa* CF infection, as during the acute stage of infection, the pilus is required for anchoring on to the host immune cells, followed by retraction to bring the bacteria into contact with the host cell to allow the delivery of toxins through the T3SS (Comolli *et al.*, 1999). After attachment to a surface, the Type IV pili is critical for the development of biofilms during the switch from acute to chronic infection, as the pili is required for aggregation of *P. aeruginosa* isolates into colonies (Burrows, 2012; Gellatly *et al.*, 2013).

Swimming motility is controlled by a single flagellum, which rotates in a corkscrew motion to allow movement through liquid environments (Taylor & Buckling, 2011; Gellatly *et al.*, 2013). Whilst studies have reported that Type IV pili play a role in surface adhesion, other studies have shown that the flagellum is also necessary for epithelial surface attachment (Tran *et al.*, 2011). Flagella can be recognised by the host immune cells and initiate the inflammatory immune response, and hence the downregulation of these genes are an effective mechanism of protection (Compodonico *et al*., 2009). Therefore, flagella are not required for biofilm formation, and after initial attachment in the lungs, the flagella proteins are often downregulated (Sauer *et al*., 2002). However, once a biofilm has matured, flagella are upregulated again, and are responsible for dispersion of *P. aeruginosa* bacteria from the biofilm (Co *et al.*, 2018).

1.3.4 Siderophores

Siderophores are excreted molecules that have affinity for iron, specifically Fe^{3+}. *P. aeruginosa* has two siderophore systems; the pyoverdine siderophore and the pyochelin siderophore (Figure 1.7), which can both sequester Fe^{3+} from the environment, and are both re-uptaken by siderophore-specific receptors in the bacterial membrane (Cornelis & Dingemans, 2013).

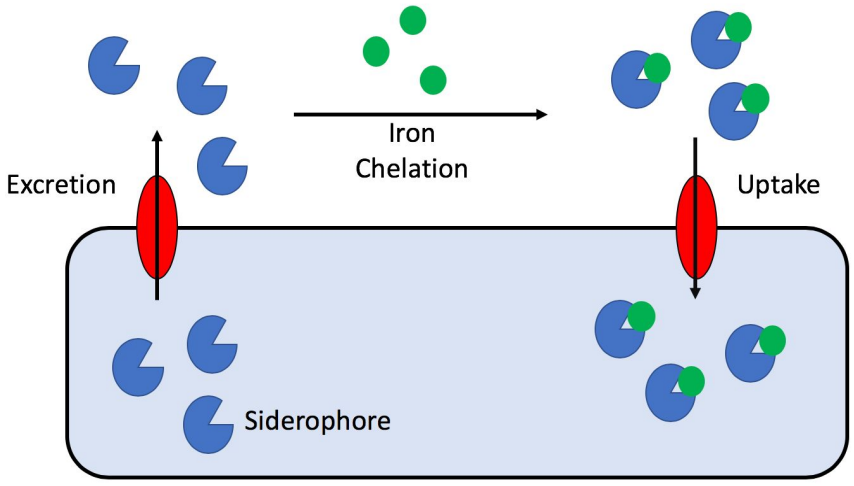


Figure 1.7 Pyochelin and pyoverdine are two siderophores that are excreted by *P. aeruginosa* to scavenge iron from the environment.

In iron-limited conditions, pyoverdine is produced by *P. aeruginosa* preferentially (Cornelis & Dingemans, 2013). Pyoverdine is a high-affinity siderophore (Visca *et al.*, 2007) that is able to displace iron sequestered by other chelating molecules in the lung, such as transferrin (Cornelis & Dingemans, 2013). Pyoverdine also acts as a signalling molecule between *P. aeruginosa*, and pyoverdine uptake initiates a cascade which ends with the production of two virulence factors; protease PrpL and the potent exotoxin ToxA (Beare *et al.*, 2002). The presence of pyoverdine is essential for acute infection, as pyoverdine knockout mutants are rendered completely avirulent (Cornelis & Dingemans, 2013). Pyoverdine has also been identified as important for the development of stable biofilms in acute lung infection (Kang & Kirienko, 2018).

Once chronic infection has established in the CF lung, *P. aeruginosa* isolates adapt to the iron-rich lung, to produce the low-affinity siderophore, pyochelin, to uptake haem and free iron, instead of relying on pyoverdine-sequestered iron (Marvig *et al.*, 2014).

1.3.5 Rhamnolipids

Rhamnolipids are glycolipid biosurfactants produced by *P. aeruginosa* that have a variety of natural functions. Their amphipathic nature facilitates the uptake of hydrophobic substrates, such as short-chain alkanes, and their subsequent degradation for use as a carbon source (Abdel-Mawgoud *et al.*, 2010). Rhamnolipids are considered to be potent virulence factors, especially in the CF lung. At high concentrations, rhamnolipids lyse immune cells, but at lower concentrations they stimulate the release of inflammatory response molecules Interleukin-6 and Interleukin-8 (Bedard *et al.*, 1993). Therefore, rhamnolipids can both protect *P. aeruginosa* from the inflammatory response, and aggravate the inflammatory response, which consequently leads to increased damage to the CF lung (Abdel-Mawgoud *et al.*, 2010).

Rhamnolipids also act directly on epithelial cells by further inhibiting ciliary function, and further preventing mucosal clearance (Zuliano *et al.*, 2006). They also display activity against other microbes in the lung by mimicking amphipathic lipid membrane molecules and causing cell permeabilisation (Sotirova *et al.*, 2008). Secretion of rhamnolipids also facilitates movement across a surface in a motion called swarming, by acting as a surface lubricant (Verstraeten *et al.*, 2008). This lubrication also increases the ability of non-motile cells to stick to the populated surface, which encourages *P. aeruginosa* to

aggregate together, and hence stimulates biofilm production (Pamp & Tolker-Nielsen, 2007). Rhamnolipid production is vital for the formation of normal biofilm, as well as motility-associated dispersal of the biofilm, which allows *P. aeruginosa* to colonise a new niche (Boles *et al.*, 2005).

1.3.6 Quorum sensing and virulence

Quorum sensing (QS) molecules are intercellular communication molecules, which help regulate gene expression based on population density (Wang *et al.*, 2014). *P. aeruginosa* secrete a variety of QS molecules, which bind to specific receptors and induce intracellular signalling cascades that end with altered gene expression. Roughly 10% of the *P. aeruginosa* genome is estimated to be regulated by quorum sensing molecules, impacting a range of virulence factors including biofilm formation, rhamnolipid production, toxin production, siderophore production, and motility (Williams & Camara, 2009; Antunes *et al.*, 2010; Hawver *et al.*, 2016).

There are four well-characterised QS mechanisms in *P. aeruginosa*; the *las*, *rhl*, *iqs* and *pqs* systems, which are all hierarchically interconnected (Lee & Zhang, 2015). At the top of the hierarchy (Figure 1.8) is the LasR regulator, which forms a complex with the QS molecule OdDHL (Latifi *et al.*, 1996). This

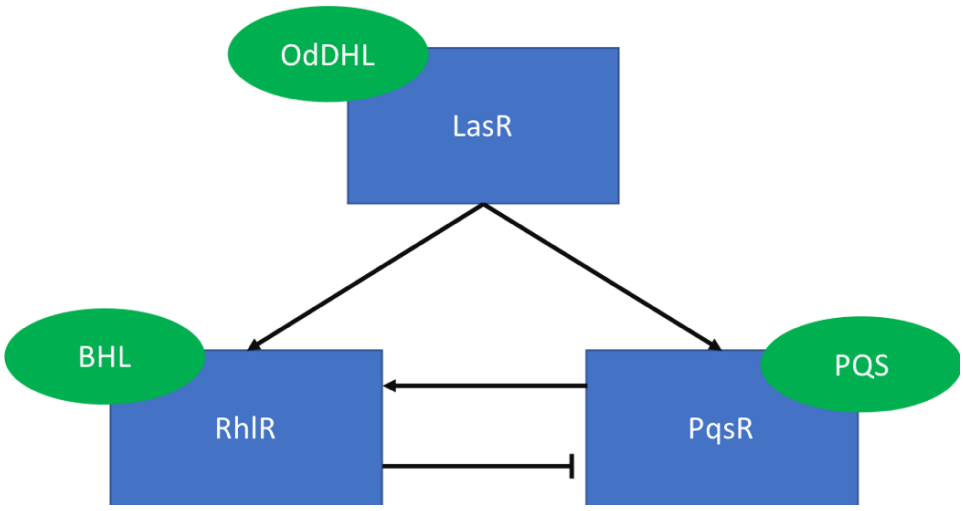


Figure 1.8 The quorum sensing network in *P. aeruginosa* is hierarchical and interconnected. Blue boxes represent the regulatory protein that the QS signalling molecules, represented as green ovals, activate.

activates the expression of 300 different genes within the *P. aeruginosa* genome, including various virulence factors such as T1SS, T2SS, ToxA exotoxin, alginate, and bacterial motility (Gilbert *et al.*, 2009). This complex also activates the *lasI* gene, which leads to the biosynthesis and secretion of the OdDHL molecule, acting as a positive feedback loop (Wargo & Hogan, 2007).

The LasR-OdDHL complex also activates the *rhl* QS system by activating transcription of *rhlR*. The *rhl* QS system is also activated by the QS molecule BHL (Boursier *et al.*, 2018). This activation leads to the expression of several virulence factors, which include the T2SS, siderophores, rhamnolipids, and the potent virulence factor hydrogen cyanide (Brint & Ohman, 1995). The activation of *rhlR* leads to the biosynthesis and secretion of the BHL QS molecule, which provides positive feedback to the *P. aeruginosa* community (de Kievit *et al.*, 2002).

The LasR-OdDHL complex also activates the PQS mechanism, by positively regulating *pqsR* (Lee & Zhang, 2015). PqsR forms a complex with the PQS QS molecule, which increases the expression of siderophores and hydrogen cyanide, and is involved in protection from environmental stressors, such as radical oxygen species and ultraviolet light (Lee & Zhang, 2015; Haussler & Becker, 2008). The activation of *pqsR* also provides a positive feedback loop, as the *pqsABCD* operon is activated to produce the PQS QS molecule (Deziel *et al.*, 2004). This negatively regulates the pqs system. The *pqsR*-PQS complex activates the *rhl* QS system, but the *rhlR*-BHL complex negatively regulated the *pqs* QS system. Therefore, the proportion of all QS molecules tightly controls the activation of each QS system, and hence the relative ratio of expression of each of the virulence factors under their control (Cao *et al.*, 2001).

1.3.7 Proteases

Around 3% of the *P. aeruginosa* genome encodes potent virulence factors called proteases (Galdino *et al*., 2017). Proteases are secreted enzymes with the ability to break down protein and peptide structures, and are considered essential for bacterial infection and pathogenicity (Culp & Wright, 2017).

Caseinolytic proteases are conserved proteins in most bacterial species, and are characterised by their ability to degrade casein; a phosphoprotein that forms 80% of cow's milk (Brown & Foster, 1970). Caseinase enzymes have many functions, including intracellular protein degradation, transcriptional

regulation, stress response, and virulence (Gersch *et al*., 2012). Several secreted proteases have caseinolytic activity in *P. aeruginosa*, including protease IV, LasB protease, and the Apr alkaline phosphatase (Caballero *et al*., 2001).

The *P. aeruginosa* genome also encodes gelatinolytic proteases, which are able to digest gelatin, a hydrolysed form of collagen. The LasA protease, LasB protease, protease IV, and Apr alkaline phosphatase have been characterised as gelatinolytic enzymes (Tange *et al*., 2009; Caballero *et al*., 2001). These proteins have been shown to degrade both host immune proteins and host surface proteins, and play an important role in facilitating bacterial surface adhesion (Smith *et al.*, 2006b).

1.3.8 Type VI secretion systems

The Type VI Secretion System (T6SS) is responsible for delivering effector proteins directly into other competing bacterial cells to aid *P. aeruginosa* survival within the current niche (Broms *et al*., 2012). Within the *P. aeruginosa* PAO1 genome, there are three evolutionarily distinct T6SS, suggesting that their acquisition occurred historically by horizontal gene transfer (Filloux *et al*., 2008). *P. aeruginosa* express immunity proteins to these three toxins, to prevent any toxic effects on itself or through accidental delivery from sister cells (Sana *et al.*, 2016).

The first T6SS, H1-T6SS, delivers up to seven different effectors to a target cell; effector proteins Tse1 to Tse7. Tse1 and Tse3 target distinct chemical bonds during peptidoglycan degradation of the target cell, which lead to cell death through the dissolution of the cell envelope (Benz *et al*., 2012; Lu *et al.*, 2014). The effector protein Tse2 is a reversible ADP-ribosylating enzyme, which posttranslationally modifies target proteins in the cell cytoplasm, inhibiting bacterial growth (Robb *et al*., 2016). This is not lethal to the target bacteria, but confers a major fitness advantage to *P. aeruginosa*, allowing it to outgrow the competition (Chen *et al*., 2015). The mechanism of action of both Tse4 and Tse5 is unknown, but they have been shown to have a bacteriostatic effect on populations of *E. coli* (Whitney *et al*., 2014). Tse6 has bacteriostatic effects by degrading essential compounds NAD^+ and $NADP^+$, preventing metabolic and catabolic processes within the target cell (Whitney *et al.*, 2015). The Tse6 crystal structure was elucidated in 2015, which showed similarity with the diphtheria toxin (Whitney *et al*., 2015). Tse7 was identified in 2018, and has been shown to have bacteriostatic activity through degradation of DNA (Pissaridou *et al.*, 2018).

The H2-T6SS and H3-T6SS systems deliver phospholipase enzymes PldA and PldB respectively (Sana *et al*., 2016). These allow *P. aeruginosa* to compete with other bacteria by degrading the target cell's membrane through enzymatic cleavage of phosphatidylethanolamine (Sana *et al.*, 2016). *P. aeruginosa* also use these T6SS to deliver these effectors into epithelial cells, disrupting cell membranes and allowing internalisation of the bacteria (Sana *et al*., 2016).

T6SS are required for biofilm formation (Mougous *et al*., 2006) and are under tight QS control. When the population density gets too high, the *las* and *pqs* QS systems negatively regulate the T6SS, to prevent delivering the effector proteins to sister cells (Balasubramanian *et al*., 2013).

Other effector proteins may be encoded for in the *P. aeruginosa* genome that may have additional functions beyond antimicrobial activity, and may also act on human immune cells and epithelial cells directly during CF infection (Sana *et al*., 2016). The T6SS is active during chronic infection, as it is negatively regulated by the global regulator RsmA during acute infection (Mougous *et al*., 2006). Mutations in the T6SS have been shown to attenuate, and even abolish, chronic CF infection in rat models, making them attractive targets for antimicrobial development (Chen *et al.*, 2015).

1.3.9 Hypermutation

In the CF lung, *P. aeruginosa* have been observed to have an increased rate of spontaneous mutation, a phenotype called hypermutation (Oliver & Mena, 2010). Hypermutator isolates have an increased mutation rate that is caused by inactivation of the DNA repair mechanisms in the bacterium (Taddei *et al*., 1997). In *E. coli*, mutation rates 1,000 and 10,000x higher than wild-type have been observed in hypermutator strains (Oliver & Mena, 2010; Ramiro *et al*., 2019), and in *P. aeruginosa,* hypermutation has been linked to the ability to rapidly adapt to hostile and new environments, such as the CF lung (Oliver & Mena, 2010). Hypermutability decreases the time required to gain AMR mutations, mutations to aid survivability in the lung, and also the loss of function mutations that facilitate the switch from acute to chronic infection (Oliver & Mena, 2010).

1.4 *P. aeruginosa* and bacteraemia

Outside of CF lung infection, *P. aeruginosa* can also cause serious bloodstream infections, associated with a mortality rate of between than 20-40% (Micek *et al*., 2005; Van Delden *et al*., 2007). *P. aeruginosa* is the third most common cause of Gram-negative bloodstream infections, behind *E. coli* and *Klebsiella spp*., and the second highest cause of mortality (Van Delden *et al*., 2007; Wisplinghoff *et al.*, 2004). Poor prognosis is explained by the rapid progression of the *P. aeruginosa* bacteraemia disease, causing 50% of the fatalities in the first few days of infection (Van Delden *et al.*, 2007).

Several risk factors are associated with increased mortality in *P. aeruginosa* bacteraemia, with poor prognosis recorded for surgery-associated, pneumonia-associated, and sepsis-associated bloodstream infection (Vidal *et al.*, 1997). AMR bloodstream infection is also associated with poor prognosis due to the ineffectiveness of initial treatment, which often cannot be rectified due to the rapid progression of the disease (Zhang *et al.*, 2016a).

1.4.1 Virulence of bacteraemia-associated *P. aeruginosa*

Albumin is the most abundant protein in human blood serum (Levitt & Levitt, 2016). Albumin can sequester quorum sensing molecules, and therefore QS-related virulence factors have been shown to be downregulated in the presence of albumin (Hickey *et al.*, 2018). Albumin is present at very low concentrations in the lung, urinary tract, and burn wounds, and hence during acute infection in these environments, QS-related virulence factors are expressed at high levels (Smith *et al*., 2017). The highly-virulent *P. aeruginosa* at these sites allows for quick dissemination into the bloodstream, causing sepsis and septic shock (Bahemia *et al*., 2015).

Transfer of *P. aeruginosa* from the initial site of injury or infection into the blood is facilitated by *P. aeruginosa* virulence factors that are encoded in the genome. The T3SS effector proteins, ExoS and ExoY, break down the integrity of the epithelial layer (Golovkine *et al*., 2018), and the LasB elastase breaks down cadherin, which is responsible for joining epithelial cells together (Hickey *et al*., 2018). Combined, these lead to regions of host cell death. Additionally, these virulence determinants stimulate

immune system inflammation, which helps the migration of *P. aeruginosa* from the site of acute infection into the bloodstream through the break down of protective barriers (Hickey *et al.*, 2018).

A 2018 study by Hickey *et al*. aimed to characterise any differences in virulence between bloodstream isolates and isolates from the site of the original acute infection (Hickey *et al*., 2018). The study found two upregulated proteins in the transcriptome of the bloodstream isolates; LecA and RpoN. LecA is an adhesion factor that binds *P. aeruginosa* to other cells; knockout mutants of which have been shown to reduce virulence (Chemani *et al.*, 2009). RpoN regulates several virulence features, including flagella-dependent motility, T6SS, and siderophore production (Hickey *et al.*, 2018). Through the upregulation of these RpoN-controlled virulence factors, there is an indirect upregulation of other virulence factors, including the ToxA toxin and Apr alkaline phosphatase, which cause damage to the surrounding tissue and induce inflammatory response from the immune system, causing further damage to the host system (Hickey *et al.*, 2018).

1.5 Antibiotic-resistant *P. aeruginosa*

Antibiotic resistant *P. aeruginosa* bloodstream infections are a major problem worldwide (Bassetti *et al*., 2018). A 2016 surveillance study of *P. aeruginosa* bloodstream infections identified a 33.9% prevalence of resistance to one or more of the antibiotics tested (Bassetti *et al.*, 2018).

1.5.1 Active efflux causes intrinsic resistance to many antibiotics

P. aeruginosa has low levels of intrinsic resistance to most antibiotics (Breidenstein *et al.*, 2011). This intrinsic resistance is conferred primarily through active efflux of antimicrobial compounds. To date, 25 different efflux complexes have been identified in the *P. aeruginosa* PAO1 and PA14 genomes, which can recognise over 50 antimicrobial compounds (Li & Plesiat, 2016).

Of the 35 *P. aeruginosa* efflux systems, eleven are broad-substrate, RND-like multidrug efflux (Mex) pumps, of which, only the MexAB-OprM efflux system is constitutively expressed (Evans & Poole, 2006). The MexAB-OprM system exports compounds from multiple antibiotic classes, including the

β-lactams, quinolones, chloramphenicol, macrolides, and tetracyclines (Li & Plesiat, 2016), and hence contributes greatly to the low-levels of resistance observed in *P. aeruginosa*. A knockout of any of the four MexAB-OprM regulator genes (*mexR*, *nalC*, *nalD*, and *cpxR*) can increase the MIC profile of *P. aeruginosa* by up to 16-times due to the overexpression of the efflux pump (Tian *et al.*, 2016).

The MexCD-OprJ efflux system can also export a range of antibiotic compounds, including members of the macrolide, chloramphenicol, novobiocin, tetracycline, trimethoprim, and 4th-generation β-lactam classes of antibiotics (Poole, 2001). In wild-type *P. aeruginosa*, the MexCD-OprJ system is not typically expressed. However, a knockout of the repressor protein, *nfxB* leads to the overexpression of the MexCD-OprJ system, leading to an overall increase in resistance (Purssell & Poole, 2013). It has also been shown that a knockout of *nfxB* leads to decreased expression of the MexXY-OprM efflux system, leading to hypersusceptibility to aminoglycoside antibiotics (Morita *et al*., 2012). Overexpression of the MexXY-OprM efflux system is the main cause of aminoglycoside resistance in *P. aeruginosa* isolates, particularly in CF patients (Armstrong & Miller, 2010). Efflux of aminoglycosides in bacterial species is relatively rare, with intrinsic resistance to aminoglycosides via efflux typically identified only in *Burkholderia pseudomallei* and *Acinetobacter baumannii* (Morita *et al*., 2012).

Downregulation of MexXY-OprM also occurs in *nfxC* knockout mutants (Morita *et al.*, 2012). The *nfxC* gene typically represses the MexEF-OprN efflux system, which, when overexpressed, leads to increased resistance to fluoroquinolones (Llanes *et al*., 2011). The efflux systems MexPQ-OpmE, MexVW-OprM and MexGHI-OpmD also export fluoroquinolones (Mima *et al*., 2005; Fabrega *et al*., 2009; Li, 2003), amongst other substrates, such as macrolides, heavy metals, tetracycline, chloramphenicol, acriflavine and ethidium bromide (Martinez *et al*., 2009). Macrolides are also the main efflux substrate of the MexJK-OprM complex, which can also export triclosan (Chuanchen *et al.*, 2002). The final Mex-style efflux pump is MexMN-OprM, which exports chloramphenicol, heavy metals and some β-lactam compounds (Mima *et al.*, 2005).

Several non-Mex efflux pumps are also encoded within the *P. aeruginosa* genome. The MuxABC-OpmB efflux pump can export monobactams, macrolides, and tetracyclines (Li *et al.*, 2014). Downregulation of the MuxABC-OpmB efflux system leads to a decrease in virulence, typically through downregulation of twitching motility through an unknown mechanism (Yang *et al.*, 2011a). Additional multi-domain efflux systems include TriABC-OpmH, the substrate for which is triclosan, and CzcCBA, which is linked to heavy metal resistance (Li *et al.*, 2014).

P. aeruginosa can also encode single-protein efflux systems, such as PmpM and EmrE. PmpM is a multi-drug efflux pump, which has substrate recognition for benzalkonium chloride, fluoroquinolones, ethidium bromide, acriflavine, and tetraphenylphosphonium chloride (He *et al.*, 2004). EmrE confers resistance to ethidium bromide, acriflavine, and EmrE-knockout leads to a decrease in MIC values for the aminoglycosides (Li *et al.*, 2003).

1.5.2 Resistance to β-lactams

The *P. aeruginosa* genome ubiquitously encodes several enzymes that confer resistance to the β-lactam class of antibiotics. These include a Class C wide-spectrum β-lactamase, *ampC*, which confers low-levels of resistance to aminopenicillins and first- and second-generation cephalosporins (Berrazeg *et al.*, 2015). High levels of resistance occur when AmpC is overexpressed, and many mutations have been identified which can extend the specificity of AmpC to the penicillins, monobactams, and third- and fourth-generation cephalosporins by altering the enzyme active site (Berrazeg *et al.*, 2015). The *P. aeruginosa* genome also encodes *poxB*, which is a Class D oxacillinase, however, full activity of PoxB has only been observed in isolates where AmpC activity is absent (Poole *et al*., 2004).

P. aeruginosa is able to acquire further resistance to the β-lactams through the acquisition of other β-lactamase enzymes on plasmids or transposons (Zhao & Hu, 2010). Examples of the acquisition of almost all classes of extended-spectrum β-lactamase (ESBL) enzymes have been recorded in *P. aeruginosa* from around the world, including TEM, SHV, CTX-M, PER, VEB, GES, BEL and OXA ESBLs (Poole, 2011). ESBL enzymes have the ability to hydrolyse compounds from several different classes of β-lactams (Bush *et al.*, 2008). Presence of these ESBLs restricts treatment options, and is of global concern due to the ease of transfer of ESBL enzymes through horizontal gene transfer, and the low fitness cost these confer (Ranjan *et al.*, 2018).

The carbapenem class of antibiotics are considered to be last-resort antibiotics for treating *P. aeruginosa* infections (Codjoe & Donkor, 2018). Resistance can be caused by the acquisition of metallo-β-lactamase enzymes of the VIM and IMP class, which are globally dispersed in *P. aeruginosa* populations (Walsh *et al.*, 2005). The metallo-β-lactamase enzymes often work in partnership with inactivation of OprD membrane porins, which are the main route of carbapenem entry into the *P. aeruginosa* cell (Trias & Nikaido, 1990). Inactivation of *oprD* by itself is not enough to cause resistance, but can lead to

carbapenem resistance when coupled with overexpression of efflux pumps, alteration of the AmpC enzyme, or acquisition of metallo-β-lactamase enzymes (Poole, 2011).

The recent increase in worldwide Gram-negative carbapenem-resistant infections has been attributed to bacterial acquisition of the broad-spectrum metallo-β-lactamase NDM-1 (Khan *et al.*, 2017). There are documented examples of *P. aeruginosa* acquisition of NDM-1 through the horizontal gene transfer of integrons and plasmids (Kazmierczak *et al*., 2016). The distribution of *P. aeruginosa* encoding for NDM-1, whilst still relatively low, is now global, resulting in one of the most concerning emerging threats of MDR *P. aeruginosa* (Kazmierczak *et al*., 2016).

1.5.3 Resistance to fluoroquinolones

Fluoroquinolone antibiotics are commonly used in the treatment of acute *P. aeruginosa* infection, and widespread use has led to the steady increase in fluoroquinolone resistant bacteria (Poole, 2011; Aldred *et al*., 2014). Common resistance determinants include changes in the amino acid residues of the quinolone-resistance-determining-region (QRDR) of both subunits of DNA gyrase (*gyrA* and *gyrB*) and topoisomerase (*parC* and *parE*) proteins (Poole, 2011). As well as modification of the antibiotic target site, common resistance determinants include overexpression of four different efflux systems; MexAB-OprM, MexCD-OprJ, MexEF-OprN, and MexXY-OprM (Zhanel *et al.*, 2004).

1.5.4 Resistance to aminoglycosides

The aminoglycoside class of antibiotics are also commonly involved in the treatment of *P. aeruginosa* infection. The acquisition of aminoglycoside modifying enzymes (AMEs) on plasmids or integrons is the major cause of aminoglycoside resistance in *P. aeruginosa* (Strateva & Yordanov, 2009). There are three classes of AMEs that can catalyse the conversion of terminal alcohol and amine groups into phosphate groups (APH-type AMEs), acetyl groups (AAC-type AMEs), or adenosine monophosphate (ANT-type

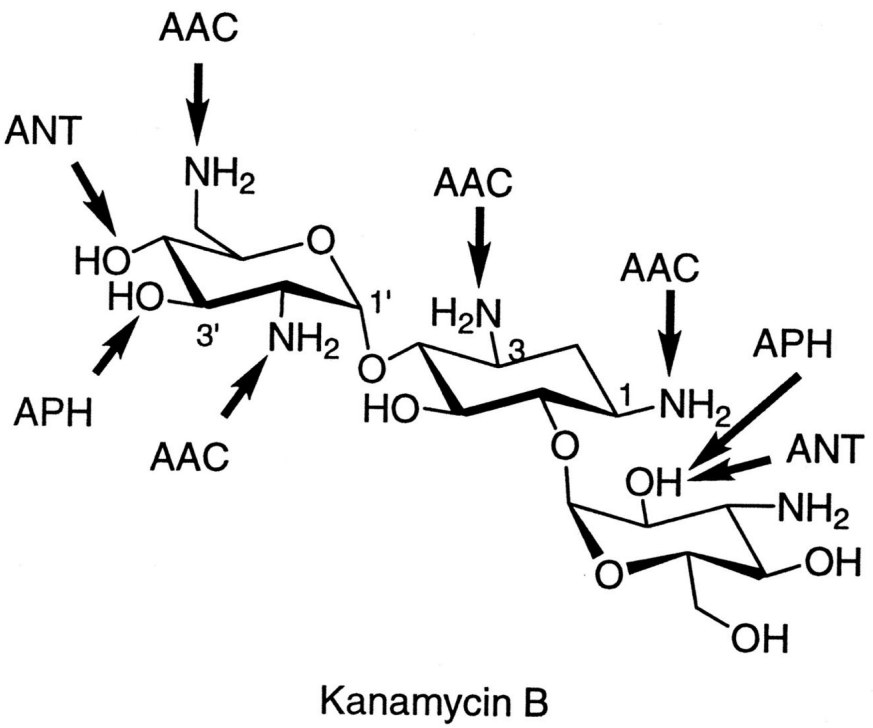


Figure 1.9 Kanamycin B is an aminoglycoside antibiotic than can be rendered inactive by all three classes of AMEs.

AMEs), which renders the antibiotic inactive (Figure 1.9) (Ramirez & Tolmasky, 2010). The presence of AMEs are rare in isolates collected from CF patients (Poole, 2011). Resistance to aminoglycosides can also occur through efflux, through the overexpression of the MexXY-OprM efflux system, which typically occurs through the inactivation of its repressor molecule (Singh *et al.*, 2017).

It is rare for *P. aeruginosa* to acquire resistance to the aminoglycoside class of antibiotics (Poole, 2011). The aminoglycoside tobramycin, in combination with either ciprofloxacin or ceftazidime, has also shown synergistic effects in preventing resistance from occurring (Ratjen *et al.*, 2009; Kapoor & Murphy, 2018), suggesting that aminoglycosides should be in regular use to treat *P. aeruginosa* infection.

1.5.5 Resistance to other classes of antibiotics

Colistin resistance in *P. aeruginosa* through mutations in *phoQ, pmrB* and *parRS* have been documented and verified (Ly *et al.*, 2011; Lee *et al.*, 2016). Due to the increased prevalence of carbapenem-resistant *P.*

aeruginosa, colistin is becoming a more viable option to treat *P. aeruginosa* infections, despite the harmful negative side-effects (Lim *et al.*, 2015).

The tetracycline class of antibiotics are rarely used in the treatment of *P. aeruginosa* due to its ability to export tetracyclines from the cell by normal levels of MexAB-OprM efflux system expression (Morita *et al.*, 2014). Similarly, glycopeptide antibiotics such as vancomycin are not used to treat Gram-negative bacterial infections, since the molecules are too large to enter through porins in the outer membrane (Yarlagadda *et al*., 2015).

1.5.6 Plasmid-mediated resistance

Plasmids are a convenient way of replicating and sharing advantageous genetic material between bacteria that are not necessarily related (Bennett, 2008). In *P. aeruginosa*, the sharing and uptake of plasmids is a common method of spreading new genetic material amongst its population. Plasmid-mediated antibiotic resistance is especially a problem in healthcare settings, but the spread of *P. aeruginosa* plasmids in healthcare is not well understood, as very few plasmids from isolates in healthcare settings have been sequenced and analysed (Strateva & Yordanov, 2009; Bi *et al.*, 2016).

In a study of pan-resistant *P. aeruginosa* by Cazares *et al*., resistance genes for eight different classes of antibiotics were identified, alongside a second copy of the MexCD-OprJ efflux pumps, on a 420 kbp megaplasmid (Cazares *et al*., 2019). The study identified the presence of these highly-resistant megaplasmids in nearly 20% of *P. aeruginosa* isolates in one hospital in Thailand, suggesting an important role of megaplasmids in spreading resistance, and suggesting that screening for these plasmids, and megaplasmids, should become routine (Cazares *et al.*, 2019). This type of surveillance would require the application of long-read DNA sequencing technology, which is currently more expensive and less high-throughput than short read sequencing technology, and hence is not used routinely in a clinical setting (Mantere *et al.*, 2019). The true impact of plasmids and megaplasmids responsible for the spread of MDR *P. aeruginosa* elsewhere around the globe is currently unknown.

1.6 *P. aeruginosa* population structure

The population structure of *P. aeruginosa* is well-defined (Figure 1.10). Historically, *P. aeruginosa* populations have been considered as split into two distinct clades, which were characterised by their ability to produce either the ExoS toxin, or the ExoU toxin (Pirnay *et al.*, 2009). These clades were named after two "gold-standard" lab reference strains; PAO1 and PA14. PAO1-like isolates produce the less-virulent ExoS toxin, and PA14-like isolates produce the more-virulent ExoU toxin (Ozer *et al*., 2019). The divergence between the two large groups was investigated in a 2019 study by Ozer *et al.* The authors identified lower recombination rates between, rather than within, the two groups, fixed polymorphic variant sites that indicate positive selection, specific gene clusters within the accessory genome, and a difference in the predominance of *O*-antigen receptors on the cell membrane (Ozer *et al*., 2019). The authors of the study concluded that the two groups diverged early in the evolution of *P. aeruginosa*, indicative of adaptation to different environmental niches, which has lead to a reduced ability for genetic exchange between the PAO1-like and PA14-like isolates (Ozer *et al*., 2019).

In 2010, a taxonomic outlier was sequenced; PA7 (Roy *et al*., 2010). The group of PA7-like isolates is distinct from the rest of the *P. aeruginosa* population, and have only 93-94% average nucleotide identity compared to the rest of the *P. aeruginosa* population (Freschi *et al.*, 2019). The PA7-like isolates contain

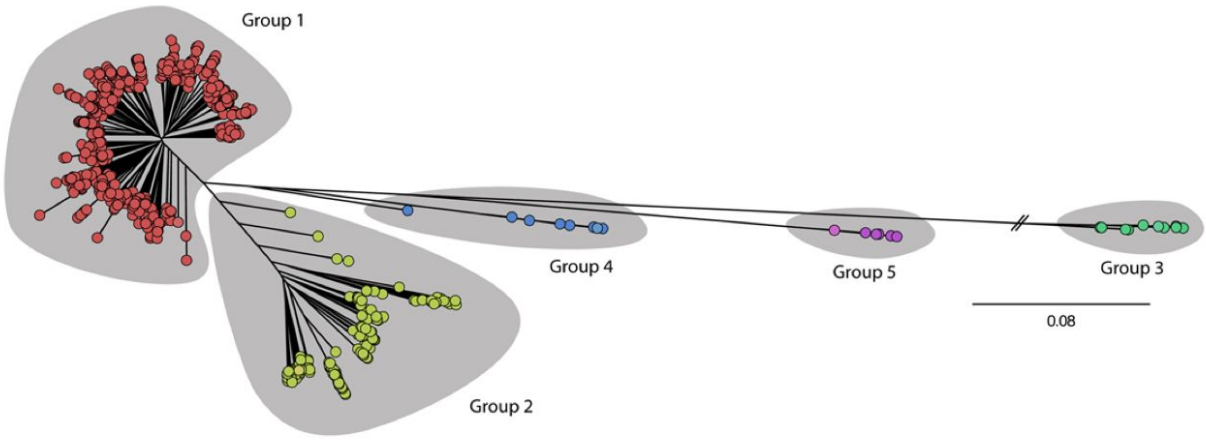


Figure 1.10 The population structure of *P. aeruginosa* resolves into well-defined groups. Group 1 isolates are PAO1-like, Group 2 isolates are PA14-like, and Group 3 isolates are PA7-like. Group 4 and Group 5 isolates were newly identified in a 2019 study (Freschi *et al.*, 2019). Image adapted with permission from Freschi *et al.*, 2019.

many AMR genes that are not found in any other *P. aeruginosa* subspecies, and they also lack key virulence factors (Freschi *et al.*, 2019). These include the absence of the T3SS and all related effector toxins, the absence of the potent *toxA* toxin, and the absence of some siderophore systems (Roy *et al.*, 2010).

Two further subgroups of *P. aeruginosa* have been identified that reside in between the PAO1/PA14-like isolates and the PA7-like isolates in the *P. aeruginosa* population structure. How these groups have diverged is not clear, and their prevalence is low within the NCBI database (Geer *et al*., 2010; Freschi *et al*., 2019).

As well as the population structure, the pangenome of 1,311 *P. aeruginosa* isolates was calculated (Freschi *et al.*, 2019). The authors identified a total of 54,272 genes, of which 27,187 genes were unique to a single isolate. Of the remaining genes, only 665 (1%) were designated as "core" genes, which are genes that were present in every isolate from all subgroups. This is only twice the 336 genes that are deemed essential for any bacterial growth (Turner *et al*., 2015). This is further evidence of the ability of *P. aeruginosa* to adapt to a range of different environments by changing their genetic content to suit their niche (Freschi *et al.*, 2019).

1.7 Associating phenotype with genotype through genome-wide association studies

One of the main goals of genome sequencing is to be able to find a particular locus within the genome to explain a particular phenotype of an organism (Visscher *et al*., 2012). In 2001 the sequence of the first human genome was determined (International Human Genome Sequencing Consortium, 2001). Since then, hundreds of thousands of human genomes have been sequenced, identifying variation in the genomes which cause genetic disease (Lappalainen *et al*., 2019). Genome-wide association studies (GWAS) are an experimental design to associate genetic variants from across the whole genome with disease states in populations of people (Visscer *et al*., 2012).

GWAS studies are unbiased, and do not make any biological assumptions of the variant with respect to the disease (Visscher *et al.*, 2012), but they can provide biologically-relevant results. In 2007, the

Wellcome Trust Case Control Consortium (WTCCC) published the first well-designed GWAS, which reported 24 independent associations with seven major diseases; bipolar disorder, coronary artery disease, Chrohn's disease, rheumatoid arthritis, type 1 diabetes, and type 2 diabetes (Figure 1.11) (Wellcome Trust Case Control Consortium, 2007). Since this first study, over 10,000 different loci have since been associated with hundreds of different disease states (Visscher *et al.*, 2017).

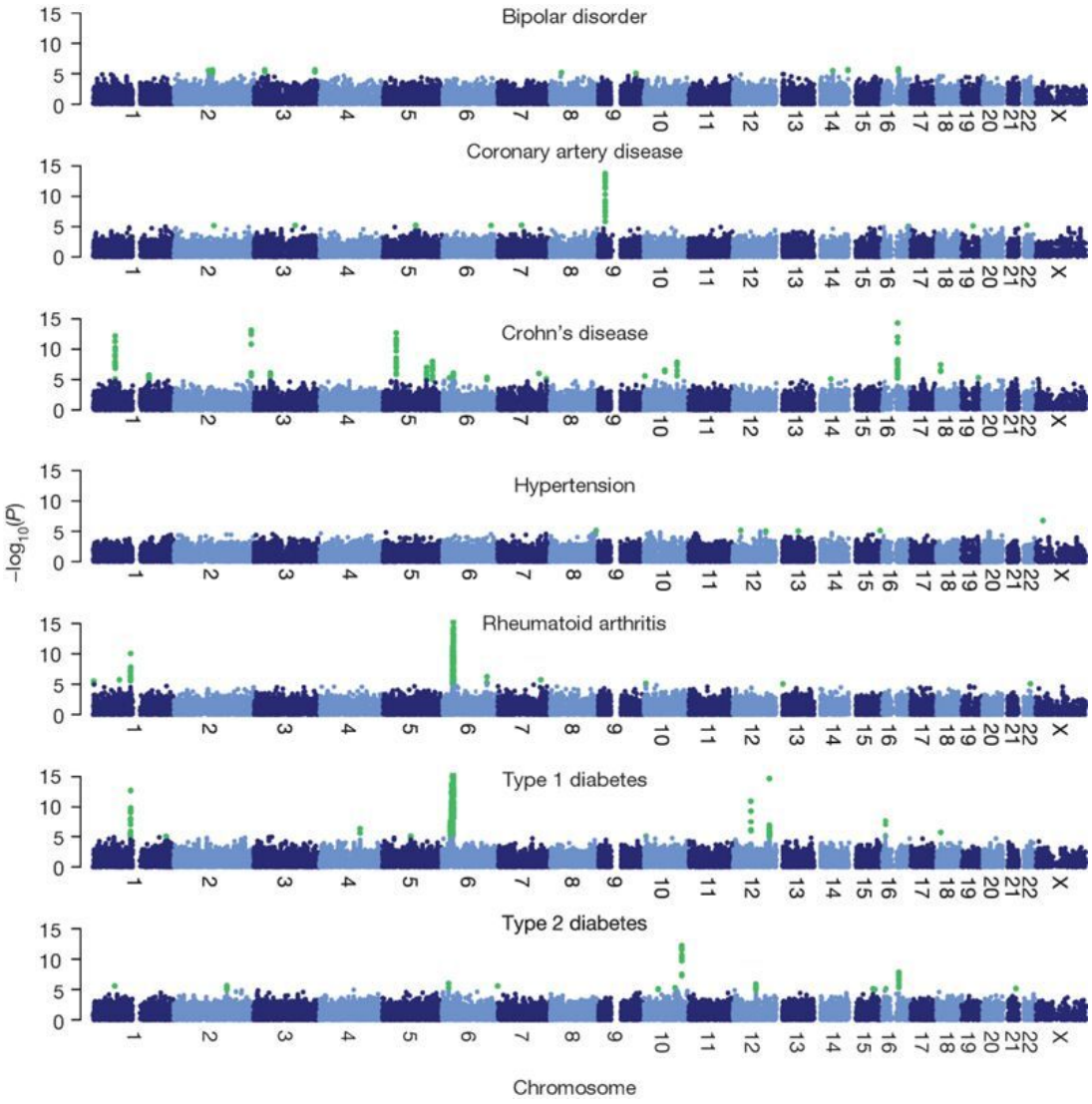


Figure 1.11 Manhattan plots from the first well-designed GWAS published in 2007, where 24 independent associations were made for 7 different diseases. Image used with permission from Wellcome Trust Case Control Consortium, 2007.

1.7.1 Challenges associated with bacterial GWAS

Applying GWAS to bacterial populations can be used to identify genomic variation associated with heritable traits of interest, such as antimicrobial resistance (Lees & Bentley, 2016). This can help increase understanding of antimicrobial resistance, virulence, and adaptive mutations which contribute to the success of a bacterial strain. The first well-powered bacterial GWAS was published in 2014 by Chewapreecha *et al*. The study identified variants within the *Streptococcus pneumoniae* population that were associated with β-lactam non-susceptibility (Chewapreecha *et al.*, 2014). Both known and novel mutations in the penicillin binding proteins, which conferred different levels of resistance to different classes of β-lactams, were identified. Since then, GWAS have been applied to bacterial studies to identify variation of pneumococcal carriage duration (Lees *et al.*, 2017), pneumococcal invasive potential (Lees *et al.*, 2019), and drug-resistance determinants in *Mycobacterium tuberculosis* (Coll *et al.*, 2018), amongst others.

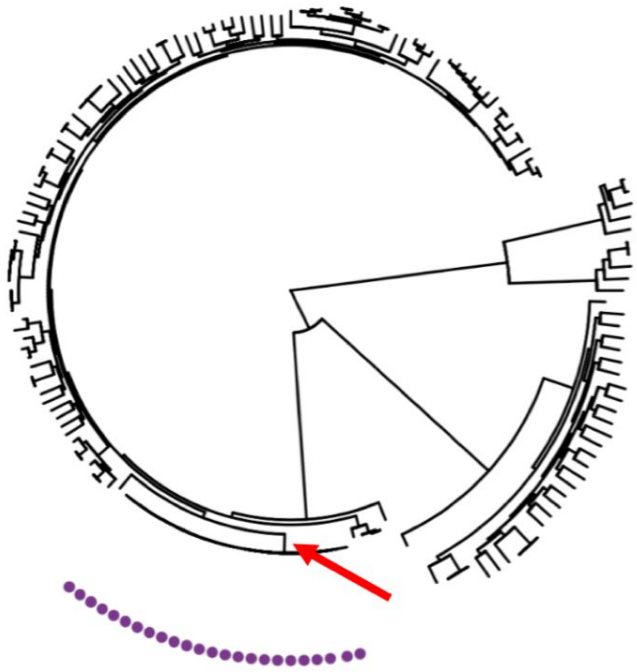


Figure 1.12 The population structure of bacteria confounds GWAS. Since bacteria reproduce clonally, the GWAS will be unable to distinguish the causal variant of the phenotype (purple dots), with all mutations that differentiate that bacterial clade from all others (red arrow). This could be hundreds or thousands of variants. Figure adapted under Creative Commons v4.0 license from Collins & Didelot, 2018.

However, there are various challenges associated with bacterial GWAS that are not present for human GWAS design. The first of these is the presence of population structure arising due to the clonal nature of bacterial reproduction (Collins & Didelot, 2018). When a set of *de novo* mutations are acquired, one of which gives rise to a phenotype, all daughter cells will inherit all mutations as well as the phenotype. Therefore, an association study will find all mutations that differentiate the population of bacteria with the phenotype from the population of bacteria without the phenotype as significant (Figure 1.12). These associations are known as "lineage effects", and could consist of hundreds or thousands of mutations which would have to be investigated experimentally (Earle *et al.*, 2016). To overcome the problems associated with population structure, GWAS should be designed to associate genetic variants with a phenotype that is distributed throughout the phylogeny to mitigate the effect of the genetic background that gives rise to population structure.

A second challenge associated with bacterial GWAS is the presence of the bacterial pan-genome. Most human variation occurs as a result of small variants that can be detected by mapping to a single reference, whereas variation due to a difference in the presence or absence of novel genes is only about 1% (Sherman et al., 2019). However, bacteria have large pan-genomes, which can range from a few genes to tens-of-thousands, which will not be accounted for when mapping all genomes of a population to a single reference (McInerney *et al.*, 2017). Therefore, associating the presence and absence of accessory genes with a phenotype should also be considered in a well-designed GWAS. However, this does not account for variation within the variable genes of the pan-genome, which may cause the phenotype.

Finally, bacterial GWAS need to consider variation from across the whole genome, facilitating the requirement for whole-genome sequencing (Power *et al.*, 2016). Currently this is expensive, and combined with difficulty in obtaining clinically-relevant bacteria with detailed phenotyping, means study sample sizes are currently limited (Power *et al.*, 2016). This is compounded by the vast amount of variation across bacterial species, increasing the multiple testing burden, necessitating large sample collections to achieve reasonable power (Power *et al.*, 2016).

1.7.2 Bacterial GWAS methodologies

Since the establishment of bacterial GWAS, several techniques have been developed to conveniently package GWAS pipelines and make them more accessible to the scientific community. The first of these

packages, called BugWAS, was developed in early 2016 by Earle *et al*., which aimed to conveniently package a pipeline to identify lineage effects, as well as specific loci such as SNPs and the presence/absence of genes (Earle *et al.*, 2016). The authors applied this technique to determinants of antimicrobial drug resistance for 17 antimicrobial compounds for four different bacteria (Earle *et al.*, 2016). This pipeline is an adaptation of existing GWAS methodology, called GEMMA, which employs both linear regression and linear mixed models for human GWAS (Zhou & Stephens, 2012).

In 2016, a novel GWAS pipeline was developed by Lees *et al.*, called SEER. Rather than associating specific SNPs with a phenotype, SEER requires the computation of kmers, which are short DNA sequences between 9 and 100 bp (Lees *et al.*, 2016). Kmers can account for different genomic variation within a DNA string, such as gene truncations, structural variations, and short repeats, as well as individual SNPs, which is a major benefit over simply considering SNP variation (Lees *et al.*, 2016). This technique was first applied to identify known resistance mutations in *S. pneumoniae* and novel kmers associated with *S. pyogenes* invasive potential (Lees *et al.*, 2016). The programme employs both linear and logistic regression to associate genotype with phenotype, requiring a relatedness matrix as a covariate to account for population structure (Lees *et al.*, 2016). The package was updated in 2018 to include a greater array of methods to test for association, including adding linear mixed models as a technique to control for population structure (Lees *et al.*, 2018).

One limitation of bacterial GWAS is the bacterial pangenome. Due to frequent recombination, infection by bacteriophage, the presence of mobile genetic elements, and acquisition of plasmids, different bacteria from the same population may have different genetic compositions (McInerney *et al.*, 2017). In 2016, a pangenome GWAS pipeline was developed, called Scoary (Brynilsdrud *et al.*, 2016). This pipeline requires a gene presence-absence matrix, and runs a self-contained GWAS pipeline, with phylogenetic population structure controls in place. However, this method is dependent on a well-calculated and robust tree to be effective (Brynilsdrud *et al.*, 2016), which may not always be possible, and does not consider variation within the genes of the pangenome. The authors validated the effectiveness of Scoray by identifying known and novel linezolid resistance genes in *S. epidermidis* (Brynilsdrud *et al.*, 2016).

In early 2018, a phylogenetic GWAS method (treeWAS) was developed by Collins & Didelot. This method identifies relevant association by mapping the evolution and acquisition of SNPs onto a phylogenetic tree, alongside the phenotype of interest, whilst accounting for both population structure and recombination (Collins & Didelot, 2018). This method was applied to a dataset of *Neisseria meningitidis*

genomes to identify known penicillin resistance variants, and to uncover novel resistant determinants (Collins & Didelot, 2018). A drawback of this method, as with the Scoary pipeline, is the requirement for a robust phylogenetic tree to correctly identify relevant associations.

DBGWAS was developed as a novel method to associate reference-free genomic variants with a phenotype (Jaillard *et al*. in 2018). This method builds upon kmer methodology, by taking all genomic variation across all genomes, and compressing this information into computationally-efficient compact De Bruijn graphs, called unitigs (Jaillard *et al.*, 2018). De Bruijn graphs are used in the *de novo* assembly of genomes and SNP calling, and connect overlapping kmers, efficiently summarising all variation over a set of genomes (Jaillard *et al.*, 2018). Variation in the genome is then associated with the phenotype of interest using the GEMMA software developed for human GWAS (Jaillard *et al.*, 2018). This technique was first used to identify known and novel antibiotic resistance determinants in *S. aureus*, *M. tuberculosis*, and *P. aeruginosa* (Jaillard *et al.*, 2018). To date, this study represents the only example of GWAS that has been applied to a *P. aeruginosa* population.

1.8 Aims and objectives

This dissertation forms part of a wider study that has been undertaken at the Laboratory for Molecular Biology, Addenbrookes Hospital, Cambridge. The wider study saw 15 patients with CF asked to undertake home-based measurement of several key physiological metrics, including lung function, physical health, and mental health over the course of six months, as well as to provide daily sputum samples. These metrics were examined for markers that may be used as predictors of APEs.

Sputum samples from nine patients who experienced a total of 18 APEs between them were selected, and 96 *P. aeruginosa* isolates were collected from each sputum sample on the day that the patient first experienced the symptoms of an APE (acute timepoint), seven days before the APE (stable timepoint), and seven days after antibiotic treatment had finished (recovery timepoint). Roughly 4,400 *P. aeruginosa* isolates were sampled, and each isolate was screened for the 10 virulence-related phenotypes listed in Table 1.2. Similar studies have previously assessed only a much more limited number of phenotypes, and therefore the ten phenotypes selected for this study address this limitation. Phenotypes of three complex

Phenotype	Description
BHL production	An autoinducer that activates the global regulator protein RhlR.
OdDHL production	An autoinducer that activates the global regulator protein LasR.
PQS production	Quinolone-based intercellular signalling which controls biofilm production and virulence factor production.
Biofilm production	A matrix of extracellular polymeric substances that hold microbial cells together to a surface.
Rhamnolipid production	Glycolipids with detergent-like activity and have cytolytic activity against macrophages.
Caseinase production	An exo-enzyme produced by bacteria that breaks down the phosphoprotein casein.
Gelatinase production	An exo-enzyme produced by bacteria that hydrolyses gelatin.
Twitching motility	The movement of *P. aeruginosa* across surfaces using the Type IV pilus.
Swimming motility	A flagellum-dependent form of movement in Gram-negative bacteria.
Siderophore production	Iron-chelating agents that *P. aeruginosa* secrete to uptake iron from the environment.

Table 1.2 A list of all of the virulence-related phenotypes that the *P. aeruginosa* isolates in this study were phenotypically screened for.

quorum-sensing systems were selected (BHL, OdDHL and PQS), which have been shown using *in vitro* studies to regulate virulence factor expression in *P. aeruginosa* (Lee & Zhang, 2015). Three well-studied virulence-associated phenotypes (gelatinase, caseinase and siderophore) were also selected, because their expression is regulated by specific, known quorum-sensing sub-systems, and significant diversity has been reported in both epidemic and non-epidemic *P. aeruginosa* isolates from patients with CF (Voynow *et al*., 2008; Tyrrel & Callaghan, 2016). The rhamnolipid production phenotype was also selected, because rhamnolipids are involved in the establishment of persistent chronic *P. aeruginosa* infection in CF patients (Zulianello *et al.*, 2006). The biofilm phenotype is an indicator of the switch from acute to chronic infection in CF patients, and encourages persistence of *P. aeruginosa* in the CF lung, as well as increasing resistance to antibiotics (Hoiby *et al.*, 2010). Finally, swimming and twitching motility phenotypes were selected, because although *P. aeruginosa* isolates from CF infections often display non-motile phenotypes, this is poorly understood (Schick & Kassen, 2018).

To date, research surrounding the genomic variation and evolution of *P. aeruginosa* infection within the CF lung has focussed on single samples from long-term and longitudinal studies. This study presents a unique opportunity to fully examine the variation of *P. aeruginosa* within the CF lungs within a cohort of patients, and determine any genetic changes which may underpin phenotypic diversity. We hypothesise that there will be greater variation of the *P. aeruginosa* population within the lungs of patients with CF

than has been previously described, and that common genetic changes within these populations will underpin phenotypic diversity.

The 4,400 phenotyped isolates were whole-genome sequenced for this dissertation, and the aims and objectives of this study are as follows:

- To quantify the inter-patient and intra-patient diversity of *P. aeruginosa* bacterial infection of the CF lung from the whole-genome sequences of ~4,400 isolates collected from nine patients over six months.
- To investigate the evolution of *P. aeruginosa* population when facing the selective pressures of the CF lung.
- To determine the genetic basis for any phenotypic diversity by using genome-wide association studies.

In addition to the 4,400 CF *P. aeruginosa* isolates, this dissertation will investigate an additional dataset of *P. aeruginosa* isolates from bloodstream infection. Prior to this dissertation, 352 bloodstream isolates were whole-genome sequenced from two separate sources. The source of 79 of these isolates was the British Society for Antimicrobial Chemotherapy (BSAC) bacteraemia antimicrobial resistance surveillance project, a national surveillance project that aims to understand how antimicrobial resistance is changing in a healthcare setting. The remaining 253 isolates were collected from Cambridge hospitals, to understand the current state of *P. aeruginosa* bacteraemia infections on the local scale. The antimicrobial resistance profiles of all of these isolates have also been determined. We hypothesise that the local population structure of *P. aeruginosa* bacteraemia is not representative of the UK-wide population structure, and local bacteraemia is driven by specialised healthcare associated strain due to factors and practices that are specific to local areas.

The additional aims of this dissertation are as follows:

- To understand the population structure and genetic diversity of *P. aeruginosa* bacteraemia on both a local and national level.
- To determine the genetic cause of antimicrobial resistance in these isolates.

1.8.1 Importance and novelty of this study

Bacterial infections of the blood, and lungs of people with cystic fibrosis, are a leading cause of morbidity and mortality, and are therefore a key research topic in healthcare. The development of comparative genomics techniques enables powerful analyses that can help to uncover how and why bacteria behave the way they do in healthcare settings, which can ultimately inform clinical decisions.

The novelty of this research is as follows:

- *P. aeruginosa* bacteraemia infections are a big problem across the UK. By placing the Cambridge isolates into the wider context of the UK, we can identify whether the local *P. aeruginosa* bacteraemia population is representative of the wider UK *P. aeruginosa* population, or whether any differences within the Cambridge isolates may be linked to differences in practice.
- In order to overcome issues such as low sampling depth, study population size and limited diversity of isolates and phenotypes that have defined most previous short-term studies of clonal populations of *P. aeruginosa* within the CF lung, 4,400 isolates from 9 patients were sampled in this study. This study is therefore the largest comparative genomics study of clonal populations of *P. aeruginisoa* within the CF lung to date. This will enable a better estimation of the true diversity of the *P. aeruginosa* populations within the CF lung, enabling more powerful analyses of the genetic changes that may underpin phenotypic diversity and evolution, and how this variation may relate to pulmonary exacerbations in CF patients.
- GWAS are an emerging and powerful tool in the field of comparative genomics. By using so many genomes and phenotypes, this study is the largest *P. aeruginosa* GWAS to date. This will enable the most powerful analysis, and test the limit of current GWAS technologies.

Chapter 2

Genomic surveillance of bacteraemia-associated *Pseudomonas aeruginosa* in the UK and Ireland

Declaration of Contributions

The study was initiated by Sharon Peacock, and Julian Parkhill supervised this work. All phenotyping and DNA extraction was carried out by the Peacock group, and the isolates were DNA sequenced by the Wellcome Trust Sanger Institute DNA Pipelines Operations. I carried out all analysis in this chapter.

2.1 Introduction

Pseudomonas aeruginosa is a Gram-negative bacterium found to be widespread in natural environments, but rarely found as part of the normal, healthy human microbiome (Yang *et al*., 2011b). Due to the genomic and metabolic plasticity of *P. aeruginosa*, it is a serious opportunistic pathogen, known for causing diseases of the urinary tract, burn sites, and cystic fibrosis lung (Yang *et al*., 2011b). Bloodstream infections (bacteraemia) caused by *P. aeruginosa* are amongst the most serious caused by Gram-negative bacteria (McCarthy & Paterson, 2017). Poor prognosis is explained by the rapid progression of the disease, causing fatalities in up to 50% of cases in the first few days of infection, and further mortality rates of 20-40% within a month of infection (McCarthy & Paterson, 2017; Van Delden *et al.*, 2007).

The British Society for Antimicrobial Chemotherapy (BSAC) bacteraemia resistance surveillance programmes were set-up in the late 1990s in response to the rising antimicrobial resistance rates seen in bacteria across the UK (Reynolds *et al.*, 2008). The first surveillance programme started in 1999 and focussed on community-acquired, lower respiratory tract infections (Reynolds *et al*., 2008). The second programme started in 2001, and covered a wide range of bacteraemia-causing pathogens, including *P. aeruginosa* (Reynolds *et al*., 2008). Surveillance of *P. aeruginosa* is particularly important due to its intrinsic resistance to a number of antibiotics, and the ability to acquire resistance genes through a variety of mechanisms (Oliver *et al*., 2015). Multi-drug resistant *P. aeruginosa* infections are typically associated with worse outcomes for patients (Hirsch & Tam, 2011). Treatment of *P. aeruginosa* bloodstream infection is required immediately, and if MDR *P. aeruginosa* are present, initial antibiotic treatment may be ineffective (Zhang *et al.*, 2016b). Due to the rapid progression of *P. aeruginosa* bacteraemia infection and the length of time required to culture and determine resistance profiles, correct treatment of MDR infections are essential (Zhang *et al.*, 2016b).

The ability of *P. aeruginosa* to acquire novel genetic material, is one of the reasons that it is highly successful at adapting to a range of environments (Battle *et al.*, 2009). It is estimated that only about 90% of the *P. aeruginosa* genome is conserved, with the remaining 10% attributed to different genomic islands that are present in some strains and not others (Battle *et al.*, 2009). Genomic islands of *P. aeruginosa* are known to carry antimicrobial resistance genes, and have been attributed to the spread of antimicrobial resistant *P. aeruginosa* around the globe (Chowdhury *et al.*, 2017). However, these MDR-conferring genomic islands are strongly associated with only a few *P. aeruginosa* sequence types (STs), which are

labelled as international and high-risk clones (Oliver *et al*., 2015). These international and high-risk STs are so-called as they cause the majority of multidrug resistant (MDR), extensively drug resistant (XDR), and pan-resistant *P. aeruginosa* infections around the globe (Oliver *et al*., 2015). Traditionally, these were considered to be *P. aeruginosa* ST111, ST175, ST235, ST253, and ST395 (Oliver *et al*., 2015; Chowdhury *et al.*, 2017; Petitjean *et al*., 2017), though a greater number of STs are being classified as high-risk as surveillance efforts are increased globally.

International and high-risk clones are associated particularly with acquisition of metallo-β-lactamase (MBLs) and extended-spectrum β-lactamases (ESBLs) on genomic islands, particularly MBLs and ESBLs that are able to break down the favoured drug-of-last-resort, the carbapenems (Wright *et al*., 2015). The carbapenems are the most effective class of antibiotics able to combat Gram-negative infections, including *P. aeruginosa* infections (Meletis *et al*., 2012), and resistance to the carbapenem antibiotics is of severe global concern (Perez & Bonomo, 2018). Due to the increase of resistance to the carbapenems, a resurgence in the use of colistin to treat MDR infection is occurring, and observed levels of resistance are currently low (Lee *et al.*, 2016). Colistin use was abandoned several decades ago, due to serious nephrotoxicity and neurotoxicity, but recent studies have shown successful treatment of MDR *P. aeruginosa* infections with colistin, particularly when used in combination with other antibiotics (Falagas *et al*., 2005; Javed *et al*., 2018).

In this chapter, the whole-genome sequences of 79 MDR *P. aeruginosa* bacteraemia isolates, collected as part of the BSAC surveillance programme, were analysed to obtain an understanding of the current MDR *P. aeruginosa* population within the UK and Ireland. Additionally, 273 isolates from a second systematic survey of *P. aeruginosa* bacteraemia infection in the Cambridge area were also whole-genome sequenced. From this data, trends in local population structure and mechanisms of antibiotic resistance were investigated. The two datasets were compared to fully understand the wider context of *P. aeruginosa* bacteraemia within the UK & Ireland.

2.1.1 Aims

Bloodstream infections caused by bacteria such as *P. aeruginosa* are a leading cause of morbidity and mortality. How and why bacteria behave the way they do in healthcare settings is not well understood,

however, comparative genomics enables powerful analyses that can improve our understanding of bacterial infection on a population level, and ultimately inform clinical decisions.

In this chapter, we aim to understand the current state of *P. aeruginosa* bacteraemia infections on a local scale, by investigating a collection of *P. aeruginosa* bacteremia isolates collected from Cambridge hospitals. The phenotypic antimicrobial resistance profiles of all of these isolates have previously been determined. We aimed to compare the population structure and genetic diversity of *P. aeruginosa* bacteraemia on the local level and a national level, in order to determine whether the local population structure of *P. aeruginosa* bacteraemia is representative of the UK-wide population structure, or whether local bacteraemia is driven by specialised healthcare-associated strains. Additionally, we aim to use genetic data to determine the genetic cause of antimicrobial resistance in these isolates.

2.2 Methods

2.2.1 Isolate selection and DNA sequencing

Between 2006 - 2013, and between 2017 - 2018, single *P. aeruginosa* samples were taken from confirmed bacteraemia patients' bloodstreams from Addenbrookes, Hinchingbrook, and Papworth Everard Hospitals, Cambridgeshire (Rachel Bousfield, Peacock Lab, University of Cambridge).

Between 2001 - 2011, *P. aeruginosa* isolates that were sampled from patients with bacteraemia and had multi-drug resistance (MDR) were collected by 25 different laboratories distributed across the UK and Ireland (Table 2.1) as part of the BSAC bacteraemia resistance programme (www.bsacsurv.org) (Reynolds *et al.*, 2008).

A local collection of isolates was selected including 224 isolates from the Cambridge samples between 2006 and 2013. A further 60 isolates were also selected from the Cambridge samples between 2017 and 2018. In order to put these samples into a nationwide context, 81 isolates were selected from the MDR BSAC bacteraemia resistance programme, from the UK and Ireland (Rachel Bousfield, Peacock Lab, University of Cambridge).

DNA was extracted using a QIAxtractor (QIAgen), according to the manufacturer's instructions. Library preparation was performed according to the Illumina protocol, and sequencing was performed on Illumina HiSeq 2000 and 2500 platforms. *De novo* genome assembly was carried out using Velvet (version 1.2, Zerbino & Birney, 2008) with a k-mer length of 31, and VelvetOptimiser (version 2.2.5) with a k-mer length of 31. Scaffolds were annotated using PROKKA (version 1.5, Seemann, 2014), using default settings and specifying the *Pseudomonas* genus.

Fourteen isolates were excluded from the study (eleven isolates from the local collection and three isolates from the BSAC collection) due to low quality sequence data (average read depth <5) or incorrect species identification, resulting in a final sample size of 352 isolates. Sequence data for all isolates have been submitted to ENA with the accession numbers included in Appendix 1.

BSAC Participating Hospitals				
Ashford William Harvey Hospital	Bristol Royal Infirmary	Cork University Hospital	Leicester Royal Infirmary	Manchester Wythenshawe Hospital
Bangor Ysbyty Gwynedd	Bury St Edmunds West Suffolk Hospital	Coventry and Warwickshire Hospitals	London Northwick Park and St Mark's Hospitals	Newcastle Freeman Hospital
Barnstaple North Devon District Hospital	Cambridge Addenbrookes Hospital	Dublin Beaumont Hospital	London St Mary's and Imperial	Shrewsbury Royal Hospital
Belfast City Hospital	Cardiff University Hospital of Wales	Dublin St Vincent's Hospital	London University College Hospital	Southampton General Hospital
Birmingham City Hospital	Chelmsford Public Health Laboratory	Glasgow Royal Infirmary	Manchester Salford Royal Hospital	Truro Royal Cornwall Hospital

Table 2.1 In total, 25 laboratories from the UK and Ireland submitted samples to this study.

All samples were mapped to the *P. aeruginosa* PAO1 reference genome (accession number: PRJNA331) using bwa mem (version 0.7.17, Li & Durbin, 2009), discounting identical fastq reads and outputting all alignments. Samtools mpileup (version 1.6, Li *et al*., 2009) was used to determine which nucleotide base occured at each position, using a quality score of 50, discarding anomalous reads and disabling read-pair-overlap detection. Bcftools call (version 1.5, Li *et al*., 2009) was used to call SNPs against the reference genome, assuming a mutation rate of 0.001. Multiple sequence alignments were generated by combining all individual consensus sequences. For analysis of variant sites, SNPs were extracted from the multiple sequence alignments using snp-sites (version 2.4.1) (Page *et al.*, 2016) using default settings.

2.2.2 MLST assignment

The MLST of all isolates were assigned using ARIBA (version 2.12.1) (Hunt *et al.*, 2017) and PubMLST (Jolley & Maiden, 2010), which contains 3,336 *P. aeruginosa* ST assignments, as of 8th August 2019.

2.2.3 Phylogeny and pangenome

A maximum-likelihood phylogenetic tree of all 352 isolates included within this analysis was generated with FastTree (version 2.1.10) (Price *et al.*, 2010) using the Jukes-Cantor + CAT model using the SNP multiple sequence alignment generated in section 2.2.1. Circular genome visualisations with metadata were visualised using iTOL (Letunik & Bork, 2019).

The pangenome was calculated with annotation files using Roary (version 1.7.1) (Page *et al.*, 2015), for every isolate within the study, using a minimum percentage identity of 95% and a gene was identified as "core" if it was present in 99% of samples or more. Pangenome visualisation was carried out using

Phandango (Hadfield *et al*., 2017). A maximum-likelihood phylogenetic tree of all 352 isolates was also produced using the SNPs of the core genome, as above, and compared to the previous phylogeny using Dendroscope (version 3) (Huson & Scornavacca, 2012).

2.2.4 Investigating overrepresented clades

The highest quality assembled genome from each of the overrepresented clades (defined as clades containing a ST with over 20 isolates) were used as an internal reference genome, and all other isolates belonging to that clade were mapped to that internal reference, as in section 2.2.1. The maximum-likelihood phylogenies of each overrepresented clade were generated from the SNP multiple sequence alignments obtained from mapping to the internal reference, using RAxML (version 8.2.8) using the GTR model with gamma correction for among site rate variation with AVX vector instructions, and replicated for 100 bootstraps (Stamatakis *et al.*, 2014).

Genomic islands were identified from the genomic assemblies using the AlienHunter (version 1.7) software (Vernikos & Parkhill, 2006) using default parameters. Putative genes within each genomic island were then extracted from the annotation files for each isolate and investigated manually.

To identify evidence of phage within the isolate genomes, all assembly files were compared with the PHASTER database (Arndt *et al.*, 2016) using the API. The record for each isolate was downloaded, and complete phage annotations were extracted.

Virulence factor genes were identified by mapping the isolate sequencing files, as in section 2.2.1, to the *P. aeruginosa*-specific virulence genes available in the Virulence Factor DataBase (Liu *et al.*, 2019). This consisted of 239 *P. aeruginosa* genes, as of April 2019.

Dating analysis was carried out using the R package BactDating (version 1.0.1), which calculates time to most-recent common ancestor using MCMC-based Bayesian dating of the nodes of the phylogeny (Didelot *et al.*, 2018). The phylogeny calculated for each overrepresented ST in section 2.2.3 was dated using this method, by calculating 10,000,000 MCMC chains so that all ESS values were over 100, for three separate evolutionary models; strict gamma, relaxed gamma, and mixed gamma. The consensus date

was defined as the range covered by every repeat of all models, and the range was defined as the oldest date to the newest date reported for every repeat of all models.

2.2.5 AMR profiling

Phenotypic antimicrobial susceptibility testing was performed using the Vitek2 instrument with the Vitek N210 card (for isolates collected between 2006-2013), and the Vitek N352 card (for isolates collected between 2017-2018) (Rachel Bousfield, Peacock Lab, University of Cambridge).

The genetic AMR profiles of each isolate were calculated using ARIBA (version 2.12.1), with reference genes and reference SNPs downloaded from the Comprehensive Antimicrobial Resistance Database (version 3.0.1) (Jia *et al*., 2017). The presence/absence of each AMR gene or gene variant was then displayed against each patient phylogeny, as in section 2.2.3.

2.3 Results

2.3.1 Isolate selection and DNA sequencing

In order to investigate the evolution of *P. aeruginosa* bacteraemia over local and national scales, a sampling strategy was devised to target one local area, which included 273 isolates from the Cambridge collection, and 79 isolates from the BSAC collection of isolates from the UK and Ireland (Table 2.2). Phenotypic AMR assays had previously been carried out, in order to understand which therapies might be suitable for these infections.

The national BSAC collection consisted of isolates submitted to a Bacteraemia Resistance Surveillance Programme (www.bsacsurv.org) between 2001 and 2011, by 25 hospitals across the UK and Ireland. Only isolates that were determined to be Multi-Drug Resistant (MDR) were included in the study. The

Year of isolation	Number of BSAC Isolates	Number of local isolates	Total Number of Isolates
2001	8	0	**8**
2002	3	0	**3**
2003	2	0	**2**
2004	14	0	**14**
2005	9	0	**9**
2006	4	1	**5**
2007	9	20	**29**
2008	10	29	**39**
2009	6	27	**33**
2010	8	41	**49**
2011	6	40	**46**
2012	0	50	**50**
2013	0	5	**5**
2017	0	19	**19**
2018	0	41	**41**
Total	**79**	**273**	**352**

Table 2.2 Year of isolation and collection from which the isolates in this study were obtained.

local collection isolates were sourced from Cambridge University Hospital NHS Foundation trust, Cambridge, UK. The first isolate from every patient that tested positive for *P. aeruginosa*-associated bacteraemia, regardless of MDR status from 2006 - 2013 and 2017 - 2018 was included in the study.

Isolates from both collections were sequenced using Illumina HiSeq 2000 and 2500 platforms. The average length of each assembled genome was 6,712,355 bp (range 6,140,300 - 7,960,112 bp), with an average number of contigs of 63 (range 18 - 1886 contigs). The mean N50 value was 420,958 bp (range: 10,925 - 1,730,489 bp). The mean largest contig was 928,087 bp (range 79,234 - 2,203,947 bp). All isolates were also mapped to the reference *P. aeruginosa* strain, PAO1. The mean mapping coverage was 96.3% (range 37.8 - 99.2%), with a mean depth of 69.5 (range 36.4 - 100.2). A mean of 34,245 SNPs per sequence were called against the PAO1 reference (range 6,583 - 71,216 SNPs per sequence).

2.3.2 MLST prevalence

Initially, genomes were classified based on MLST, in order to investigate how population structure differs between local and national isolates. MLST assignment is a method used to distinguish different subtypes of the same bacterial species, by comparing the different allelic profiles of seven housekeeping genes. For *P. aeruginosa*, these housekeeping genes are *acsA*, *aroE*, *guaA*, *mutL*, *nuoD*, *ppsA* and *trpE* (Jolley & Maiden, 2010). Each combination of alleles results in a numbered ST which is unique to that allelic profile (Jolley & Maiden, 2010). The MLST of all isolates from both collections were determined (Figure 2.1). MLST profiles that occurred once within the dataset were recorded as part of the "singletons" group.

In total, 134 unique MLST profiles were identified from the 352 isolates of the study. Of these, 92 MLST profiles occurred only once in the dataset. Five of the MLST profiles were present in ten or more of the isolates in this study. These were ST395 (n = 32), ST253 (n = 30), ST175 (n = 22), ST244 (n = 12), and ST207 (n = 10).

The three most prevalent STs in the dataset, ST395, ST253, and ST175, are international and high-risk clones, which are frequently associated with multi-drug resistance and epidemics around the world (Petitjean *et al.*, 2017). Hospital-acquired ST395 infections are particularly prevalent in France and the UK, and can be transmitted to at-risk patients through contaminated water supplies (Quick *et al.*, 2014).

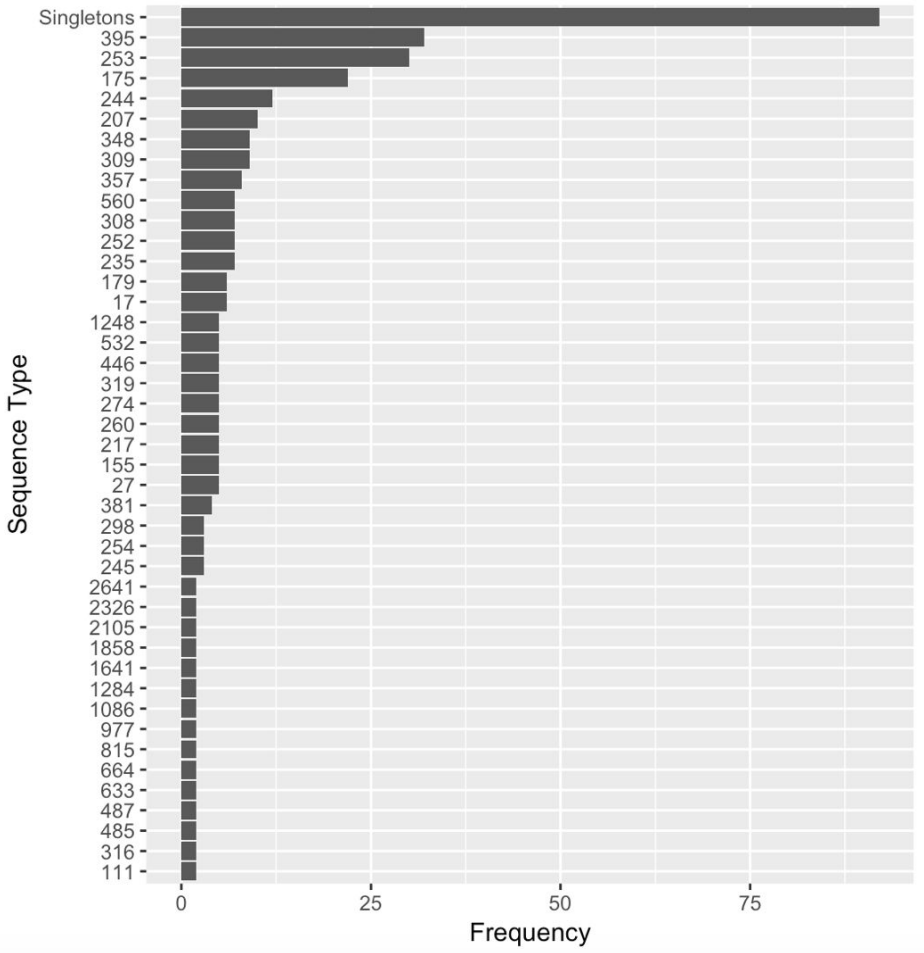


Figure 2.1 Frequency of MLST profiles for the 352 *P. aeruginosa* isolates. In total, 134 unique MLST profiles were observed, and ST profiles that were only observed once were recorded as part of the "singletons" group.

ST175 is an important European strain that has been known to cause frequent MDR hospital-acquired infections in Spain and France. ST175 has acquired AMR through mutational events, rather than through the acquisition of AMR genes, which is unusual for international and high-risk clones (Cabot *et al.*, 2016). ST253 is the ST to which the *P. aeruginosa* reference strain PA14 has been assigned (Treepong *et al*., 2018). ST253 has been associated with very virulent CF infections, global animal infections, and is frequently sampled from the environment sources (Ruiz-Roldan *et al.*, 2018). Four other international and high-risk clones, ST244, ST111, ST235, and ST357, are also present in this study (Treepong *et al*., 2018). ST244 is a globally spread sequence type that is often associated with extensive drug resistance (Fan *et al*., 2016; Barrio-Tofino *et al*., 2017). ST111 is well dispersed across Europe, and is a prevalent epidemic strain in the UK. The low frequency of ST111 (n = 2) in this dataset is fortunate, as ST111 is associated with the presence of metallo-β-lactamase enzymes that lead to resistance to the antibiotics-of-last-resort,

the carbapenems (Golle *et al.*, 2017). The ST357 strain has also become a major European source of carbapenem-resistant infections, and has now spread all over the world (Papagiannitsis *et al.*, 2017; Kainuma *et al.*, 2018). The ST235 strain emerged in Europe in the early 1980s, and is associated with the introduction of fluoroquinolones (Treepong *et al.*, 2018). It has since developed resistance to many other common antibiotics and has spread around the globe. The ST207 strain, which is present as the fifth most frequently-occurring ST (n = 10) in this dataset, is a globally dispersed strain, but is not typically MDR (Gomila *et al.*, 2013).

The frequency of isolates belonging to the five most prevalent STs is plotted against time in Figure 2.2. The ST175 isolates form the majority of the collection between 2002 and 2005. Subsequently, the rate of isolation of ST175s declines until 2011, after which no further ST175 isolates were collected. The majority of ST175 isolates were observed as part of the BSAC collection, with only one ST175 isolate observed as part of the local collection.

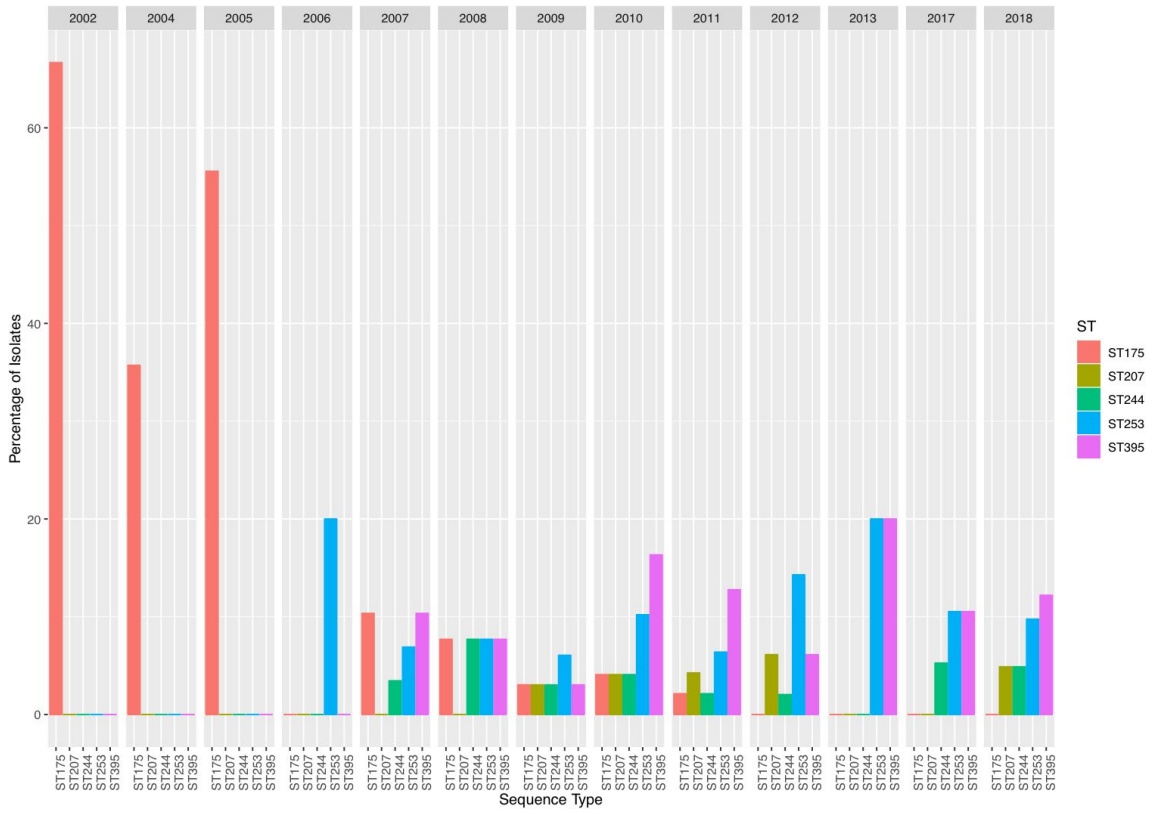


Figure 2.2 The five most prevalent STs, as percentage of isolates from each year when at least one of the STs were present.

Between 2006 and 2018, the epidemic strains ST253 and ST395 become the dominant clones in the CUH collection. These are the only STs that were consistently collected through every year across the CUH collection period, which suggests that reservoirs of these STs contribute to patient infection within the hospital, or that they are particularly associated with community-acquired infections within the Cambridge area.

After 2005, no single high-risk clone is present as more than 20% of the collection per year. Combined, the top five most prevalent STs form 15 to 40% of the collection per year, suggesting diverse sources of infection that are not dominated by high-risk clones.

2.3.3 Population phylogeny

The SNPs called against the *P. aeruginosa* PAO1 reference genome were used to investigate the population structure of isolates within the study (Figure 2.3), in order to provide a visual representation of the population structure of the local isolates within the context of the national isolates. The population of the isolates split at a basal node into two distinct clades; PAO1-like and PA14-like. This broad population structure is typical of *P. aeruginosa*, which resolves into a non-clonal epidemic structure, in which successful clones sometimes emerge (Pirnay *et al*., 2009). This dataset supports this observation, and indicates that the infections investigated in this study were not caused by a single strain.

The structure of the phylogenetic tree clearly identifies three major clonal expansions in this study. The ST175 clade consists primarily of isolates from the BSAC collection, with only one isolate from the local collection. High prevalence of ST175 in the UK-wide MDR surveillance collection indicates that ST175 is circulating as a high-risk MDR epidemic clone in many UK hospitals. The ST253 and ST395 clones were found to be a source of major infection within the CUH collection, with only one isolate of both STs identified as part of the BSAC collection. In contrast to ST175, this indicates that the ST253 and ST395 isolates are not associated with UK-wide MDR infection, but with local infection, suggesting a reservoir within the local environment. Other than the ST175 isolates, there is one additional BSAC-specific clade, which consists of eight isolates belonging the ST357 strain. One previous study has identified ST357 as a major European source of carbapenem-resistant infections between 2011-2017 (Papagiannitsis *et al.*, 2017). In this dataset, ST357 appears sporadically across the UK and Ireland only between 2001-2009,

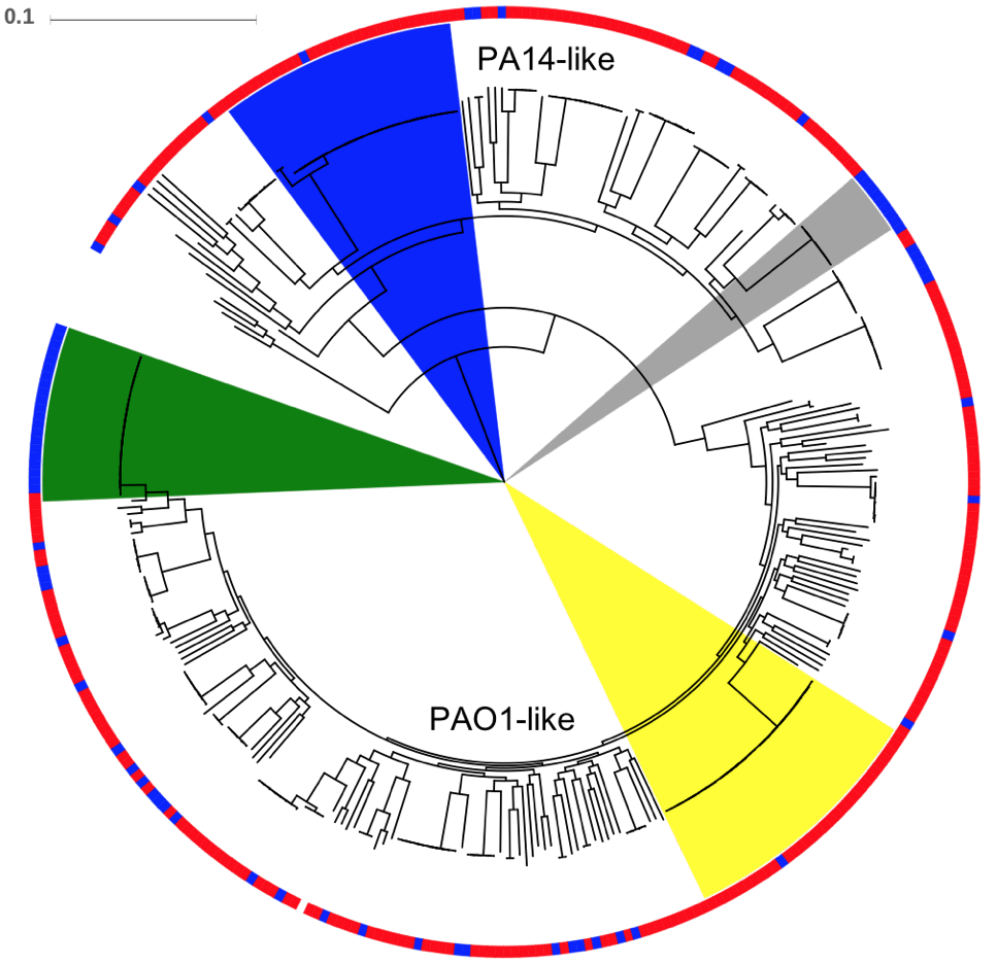


Figure 2.3 A Maximum-likelihood tree of 352 isolates in this dataset. The overrepresented STs are highlighted in yellow for ST395, green for ST175, blue for ST253, and grey for ST357. The outer ring indicates the source of the isolates, with red isolates from the local collection, blue from the BSAC collection, and white for the *P. aeruginosa* PAO1 reference genome.

and isn't present in the local collection. However, they are associated with carbapenem resistance, with 75% of ST357 isolates phenotypically non-susceptible to either Imipenem or Meropenem (Section 2.3.5). Other than the ST175 and ST357 strains, the isolates from the BSAC collection are spread sporadically throughout the population, indicating that the UK state of *P. aeruginosa* bacteraemia infections are diverse, and not limited to one UK epidemic strain.

Pairwise SNP distances were plotted for the isolates in the PAO1-like and PA14-like clades (Figure 2.4a). The majority of PAO1-like isolates (93%) are separated by 15,000 to 25,000 SNPs, whilst the majority of

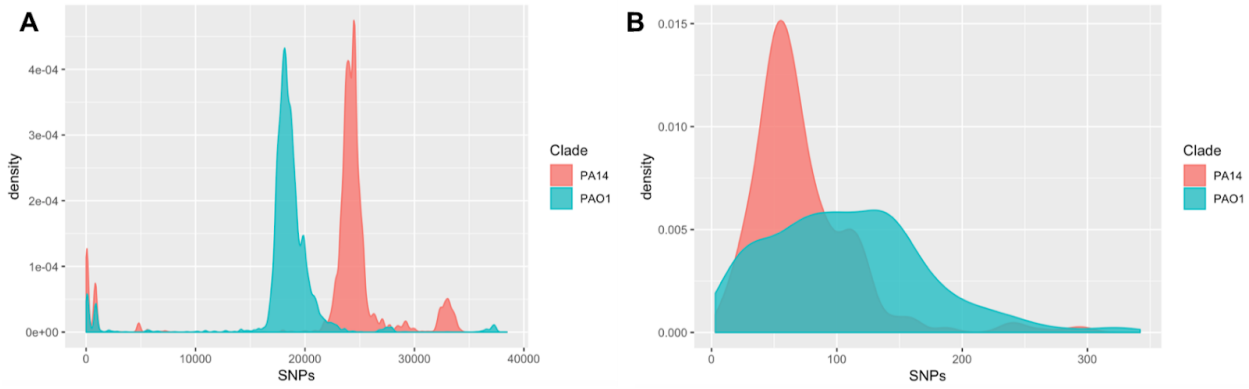


Figure 2.4 a) Histogram of pairwise SNP matrix separating the PA14-like isolates (red) and the PAO1-like isolates (green). b) Isolates sharing fewer than 400 pairwise SNPs were defined as belonging to the same clade, which shows that PA14-like isolates that form part of the same clade (red) are more closely-related to each other than the PAO1-like isolates that form part of the same clade (green).

PA14-like isolates (84%) are separated by 20,000 and 30,000 SNPs. Closely-related isolates were defined as belonging to the same clade if they were related by 400 SNPs or less (Figure 2.4b). The closely-related PA14-like isolates have a mean pairwise SNP distance of 72 SNPs, and the closely-related PAO1-like isolates have a mean of 108 SNPs. A greater proportion of the PA14-like isolates belong to closely-related clades of two or more isolates (86% compared to 66% for the closely-related PAO1-like isolates). This suggests that there have been a greater number of more-recent outbreaks and clonal expansions of the PA14-like isolates than of the PAO1-like isolates.

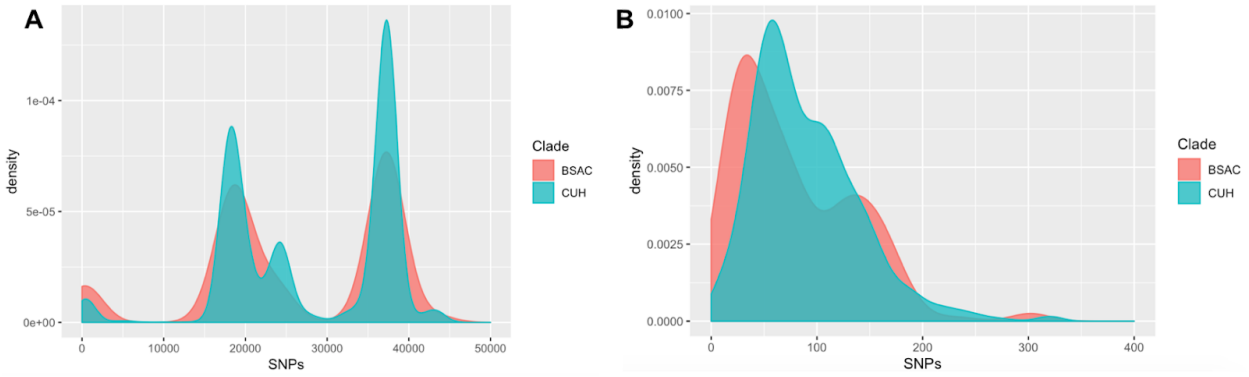


Figure 2.5 a) Pairwise SNPs of each of the BSAC collection and each of the CUH collection isolates. b) Isolates sharing fewer than 400 pairwise SNPs were defined as belonging to the same clade. This shows the BSAC isolates are more closely-related to each other than the CUH isolates.

The same pairwise analysis was repeated to compare the CUH and BSAC datasets (Figure 2.5). This shows clearly the split into PAO1 (15,000 to 30,000 SNPs) and PA14-like (30,000 to 50,000 SNPs) clades (Figure 2.5a). In general, the BSAC isolates form more-related clades (mean = 79 SNPs) than the CUH isolates (mean = 90 SNPs) (Figure 2.5b).

The pangenome of the dataset demonstrated an open pangenome containing 32,818 genes, which did not reach a plateau (Figure 2.6). A total of 4,453 core genes were present in 99% or more of the isolates (13.4%), and 2,505 genes were found to be present in all 352 isolates in the dataset (7.6%). Of the 28,365 genes in the accessory genome, 11,667 genes were found in only one isolate. The large number of accessory genes indicate high diversity and flexibility of the *P. aeruginosa* genome. The 2,505 genes present in every isolate is almost identical to the Mosquera-Rendon *et al.*, 2016 study of the *P. aeruginosa* pangenome, in which 2,503 of the genes were present in all 181 isolate. However, an updated 2018 *P. aeruginosa* pangenome study of 1,311 isolates identified only 665 genes shared across all isolates in the population (Freschi *et al.*, 2019). The study by Freschi *et al*. included a group of PA7-like strains of *P. aeruginosa*, which were not identified in either the Mosquera-Rendon *et al*. study, or found as part of this study. *P. aeruginosa* PA7 is a taxonomic outlier of *P. aeruginosa*, which diverged early in the evolution of *P. aeruginosa* and has been shown to cause acute infection in rare cases, but lacks many virulence genes present in PAO1-like and PA14-like strains (Roy *et al.*, 2010).

Figure 2.6 The number of unique genes in the dataset plotted against number of genomes. This indicates an open pangenome with no signs of reaching a plateau.

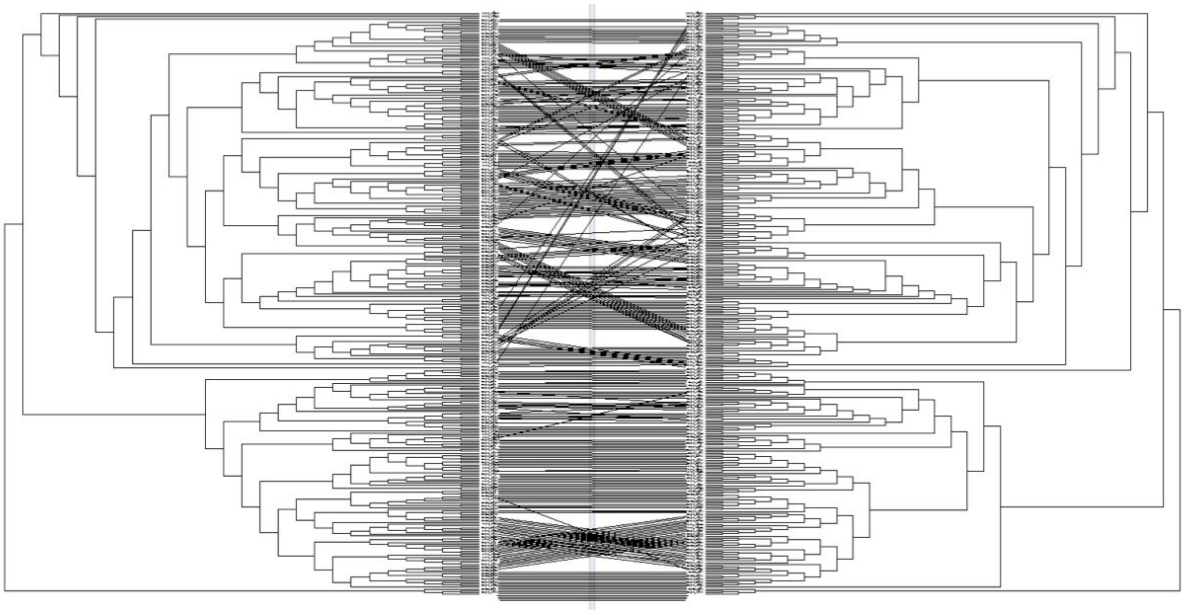


Figure 2.7 A dendroscope comparison of the cladograms made by mapping the 352 isolates to PAO1 (left) and the core genome (right), which shows good congruence.

A core genome phylogeny was estimated from the 4,453 genes that made up the core genome. The core-genome phylogeny was compared with the phylogeny generated from SNPs called against PAO1, which showed a good correlation between the two techniques (Figure 2.7). The presence or absence of genes in the accessory genome was compared against the core genome phylogeny. There was found to be significant congruence between the phylogenies of the core genome and the accessory genome (Mantel test, $p < 0.001$), suggesting that there has been limited exchange of accessory genes between STs.

Distinct patterns of gene presence in the accessory genome were observed for the high-risk and international STs in the dataset, ST175, ST253, and ST395. These patterns of gene presence were attributed to the presence of phage and genomic islands within each ST that are not present in the rest of the dataset (see section 2.3.4.2 and section 2.3.4.3).

2.3.4 Investigating overrepresented clades

In order to describe the three clades that were over-represented, and therefore dominant, in the Cambridge and national collections, the MLST of the over-represented strains was determined. The three overrepresented strains were international and high-risk STs ST395, ST253 and ST175. These

globally-disseminated strains are considered epidemic, and have been associated with multi-drug resistance (MDR) and extensive drug resistance (XDR) in healthcare settings (Treepong *et al.*, 2018), and are often resistant to the carbapenem antibiotics of last resort (Buerhle *et al.*, 2016).

2.3.4.1 Sequencing statistics

The assembly lengths of the overrepresented STs were compared with the assembly lengths of the rest of the isolates in the study (Figure 2.8). For all three overrepresented STs, the average genome length was significantly higher than the average genome length of the other isolates in the dataset, and greater than the average genome length of the dataset as a whole (Mann-Whitney test for independence, $p < 0.05$).

The core genomes of the overrepresented STs were also calculated (Table 2.3) and the number of core genes were not found to be correlated with increased assembly length. The ST395 isolates, which have the longest genomes, have 5,542 core genes. The ST253 isolates, which have the shortest genomes, have 5,395 core genes. The ST175 isolates have the most core genes, with 5,913 core genes. This suggests that

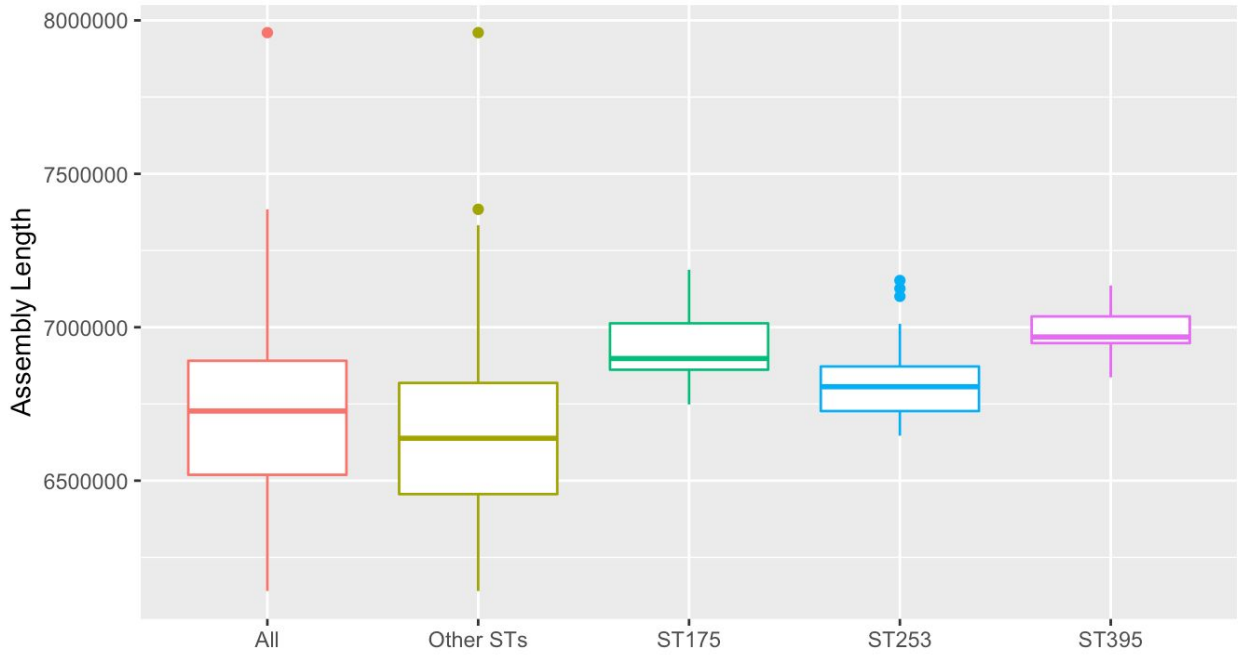


Figure 2.8 Assembly lengths of the genomes for all STs in the dataset, all STs minus the overrepresented clades, and the overrepresented clades ST175, ST253 and ST395.

	ST175	ST253	ST395
Core genes	5,913	5,395	5,542

Table 2.3 Core genomes of the overrepresented isolates.

the ST253 collection is more diverse than the ST395 and ST175 collections, whilst the ST175 isolates are closely related.

The average GC content for the dataset was 65.8%, which is similar to published *P. aeruginosa* GC content of 65-67% (Klockgether *et al.*, 2011). The range within this dataset is 6.0%, with a minimum of 62.3% to a maximum of 68.3%. The mean GC content for ST175, ST253, and ST395 differ from that of the overall dataset by -0.7%, -0.9% and -0.2% respectively (Figure 2.9). This difference was significant ($p < 0.05$) between the ST175 isolates and the whole dataset, and between the ST253 isolates and the whole dataset. The ST395 GC content was found to be not significantly different from the whole dataset. Lower GC content can be an indicator of a large number of genomic islands (Feil *et al.*, 2005) or the presence of phage within the genomes (Kwan *et al.*, 2006).

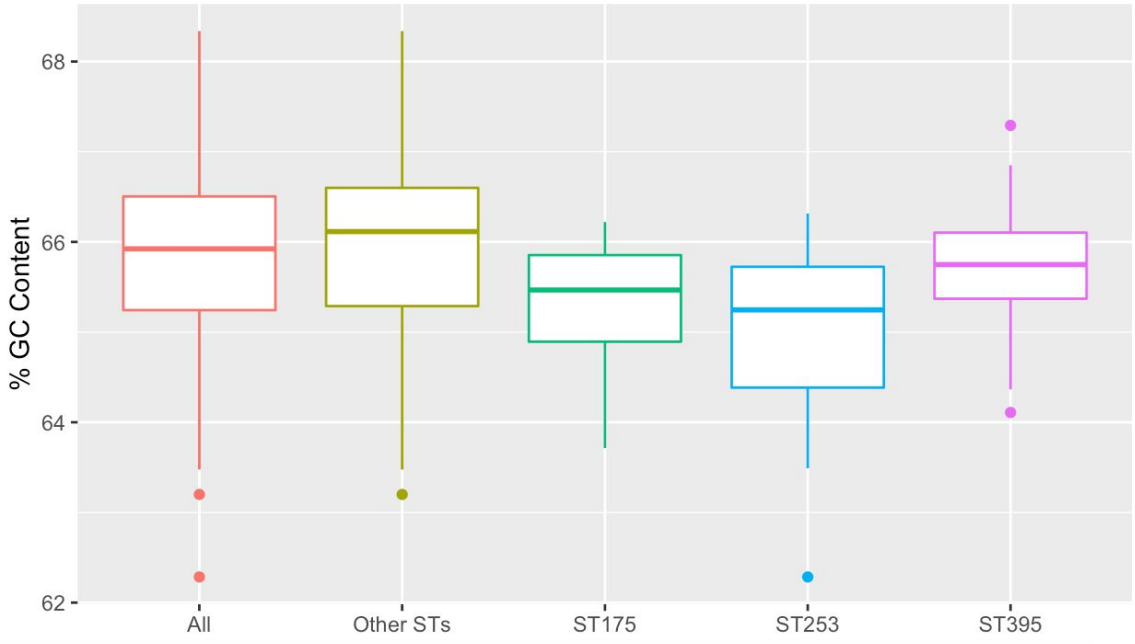


Figure 2.9 Mean GC content of the genomes for all STs in the dataset, all STs minus the overrepresented clades, and the overrepresented clades ST175, ST253 and ST395.

2.3.4.2 Genomic islands

The higher assembly lengths and the lower GC content of the three overrepresented STs were significantly correlated ($p < 0.05$) with an increased prevalence of genomic islands within the isolates (Figure 2.10). The average number of genomic islands was 61, 60 and 64 for ST175, ST253 and ST395, respectively. This was higher than the average number of genomic islands present in all other STs, which was 45. International and high-risk clones typically contain a higher number of genomic islands, which can encode for a large number of virulence and resistance genes (Jani *et al.*, 2016).

A total of 1,495 genes within genomic islands were identified within the three overrepresented STs. Of these, 950 genes were unique to one of the three STs. The ST395 isolates had the highest prevalence of genomic islands on average but had the least diversity amongst genomic islands, with 267 unique genes and 392 hypothetical proteins of unknown function, across all isolates. Petitjean *et al.* also found that ST395 harboured more genomic islands than other international and high-risk clones (Petitjean *et al.*, 2017). The genomic islands within the isolates of the Petitjean *et al*. study were linked to copper resistance, which was confirmed phenotypically. The authors suggested that the presence of these

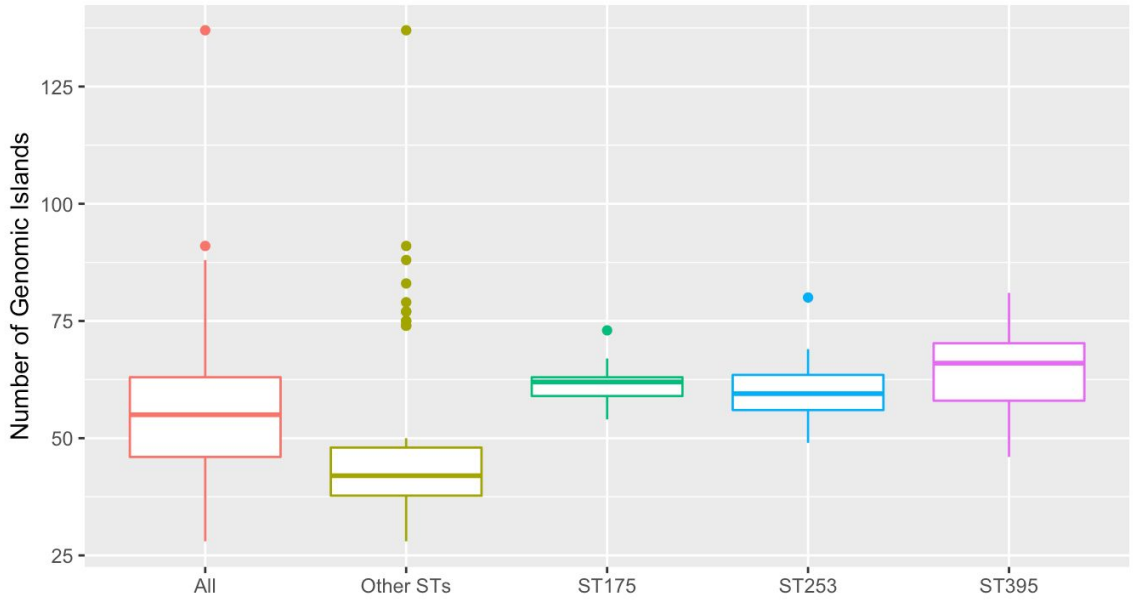


Figure 2.10 Number of genomic islands identified in the genomes of all STs in the dataset, all STs minus the overrepresented clades, and the overrepresented clades ST175, ST253 and ST395.

genomic islands accounted for the frequent hospital-related ST395 outbreaks in their study, where copper-lined water networks have been compromised and favoured the spread of ST395 in hospital settings. Copper resistance genes were not identified in any of the ST395 genomic islands within this study.

The ST253 isolates harboured the fewest number of genomic islands on average, but the greatest diversity within the genomic islands, with 804 unique genes and 458 hypothetical proteins across all isolates. The ST175 isolates harboured 424 unique genes and 468 hypothetical proteins across all isolates. This greater diversity of genomic islands in the ST253 isolates may reflect a greater sampling diversity of unrelated sources of infection for these STs.

The majority of genomic island gene content found within all three of the STs falls into two distinct gene functions; metabolism, and survival. As a whole, genomic islands were found to be related to iron uptake, phage defence, Type VI secretion, and motility. A high number of alcohol dehydrogenase enzymes, and a high number of genes with predicted virulence functions were also present in the genomic islands of all isolates of the three STs.

Of the genomic islands that are unique to ST253, which has the most variation in genomic island gene content from the three STs, several imported resistance genes from other species were found to be present. These included chloramphenicol-resistance variants, *catI* and *catM*, which are transposon-encoded resistance genes typically found in *E. coli* and *Acinetobacter baumannii* (van Hoek *et al.*, 2011). Additionally, a homologue of the *nylB*' beta-lactamase enzyme, and a copy of the *rcn* efflux system that provides resistance to Nickel commonly found in *E. coli* and *A. baumannii*, were also identified in these isolates (Bleriot *et al*., 2011). Other genes related to the transport, efflux, and resistance of other heavy metals including mercury, cobalt, zinc and cadmium were also present in the ST253 isolates.

Some of the genomic islands unique to ST175 were similarly found to confer resistance to heavy metals. Most notably present are the *copA*, *copB*, *copC*, *copK*, *copR*, and *copZ* genes, which confer resistance to copper. Additional genes related to copper storage and copper tolerance were also identified, which were absent in the other STs. There was no evidence of horizontal gene transfer from distantly-related species, and only one unique genomic island originating from the closely-related *Pseudomonas putida* (Lee *et al.*, 2001), in which three resistance genes to cadmium were obtained.

Only five genes were found to be unique to the ST395 genomic islands, which include an aldehyde dehydrogenase, a glucose dehydrogenase, two genes identical to *Rhizobium meliloti* secretion enzymes that are homologous to the *P. aeruginosa* Type I secretion system genes (Russo *et al*., 2006), and a gene which is predicted to be a regulator of stress response. The genomic islands unique to ST395 did not encode resistance to antimicrobial compounds or heavy metals.

2.3.4.3 Phage identification

The higher assembly lengths and the lower GC content of the three overrepresented STs are an indicator of phage within the genomes (de Brito *et al.*, 2016). The greatest diversity of phage was found in the ST253 isolates, in which 13 different phage were present in at least one isolate (Figure 2.11). Seven different phage were found in the ST395 isolates, and five different phage in the ST175 isolates.

YMC11/02/R656 is a previously described, but as yet uncharacterised, *P. aeruginosa* bacteriophage that was present in 100% of the ST253 isolates. All other phage appeared sporadically throughout the ST253 phylogeny, with the exception of the *P. aeruginosa* bacteriophage JBD88a, which is associated with one

Figure 2.11 The presence of each phage is displayed next to the ST253 phylogeny. Purple indicates phage presence, white indicates phage absence.

Figure 2.12 The presence of each phage is displayed next to the ST395 phylogeny. Purple indicates phage presence, white indicates phage absence.

subclade of ST253. The JBD88a phage has anti-CRISPR capabilities in *P. aeruginosa*, to ensure the phage survival within a bacterium (Bondy-Denomy *et al.*, 2016).

Interestingly, two phage typically identified outside of the *Pseudomonas* genus were present within the ST253 isolates; Enterobacteria lambda phage mEp235 was present in two isolates, and *Stenotrophomonas* siphovirus phage S1 was present in one isolate. *Stenotrophomonas* is a bacterium that frequently co-occurs with *P. aeruginosa*; it forms multi-species biofilms with *P. aeruginosa*, in which *Stenotrophomonas* substantially influences the architecture of the *P. aeruginosa* (Ryan *et al.*, 2008), suggesting a plausible source of acquisition for this phage.

No phage were found in every ST395 isolate (Figure 2.12). The phage with the highest prevalence was the *Pseudomonas* phage PMG1, which was found in 88% of ST395 isolates. The phi-CTX phage encodes a pore-forming toxin, which enhances virulence in a number of infection models, was present in 69% of isolates (Kung *et al.*, 2010). The Pf1 phage is present in 56% of the ST395 isolates, the presence of which has been shown to result in a less virulent *P. aeruginosa* more able to evade the immune response, promoting chronic infection (Secor *et al.*, 2016).

Figure 2.13 The presence of each phage is displayed next to the ST175 phylogeny. Purple indicates phage presence, white indicates phage absence.

A much less diverse range of phage was present in the ST175 genomes, with only five different phage present within the isolates (Figure 2.13). Three of these phage, PMG1, YMC11/02/R656, and phi-CTX, are present in the majority of ST175 isolates (86%, 91% and 73%, respectively) and have been previously described in the ST253 and ST395 isolates. There are two occurrences of *P. aeruginosa* phage D3, which is genetically similar to the PMG1 phage and only occurs in isolates that do not contain the PMG1 phage. There is one occurence of the *Pseudomonas* phage phi-2, which has been shown to co-evolve with the virulent mucoid phenotypic state of *Pseudomonas* (Scanlan & Buckling, 2012).

2.3.4.4 Virulence factors

A total of 239 *P. aeruginosa* virulence genes from the Virulence Factor Database (Liu *et al.*, 2019) were screened against the isolates in this dataset. These genes encode for virulence phenotypes such as motility, secretion systems, and toxin production. The mean number of virulence genes found within the genomes of the overall dataset was 219 (range: 203 - 235) (Figure 2.14).

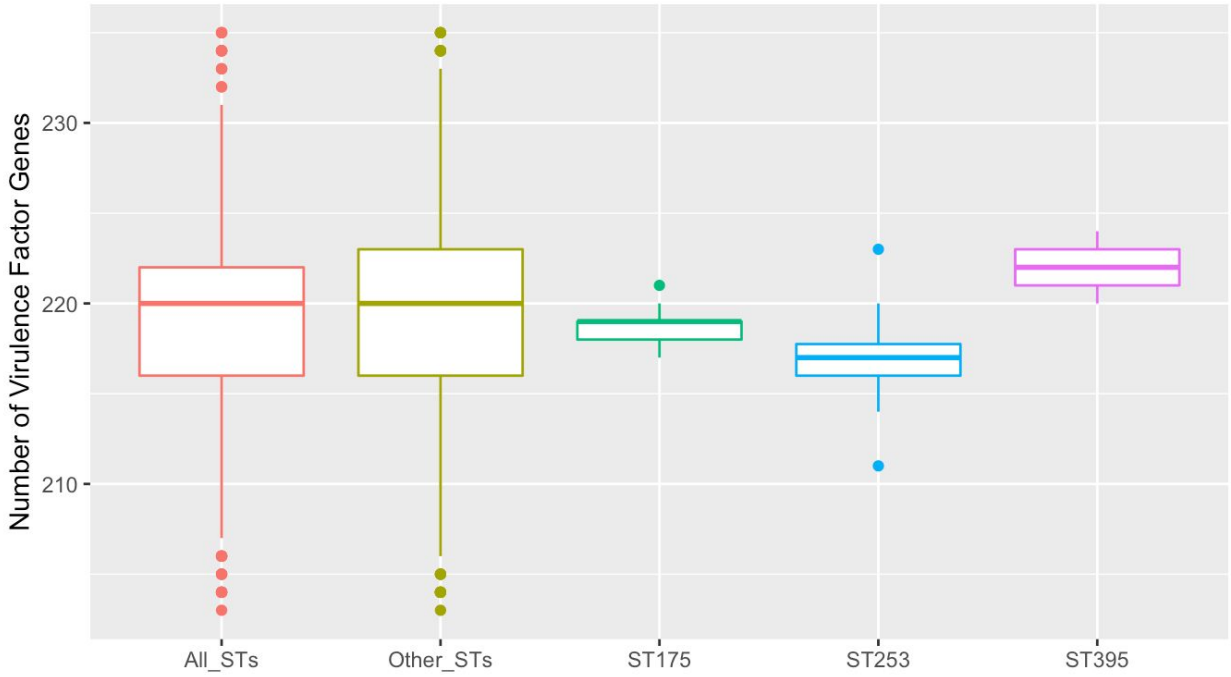


Figure 2.14 The number of virulence genes identified in the overrepresented STs, compared to the number of virulence genes in all other STs, and the whole dataset.

Of the three overrepresented STs, the ST253 isolates had the lowest number of virulence genes on average, which was significantly lower than the overall dataset (Wilcoxon ranked sum, $p = 0.002$). The ST395s had a slightly higher number of virulence genes compared to the overall dataset (Wilcoxon ranked sum, $p = 0.001$), and the ST175 isolates did not have a significantly different number of virulence genes compared to the overall dataset (Wilcoxon ranked sum, $p > 0.05$).

Within the ST175 isolates, 213 (89%) of the identified virulence genes were present in every isolate. These include the Type I, Type II and Type III secretion systems, responsible for cell invasion and survival in the host (Filloux, 2011). The Type VI secretion system, which delivers protein effectors directly into neighbouring pathogens and host cells (Chen *et al.*, 2015), is only completely present in 3 of the isolates (13.6%), and partially present in a further 13 isolates (59.1%). All ST175 isolates encode ExoS, but not ExoU, which is typical of PAO1-like isolates (Pazos *et al.*, 2017). The gene for *lasB* protease, which is secreted via the Type II secretion systems (Saint-Criq *et al.*, 2017), is present in 21 of the 22 isolates (95%). The remaining *las* genes, part of the quorum sensing pathway, are also present in 95% of the isolates, with the isolate missing *lasB* also missing *lasI*. Within the ST175 dataset, only 14 of the identified virulence genes (6%) were completely absent. Of the overrepresented STs, ST175 is the

only ST where the rhamnolipid production system, required for biofilm production and for invasion through the airway, is absent (Davey *et al.*, 2003). This leads to a *P. aeruginosa* bacterium with reduced virulence compared to wild-type (Wittgens *et al.*, 2017). All of the ST175 isolates contain all of the genes responsible for the twitching motility, but were missing 7 out the 44 genes screened for flagella-dependent swimming motility.

Of the ST253 isolates, 201 (84%) of the identified virulence genes were present in every isolate. As with the ST175 isolates, Type I, II and III secretion systems are present in 100% of the ST253 isolates, and the Type VI secretion system is partially present in some isolates. Unlike ST175, the ST253 isolates are located within the PA14-like clade of the *P. aeruginosa* phylogeny, which typically encode ExoU, but not ExoS (Pazos *et al*., 2017). Fittingly, 28 of the ST253 isolates encode the ExoU toxin. However, one isolate encodes both ExoS and ExoU, and one isolate did not encode either ExoU or ExoS. In contrast to the ST175 isolates, 100% of the ST253 isolates encode for the quorum sensing regulators *lasI, lasR, rhlI* and *rhlR*. The ST253 isolates encode for more of the flagella-dependent motion genes, with only 4 out of the 44 genes absent from the isolates. However, the twitching motility machinery was only partially present in the ST253 isolates; of the 24 genes of the Type IV pili system, eleven of the genes were absent. Non-motile bacteria have previously been shown to have attenuated virulence (Li *et al.*, 2015a). All genes required for alginate and rhamnolipid production were present, although three of the isolates are missing the alginate transcriptional regulator *algP*. Only nine of the identified virulence genes (4%) were completely absent from all ST253 isolates.

For the ST395 isolates, 214 (90%) of the identified virulence genes were present in every isolate. The most striking observation within the ST395 dataset was a 15kbp deletion in the flagella-dependent motion region, which covers the virulence genes *flgK*, *flgL*, *fgtA*, *fliC*, *flaG*, *fliD*, *fliS* and *fleP*, responsible for flagella-dependent motion. This deletion was also present in the study of *P. aeruginosa* ST395 by Petitjean *et al*., but the isolates in that study also had an additional 131 kbp deletion of secreted proteases and associated secretion machinery, as well as biofilm dispersion regulatory genes (Petitjean *et al.*, 2017), all of which are present in the isolates in this collection. Absence of flagella proteins has been linked to early biofilm defects (Leid *et al.*, 2009), but also a decrease in mortality in acute infections (Campodonico *et al.*, 2009).

The pattern of other gene presence/absence for ST395 is similar to the ST253 isolates. The Type I, II and III secretion systems are present in 100% of the isolates. Six isolates also contain complete Type VI

secretion systems, and 14 of the isolates contain partial Type VI secretion system genes. There are also three Type IV pili genes required for twitching motility that are absent, however, the alginate and rhamnolipid pathways are completely present. The ST395 isolates reside within the PAO1-like clade of the *P. aeruginosa* population structure, and therefore only the ExoS toxin is present in 100% of the isolates.

2.3.4.5 Pairwise SNP distances and transmission

The pairwise SNP distances between the isolates belonging to each of the overrepresented STs were calculated to investigate the relatedness of the isolates within each ST (Table 2.4, Figure 2.15). Three of the ST253 were significantly different to the rest of the isolates (>5,000 SNPs) and were therefore removed from this analysis.

Within the ST395 isolates, there are 14 pairs of isolates that are separated by 5 SNPs or less. Four of the pairs of isolates are separated by 0 SNPs. Of these, three of the pairs are isolates sampled within 3 months of each other in two separate areas of the hospital, and form part of a closely-related clade (Clade A, Figure 2.16). However, without more detailed epidemiological data it will be difficult to identify whether this was the result of direct transmission, or a persistent reservoir of isolates within the hospital environment. The fourth pair of isolates that are separated by 0 SNPs (Clade C, Figure 2.16) were sampled two years prior to the Clade A isolates, and are unlikely to have come from the same source (pairwise difference of 80 SNPs), indicating a second introduction of ST395. Of the remaining ten pairs of isolates separated by five or fewer SNPs, six belong Clade A. These isolates were sampled from the same location within the hospital, and within 6 months of the rest of Clade A, potentially indicating they originated from the same, source. The remaining Clade A isolates were separated from other Clade A isolates by 3 SNPs, but were sampled 2 years later. It is likely that these could also have come from the

	Minimum Pairwise SNP distance	Mean Pairwise SNP distance	Maximum Pairwise SNP distance	Number of pairs of isolates separated by ≤ 5 SNPs
ST175	2	29	90	7
ST253	0	61	153	2
ST395	0	114	219	14

Table 2.4 Pairwise SNP distances between the isolates within the three overrepresented STs.

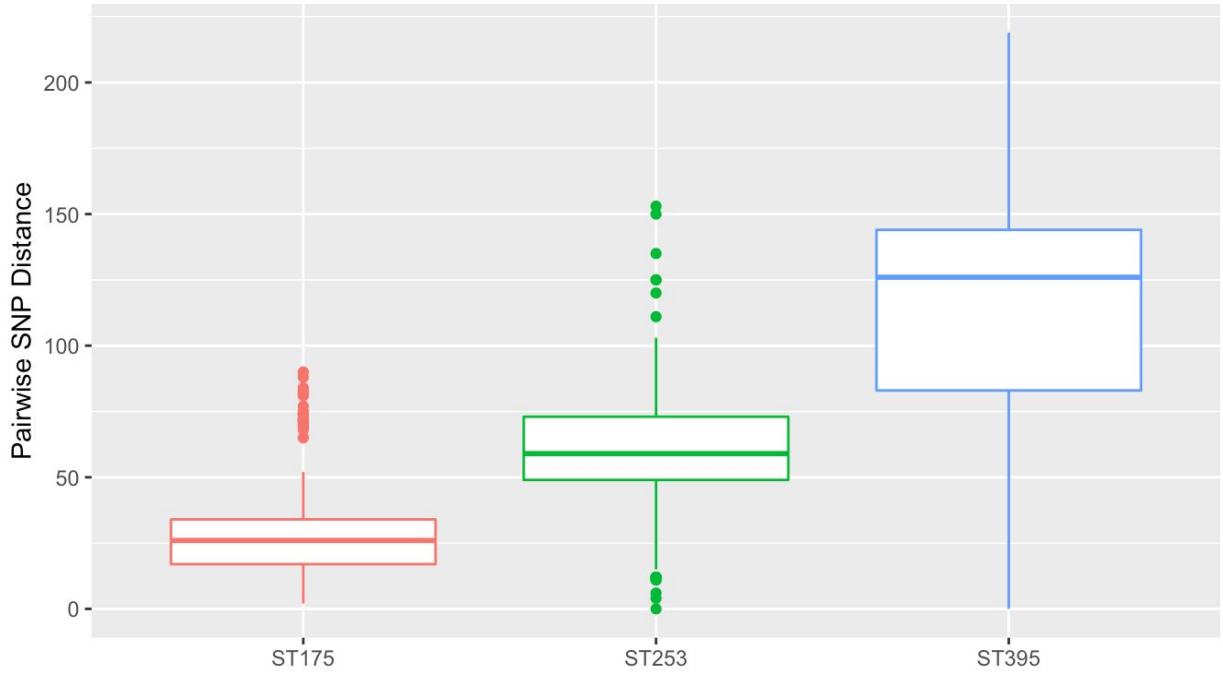


Figure 2.15 Pairwise SNP distance distribution between the isolates within the three overrepresented STs.

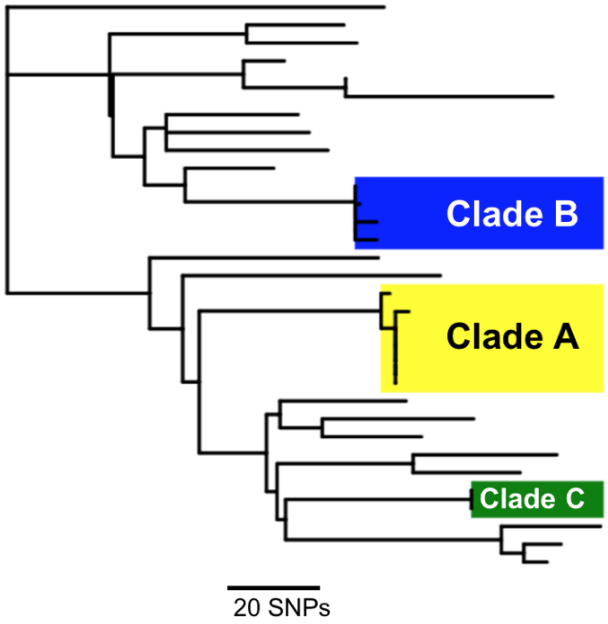


Figure 2.16 Maximum-likelihood tree of ST395 isolates. Isolates which differ by 5 SNPs or less are indicated in the coloured boxes: yellow is Clade A, blue is Clade B, green is Clade C.

same source (Marvig *et al.*, 2013). The remaining four pairs of ST395 isolates that were separated by 5 SNPs or less form their own clade (Clade B, Figure 2.16). These isolates were sampled from different sites within the hospital, over a span of two years, and differ from each other by 1-5 SNPs. Assuming normal rates of *P. aeruginosa* evolution of 1.5 SNPs per genome per year, these isolates could therefore also have come from the same source. However, they differ from Clade A and Clade B isolates by 79 SNPs and 142 SNPs respectively, indicating different reservoirs for all three clades.

Only two pairs of ST253 isolates were separated by 5 SNPs or less. One pair differed by 0 SNPs, which were isolated from different wards 1.5 years apart from each other. The second pair were separated by 4 SNPs, and were isolated over 6 years apart from each other. This suggests that there is a ST253 reservoir that may be persisting in the environment. However, it's difficult to identify any potential transmission without more detailed epidemiological data.

Seven pairs of ST175 isolates differed by 5 SNPs or less. The minimum distance between any pair of isolates was 2 SNPs, between two isolates that were collected from Bristol Royal Infirmary in 2008. This

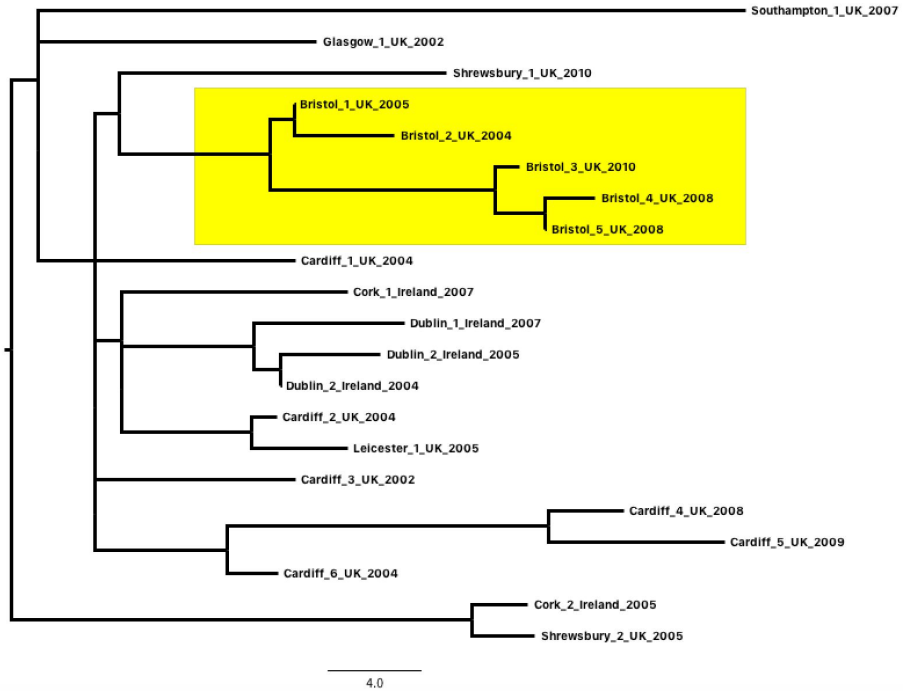


Figure 2.17 Maximum-likelihood phylogeny of the BSAC ST175 isolates from the BSAC collection, labelled by location. The yellow box highlights a local Bristol outbreak that is present in the dataset between 2004 and 2010.

suggests that these two isolates form part of the same clone in Bristol. A third isolate from Bristol Royal Infirmary in 2010 is related to these two isolates by 3 and 5 SNPs, suggesting that this sub-clade has persisted in the area between 2008 and 2010. Two further ST175 isolates from Bristol were isolated in 2004 and 2005, and differ by 3 SNPs. These two isolates share a recent ancestor with the two isolates from Bristol in 2008 and 2010, suggesting the presence of this sub-clade has persisted in Bristol since at least 2004 (Figure 2.17).

2.3.4.6 Temporal Analysis

Root-to-tip regression analysis of the ST253 isolates indicated presence of temporal signal ($R^2 = 0.22$, $p = 0.0056$), suggesting a time to MRCA of these isolates as 1985. This corresponds to the time when ceftazidime, imipenem (the first carbapenem), and aztreonam were introduced into the clinic. Bayesian dating indicated a consensus time to MRCA of 1830-1962 (median: 1572, range: 205-1970) of three evolutionary models (Figure 2.18), which is significantly earlier than the root-to-tip regression estimates. However, because the error margins are so broad, the Bayesian estimate of ST253 emergence is unlikely

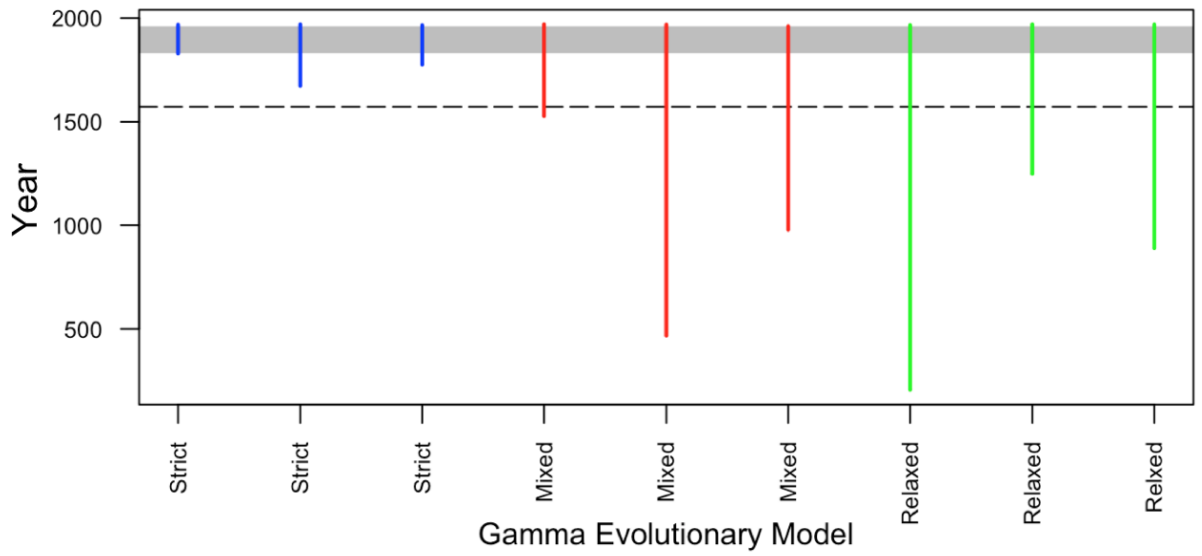


Figure 2.18 Three evolutionary models were used to estimate the time to most recent common ancestor of the ST253 isolates. The dashed line indicates the mean estimated year of emergence, whilst the grey box indicates the range of years that confidence intervals from all repeats of all models are contained within.

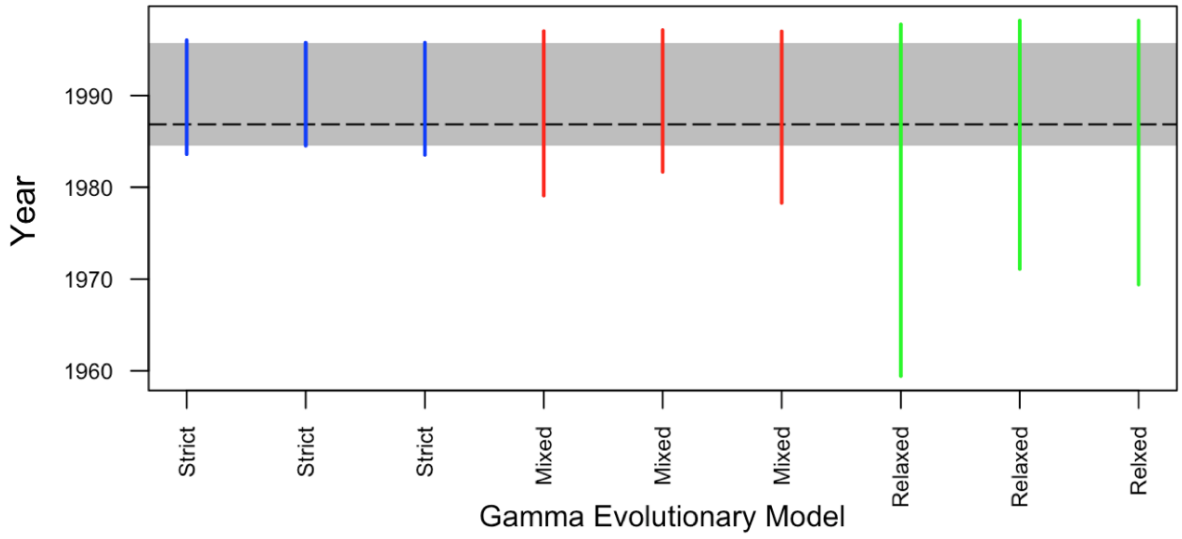


Figure 2.19 Three evolutionary models were used to estimate the time to most recent common ancestor of the ST175 BSAC isolates. The dashed line indicates the mean estimated year of emergence, whilst the grey box indicates the range of years that confidence intervals from all repeats of all models are contained within.

to be correct. It is unlikely that ST253 has been circulating in the UK since the 1830s, and if this was the case, then it would be expected that the ST253 isolates would be more represented in the UK-wide BSAC collection. In order to narrow the confidence intervals and increase the temporal signal to identify the true date of ST253 emergence, more isolates from a range of dates would need to be included.

Root-to-tip regression analysis of the ST175 clade indicated low temporal signal ($R^2 = 0.02$, $p = 0.013$), and a time to most recent common ancestor of 1921. Bayesian dating indicated a consensus time to MRCA of 1928-1975 (median: 1937, range: 1864 - 1986) of three evolutionary models. However, the temporal signal was improved significantly when the outlier ST175 isolate from the Cambridge collection was removed ($R^2 = 0.35$, $p = 0.0026$), leaving only the ST175 isolates from the BSAC collection. The root-to-tip analysis indicated a MRCA of 1995, and Bayesian dating estimated a consensus time to MRCA of 1985-1996 (median: 1987, range: 1959-1998) (Figure 2.19). This suggests that the ST175 isolate that formed part of the CUH collection is related to an ancestor of the BSAC ST175 strains, whereas the BSAC isolates form part of a very recent expansion within the UK. Between the years 1985 and 1996, several clinical antibiotics still in use today were introduced. These include ampicillin, ciprofloxacin, piperacillin/tazobactam, and meropenem, which suggests new antibiotic treatments were a driving force for the emergence of ST175.

No temporal signal was present within the ST395 isolates when the collection was mapped to a local reference. No temporal signal was present when the ST395 isolates were mapped to the PAO1 reference strain, or when the core genome was calculated. Therefore, there is no statistical relationship between the genetic diversity and temporal diversity within this ST395 dataset (Rambaut *et al*., 2016).

2.3.4.7 European context of ST253 and ST175

ST253 is a major cause of *P. aeruginosa* infections (Treepong *et al.*, 2018) and several studies have identified ST253 isolates within their collections (Hilliam *et al.*, 2017, Kos *et al.*, 2014, and Barrio-Tofino *et al*., 2017). Hilliam *et al*. identified and sequenced 20 ST253 isolates from bronchiectasis lungs from nine centers across the UK; one ST253 isolate from France was sequenced by Kos *et al*; and one ST253

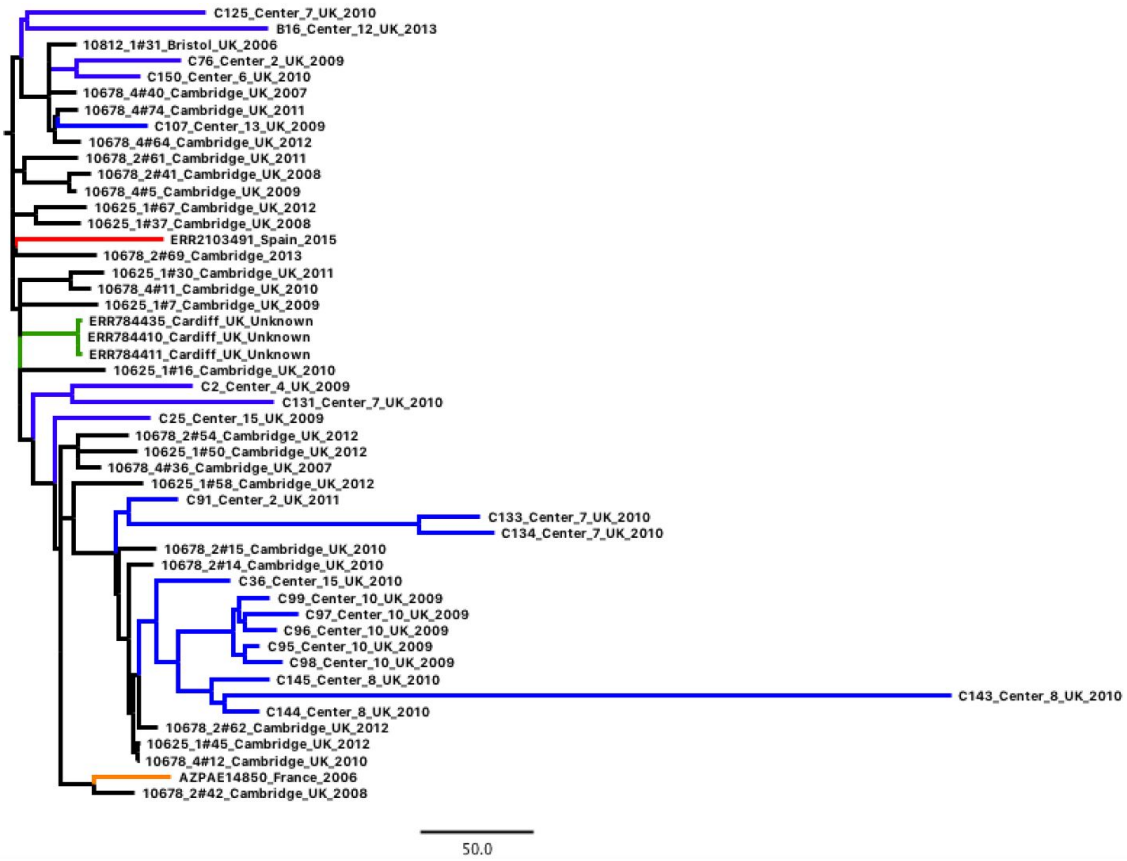


Figure 2.20 Outgroup-rooted, recombination-removed, maximum-likelihood tree of the ST253 isolates, including 25 ST253 isolates from additional studies/sources (blue = Hilliam *et al*., red = Barrio-Tofino *et al.*, green = PubMLST, orange = Kos *et al*.)

isolate from Spain was sequenced by Barrio-Tofino *et al*. Three additional ST253 isolates of unknown isolation date from Cardiff were also available from the PubMLST database (Jolley & Maiden, 2010). These 25 additional ST253 sequences were mapped to the ST253 local reference and combined with the ST253 isolates identified in this study to produce a combined phylogeny (Figure 2.20).

The structure of the phylogeny indicated that the isolates from the local collection are not part of a single local outbreak, as they are interspersed within the isolates from the external studies. This suggests multiple introductions of ST253 infections at the local hospital sites. The two European isolates are also interspersed within the ST253 isolates, which suggests that the ST253 isolates from the local collection are part of a wider, international circulation, and not a UK-specific ST253 clone. Bayesian dating suggests that the emergence date of this group of isolates is 1845-1981 (range: 1715-1985). Similarly to the CUH-only ST253 Bayesian dating analysis, these confidence intervals are very large, and it is unlikely that the ST253 isolates have been present in Europe since the 1800s, since they are not represented in other surveillance collections. The confidence intervals could be reduced by using a greater variety of isolates from a range of sources and isolation dates.

Hilliam *et al.* also identified one patient with bronchiectasis that was infected with an ST175 strain, and sequenced eleven isolates from the lungs of that patient (Hilliam *et al.*, 2017). Cabot *et al.* also sequenced eleven ST175 isolates from a study of XDR *P. aeruginosa* ST175 in Spain and France (Cabot *et al.*, 2016). Two additional ST175 isolates from the USA and one isolate from Italy were also available from the PubMLST database. These 25 additional ST175 sequences were mapped to a local reference and combined with the ST175 isolates identified in this study to produce a combined phylogeny (Figure 2.21).

The results suggest that the isolates collected from the BSAC hospitals in this study are part of a contained expansion, and are closely related to the isolates from Spain, France and Italy. This suggests that a prominent European clone has been introduced to the UK in the recent past, and spread throughout the hospitals in the UK. The isolates identified by Hilliam *et al*., which were also isolates in a UK study, are not part of this outbreak, and form a clade more similar to the ST175 isolate identified as part of the local collection. There is as much diversity within the isolates from the Hilliam *et al.* patient as in the rest of the European dataset combined. This large within-patient diversity was noted by the authors of the study. The Hilliam *et al*. isolates were collected from patients with non-CF lung infection, whilst the

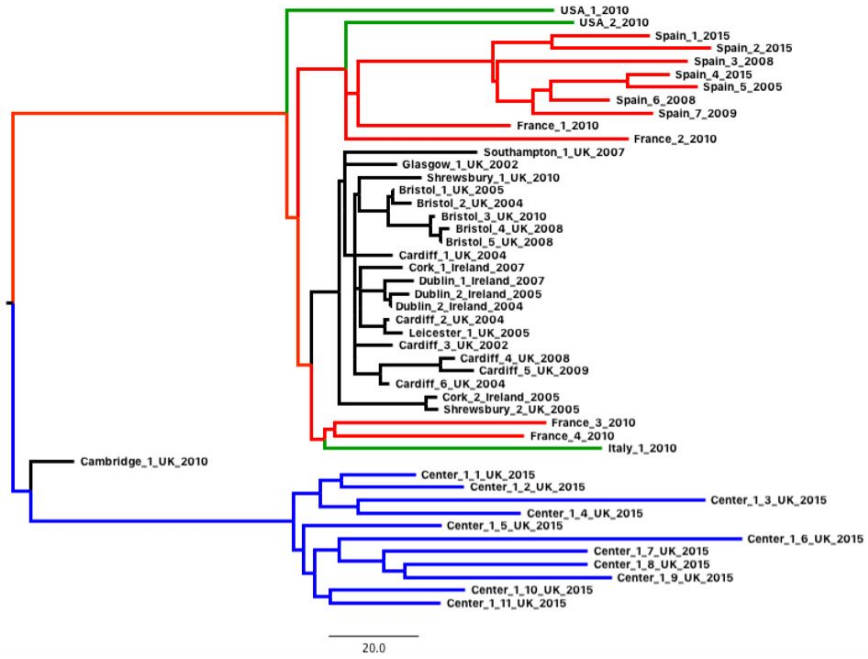


Figure 2.21 Outgroup-rooted maximum-likelihood tree of the ST175 isolates from this study (black), including 25 ST175 isolates from additional studies/sources (blue = Hilliam *et al*., red = Cabot *et al.*, green = PubMLST).

rest of the European dataset were collected from bloodstream infections, which suggests that *P. aeruginosa* behaves differently within different human niches. Large within-patient diversity has been noted in other non-bacteraemia studies, mostly due to long-term persistence of *P. aeruginosa* within the patients of those studies (Sherrard *et al.*, 2017; Williams *et al.*, 2015; Schick & Kassen, 2018).

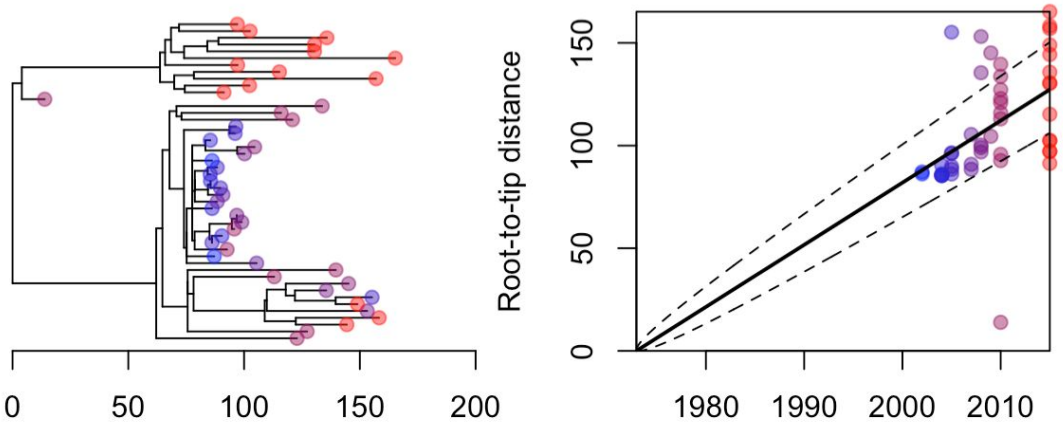


Figure 2.22 Root-to-tip regression of the ST175 isolates, including 25 ST253 isolates from additional studies/sources, suggests that the time to MRCA of this dataset is the early 1970s.

There is geographical clustering of the ST175 isolates, particularly of the isolates from Bristol and Dublin, suggesting some local expansion within geographical areas of the UK (refer back to Figure 2.17). Root-to-tip regression (Figure 2.22) and Bayesian dating suggested that the MRCA of these isolates is 1922-1974 (range: 1915-1978). This is 16-68 years prior to the MRCA of BSAC isolates, and suggests that the ST175 clone has been circulating in Europe for slightly longer than the ST253 isolates. This also coincides with the introduction of amoxicillin and fosfomycin as antibiotics, suggesting a driving force for the emergence of ST175.

2.3.5 AMR surveillance

Phenotypic information on antimicrobial resistance was collected for both the local collection isolates and the BSAC collection isolates. Resistance to a range of antibiotics and antibiotic classes were phenotypically tested for, in order to better understand the implications of *P. aeruginosa* resistance in a clinical bacteraemia setting, and provide information to clinicians about which antibiotics might be most effective to treat patients presenting with *P. aeruginosa* bacteraemia.

2.3.5.1 Local collection phenotypic AMR profiles

For the local isolates, phenotypic data for susceptibility, intermediate resistance, or resistance to 14 different antibiotics was collected (Figure 2.23). In some individual cases, not all antibiotic resistances were tested for. For isolates collected between 2017-2018, colistin and piperacillin resistance data were not collected, and for isolates between 2006-2013, cefepime and levofloxacin resistance phenotypes were not collected.

In total, 156 (57.1%) of the local isolates were phenotypically resistant to at least one antibiotic, 33 isolates (12.1%) were resistant to 5 or more antibiotics, and 17 isolates (6.2%) were resistant to over half of the antibiotics tested for. One ST217 isolate from 2018 was defined as pan-resistant, meaning that the isolate was resistant to all of the antibiotics (colistin and piperacillin was not tested for).

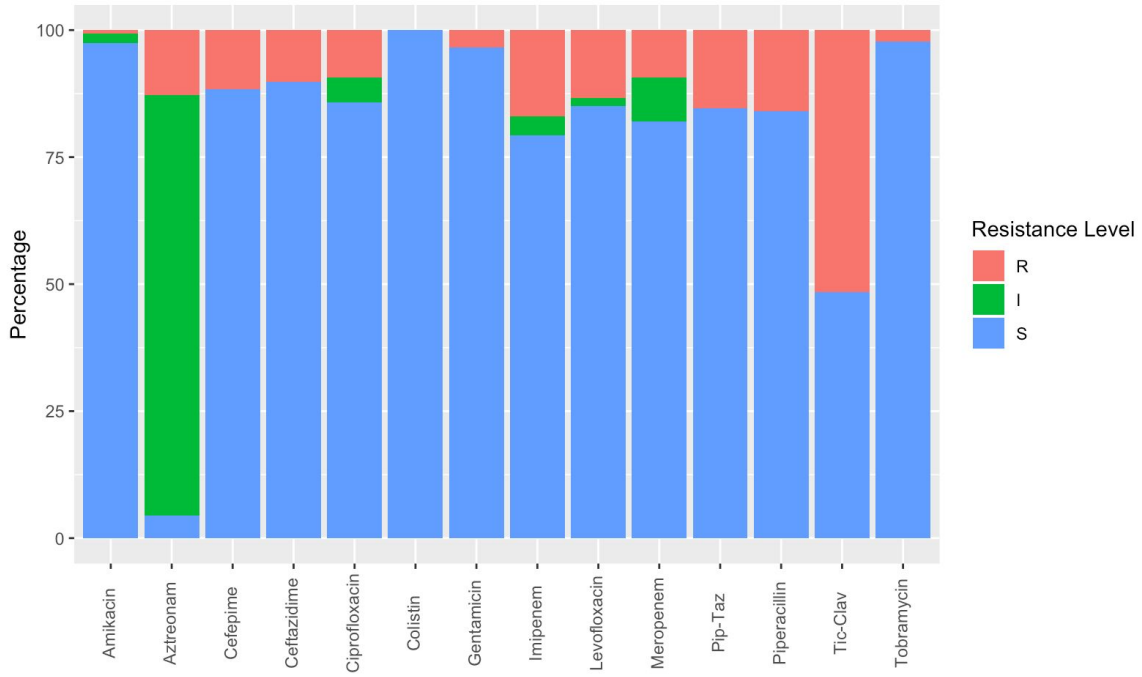


Figure 2.23 Phenotypic antimicrobial resistance for the 273 isolates from the local collection, minus any missing data.

No isolates were recorded as resistant to colistin (although the 2017-2018 isolates were not tested for), suggesting that colistin is a viable choice as a drug of last resort to treat *P. aeruginosa* bacteraemia infections (Javed *et al.*, 2018). The local isolates also displayed very little resistance to the aminoglycoside antibiotics amikacin and tobramycin, with only 6 isolates resistant to either one. Amikacin or tobramycin, as part of a combination therapy, are currently favoured treatment options for CF patients, as *P. aeruginosa* tends to be susceptible to amikacin or tobramycin, even if resistance to other aminoglycosides are present (Olivares *et al.*, 2017). The antibiotic with the highest resistance was for the combination therapy ticarcillin-clavulanate, and the antibiotic with the highest prevalence of non-susceptible isolates (including intermediate resistant isolates) was aztreonam, a member of the β-lactam class of antibiotics.

Figure 2.24 shows the level of antibiotic resistance per year for the local dataset. The number of isolates with resistance to ticarcillin-clavulanate is consistently higher than any of the other antibiotics throughout the study period, ranging from a minimum of 31% of isolates in 2008, and a maximum of 68% of isolates in 2017. Ticarcillin is a β-lactam antibiotic, and as a member of the penicillin class of antibiotic, can be

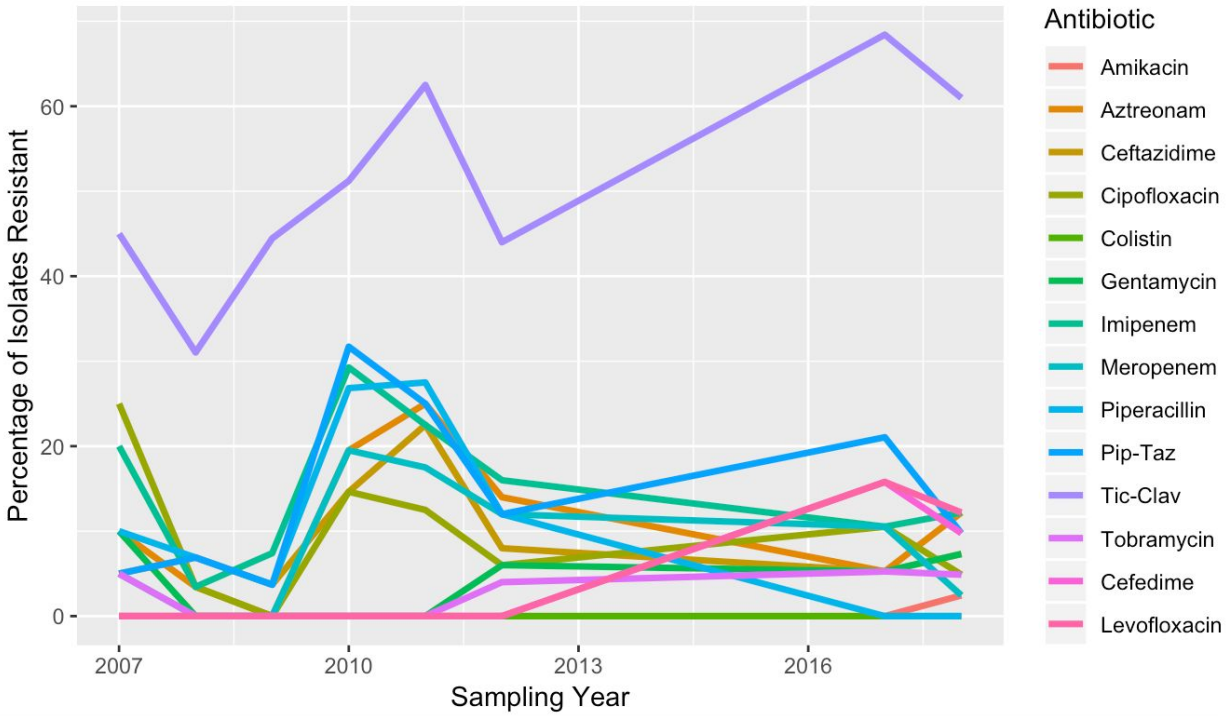


Figure 2.24 Percentage of the 273 local collection isolates with phenotypic resistance to each antibiotic over time.

easily hydrolysed by β-lactamase enzymes. Therefore, to combat the action of the β-lactamase enzymes, ticarcillin is paired clinically with clavulanate, a β-lactamase inhibitor. However, clavulanate can induce the expression of the AmpC β-lactamase enzyme in *P. aeruginosa*, which has activity against ticarcillin and is not inhibited by clavulanate (Lister *et al.*, 1999). High-levels of resistance to ticarcillin-clavulanate are therefore observed naturally in *P. aeruginosa* populations (Lister *et al.*, 1999).

The number of isolates with resistance to the other antibiotics generally followed a similar trend; the number of resistant isolates decrease from 2007, and then rise to a maximum value between 2010-2012. After 2012, resistance to half of the antibiotics increased (Amikacin, Ciprofloxacin, Piperacillin-Tazobactam, Tobramycin), and the other half decreased (Aztreonam, Ceftazidime, Gentamycin, Imipenem, Meropenem, Piperacillin).

The peak in AMR resistance between 2010-2012 is associated with the highest prevalence of ST395 and ST253 isolates, which are international and high-risk clones that are often associated with MDR (Petitjean *et al.*, 2017). Isolates from the ST560, ST309 and ST319 STs, which are all associated with MDR (Kos *et al.*, 2016; Morales-Espinosa *et al.*, 2017; Ruiz-Roldan *et al.*, 2018), were also obtained between

2010-2012. These MDR-STs reappear in 2017 and 2018, but are not present between 2013-2017. Combined, presence of high-risk MDR-associated STs are the cause of the rise, and subsequent fall, of phenotypic resistance observed in the dataset over time.

2.3.5.2 BSAC phenotypic AMR profiles

For the BSAC isolates, phenotypic data for susceptibility, intermediate resistance, or resistance to 11 different antibiotics was collected, which did not include cefepime, levofloxacin, or piperacillin-tazobactam (Figure 2.25). A much higher proportion of antibiotic resistance was observed amongst the BSAC isolates than the local isolates, as the BSAC isolates were selected based on their MDR status.

In total, 66 (84%) of the BSAC isolates were phenotypically resistant to five or more antibiotics, and 52 isolates (66%) were resistant to over half of the antibiotics tested for. No isolates were pan-resistant, but ten of the isolates (13%) were resistant to ten of the antibiotics. All of the ten isolates were susceptible to colistin, suggesting that therapeutic options still remain for the most MDR *P. aeruginosa* infections.

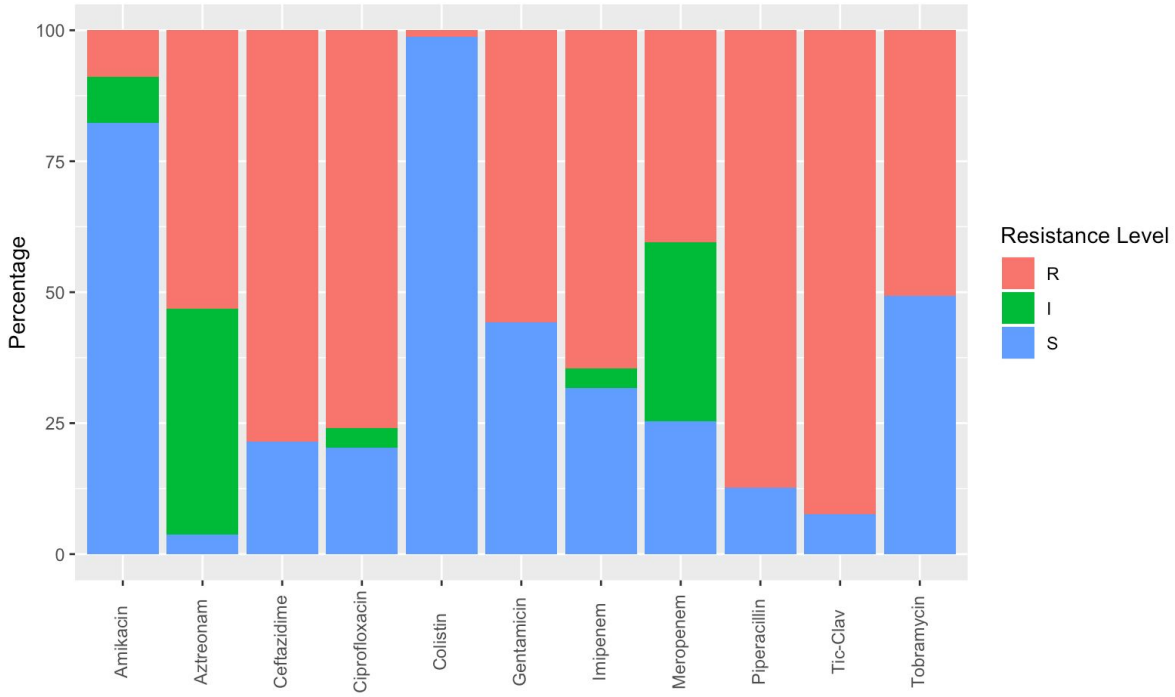


Figure 2.25 Phenotypic antimicrobial resistance for the BSAC collection.

Similarly to the local collection, the antibiotic with the highest prevalence of resistance (92%) was for the ticarcillin-clavunate antibiotic combination. The antibiotic with the highest prevalence of non-susceptible isolates (including intermediate resistant isolates) was also aztreonam (96%). Only one isolate was resistant to colistin, the antibiotic-of-last-resort, whereas the second-lowest prevalence of resistance was for the aminoglycoside antibiotic amikacin (18%).

Figure 2.26 shows the level of antibiotic resistance per year for the BSAC isolate collection. Antibiotic resistance was more variable for the BSAC study than for the local study. The proportion of ticarcillin-clavulanate resistant isolates was consistently high throughout the whole study period, ranging from 89 - 100%. Consistently high levels of piperacillin, which is also a β-lactam antibiotic, was also observed. The only colistin resistant isolate was sampled in 2004, and no other colistin-resistant isolates were identified from any other year.

A drop in the number of ciprofloxacin-resistant isolates was observed over the study period. The proportion of ciprofloxacin resistant isolates between 2001 - 2010 was consistently high, with an average

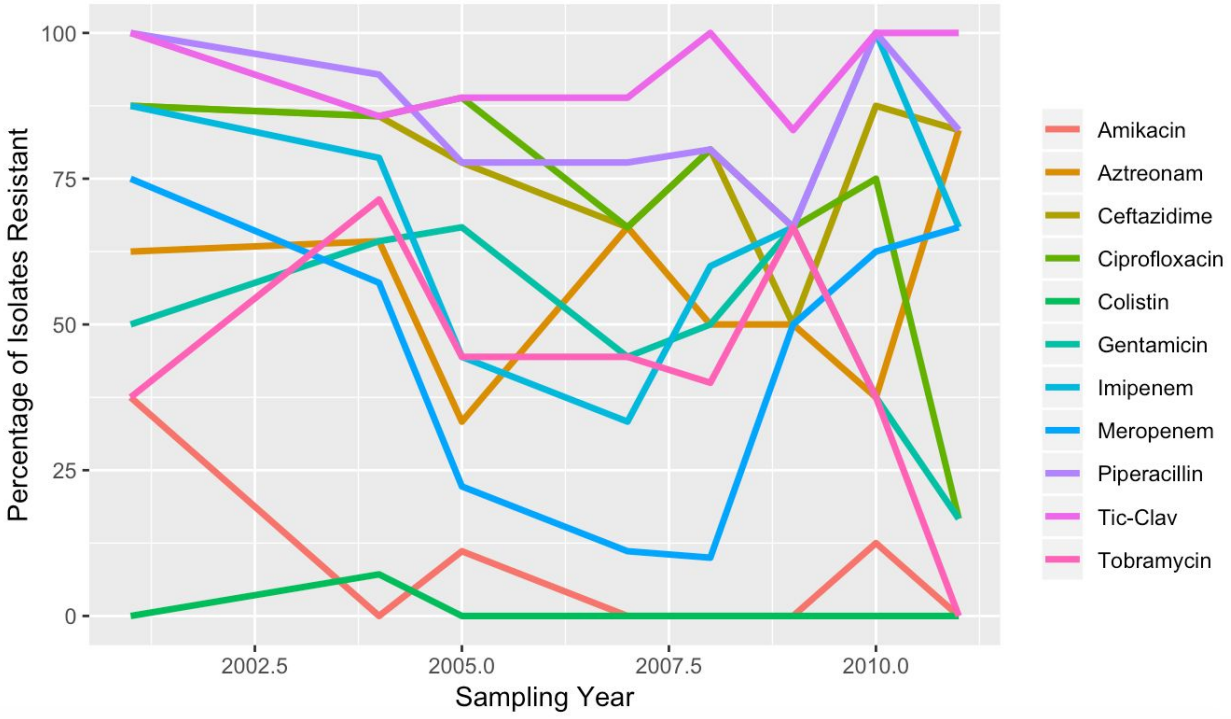


Figure 2.26 Percentage of the isolates 79 BSAC collection isolates with phenotypic resistance to each antibiotic.

of 79% of isolates displaying resistance. However in 2011, this fell to just 17% of isolates. A similar trend was observed for gentamicin-resistant isolates. Between 2001 and 2010, gentamicin resistance was observed in an average of 54% of isolates, which fell to just 17% of isolates in 2011.

Two of the most variable AMR over the timeframe of the study were for meropenem and imipenem, which are both members of the carbapenem class of antibiotics. Carbapenem-resistant *P. aeruginosa* is currently of critical priority for the World Health Organisation as they cause severe, and often deadly, infections, and are becoming increasingly resistant to the best antibiotics available (World Health Organisation, 2017). At the beginning of this study period, in 2001, meropenem resistance was prevalent, and found in 75% of isolates. The prevalence of meropenem-resistant isolates decreased every year until it reached its lowest level in 2008, where just 10% of isolates were resistant. However, meropenem resistance then began to rise, and by 2011, 67% of isolates were meropenem resistant. Imipenem resistance followed a similar trend, with 88% of isolates displaying resistance in 2001, dropping to 33% of isolates in 2007, and then increasing to 100% of isolates in 2010. This then fell to 67% of isolates in 2011. High-levels of resistance to the last-resort carbapenem class of antibiotics is concerning, as they are the most effective antibiotics to treat *P. aeruginosa* infections, suggesting that new therapeutic options need to be considered (Wi *et al*., 2017).

2.3.5.3 Phylogenetic correlation with phenotypic AMR

The phenotypic AMR profiles of the isolates from both the local and BSAC collections are displayed next to the phylogeny in Figure 2.27.

The ST395 clade was associated with high AMR prevalence. The ST395 clade is made up of 31 isolates from the local collection and one isolate from the BSAC collection. Ten of the isolates displayed resistance to seven or more antibiotics. A closer inspection of the AMR distribution within the ST395 phylogeny showed that the MDR isolates are associated with the closely-related clades identified in section 2.3.4.5 (Figure 2.28). The distribution suggests that there was simultaneous acquisition of multiple AMR, followed by expansion of these lineages, as the rest of the ST395 isolates show susceptibility to the majority of antibiotics. Closer inspection of resistance determinants within the genomes of the ST395

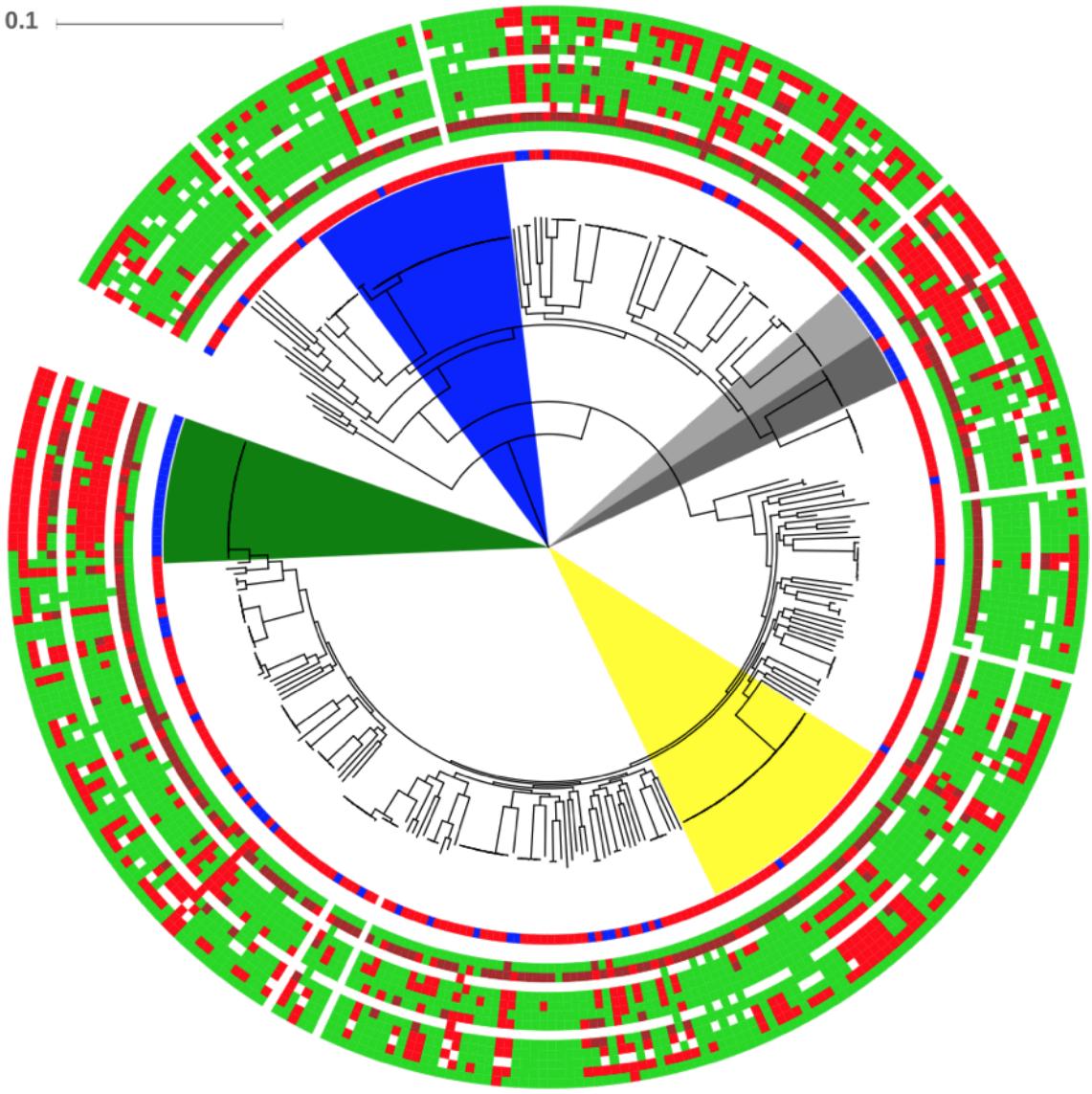


Figure 2.27 Maximum-likelihood phylogeny with phenotypic AMR data displayed around the outside. The cones indicate the high-risk and international clones, and other clones associated with high phenotypic AMR; yellow = ST395, green = ST175, blue = ST253, light grey = ST357, dark grey = ST235. The inner ring indicates the isolate collection; red = local collection, blue = BSAC collection. The outer rings indicate AMR profiles; green = susceptible, red = resistant, brown = intermediate resistant, white - missing data. The order of antibiotic resistance groups from inside to outside is amikacin, aztreonam, cefepime, ceftazidime, ciprofloxacin, gentamicin, imipenem, levofloxacin, meropenem, piperacillin, piperacillin-tazobactam, ticarcillin-clavunate, tobramycin.

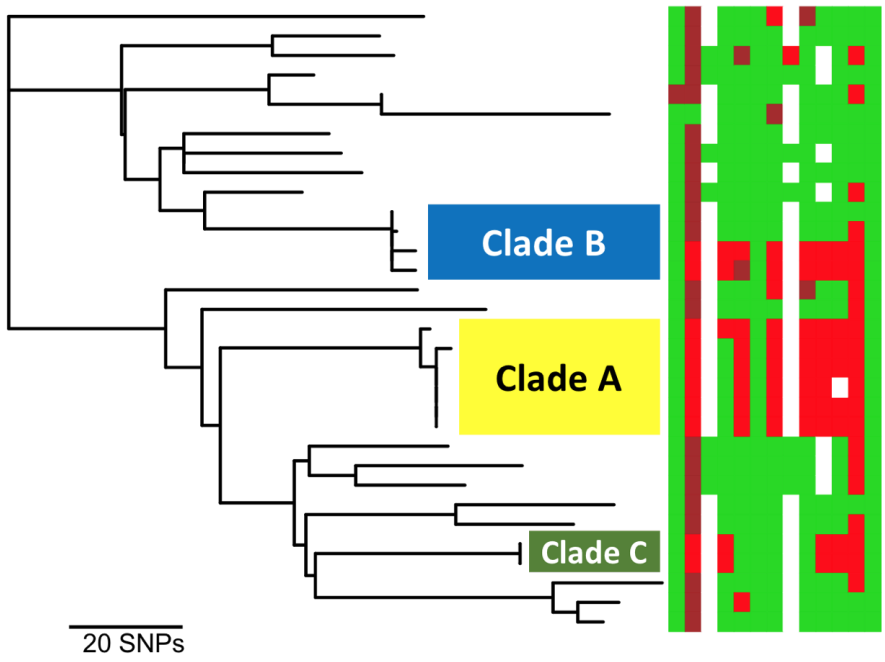


Figure 2.28 The ST395 tree with the AMR phenotypes displayed next to it. The MDR phenotype is associated with the clades identified in section 2.3.4.5. The outer bands indicate AMR profiles; green = susceptible, red = resistant, brown = intermediate resistant, white - missing data. The order of antibiotic resistance groups from inside to outside is amikacin, aztreonam, cefepime, ceftazidime, ciprofloxacin, gentamicin, imipenem, levofloxacin, meropenem, piperacillin, piperacillin-tazobactam, ticarcillin-clavunate, tobramycin.

isolates shows that only Clade A harbours a resistance-causing mutation that is not found within the rest the ST395 isolates. This mutation was within the Quinolone Resistance Determining Region (QRDR) of *gyrA*, which is linked to fluoroquinolone resistance, but does not explain resistance to the other antibiotics. Therefore, the resistance associated with these MDR clades is probably due to an increase in the expression of multidrug efflux systems that can cause resistance to almost every class of antibiotics (Poole, 2011). However, transcriptomic studies would be necessary to confirm this.

The international and high-risk ST175 clade, which is made up of one isolate from the local collection and 21 isolates from the BSAC collection, was also associated with high AMR prevalence; 14 isolates (66%) isolates were resistant to seven antibiotics or more. The ST175 isolate that was from the local collection was only resistant to one antibiotic, ticarcillin-clavulanate. All BSAC ST175 isolates were resistant to ciprofloxacin, and over 85% of isolates were phenotypically resistant to ceftazidime, gentamicin, piperacillin, ticarcillin-clavulanate, and tobramycin. No resistance was also observed for

colistin, and roughly half of isolates were susceptible to either imipenem or meropenem. Whilst resistance was observed for the aminoglycosides gentamicin and tobramycin, no resistance was identified to the aminoglycoside amikacin, suggesting a viable therapeutic option to treat this high-risk MDR clade.

High levels of AMR were not associated with the third overrepresented clade, ST253, which is made up of 29 isolates from the local collection and one isolate from the BSAC collection. Only three ST253 isolates were non-susceptible to four or more antibiotics.

Two additional clades, made up of eight ST357 isolates from the BSAC collection, and seven ST235 isolates (five from the BSAC collection and two from the local collection) were associated with high AMR prevalence. Six ST357 isolates (75%) and two ST235 isolates (29%) were resistant to 7 or more antibiotics.

2.3.5.4 Genetic mechanisms of AMR

All 352 isolates in this study contained the genes for the aminoglycoside modifying enzyme APH(3')-IIb, which confers resistance to kanamycin, neomycin and paromycin (Zeng & Jin, 2003). All isolates phenotypically resistant to either amikacin or gentamicin contained the MexXY-OprM efflux system, which is efficient at conferring low level resistance to aminoglycosides (Sobel *et al.*, 2003). High level resistance to aminoglycosides can be conferred by the knock-out of the *mexZ* repressor, which leads to overexpression of the MexXY-OprM efflux system. Known mutations in *mexZ* were identified in 18 isolates (of which 14 were frameshift mutations or indels), as well as 29 isolates with novel non-synonymous mutations in *mexZ*. Five phenotypically resistant isolates also encoded relevant aminoglycoside modifying enzymes that are specific to the aminoglycosides amikacin or gentamicin. AAC(3')-Id, AAC(6')-31, and APH(3')-VI, were each present in one isolate, and ANT(2")-Ib was present in two of the isolates (Ramirez & Tolmasky, 2010).

Fluoroquinolone resistance can be caused by mutations in the Quinolone Resistance Determining Region (QRDR) of either *gyrA*, *parC*, *gyrB* or *parE* (Maeda *et al.*, 2010). Of the fluoroquinolone resistant isolates in both collections, 85% harboured the well-characterised T83I mutation in *gyrA*, which causes a known decrease in fluoroquinolone susceptibility (Cabot *et al*., 2016). Alongside this, D87N/Y mutations (Telling *et al*., 2018) were present in 12% of the isolates, including four isolates that did not contain the

T83I mutation. The common S87L/W mutation (Cabot *et al.*, 2016) in *parC* co-occurred with the T83I mutation in *gyrA* in half of the isolates. Fluoroquinolone resistance could not be explained by *gyrA* and *parC* mutations in 12% of cases. Five of these isolates contained M437I and A473V mutations in *parE*, which causes known resistance to ciprofloxacin (Bruchmann *et al*., 2012), one isolates had a frameshift mutation in *gyrA*, and one isolate contained an I57V mutation in *parE*. Other genes that have been associated with fluoroquinolone resistance, *qnr*, *qep* and *oxq*, were not present within any of the isolates in this study (Redgrave *et al*., 2014). Only two isolates that were phenotypically resistant to fluoroquinolones did not harbour known resistance mechanisms.

One isolate in this dataset was resistant to colistin. Colistin resistance can be conferred through lipid A modifications, which are caused through mutations in *phoQ* gene, and the colistin resistant isolate contained the known polymyxin resistance marker V260G in the *phoQ* gene (Jochumsen *et al.*, 2016).

A high prevalence of carbapenem-resistant isolates was present in this dataset (n = 96 for Imipenem, n = 57 for meropenem). The carbapenems are currently an antibiotic-of-last resort for *P. aeruginosa* bacteraemia and CF infections (Fowler & Hanson, 2014). Deficiencies in the OprD porin has been linked to a basal level of carbapenem resistance in previous studies (Li *et al.*, 2012), as it decreases the outer-membrane permeability. Within the 352 isolates in this dataset, only 169 (47.9%) isolates contain complete copies of the *oprD* gene. In total, 46 different frame-shift or truncation mutations were identified, and the gene was completely absent in one isolate. However a second resistance mechanism is typically necessary to confer clinical levels of resistance to carbapenems (Li *et al*., 2012). For example, two isolates (one ST235, and one ST2613) encoded the VIM-1 and VIM-2 class of carbapenem-hydrolyzing enzymes. Additionally, when efflux pump over-expression occurs alongside low outer membrane permeability, resistance to carbapenems can develop (Meletis *et al.*, 2012).

Efflux systems within *P. aeruginosa* are an incredibly important mechanism that can confer resistance to several classes of antibiotics (Rampioni *et al*., 2017). For example overexpression of the MexAB-OprM efflux pump can lead to resistance to β-lactams and β-lactamase inhibitors, monobactams, fluoroquinolones, and aminoglycosides (Masuda *et al.*, 2000). The expression of the MexAB-OprM efflux system is regulated by *mexR*, *nalC* and *nalD*, and mutations in these regulatory genes can lead to overexpression of the efflux system, resulting in antibiotic resistance (Suresh *et al*., 2018). Within this dataset, there were 76 unique mutations within *mexR* (n = 32 unique mutations, including 7 frameshifts and INDELs), *nalC* (n = 23 unique mutations, including 3 frameshifts and INDELS) and *nalD* (n = 21

unique mutations, including 5 frameshifts and INDELs). Additionally, the *nalD* gene was completely absent in five isolates. These mutations could result in overexpression of the MexAB-OprM efflux system, and aid non-susceptibility to the β-lactam antibiotics aztreonam, ceftazidime, piperacillin and ticarcillin.

Many resistance-related genes to antibiotics that weren't phenotypically tested for were present in the genomes. For example, 95% of the isolates in this study contained a copy of the *fosA* gene, which is linked to fosfomycin resistance caused by the conjugation of glutathione to the antibiotic (Klontz *et al.*, 2017). The *fosA* gene is present amongst almost 99% of all published *P. aeruginosa* genomes, suggesting that this gene is chromosomally encoded and that resistance is widespread (Ito *et al.*, 2017).

The $_{bla}$*OXA*-50 gene, which is a chromosomally-encoded beta-lactamase (Kos *et al*., 2014), was present within 98% of the isolates in this study. This gene confers slight non-susceptibility to ampicillin, ticarcillin, moxalactam and meropenem, and increases MIC values to common β-lactams. AmpC, which is present in 100% of isolates in this dataset, also increases MIC values to common β-lactam antibiotics. The six isolates that did not contain $_{bla}$*OXA*-50 also did not contain any other $_{bla}$*OXA* type genes. $_{bla}$*OXA*-1 was present in 3 isolates, $_{bla}$*OXA*-15 in one isolate, and $_{bla}$*OXA*-6 in three isolates.

Other resistance conferring genes include the chloramphenicol acetyltransferase variant *catB7* (Murugan *et al.*, 2014), which was found to be present in 90% of the isolates, the *bcr-1* gene conferring bicyclomicin resistance (Fonseca *et al.*, 2015), which was present in 100% of the isolates, and the *sul1* gene that confers resistance to sulfonamides (Poirel *et al.*, 2001), which was present in 9% of the isolates.

Efflux systems were also highly prevalent within the genomes. MexVW-OprM, TriABC-OpmH, and MexJK-OprM, were present in 99% of isolates, and MexCD-OprJ, MuxABC-OpmB, and MexGHI-OpmD were found in 98% of isolates. PmpM and EmrE, which are small molecule transporters, are also present in 100% of isolates (Li & Plesiat, 2016).

In some cases, no genetic association with phenotypic susceptibility could be determined. Attempting to relate resistance genes to phenotypic susceptibility data for *P. aeruginosa* is difficult, due to the non-specific nature and expression-dependence of efflux systems, as well as uncharacterised mutations in known genes (Freschi *et al.*, 2018). Porin loss, upregulation of non-specific efflux pumps, and permeability changes may also be responsible for the unaccounted phenotypic resistance.

2.4 Discussion

In this Chapter, 352 *P. aeruginosa* isolates collected from two different bacteraemia surveillance programmes were investigated to gain an understanding of the population structure and AMR mechanisms of *P. aeruginosa* in bacteraemia. This collection consisted of 79 isolates from the BSAC surveillance dataset of MDR bacteraemia infection. The remaining 273 isolates were from a local collection as part of the surveillance of *P. aeruginosa* bacteraemia infections in Cambridge hospitals, but were not selected for MDR status. This is the first comprehensive study of *P. aeruginosa* bacteraemia isolates from Cambridge, and by putting these into the context of the wider UK bacteraemia *P. aeruginosa* population, we have been able to determine which high-risk and epidemic clones are most prevalent within these hospitals, and the resistance genes that they carry. Knowledge of both local and national population structures of *P. aeruginosa* bacteremia, as well as antimicrobial resistance profiles and genetic determinants, could be useful to inform clinical treatment. This comprehensive study has only been possible due to recent advancements in computational biology and genomics, which have led to substantial decreases in the cost and availability of DNA sequencing the whole genomes of such a large number of isolates.

The population structure of the 353 isolates represent the well-characterised non-clonal *P. aeruginosa* population structure, with sporadic emergence of highly-successful epidemic clones (Treepong *et al.*, 2018). This indicates that a *P. aeruginosa* isolate does not have to be specialised in order to cause bacteraemia infection, and reinforces the widely-held view that *P. aeruginosa* is well-adapted to being an opportunistic pathogen due to its genetic and metabolically diversity, able to inhabit a wide range of niches (La Rosa *et al.*, 2018). However, including non-bacteraemia *P. aeruginosa* isolates in the study would confirm this hypothesis. The MDR isolates from the BSAC surveillance collection were also widely distributed across the *P. aeruginosa* population, suggesting that the ability to acquire AMR determinants is retained by most, if not all, of the *P. aeruginosa* isolates.

Within this dataset, there is evidence of successful epidemic STs. Two of these STs were associated with local infection in Cambridge hospitals; ST395 and ST253. These are both well-known, international high-risk clones which are frequently associated with multi-drug resistance (Treepong *et al*., 2018). MDR was not associated with the ST253s collected as part of the local surveillance, but MDR was associated with two clades within the ST395 population. This showed simultaneous acquisition of multiple AMR

variants. However, it is most likely that this is due to a single mutation event which leads to the overexpression of non-specific multi-drug efflux pumps, though transcriptomic studies would be required to confirm that (Kiser *et al*., 2010). The BSAC surveillance collection was associated with the high-risk and international clone ST175. The ST175s were associated with high-levels of MDR, though susceptibility to amikacin and colistin was identified in all isolates. Roughly half of ST175 isolates were resistant to the carbapenems, but the isolates were not associated with the presence of MBLs or ESBLs. Two other heavily-MDR clades were associated with the BSAC; the ST357 and ST235, both also well-known, high-risk, international clades known to exhibit carbapenem resistance through the acquisition of several carbapenemase enzymes (Pragasam *et al.*, 2018).

The genomic islands of the three overrepresented high-risk and international STs were characterised. The high-risk and international STs had a higher-than-average number of genomic islands compared to the rest of the STs within the dataset. Most of the genes within the genomic islands were related to iron uptake, phage defence, Type VI secretion and motility. Many of the hypothetical proteins within the genomic islands were predicted to be alcohol dehydrogenase enzymes or related to virulence. This acquisition of genes via horizontal gene transfer contributes to the success of these high-risk and international clones (Silveira *et al*., 2016).

Strong temporal signal was present for the high-risk and international clade ST175. These isolates were put in context of an international dataset, and indicated that ST175 arrived in the UK sometime between the mid-1980s and mid-1990s. Temporal signal was also present for the ST253 clade, but the isolates were dispersed amongst a European dataset of ST253, suggesting that the isolates as part of this study do not capture enough diversity to identify when the ST253s may have first been introduced to the UK.

As well as whole-genome sequencing, the AMR profiles of all 353 isolates were determined. Within the local collection, 42% of isolates were susceptible to every antibiotic tested. Very little resistance (< 3%) was identified for the aminoglycoside antibiotics tobramycin, gentamicin and amikacin, and no resistance was identified to the drug-of-last resort, colistin. Resistance rates were highest (~50%) for the ticarcillin-clavunate combination therapy, whilst resistance to the carbapenem drugs of last resort was at 13%. Widespread MDR was observed for isolates collected as part of the BSAC surveillance programme, which is consistent with the criteria for inclusion into the study. Only one isolate showed resistance to colistin, indicating that it is a viable therapeutic option to treat MDR infections currently, as resistance has yet to develop. Resistance to the aminoglycoside amikacin was also present in less than 9% of isolates,

which suggests that resistance determinants to this antibiotic are more difficult for the *P. aeruginosa* population to obtain, an observation that has been previously described (Olivares *et al.*, 2017). Levels of resistance to amikacin was lower than resistance to the carbapenems in the BSAC collection (~53%), suggesting that these antibiotics should no longer be considered as last-resort against MDR infections, as resistance rates are comparable to those of other first-line antimicrobial compounds, such as aztreonam (53%), gentamicin (55%), and tobramycin (51%).

Apart from instances of MDR concentrated within the ST175, ST235, and ST357 BSAC clades, MDR resistance is distributed across the whole *P. aeruginosa* population, suggesting that acquisition of variants conferring resistance to multiple antimicrobial compounds is not limited to a subset of strains.

Chapter 3

Genomic diversity of exacerbation-related *Pseudomonas aeruginosa* from nine patients with cystic fibrosis

Declaration of Contributions

The study was initiated by Andres Floto. Julian Parkhill, Martin Welch and Josephine Bryant supervised this work. Emem Ukor collected and processed all clinical data, sputum samples, and selected all *P. aeruginosa* isolates for analysis. DNA was extracted by Beth Blane. The isolates were sequenced by the Wellcome Trust Sanger Institute DNA Pipelines Operations. I carried out all analysis in this chapter.

3.1 Introduction

Cystic fibrosis (CF) is caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) protein (Mall & Hartl, 2014). The CFTR protein is responsible for transporting chloride ions across the epithelial cell layer and into the extracellular space, however, when the CFTR protein is defective, an ion imbalance across the membranes results (Puchelle *et al*., 2002). This ion imbalance draws water out of the extracellular space by osmosis, dehydrating the mucus in the lung and gut, causing the thick and sticky mucus that is characteristic of CF (Puchelle *et al*., 2002). This thick and sticky mucus supports bacterial colonisation and growth within the CF lung, since mucus clearance is impaired (Hill *et al.*, 2018), and the mucus provides a good source of carbon and nitrogen to aid bacterial growth (Palmer *et al*., 2007). From infancy, the CF lung is often colonised by commensal bacteria from the upper respiratory system, such as *Staphylococcus aureus* and *Haemophilus influenzae*, but replacement with *Pseudomonas aeruginosa* occurs in 80% of patients by early adulthood (Lyczak *et al.*, 2002).

Transmission of acquired *P. aeruginosa* strains from one patient to another has been well documented (Williams *et al*., 2018). The ability of a strain to rapidly colonise and adapt to the CF lung, whilst retaining the ability to transmit from patient-to-patient, is a hallmark of a few specialised strains of *P. aeruginosa* (Williams *et al.*, 2018). These transmissible strains, called epidemic strains, have been isolated in studies from the UK, Denmark, Belgium, Canada, Australia, and the USA (Williams *et al.*, 2018). These epidemic strains are associated with high virulence and worse patient outcomes (Al-Aloul *et al*., 2004), and are considered to be very well adapted to the CF lung and highly transmissible (Salunkhe *et al*., 2005).

After the introduction of *P. aeruginosa* into the CF lung, the bacteria can undergo a transcriptional switch from acute infection, which can be cleared by aggressive antibiotic therapy, to chronic infection, in which they remain in the patient's lung for the rest of their life (Hogardt & Heesemann, 2013). The switch from acute infection to chronic infection, and consequential selection of adaptive mutations to the CF lung, can be accelerated with an increased mutation rate; a phenotype called hypermutation (Rees *et al*., 2019). This hypermutation phenotype emerges due to defects in the DNA mismatch repair systems, and can lead to better adaptation to the CF lung and increased antimicrobial resistance. Hypermutator strains are traditionally identified through phenotypic assays (Mather & Harris, 2013). However genomic signatures can be identified as a consequence of this increased mutation rate, which can also be used to identify

hypermutators (Mather & Harris, 2013). For example, a bias towards transition mutations over transversion mutations are observed in isolates that hypermutate (Dettman *et al*., 2016). Transitions comprise 98.8% of all mutations generated by hypermutator strains, a rate that is over 900x higher than wild-type (Dettman *et al*., 2016). The increased mutation rate caused by hypermutation can be identified phylogenetically, as higher mutation rates result in longer branch lengths on a phylogeny, compared to wild-type strains (Marvig *et al*., 2013).

Whole-genome sequencing of hypermutator strains has previously identified non-synonymous mutations in several genes that may be linked with the hypermutator phenotype, with the highest mutation rates associated with mutations in the *mutLMSTY* mismatch repair genes (Lopez-Causape *et al*., 2017). Exploring and correctly characterising these genomic signatures could replace the need for carrying out phenotypic assays to identify hypermutator lineages in a population of isolates.

Antibiotics are used to control and prevent negative outcomes associated with chronic *P. aeruginosa* lung infection, and as a consequence, resistance to antibiotics arises frequently (Macia *et al*., 2005). Hypermutation of *P. aeruginosa* isolates is associated with an increase in resistance to antimicrobial compounds used to treat CF lung infections (Oliver *et al.*, 2000). This is because resistance to many antibiotics often develop through accumulation of point mutations within the *P. aeruginosa* genome (Oliver, 2015). Resistance to the fluoroquinolones through point mutations in *gyrA*, *gyrB*, *parC*, and *parE* (Lopez-Causape *et al.*, 2017), reduce susceptibility to the carbapenems through inactivation of *oprD* (Henrichfreise *et al*., 2007), and mutations in repressor genes for multi-drug efflux pumps, have been identified in hypermutator strains (Rees *et al*., 2019).

P. aeruginosa within the lungs of CF patients are the main cause of lung function decline (Bhagirath *et al.*, 2016). This is caused primarily by epithelial surface damage caused by *P. aeruginosa* virulence factors and the associated host immune response (Gellatly *et al*., 2013). Acute pulmonary exacerbations (APEs) are periods of sudden and rapid worsening symptoms that patients with CF often undergo (Bhatt, 2013). The symptoms of an APE include a worsening cough, chest pain, shortness of breath, weight loss, and lung function decline, and are characterised by increased lung inflammation and damage, from which lung function never fully recovers (Cogen *et al.*, 2017). Currently, APEs are treated with antibiotics, usually by intravenous therapy for between 7 and 10 days, and are effective at clearing the symptoms, suggesting a bacterial infection could be the cause of APEs (Bhatt, 2013).

Two previous studies have shown that variation in the *P. aeruginosa* population in the CF lung remains stable over the course of an APE (Fothergill *et al*., 2010; Mowat *et al*., 2011). However, these studies focussed only on variation in PCR products (Fothergill *et al*., 2010) and haplotype composition (Mowat *et al*., 2011). In this study, 4,094 *P. aeruginosa* isolates were whole-genome sequenced from nine patients who experienced 18 APEs between them. The isolates were collected over the course of three APE-related timepoints: 96 isolates from each acute timepoint per patient (collected on the day of an APE); 96 isolates from each stable timepoint per patient (collected seven days before the APE); and 96 isolates from each recovery timepoint per patient (collected seven days after antibiotic treatment had finished). This provides a unique insight into how the *P. aeruginosa* population changes over the course of an APE in these patients. This whole-genome sequence approach also enables the development of a better understanding of genetic diversity and adaptation of *P. aeruginosa* to the CF lung, and a better understanding of the population structure that affects each individual patient.

3.1.1 Aims

APEs account for much of the morbidity and mortality associated with CF. The events that trigger these APEs are not currently understood, although they are typically treated using antibiotics, which significantly alters the *P. aeruginosa* population within the CF lung. In this study, we aim to examine the variation of 4,400 *P. aeruginosa* isolates collected from the lungs of a cohort of 9 patients with CF, and determine any genetic changes which may underpin phenotypic diversity and evolution, and how this variation may relate to APEs in CF patients.

3.2 Methods

3.2.1 Sample selection

3.2.1.1 Patient recruitment

To investigate the genotypic composition of *P. aeruginosa* clones within the CF airway that may be correlated with acute pulmonary exacerbation, a clinical trial (TeleCF) was carried out, where *P. aeruginosa* isolates were sampled from a pre-defined population of adult patients with CF through periods of acute pulmonary exacerbation (Emem Ukor, Floto Lab, University of Cambridge). Nine participants (Table 3.1) collected expectorated sputum samples following their usual chest clearance routine over a period of six months. Patients were 17 years of age or above, had a confirmed diagnosis of CF, were willing to provide daily home monitoring data, had a history of at least two acute pulmonary exacerbations within 12 months of study enrolment, and had evidence of chronic airway infection with *P. aeruginosa*. Patients were excluded from the study if they had evidence of airway co-infection with non-tuberculous mycobacteria, were unable to provide written informed consent, or were unable to provide regular sputum samples (Emem Ukor, Floto Lab, University of Cambridge).

Patient ID	CF gene 1 mutation	CF gene 2 mutation	FEV1 (%)
K1	R117H	4326delTC	45.6
K3	G85E	Q207X	50.4
K4	F508del	c.3889dupT	34.7
K6	F508del	G542X	47.9
K7	F508del	F508del	72.9
K9	F508del	c.1766+1A->G	14.1
K11	F508del	F508del	32
K14	F508del	3752G/A	59.7
K15	F508del	F508del	53.3

Table 3.1 The clinical metadata available for each patient in the TeleCF study. FEV1 (force expired volume) measurements are the scores recorded at the start of the study period.

3.2.1.2 Sample growth

Clinical data was used to identify periods of acute pulmonary exacerbation in the nine patients during the 6-month TeleCF clinical trial (Emem Ukor, Floto Lab, University of Cambridge). Acute pulmonary exacerbation was defined as a clinical decision to initiate antibiotic treatment by a CF physician, due to changes in a patient's respiratory status.

To identify any genetic changes in the *P. aeruginosa* population within the CF lung that may be associated with a cause of APE, sputum samples from before, during and after each APE that occurred during the study period were selected. Where possible, for each period of acute pulmonary exacerbation, three frozen sputum samples were selected; 1) a non-exacerbation sample that was collected 7 - 14 days before an APE; 2) an exacerbation sample that was collected on day 0 of an APE (+/- 5 days); 3) a recovery sample that was collected 7 - 14 days after an APE. In total, 14 non-exacerbation, 19 exacerbation and 14 recovery samples were selected, from a total of 18 APE events, totalling 47 sputum samples.

Each of the 47 sputum samples was grown on *Pseudomonas* selective agar containing cetrimide and sodium nalidixate and incubated at 37^{O}C for 48 - 72 hours, then streaked on *Pseudomonas* selective agar to form single colonies. In order to try to capture the diversity of the *P. aeruginosa* population within the CF lung at each time point, where possible, 95 single colonies were selected from each of the 47 streaked sputum plates in order to proportionally represent all morphotypes that were present. In total, 4,408 *Pseudomonas* strains were selected from the initial 47 sputum samples. Single colonies of each of the 4,408 *Pseudomonas* strains were incubated for a further 6 hours at 37^{O}C in 1 mL cetrimide broth to increase cell density. To confirm that each isolate was *P. aeruginosa*, each of the 4,408 strains were re-streaked on *Pseudomonas* selective agar containing cetrimide and sodium nalidixate and incubated at 37^{O}C for 48 - 72 hours. A single colony was selected from each plate and incubated for a further 6 hours at 37^{O}C in 1 mL cetrimide broth, and stored at -80^{O}C in 25% glycerol solution.

3.2.2 DNA sequencing and quality checks

DNA was extracted from the 4,408 CF *P. aeruginosa* isolates using a QIAxtractor (QIAgen) instrument according to the manufacturer's instructions. Library preparation was performed according to the Illumina

protocol, and sequencing was performed on the Illumina HiSeq 2500 and X10 platforms. A total of 4,168 (95%) *P. aeruginosa* isolates passed sequencing QC and were initially included in the study.

All 4,168 sequences were mapped to the *P. aeruginosa* PAO1 reference genome (accession number: PRJNA331) using bwa mem (version 0.7.17, Li & Durbin, 2009), discounting identical fastq reads and outputting all alignments. Samtools mpileup (version 1.6, Li *et al*., 2009) was used to determine which nucleotide base occured at each position, using a quality score of 50, discarding anomalous reads and disabling read-pair-overlap detection. Bcftools call (version 1.5, Li *et al*., 2009) was used to call SNPs against the reference genome, assuming a mutation rate of 0.001. Short indels were called using GATK IndelRealigner (version 3.4.46, McKenna *et al.*, 2010), using default settings. Multiple sequence alignments were generated by combining all individual consensus sequences.

A base was called as heterozygous if the consensus base call was supported by less than 90% of all mapped fastq reads. The number of heterozygous base calls for every isolate were plotted as a distribution, and outliers outside of the 95% confidence interval were identified and masked from the consensus sequence. In total, 17 outliers were excluded in this analysis due to high numbers of heterozygous base calls, which is an indicator of contamination.

Mapping coverage and divergence from the *P. aeruginosa* PAO1 reference genome were calculated using samtools depth (version 1.6), with default parameters. There were 57 sequences that mapped to PAO1 with 100% coverage and 0 SNP divergence. These isolates were all sequenced from the same growth plate, and hence were determined to be the result of a contamination of *P. aeruginosa* PAO1 lab reference strain. These isolates were removed from the analysis.

In total, 4,094 *P. aeruginosa* DNA sequences were carried forward for further analysis. *De novo* genome assembly was carried out using Velvet (version 1.2, Zerbino & Birney, 2008) with a k-mer length of 31, and VelvetOptimiser (version 2.2.5) with a k-mer length of 31. Scaffolds were annotated using PROKKA (version 1.5, Seemann, 2014), using default settings and specifying the *Pseudomonas* genus.

3.2.2.1 Mapping to internal references and specific genes

For all analyses where mapping to internal references was required, the highest quality assembled isolate from each patient was selected, based on N50 and total number of contigs. After identifying K7, K9, K11, and K15 isolates as the Liverpool Epidemic Strain (LES), these isolates were re-mapped to the *P. aeruginosa* LESB58 reference genome (accession number: PRJEA31101) for further analysis. Details of these internal references are provided in Table 3.2.

Patient ID	Reference ID	Total length	Number of contigs	Average contig length	Longest contig	N50
K1, K14	21320_8#5	6,241,677	18	346,760	1,673,009	768,845
K3, K6	21513_8#29	6,248,324	12	520,694	3,598,558	3,598,558
K4	21931_2#284	6,208,894	18	344,939	1,329,670	831,824
K7, K9, K11, K15	*P. aeruginosa* LESB58	6,601,757	1	6,601,757	6,601,757	6,601,757

Table 3.2 Assembly statistics for internal reference sequences.

Contig breaks were removed and mapping was carried out as in section 3.2.2. Any SNPs within 150 bp of the contig breaks were excluded from all further analysis to avoid structural errors.

3.2.2.2 Sequence manipulation

For analysis of variant sites within the genomes, SNPs were extracted from the multiple sequence alignments generated in section 3.2.2 using snp-sites (version 2.4.1, Page *et al.*, 2016) with default parameters.

3.2.3 Phylogeny

The maximum-likelihood phylogenetic tree of all 4,094 isolates was generated with FastTree (version 2.1.10, Price *et al.*, 2010), using the Jukes-Cantor + CAT model using the SNP multiple sequence alignment generated from mapping all isolates to *P. aeruginosa* PAO1 in section 3.2.2. Phylogenies for each individual patient, or groups of patients, were generated from the SNP multiple sequence alignment produced by mapping to an internal reference as in section 3.2.2.1, using RAxML (version 8.2.8, Stamatakis *et al.*, 2014) with AVX vector instructions, and replicated for 100 bootstraps.

Circular genome visualisations with metadata were visualised using iTOL (Letunik & Bork, 2019). All other metadata were displayed next to phylogeny using phandango (Hadfield *et al*., 2017).

3.2.4 Genomic clustering

PCA was carried out using smartpca, as part of the EIGENSOFT (version 5.0, Patterson *et al.*, 2006) software package. The SNP alignments from section 3.2.2 were manually curated into the smartpca-accepted "snp" and "geno" file format from the vcf file. The isolate name and patient ID were combined to make the "ind" file format. The first 30 PCs were calculated, and all graphs were created using ggplot2 (version 3.1.0, Wickham, 2016) R package.

The MLST of all isolates were assigned using ARIBA (version 2.12.1, Hunt *et al.*, 2017) and PubMLST, which contained 3,308 sequence type assignments (as of June 2019). Any initially unassigned alleles were submitted to the PubMLST database for novel ST assignment.

The SNP alignment file generated from mapping to the *P. aeruginosa* PAO1 reference genome in section 3.2.2, and the corresponding phylogenetic trees generated in section 3.2.3, were analysed for phylogenetic clustering by FastBaps, using the R packages Fastbaps (version 1.0.0, Tonkin-Hill *et al.*, 2019), and all required dependencies.

3.2.5 Recombination analysis

Gubbins (version 1.4.6, Croucher *et al.*, 2015) was used to analyse recombination within the isolates, using the multiple sequence alignment files generated for each patient in section 3.2.2. Clustering of non-synonymous SNPs in a small window could be indicative of recombination. Therefore, to identify even the smallest trace of recombination, the default parameters were altered to find a density of 3 SNPs within a sliding window of 1,000 bp. Any reported results that were within 75 bp of contig breaks in the reference were excluded. Any genes within the recombination window were compared to the *P. aeruginosa* PAO1 reference genome using BLAST (version 2.7.1, Camacho *et al*., 2009) to identify possible recombination donors.

3.2.6 Pangenome analysis

The pangenome was calculated from the annotation files generated in section 3.2.2 using Roary (version 1.7.1, Page *et al.*, 2015) for every isolate within each patient and patient group, using a minimum percentage identity of 95% and a gene was identified as "core" if it was present in 99% of samples or more. The gene presence and gene absence data was displayed against the phylogeny as in section 3.2.3, and assignment of function of blocks was based on the majority function of the non-hypothetical protein genes from the PROKKA annotation files. All correlations and linear regression statistical tests were calculated in the R environment 3.5.0. In order to identify donors of putative acquired genetic regions, each region was extracted from assemblies where the genes are present, and compared to nucleotide sequence databases using BLAST.

3.2.7 Temporal analysis

The dating analysis was carried out using the R package BactDating (version 1.0.1, Didelot *et al.*, 2018), which calculates time to most-recent common ancestor using MCMC-based Bayesian dating of the nodes of the phylogeny. The phylogeny calculated for each patient in section 3.2.2.1 were used as the input, with the date of sampling of each isolate expressed in days since January 1st 2013. The analysis was calculated over 100,000,000 MCMC chains so that all ESS values were over 100, and three separate evolutionary models were used; strict gamma, relaxed gamma, and mixed gamma to estimate a consensus time to MRCA.

3.2.8 Transmission analysis

To investigate transmission, the pairwise SNP distances from the multiple sequence alignments generated in section 3.2.2 were calculated for every pair of isolates within each patient group. All phylogenies were generated and displayed as in section 3.2.3. All correlation and linear regression analyses were calculated in the general R environment 3.5.0.

3.2.9 Phylogenetic clustering

Individual patient phylogenies were calculated and displayed as in section 3.2.3. The exacerbation states (non-exacerbation, exacerbation or recovery) were assigned to each isolate from the anonymised patient data, and displayed against the phylogeny as in section 3.2.3. To test for phylogenetic signal and clustering of isolates on the tree, Pagel's lambda and Fritz and Purvis' D values were calculated as part of the Phytools (version 0.6.60; Revell, 2011) and Caper (version 1.0.1) R packages, respectively.

3.2.10 *P. aeruginosa* population context

To investigate where the isolates sit relative to PAO1, one representative isolate was selected from within each patient's phylogeny, and mapped to PAO1 alongside the 352 isolates from Chapter 2. The phylogeny was then calculated and displayed as in section 3.2.3.
The ExoS and ExoU status of each isolate was confirmed by mapping the fastq files of every isolate to the ExoS and ExoU genes, as in section 3.2.2.1.

3.2.11 Genetic AMR profiling

The genetic AMR profiles of each isolate were calculated using ARIBA (version 2.12.1), with reference genes and reference SNPs downloaded from the Comprehensive Antimicrobial Resistance Database (version 3.0.1, Jia *et al*., 2017). The presence/absence of each reference gene or SNP was then displayed against each patient phylogeny, as in section 3.2.3.

3.3 Results

3.3.1 DNA sequencing and quality checks

In order to investigate whether genetic variation of *P. aeruginosa* isolates is associated with APEs, DNA was sequenced and analysed from 4,400 *P. aeruginosa* isolates from the lungs of nine patients with CF before, during and after an APE. Of the 4,400 *P. aeruginosa* isolates from nine patients with CF that initially underwent the DNA extraction procedure, 4,168 isolates had sufficient DNA for sequencing using Illumina HiSeq 2500 and X10 machines. In order to identify and remove contamination, the isolates were mapped to the reference *P. aeruginosa* strain PAO1, the number of heterozygous base calls for each isolate were calculated and 17 isolates that lay outside of the 95% confidence intervals were removed. A further 57 sequences were removed from the study as contamination of the *P. aeruginosa* PAO1 lab reference strain. In total, 4,094 sequences were carried forward for further analysis.

The raw sequence data for each of the 4,094 *P. aeruginosa* isolates was assembled, with an average average genome length of 6,342,381 bp, ranging from 6,154,712 - 6,519,436 bp. The average number of contigs was 34, ranging from 12 - 1,902 contings, and the mean N50 value was 602,740 bp, ranging from 6,147 - 3,598,558 bp. The mean largest contig was 1,088,412 bp, ranging from 28,943 - 3,598,558 bp. All isolates were also mapped to the reference *P. aeruginosa* strain, PAO1. The mean mapping coverage was 93.7% (ranging from 84.1% - 95.9%), with a mean depth of 44x (ranging from 15x - 77x). A mean of 23,864 SNPs per sequence were called against the PAO1 reference (ranging from 19,441 - 24,867 SNPs).

3.3.1.1 Isolate metadata

A breakdown of the exacerbation-associated data from each of the nine patients, and the corresponding number of sequenced *P. aeruginosa* isolates, are outlined in Table 3.3. A complete exacerbation is defined as when isolates from all three defined exacerbation timepoints were sequenced; stable timepoint - seven days before an exacerbation; acute timepoint - the day the patient first received antibiotics; recovery timepoint - seven days after antibiotic treatment finished. In total, ten complete exacerbations

Patient ID	Age during study	Complete Exacerbations	Incomplete Exacerbations	Total Exacerbations	Isolates in Study
K1	31	2	1	3	639
K3	32	4	0	4	1013
K4	21	1	2	3	654
K6	24	0	1	1	183
K7	52	1	0	1	373
K9	36	1	2	3	637
K11	34	0	1	1	187
K14	24	1	0	1	219
K15	19	0	1	1	189
Total	-	**10**	**8**	**18**	**4094**

Table 3.3 *P. aeruginosa* isolates were collected for sequencing from nine patients with CF, aged 19 - 52, during different exacerbation timepoints. Complete exacerbations are defined as the presence of isolates from the acute, recovery and stable timepoints. Incomplete exacerbations are defined as the presence of isolates from only two out of the three exacerbation timepoints.

were recorded from six patients. Only two patients, K1 and K3, provided isolates during more than one complete exacerbation. Eight incomplete exacerbations were also recorded, defined as when isolates from only two out of the three exacerbation state timepoints were sequenced. For three patients (K6, K11, and K15), isolates were only provided from incomplete exacerbations.

3.3.2 Genome clustering

Principal Component (PC) analysis was performed on the genotype of each isolate to visualise the highest level of genetic variation within the 4,094 *P. aeruginosa* isolates. The first two principal components separate the isolates into four distinct clusters (Figure 3.1); isolates from patients K1 and K14; K3 and K6; K7, K9, K11 and K15; and K4 only. Every isolate from each patient was contained within each cluster.

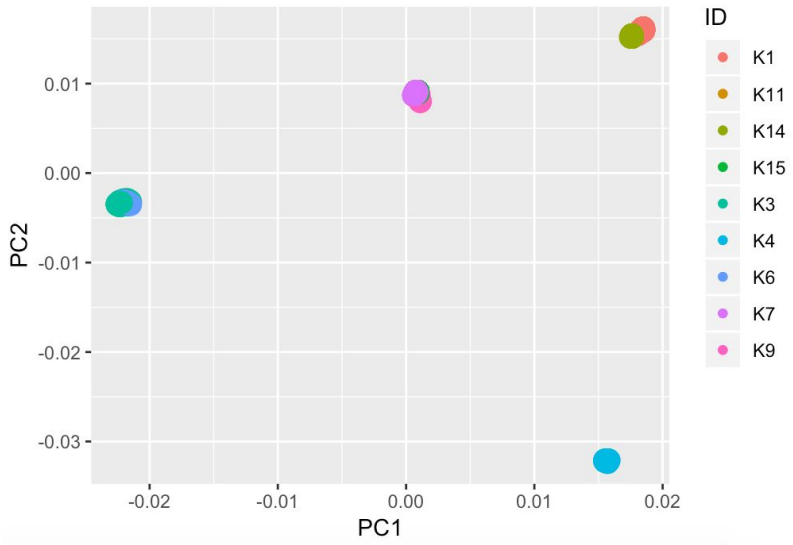


Figure 3.1 The first two Principal Components of patient genotype are split into four distinct clusters, which contain all isolates of the patients contained within the cluster.

The same clustering patterns are retained when plotting PC2 against PC3 (Figure 3.2a). However, when PC3 is plotted against PC4 (Figure 3.2b), the isolates from patients K1 and K14 diverge into two separate clusters. In total, 93.4% of the genetic variance can be explained by the first three PCs; 33.8%, 30.6% and 29.0% respectively, after which point there is an elbow (Figure 3.3). This elbow indicates deep divergence and separation of each cluster within the dataset. The following PCs account for a very small amount of variance, with PC 4, 5 and 6 only accounting for 3.7% of the total variance, which suggests that the isolates within each cluster are very closely related to each other.

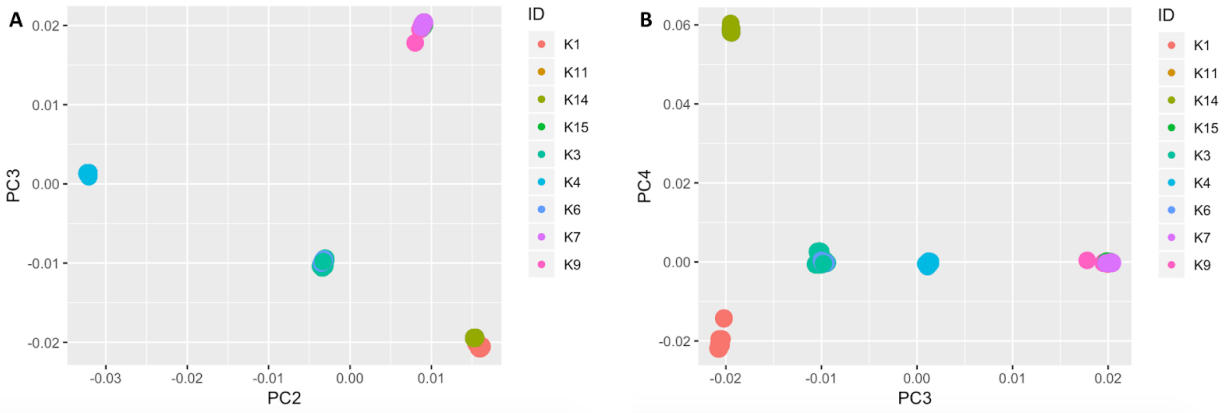


Figure 3.2 Principal Component analysis of isolate genotype for a) PC2 vs PC3 and b) PC3 vs PC4.

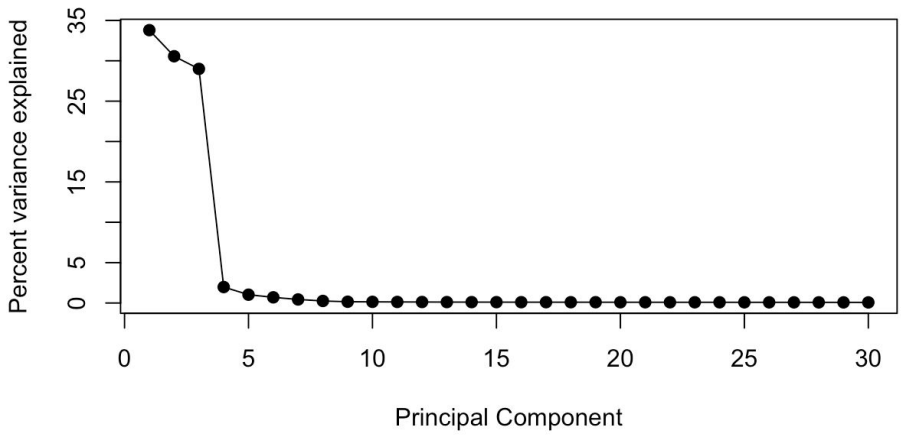


Figure 3.3 Percent of variance that can be explained by each principal component, for the first 30 PCs of every isolate genotype.

The multi-locus sequence type (MLST) profiles of all 4,094 isolates are summarised in Table 3.4. MLST is a method of distinguishing different subtypes of the same bacterial species by comparing the different allelic profiles of seven housekeeping genes. In *P. aeruginosa*, these housekeeping genes are *acsA*, *aroE*, *guaA*, *mutL*, *nuoD*, *ppsA* and *trpE*. Each combination of alleles results in a numbered sequence type which is unique to that allelic profile (Jolley & Maiden, 2010). The MLST profiles show that during the study duration, each patient was infected with only one *P. aeruginosa* ST. However, some patients are infected with the same ST as other patients. The distribution of the STs match those of the clusters identified for the first two PCs in Figure 3.1.

Patient IDs	Sequence Type	Alleles of Housekeeping Genes						
		acsA	*aroE*	*guaA*	*mutL*	*nuoD*	*ppsA*	*trpE*
K1, K14	ST217	28	5	11	18	4	13	3
K3, K6	ST3307*	23	143	190	11	3	15	7
K4	ST3308*	36	5	12	3	4	4	52
K7, K9, K11, K15	ST146	6	5	11	3	4	23	1

Table 3.4 MLST allelic profiles of the isolates from each patient collected during the study. * indicates a novel ST identified in this study.

The MLST allelic profiles show that patients K1 and K14 were both infected with ST217, more commonly known as Manchester Epidemic Strain (MES) (Syrmis *et al*., 2014). The MES was first identified in Manchester, and was found to have spread from patient-to-patient within the original clinic, and eventually across the UK (Fothergill *et al.*, 2012). This high transmissibility was linked to enhanced virulence and high antibiotic resistance.

Patients K7, K9, K11, and K15 were infected with ST146, also known as the Liverpool Epidemic Strain (LES) (Syrmis *et al*., 2014). The LES isolates are amongst the most virulent and transmissible strains of *P. aeruginosa* that infect the CF lung. LES isolates harbour almost all known *P. aeruginosa* virulence genes, and as a consequence are associated with more aggressive infection, which leads to a greater and more rapid loss of lung function in infected patients (Jani *et al*., 2016).

Patients K3 and K6 were both infected with the same ST, a previously undescribed ST in the PubMLST database (Jolley & Maiden, 2010). Patient K4 was also found to be infected with a previously undescribed ST, but separate to the patient K3 and K6 ST. A representative sample was submitted to the PubMLST database to assign a ST to these isolates. Two novel STs were assigned; ST3307 for the patient K3 and K6 isolates, and ST3308 for the patient K4 isolates.

Variant sites of all 4,094 isolates mapped against the *P. aeruginosa* PAO1 reference strain were used to estimate a maximum-likelihood tree, to investigate the population structure of the dataset. The same clustering patterns that match the PC and MLST analyses were observed in the overall phylogenetic structure. Phylogenetic clustering was confirmed using FastBAPs (Figure 3.4).

The PC clusters, FastBAPs clusters, and MLST analysis all show that each patient was only infected with one strain of *P. aeruginosa* during the study period, and that some patients were infected with the same strain as other patients. This is typical of *P. aeruginosa* lung colonisation, where patients with CF are found to be infected with only one clone of *P. aeruginosa* at any point in time (Workentine *et al.*, 2011). The presence of two different strains has been observed in patients during the early years of their *P. aeruginosa* colonisation, but one strain eventually is able to outcompete the other (Fluge *et al*., 2001). Once an infection has established and become chronic, that patient ends up harbouring that infection for the majority of their lifetime (Cramer *et al.*, 2010).

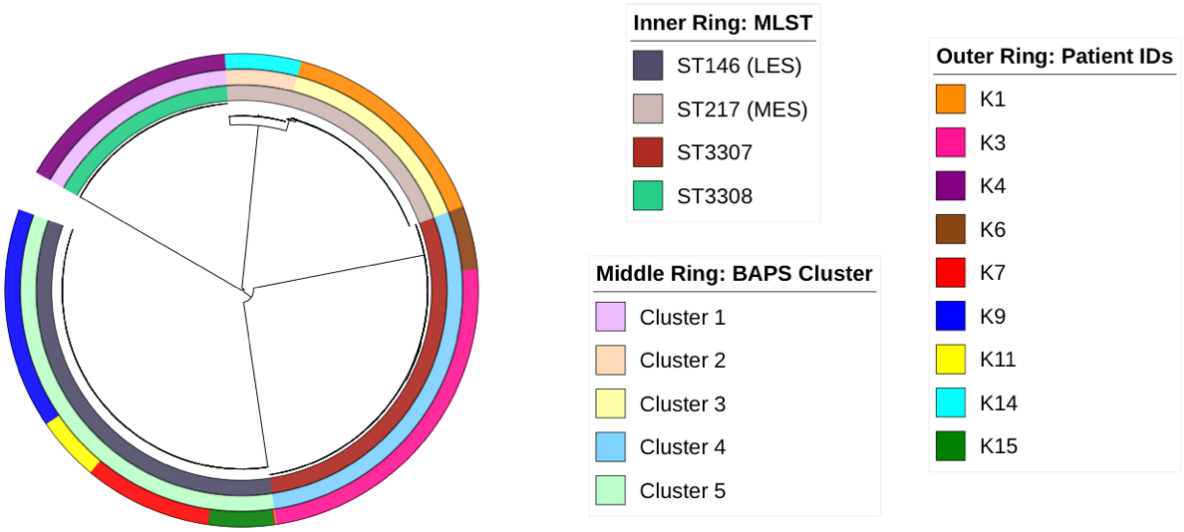


Figure 3.4 A maximum-likelihood tree of the *P. aeruginosa* isolates, with the clustering for each analysis displayed around the outside. Outer ring: patient ID; middle ring: BAPs Cluster; inner ring: MLST.

3.3.3 Individual patient phylogenies

To understand the extent of the diversity had been captured by the sampling of isolates from each individual patient, the maximum-likelihood phylogenies were calculated for each group of isolates from each individual patient (Figure 3.5). As all of the patients are colonised with only one *P. aeruginosa* ST, it follows that each infection is likely to have developed from a single introduction of that ST into the patient lung, and therefore all isolates within one patient should be clonally related to one another. The maximum-likelihood phylogenies for the isolates from patients K6, K7, K9 and K15, are all relatively unstructured, which indicates the presence of a single, diverse infection.

However, the patient K14 phylogeny indicates two separate populations of isolates, made up primarily of one population with a short root-to-tip distance, and a second, smaller population with a long root-to-tip distance. These long branches may indicate hypermutation within some of the isolates. Hypermutation is characterised by an increased rate of DNA mutation within the lineages, primarily due to the knockout of DNA mismatch-repair proteins. This has been linked to an increase in antibiotic resistance and adaptation to the CF lung (Mena *et al.*, 2008), and has been found to be present in roughly half of CF patients (Feliziani *et al.*, 2014).

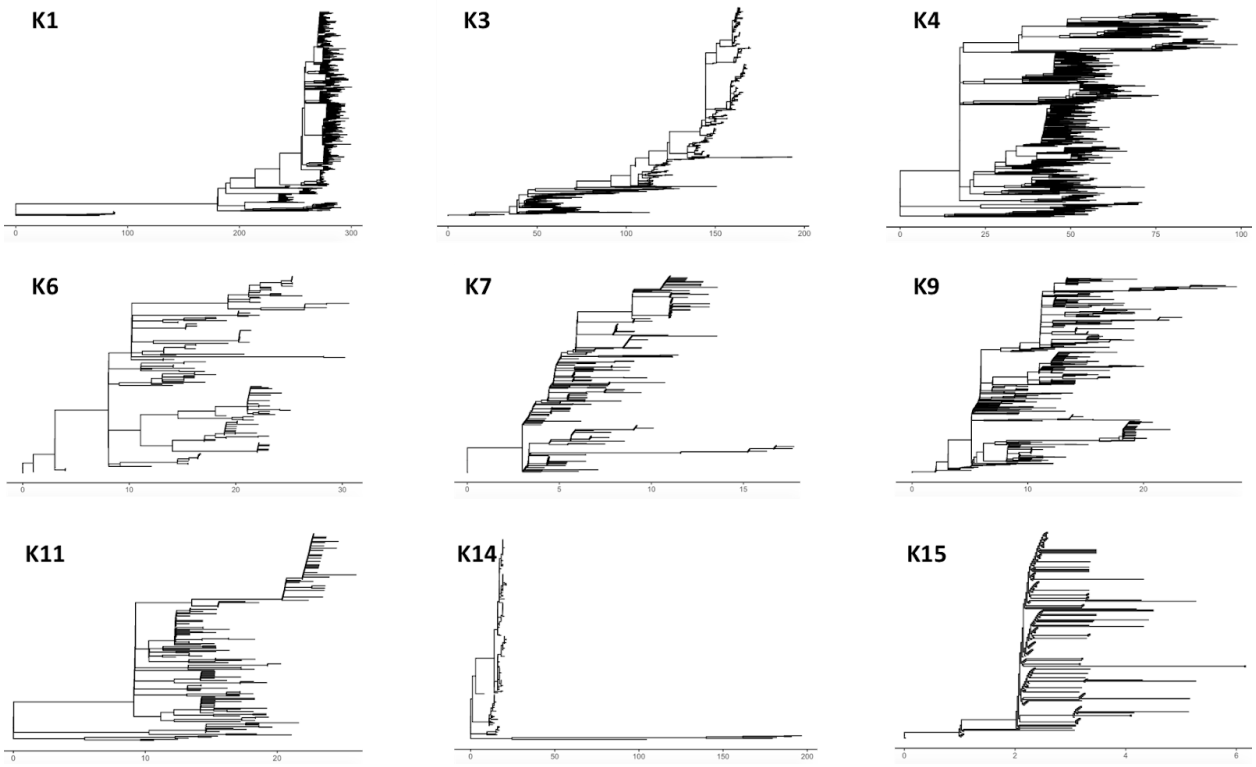


Figure 3.5 Outgroup-rooted maximum-likelihood phylogenies for the isolates from each individual patient. The x-axis indicate the number of SNPs.

The patient K1 phylogeny also indicates two separate populations within the lungs. The smaller population closer to the root may indicate non-hypermutation. The second population of isolates do not have long branches but are branched off a long branch, which could indicate that this second population of isolates are descended from a hypermutator lineage that has since reverted.

The patient K3 isolates are characterised by longer branches in the isolates towards the root of the phylogeny, and a more structured expansion further away from the root. This could indicate selection pressures within the patient's lungs, showing lineage replacement and expansion of newer clones.

The tips of the patient K4 phylogeny are all separated by long branches, with no evidence of a recent clonal expansion. This could be indicative of all of the isolates currently hypermutating. It is also possible that the infection within the lungs is simply diverse, without there being an expansion of any particular lineage.

Both the K4 and K11 phylogenies contain two divergent lineages within the phylogeny that are not indicative of hypermutation, but could instead indicate two separate evolutionary paths brought on by different selection pressures within different niches of the CF lung. Divergent, co-existing lineages have been identified in a study by Williams *et al.* in 2015, in eight CF patients infected with the LES (Williams *et al.*, 2015).

3.3.4 Intrapatient diversity

The diversity of *P. aeruginosa* isolates from within each patient's lungs may be correlated with the time of sample collection, e.g. before, during or after an APE. Therefore, the intrapatient diversity was compared with the stage of APE. The intrapatient diversity for all isolates mapped to a patient-specific internal reference is listed in Table 3.5. A number of isolates from every patient were identical, which corresponds with a minimum pairwise SNP distance of 0. However, the mean and maximum pairwise SNP distances varied within each patient. The greatest diversity was observed in the isolates from patient K1, with a mean pairwise SNP distance of 84, and a maximum of 361. High-diversity was identified in K3, K4 and K14, with a maximum pairwise SNP distance of over 150 within each of those patients. High levels of within-host genotypic diversity is typical of *P. aeruginosa* chronic lung infections. In one previous study, a maximum of 130 SNPs separated 40 isolates collected over the course of a year (Williams *et al.*, 2015), and a second study found a maximum of 121 SNPs separating 22 isolates collected from one sputum sample (Darch *et al.*, 2015). The high levels of diversity that were identified in this chapter suggests that for patients observed with lower within-host diversity, the infection may be more recent.

Patient ID	Pairwise SNPs			% isolates with 0 SNPs difference
	Minimum	Mean	Maximum	
K1	0	84	361	0.03
K3	0	48	190	1.29
K4	0	72	159	0.01
K6	0	16	35	1.35
K7	0	5	16	3.13
K9	0	10	30	0.90
K11	0	13	39	2.07
K14	0	19	232	4.21
K15	0	1	8	27.76

Table 3.5 Pairwise SNP distances for all isolates within each patient.

The number of isolates sampled from each patient was plotted against the mean and maximum pairwise SNP distances (Figure 3.6), which gave a general correlation ($r = 0.62$ and $r = 0.45$ respectively), but was statistically insignificant when tested under linear regression ($p = 0.073$ and $p = 0.224$ respectively). This suggests that the intrapatient diversity was not greater within patient groups where a higher number of isolates had been sequenced, which means that sampling diversity is independent of genetic diversity.

There was no statistically significant correlation (linear regression, $p = 0.548$) between the mean pairwise SNP distance and patient age, and the number of samples from a patient were also not correlated with patient age (linear regression, $p = 0.690$) (Figure 3.7).

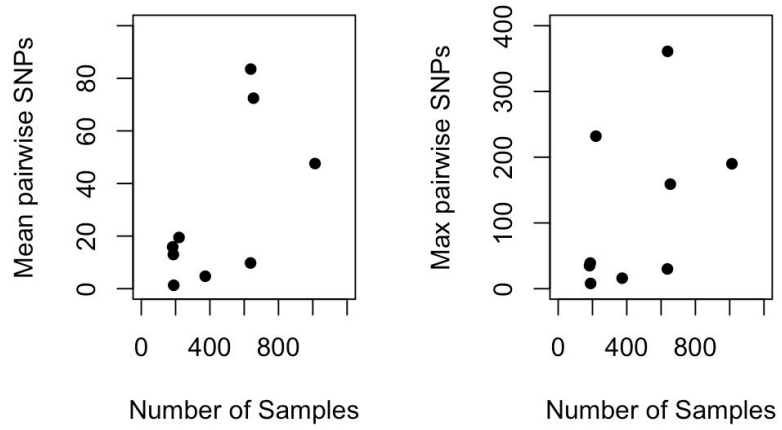


Figure 3.6 The correlation between the number of isolates from each patient and the mean and maximum pairwise distances within those isolates are not statistically significant.

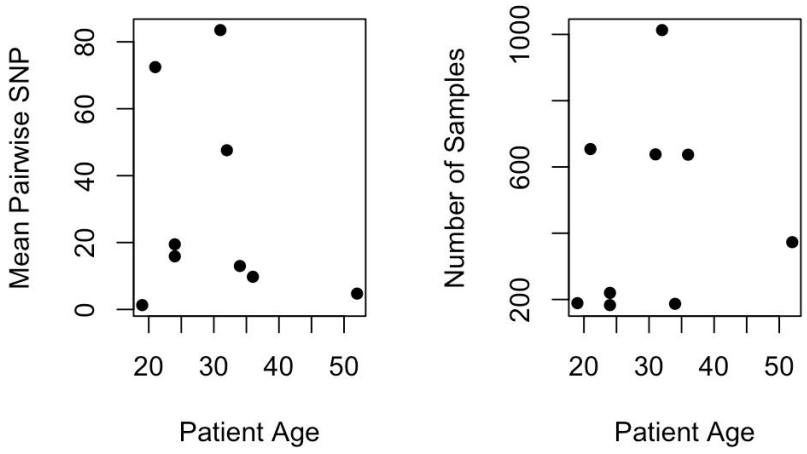


Figure 3.7 Patient age was not significantly correlated with mean pairwise SNP distance, and number of samples.

To investigate whether the pairwise SNP diversity varied across different exacerbations, the pairwise SNP distributions were plotted for each exacerbation for each patient (Figure 3.8). The range of pairwise SNPs was found to not vary between different exacerbations from the same patient. This suggests that the population structure and diversity within a single patient does not change from one exacerbation to another.

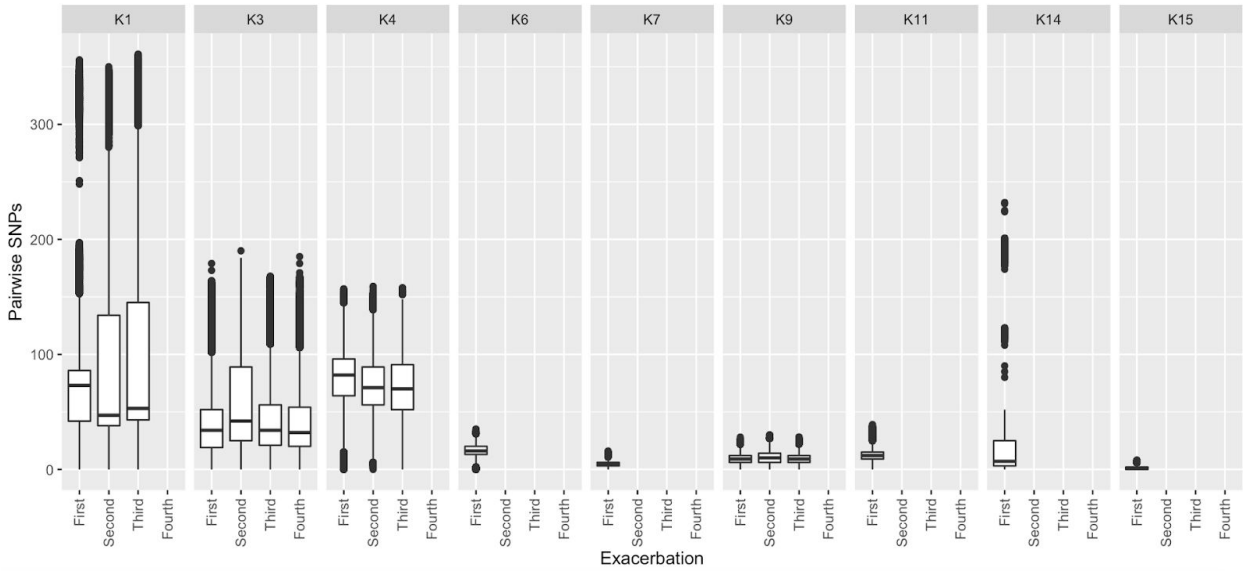


Figure 3.8 The pairwise SNP distances for each exacerbation that each patient experienced.

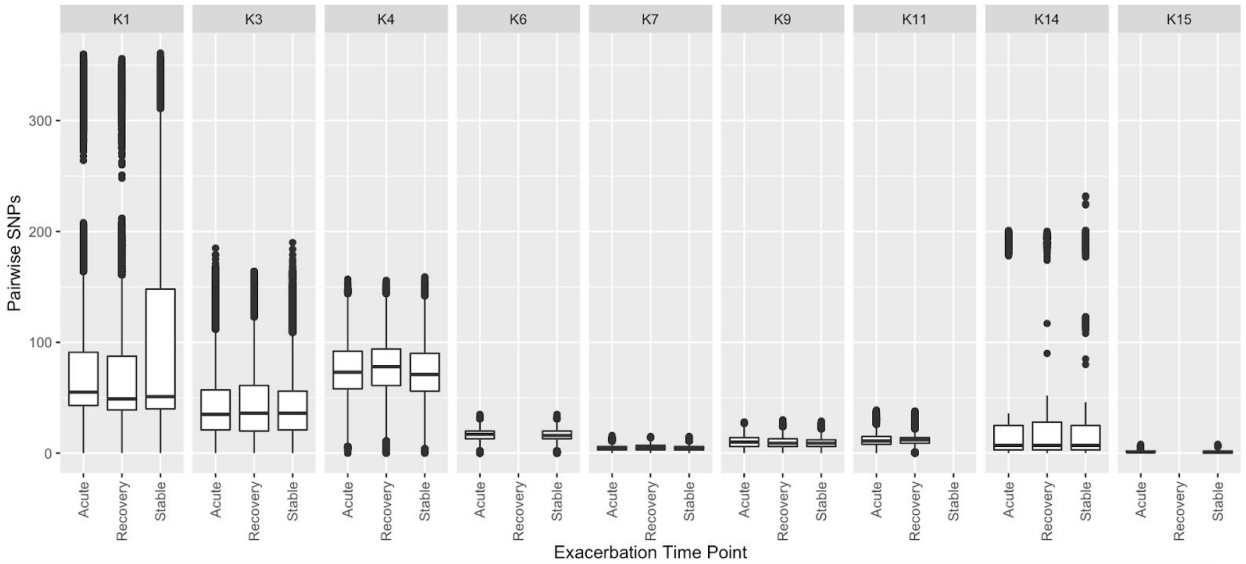


Figure 3.9 The pairwise SNPs for each exacerbation timepoint that each patient experienced are plotted.

However statistically significant variation in diversity was observed in isolates collected from each of the different exacerbation timepoints from patients K1, K3, and K14 ($p < 0.05$) (Figure 3.9). This significance is caused by a difference in the number of isolates collected from each of the acute, stable, and recovery time points within those exacerbations (refer to Table 3.3 which outlines the number of complete and incomplete exacerbations). This observed significance disappears when the distributions of isolates are normalised by the number of each exacerbation timepoints experienced. This indicates that *P. aeruginosa* diversity within the lung of CF patients does not vary with different exacerbation timepoints, or between different exacerbations experienced by the same patient. This suggests that the *P. aeruginosa* diversity is not altered by the physiological changes and medical treatment associated with an exacerbation, or with any treatment received during the time between exacerbations.

This finding is supported by a 2010 study by Fothergill *et al.* who identified very little genetic variation in PCR assays in 60 isolates from sputum samples taken before an exacerbation, during treatment of an exacerbation, and after recovery from an exacerbation (Fothergill *et al.*, 2010). A follow up study in 2011 by Mowat *et al.* of 1,720 *P. aeruginosa* isolates from 43 samples associated with the three stages of exacerbation also identified very little change of haplotype composition within the CF lung over the course of exacerbation (Mowat *et al*., 2011). The lack of variation in SNP diversity identified by WGS undertaken in this study, also shows that the population of *P. aeruginosa* isolates within the CF lung remain stable over the course of an exacerbation.

3.3.4.1 Early adaptation to the CF lung

The patient K15 isolates are separated by very few SNPs, with a high proportion of isolates sharing no SNPs, suggesting that the strain has not had time to diversify within the host. This, combined with clinical data indicating that the patient developed chronic infection during their treatment regime at Papworth hospital (data not shown), indicates a recent infection. Therefore, these isolates could provide an insight into the early development of the *P. aeruginosa* infection within the lung of patient K15.

The 189 K15 isolates contain just 79 variant sites when compared with the *P. aeruginosa* LESB58 reference genome, only one of which is a fixed variant present in all isolates. This suggests that patient K15 is infected with the *P. aeruginosa* LESB58 strain, and the 78 non-fixed variant sites have occured within the lungs of the patient. Of these 78 non-fixed variant sites, 52 of the mutations occurred in

hypothetical proteins, 8 occurred in intergenic regions, and 18 occurred in gene regions of known function. All of the mutations that appeared in more than one isolate appeared along a shared branch between the isolates, and none of the variant sites appeared twice or more independently. This is evidence that recombination between this group of isolates has not yet occurred.

Mutations in two of the hypothetical proteins introduced early stop codons. In the hypothetical protein PALES_24321, the stop mutation R5* has been introduced. This protein is predicted to be a part of the T2SS machinery, responsible for secreting virulence factors. Alongside this, a non-synonymous mutation was also introduced in the *phoA* gene, which is responsible for biosynthesis of an alkaline phosphatase that is secreted by the T2SS (Liu *et al.*, 2016). Loss of key virulence factors are an indicator of the development of chronic infection (Winstanley *et al.*, 2016), and so this loss of T2SS may be a signal of early chronic development in the lung of patient K15. Additionally, there is also a non-synonymous mutation in the *pcrD* gene, which is a part of the T3SS responsible for secreting protease enzymes (Galle *et al.*, 2012). Loss of the T3SS is a common phenotype observed in the switch from acute to chronic infection (Jain *et al.*, 2004). The second hypothetical protein with an early stop codon is PALES_24761, in which the mutation Q117* has been introduced. This hypothetical protein has an unknown function, but is predicted to be an ATPase, with membrane secretion function. These mutations are all indicative of early loss of virulence, and hence adaption to the CF lung.

Four different non-synonymous mutations occurred in one gene, *tpbB*. The *tpbB* gene is also known as *yfiN*, an activator protein which produces c-di-GMP and initiates the pathways required for biofilm production. The four non-synonymous mutations within *tpbB* could knockdown or knockout the protein, which could reduce the ability of the K15 isolates to produce biofilm. Transition from non-biofilm-producing to biofilm-producing phenotypes are typical of development of chronic infection, and therefore these mutations in *tpbB* may be selected against over time, as biofilms help protect the bacteria from antibiotics as well as host defense (Winstanley *et al.*, 2016). One of the non-synonymous mutations in *tpbB* appears in three related isolates, and the other three non-synonymous mutations only appear in one isolate.

Two genes that contain non-synonymous mutations are responsible for fatty acid catabolism; *faoB* and *lpxK*. Previous studies have shown that *P. aeruginosa* fatty acid metabolism genes are upregulated during acute infection, and downregulated during chronic infection. However the *P. aeruginosa* genome encodes

seven similar dehydrogenase enzymes, suggesting that there is redundancy within the genome (Crousilles *et al.*, 2015).

Non-synonymous mutations in other genes responsible for a variety of cell mechanisms were also observed; *pchA*, which is involved in converting chorismate to isochorismate (Meneely *et al*., 2013); *thiD*, which is involved in thiamine biosynthesis (Poulsen *et al*., 2019); *phnC*, which is involved in phenazine biosynthesis (Zaborin *et al*., 2012); *ambE*, which is involved in the biosynthesis of the potent antibiotic *L*-2-Amino-4-methoxy-*trans*-3-butenoic acid (Murcia *et al.*, 2015); *recR*, which is a recombination regulation protein (Che *et al*., 2018); and *atpD*, which is an ATP synthase (Jorth *et al*., 2017). Transcriptional knockout studies of these genes would help uncover the relevance of these mutations to the adaptation of *P. aeruginosa* to the CF lung.

3.3.5 Recombination

Genetic diversity can be attributed to linear evolution through the accumulation of SNPs, or horizontal gene transfer such as recombination. Therefore, to determine the major cause of diversity within the 4,094 *P. aeruginosa* isolates, evidence of recombination was explored.

Very little structure was identified when the frequency of SNPs were compared against the location within the genome. In most populations, low frequency, random mutations are the most common variation present in the population. This suggests that all isolates from the same patient have an almost-identical genetic background, with mutations occurring across the genome randomly. The majority of the mutations are spread diffusely across the entire genome, suggesting that there are no high-frequency mutational hotspots. It is also unlikely that any variation has been introduced to these populations by recombination from external sources during the study period, as the intrapatient pairwise SNP distances were relatively small (section 3.3.4). There is debate in the literature about the role of recombination in the evolution of *P. aeruginosa* within the lungs (Winstanley *et al.*, 2016), but any recombination introduced prior to the last common ancestor would be undetectable using this analysis.

To investigate any potential signs of recombination or mutational hotspots within the isolates from each patient, dense regions of SNPs, which were defined as more than 3 SNPs within a 100-10,000 bp sliding

Patient ID	Region of genome	Base pair window	Number of SNPs	Number of isolates	Genes affected
K4	2790762 - 2791738	976	6	2	PA3965 – PA3966
K1	4073544 - 4075819	2275	8	11	PhuR - PhuS
K1	5281268 - 5281540	272	5	3	PctA
K3	6047602 - 6048600	998	5	1	PA2439
K3	5950589 - 5951246	657	5	1	PA4140
K3	4502640 - 4502856	216	5	3	PA0257
K3	1903948 - 1904202	254	7	10	PA0095
K3	69372 - 69976	604	5	5	PA2526

Table 3.6 Putative recombination windows in patients K4, K1 and K3.

window, were identified (Table 3.6). In three of the nine patients' isolates, dense regions of SNPs were identified. In the remaining six patients' isolates, no evidence of recombination or mutational hotspots were identified.

Six SNPs were identified within a 976 bp window within two isolates from patient K4. This window covers the genes *PA3965* and *PA3966*, which both encode hypothetical proteins and are not within a known *P. aeruginosa* recombination hotspot. There are no records in the NCBI database which have a 100% match with these recombined alleles, however, there are nine isolates from the bacteraemia dataset (Chapter 2), which do match 100% with these alleles. These are distributed across sequence types from across the whole population, suggesting that this is a frequently mutating or frequently recombining region.

Eight SNPs within a 2,275 bp window were identified in eleven isolates from patient K1. These mutations are in the *phuR* and *phuS* genes, which are responsible for haem uptake in *P. aeruginosa*. BLAST analysis of both the NCBI collection and the BSAC collection indicate that this is a unique combination of alleles. The eleven isolates that contain these alleles are all clustered in a region of the phylogeny that is suspected to be descended from a hypermutator. Therefore, it's possible that this combination of alleles in this 2,275 bp window have arisen due to hypermutation of that region. Knockout of the *phuRS* operon is not fatal to *P. aeruginosa*, as compensation can occur by increasing pyocyanin siderophore production to

scavenge for extracellular free iron in place of iron used by haem uptake (Kaur *et al.*, 2008). This suggests very few detrimental effects due to mutating this region.

Three isolates from patient K1 also contain five SNPs within a 271 bp window that are contained within the chemotaxic transducer gene *pctA*. This very small region of potential recombination occured in three isolates related to the acute-APE timepoint of the third exacerbation experienced by the patient. This 271 bp DNA region also appears in the bacteraemia collection a total of 17 times. Eight of these isolates belong to the international and high-risk ST175 clade, appearing in all of the isolates associated with the Bristol hospital. The rest of the nine isolates that contain these SNPs appear across the *P. aeruginosa* population. The chemotaxic transducer *pctA* is responsible for the detection of extracellular amino acids, which starts a cascade that leads to the motility of the bacterium towards that source (Reyes-Darias *et al.*, 2015). However, as *P. aeruginosa* develops into chronic infection, motility is often lost, suggesting that this may be an accumulation of mutations in a gene which is functionally redundant in these isolates (Lozano *et al.*, 2018).

Five SNP-dense regions were identified within the patient K3 isolates. Two of these recombination regions only appear on one terminal branch of the K3 phylogeny. These recombination regions affect PA2439, and also appear in 14 bacteraemia isolates. Additionally, the regions affect PA4140, and also appear in seven bacteraemia isolates. These alleles of both genes are not found in the NCBI database. They encode hypothetical proteins with no known function associated with them.

A region of recombination is associated with ten isolates on one branch of the K3 phylogeny. This is associated with the conserved hypothetical protein PA0095, which does not have a predicted function. Isolates that contain SNPs in this region appear at least once in every exacerbation the patient underwent, meaning that this region of DNA was present both at the start and end of this study. These recombined alleles appear with 100% identity in 185 complete *P. aeruginosa* genomes in the NCBI database, including the *P. aeruginosa* PAO1 reference genome, and with 100% identity in 80% of the bacteraemia isolates, suggesting that the rest of the K3 isolates have the variable PA0095 gene. The alleles present in the rest of the K3 isolates have 100% identity with the *P. aeruginosa* LESB58 reference genome, a highly virulent LES strain. This suggests that the ST3307 isolates in K3 may have obtained DNA from a LES strain.

The region of recombination within the PA2526 gene is associated with a clade of five isolates in the centre of the patient K3 phylogeny. PA2526 is also a conserved hypothetical protein with no predicted function. BLAST analysis of these alleles confirms their presence within the *P. aeruginosa* LESB58 reference strain. The alleles belonging to the rest of the K3 isolates have 100% identity with the *P. aeruginosa* DK2 reference genome. This is a highly virulent and highly transmissible strain of *P. aeruginosa* which has infected patients in many CF centres in Denmark (Rau *et al.*, 2012). The DK2 strain of *P. aeruginosa* is closely related to the LESB58 strain, but has an additional 195 genes found across 3 genomic islands (Rau *et al.*, 2012).

The region of recombination within the PA0257 gene is associated with a clade of three isolates situated towards the root of the patient K3 phylogeny. This is also a hypothetical protein, and is predicted to have integrase functionality. The gene sequence containing the 5 SNPs identified in the recombination area, is present with 100% identity in three NCBI whole-genome sequences. All three sequences were submitted to the NCBI database in 2019, and are all currently unpublished. However, the metadata associated with the submissions suggest these isolates were collected in North America, Africa, and Europe. These alleles are also found in 13 spatially and temporally diverse isolates from the bacteraemia dataset.

Presence of hypothetical recombination regions from donors spread across the *P. aeruginosa* population, suggests that the chronic strains have been exposed to DNA from other *P. aeruginosa* since infection. The subsequent lack of coinfection within the patient lungs suggests that there has been absolute exclusion of the donor *P. aeruginosa* within the lungs, and supports the current theory that only one strain of *P. aeruginosa* establishes chronic infection.

3.3.6 Gene presence-absence within the pangenome

The SNP distribution in section 3.3.5 only accounts for diversity amongst the genetic material shared between each isolate and the corresponding reference sequence. Additional diversity can be identified by analysing the pangenome from within the genomic assemblies, which identifies variation between the gene content of each individual isolate. Very few studies have defined a pangenome within a CF patient, and those that have, have defined the pangenome as the change in gene presence-absence over time (Bianconi *et al.*, 2019).

The study by Bianconi *et al* in 2019 identified a core genome of 5,608 genes from 40 isolates collected over a period of 8 years from one patient with CF (Bianconi *et al.*, 2019). The Bianconi study identified several blocks of variable gene content in the isolates over the 8 years. In this chapter, 10,861 genes were identified within the collection of 4,094 isolates. Of these, 4,734 genes (43.6%) were identified as core genes, which are present in 99% or more of the isolates across all patients. The remaining 6,127 genes were identified as accessory genes. The pangenome for the isolates from each individual patient was then calculated (Table 3.7). This resulted in a large core genome, and a small accessory genome, which is consistent with a relatively clonal population within each individual population. The core genome composition ranged from 5,391 core genes (75.0%) in the isolates from patient K3, to 5,885 core genes (95.9%) in the isolates from patient K15.

Patient ID	Pangenome Composition		
	Core Genes (%)	Accessory Genes (%)	Total Genes
K1	5530 (88.0)	755 (12.0)	6285
K3	5391 (75.0)	1798 (15.0)	7189
K4	5478 (83.1)	1116 (16.9)	6594
K6	5439 (86.2)	870 (13.8)	6309
K7	5795 (91.1)	564 (8.9)	6359
K9	5826 (87.0)	874 (13.0)	6700
K11	5669 (85.3)	979 (14.7)	6648
K14	5767 (92.5)	468 (7.5)	6235
K15	5885 (95.9)	251 (4.1)	6136

Table 3.7 The components of each pangenome, calculated for each individual patient's isolates.

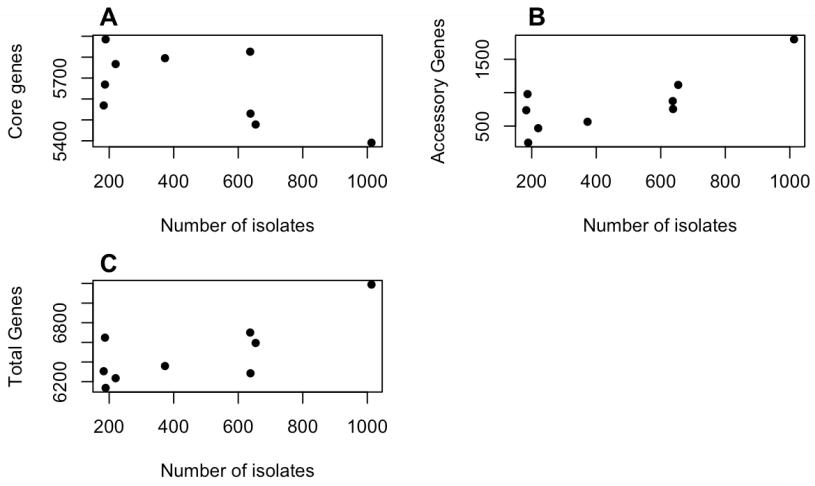


Figure 3.10 The correlation of each component of the pangenome with sample size for a) core genome, b) accessory genome, c) pan genome.

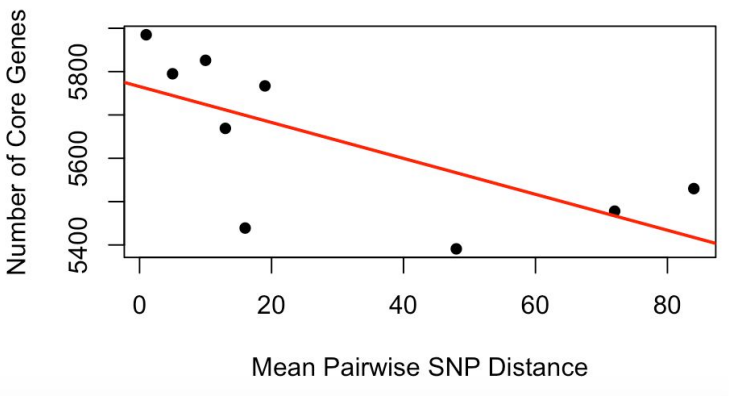


Figure 3.11 The number of core genes was negatively correlated with the mean pairwise SNP distance.

Figure 3.10 shows the correlation between each component of the pangenome with sample size. There was no statistically significant correlation between the number of core genes and the number of isolates (linear regression, $p = 0.0734$). There was a strong positive correlation between the number of accessory genes and the total number of isolates (linear regression, $p = 0.0104$), and between the number of total genes and the number of isolates (linear regression, $p = 0.0178$). There was no correlation between the number of accessory genes and the mean pairwise SNP distance (linear regression, $p > 0.05$). However there was a significant negative correlation between the number of core genes and the number of pairwise SNPs (linear regression, $p = 0.043$) (Figure 3.11), suggesting that diversification occurs via both point mutation and the acquisition or loss of genes.

3.3.6.1 Accessory genome

Accessory genes were defined as genes present in ≤ 99% of isolates. The median number of novel genes per isolate is reported in Table 3.8. On average, 1 - 2.5 novel genes are introduced per isolate per patient. These novel genes could be evidence of variation introduced by genomic islands, phage, or plasmids, or could be due to noise as a result of assembly error.

If all of the isolates from each patient infection were clonal, then all genetic variation within the population should be as a result of SNPs. When defining the pangenome, every gene sequence that is interrupted by a contig break is defined as a novel gene. Therefore, in order to investigate whether the presence of accessory genes is due to contig breaks, the number of contigs were plotted against the number of novel genes for each patient (Figure 3.12). Significant correlations (linear regression, $p < 0.05$)

were recorded for eight out of the nine patients, which suggests that much of the variation in the accessory genome is down to assembly error. This is the case for the outlier isolates in patients K3, K4, K6 and K11, which, due to the large number of contigs, have recorded a large number of accessory genes.

For the isolates from patient K14, no significant correlation was identified. There also was no significance recorded between the number of contigs and the total number of accessory genes per patient.

Patient ID	Median Novel Genes	95% Confidence Interval	Range
K1	1.00	1.00 – 1.50	0 – 7
K3	1.50	1.50 – 1.50	0 – 254
K4	1.50	1.00 – 1.50	0 – 44
K6	1.50	1.50 – 2.00	0 – 190
K7	1.50	1.00 – 1.50	0 – 12
K9	1.50	1.50 – 1.50	0 – 9
K11	2.50	2.50 – 3.50	0 – 30
K14	1.50	1.50 – 2.00	0 – 88
K15	1.00	1.00 – 1.50	0 – 6

Table 3.8 Novel genes per isolate for each patient group.

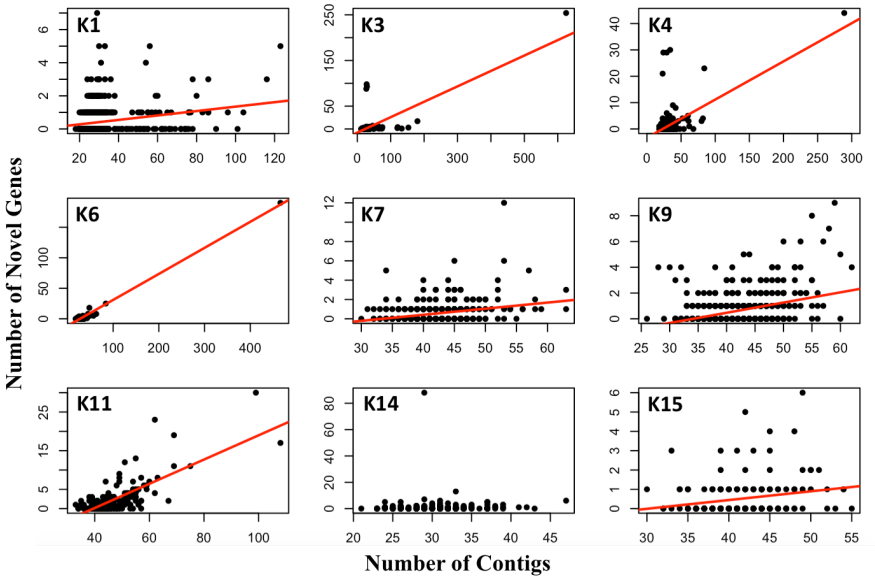


Figure 3.12 The number of contigs plotted against the number of novel genes in the isolates from each CF patient. The red line is the line of best fit, where the association was significant.

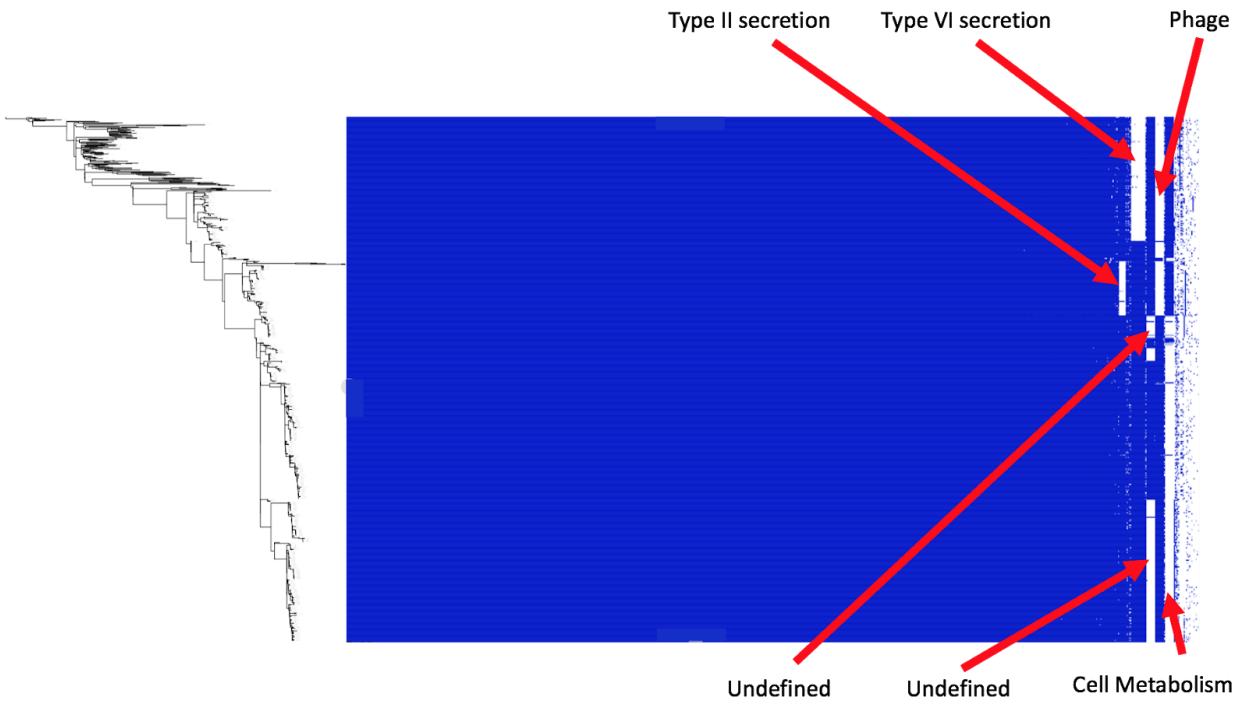


Figure 3.13 Six major blocks of gene presence-absence are present in the pangenome of the isolates in patient K3, which can distinguish between sub-populations. Blue = gene presence, white = gene absence.

A single isolate from the patient K14 lung had a much larger number of novel genes that the other isolates, even though the genome sequence was made up of an average number of contigs. These 88 novel genes are related to a genomic island carrying efflux systems related to the Type IV efflux system in *P. aeruginosa*. BLAST analysis of this genomic island indicated eight isolates in the sequencing databases that have over 90% query cover and over 99% identity with this region. These are all unpublished *P. aeruginosa* genomes, which are globally spread. The earliest isolate was a bacteraemia isolate from the USA in 2001, and the other isolates included two sputum samples from the USA in 2006, three isolates with no metadata, and two isolates collected from ocean samples in 2003 from Japan. The genomic island also shares 73% query cover and 97% identity with a bacteraemia isolate found in Germany in 1985. This suggests that this genomic island is globally spread and has persisted in *P. aeruginosa* for a long time. BLAST analysis also identified several non-Pseudomonad matches with 67% query cover and 97% identity, from *Delftia tsuruhatensis*, *Pandoraea apista*, and *Burkholderia multivorans*.

Plotting the gene presence and absence data for each isolate from patient K3 against the phylogeny identified six regions of contiguous gene presence or absence (Figure 3.13). These blocks could be

attributed to a particular lineage within the patient, suggesting that, whilst most of the accessory genome can be attributed to assembly error, there exists definite variation in the gene content of the isolates. Three of the blocks of genes that were absent from the genome could be ascribed a function (type II secretion, type VI secretion, and cell metabolism), and one block contained phage-related proteins that share 100% identity with the currently-unpublished *P. aeruginosa* L10 genome. The other two blocks encoded an assortment of hypothetical proteins.

The Type VI secretion block of genes, and the phage proteins, appear to have been acquired, as the genes are absent towards the root of the phylogeny. The area related to cell metabolism appears to have been lost from the isolates, as they are present towards the root of the phylogeny, and absent as the phylogeny develops.

The cell metabolism block represents a loss of 48 kbp, spanning PA1957 - PA2002 in the *P. aeruginosa* PAO1 reference genome. In this region is the well-characterised and highly-conserved PQQ pathway, which produce redox cofactors that allow the non-glycolytic production of ATP in bacteria, most notably from alcohol dehydrogenase enzymes (Barr *et al.*, 2016; Puehringer *et al.*, 2008). This region also encodes the *exaABC* operon, which is responsible for the oxidation of ethanol (Crocker *et al.*, 2019), and the corresponding regulatory network (Mern *et al.*, 2010). As well, this region encodes the *dhcABR* operon and *atoBE* operon, a knockout of which prevents growth of *P. aeruginosa* on the carbon source *L*-carnitine, and leads to reduced growth when *P. aeruginosa* uses *L*-phenylalanine as a source of acetyl-CoA (Palmer *et al.*, 2010). Within the CF lung, *L*-carnitine and choline, and their subsequent alcohol catabolites, leads to the induction of certain virulence factors (Wargo, 2013). A knockdown of the ability to produce virulence factors is an indicator of a switch to chronic infection.

The region labelled as Type II secretion involves a localised 41 kbp deletion, spanning PA0685 - PA0710. The start of this region involves a deletion of the majority of the *hxc* Type II secretion system, along with the secreted substrate. The *hxc* system was the second T2SS identified in *P. aeruginosa*, and has been found to secrete the *lapA* and *lapB* alkaline phosphatase enzymes under phosphate-limiting growth conditions (Cadoret *et al.*, 2014). This operon is located next to a region containing several secreted virulence factors, such as the *pdt, phb,* and *exb* exoenzymes (Quesada *et al.*, 2016). Finally, the other deleted genes include *toxR*, a regulator of secreted virulence exoenzymes (Wozniak *et al.*, 1987), *migA*, involved in the biosynthesis of outer-core oligosaccharides (Poon *et al.*, 2008), and the *cat*

chloramphenicol resistance gene (Beaman *et al*., 1998). As previously, the loss of virulence factors is typical of the *P. aeruginosa* switching from acute to chronic infection.

The region labelled Type VI Secretion System indicates an acquisition of a 43 kbp region, which corresponds to PA2341 - PA2376 in the *P. aeruginosa* reference genome. This area includes 13 genes of one of the T6SS in *P. aeruginosa* (Sana *et al*., 2013). Also in this region are the *mtl* genes, which are responsible for mannitol utilisation (Bruker *et al.*, 1998), and the *msu* genes responsible for the utilisation of organosulfonates (Tata *et al.*, 2016). In order to identify a donor strain of this region of DNA, BLAST was used to compare the acquired DNA sequences with the nucleotide sequence database. The nucleotide sequence was found to have a 100% query cover and a 99.77% identity with the *P. aeruginosa* LESB58 reference genome. This, along with the LESB58-like blocks identified in the recombination analysis, suggests that the ST3307 isolates have come into contact with the LES strain within the lung of patient K3, before outcompeting the strain to become the dominant, chronic clone. Replacement of established chronic infection by LES isolates have been documented in the literature due to their high virulence and transmissibility (McCallum *et al.*, 2001; Fothergill *et al.*, 2012). It is therefore possible that patient K3 has been chronically infected by the ST3307 strain, subsequently come into contact with the LES strain, but the LES strain has failed to replace ST3307 as the dominant strain within the lungs. It is during this process that genetic material could have been passed between the two isolates, the signatures of which have been observed in this analysis.

Two blocks of gene presence-absence were also associated with a lineage in the isolates from patient K9 (Figure 3.14). The smallest block of 15 kbp covered genes PA3453 - PA3462 in the *P. aeruginosa* PAO1 reference genome. These encode hypothetical proteins with no known function. The other block is comprised of 21 kbp, and covers genes PA0287 - PA0303 in the *P. aeruginosa* PAO1 reference genome. These regions include the *gpuARP* operon, which converts 3-guanidinopropionate into acetyl-CoA to be used in the TCA cycle (Nakada & Itoh, 2005). The *aguABR* operon is also in this deleted regions, which is responsible for converting agmatine into putrescine (Nakada & Itoh, 2003). Also in this region is also part of the *spu* operon (*spuA, B, C, D, E, and I*), which, when deleted, prevents uptake and utilisation of the polyamine spermidine (Lu *et al.*, 2002). This has been shown to decrease virulence and decrease the utilisation of T3SS under lab conditions, which is typical of a switch from acute to chronic infection (Wang *et al.*, 2016).

Figure 3.14 Two major blocks of gene presence-absence which are related to specific lineages in the K9 patient are highlighted. Blue = gene presence, white = gene absence.

No further patterns of gene presence-absence occurred with any particular lineage in the remaining patients. In each case, the apparent accessory genome was randomly dispersed throughout the phylogeny, suggesting that the presence of the observed accessory genome is due to noise from errors in genome assembly or from random loss of single genes.

3.3.7 Context of CF isolates within the *P. aeruginosa* population

PA14-like isolates are typically characterised by the presence of the ExoU virulence protein, and PAO1-like isolates are characterised by the presence of the ExoS virulence protein. ExoU-positive isolates are often selected against in the airways of CF patients because the ExoU virulence protein is more toxic than ExoS, and has been linked to more acute organ damage, leading to an increase in morbidity (Feltman *et al.*, 2001). The ExoS and ExoU status of all of the isolates in this chapter were determined, to confirm where the isolates from this study fit within the global *P. aeruginosa* population structure. Both ExoS and ExoU proteins were absent in one isolate from patient K6, however, all other isolates were ExoS positive and ExoU negative suggesting that all of the isolates are PAO1-like, and that no PA14-like isolates were found within these patients.

The bacteraemia *P. aeruginosa* isolates from Chapter 2 were found to be a good representation of the population structure of *P. aeruginosa* (Chapter 2, section 2.3.3). Therefore, a maximum-likelihood tree was calculated using one representative isolate from each patient and the 352 isolates of the bacteraemia collection in Chapter 2 (Figure 3.15). This confirmed that the isolates from the patients in this chapter sit within the PAO1-like clade. These PAO1-like isolates are well dispersed within the PAO1-like clade, suggesting that they do not form part of a local epidemic clade.

The isolates from the two patients harbouring the MES strain, patients K1 and K14, appear to be interspersed within the five MES isolates that were identified within the bacteraemia dataset (Chapter 2, section 2.3.2). In order to investigate this further, a local tree was calculated from the five bacteraemia isolates and ten diverse isolates from each of the patients infected with MES (Figure 3.16). The resulting phylogeny indicates that the separate patient infections are distinct from each other. The K14 MES isolates are related to the bacteraemia isolates by a minimum of 2,271 SNPs, and to the K1 isolates by a minimum of 2,680 SNPs. This suggests that the two MES clones infecting patients K1 and K14 have been

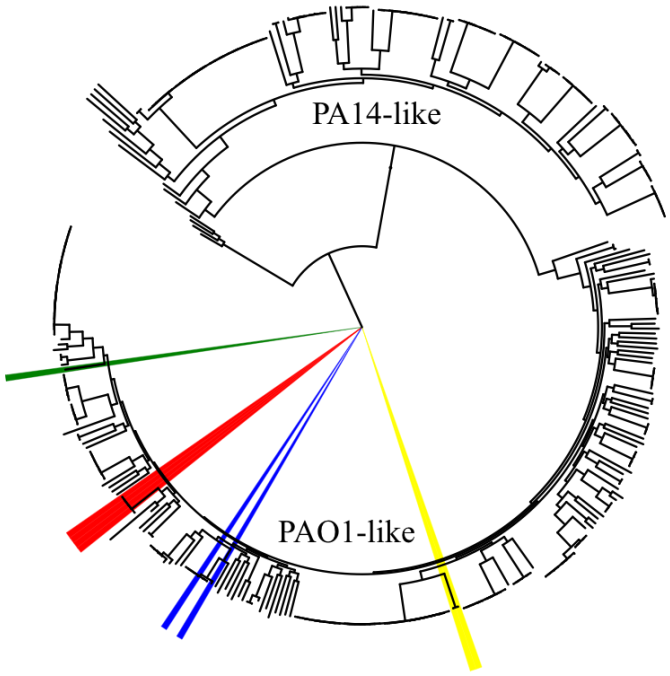


Figure 3.15 A maximum-likelihood tree showing how each representative isolate from each patient with CF patient sits within the PAO1 population structure of *P. aeruginosa*. Blue = ST217 MES isolates (patients K1 and K14), red = ST146 LES isolates (patients K7, K9, K11 and K15), yellow = ST3307 isolates from patients K3 and K6, green = ST3308 isolates from patient K4.

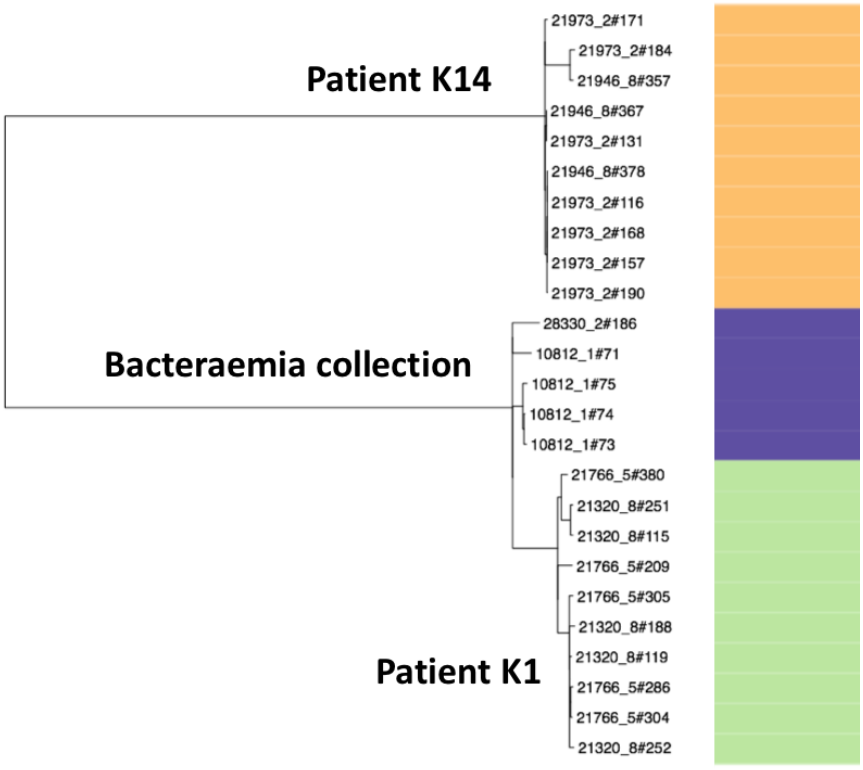


Figure 3.16 A MES phylogeny was calculated from the five MES bacteraemia isolates, and ten MES isolates from each of patients K1 and K14. The colour indicates the source: Orange = CF patient K14, Purple = bacteraemia study, Green = CF patient K1.

evolving separately over a long time. The K1 isolates were related to the BSAC isolates by a minimum of 408 SNPs, suggesting that there has been no recent transmission between patients with bacteraemia in the Chapter 2 dataset and patient K1, and that all infections were distinct and separate.

Comparing accessory genomes of the three different groups of isolates indicate the presence of 201 genes in the K14 isolates that are not present in the BSAC or K1 isolates. These genes are related to mobile DNA sequences; phage, Type II Secretion Systems, integrons, and transposons, suggesting that the majority of the variation separating patient K14 from the BSAC and patient K1 isolates has been externally acquired. The patient K1 and BSAC isolates were found to harbour a bacteriophage which is not present in the K14 isolates. There was no obvious variation in the accessory genomes of the BSAC and K1 isolates, suggesting that the variation between these isolates has been solely caused by the accumulation of SNPs over time.

3.3.8 Patient-to-patient transmission

There are three patient groups that share similar closely-related isolates, as identified in section 3.3.2. The first group consists of patients K1 and K14, who are both infected with the MES strain, ST217. In order to investigate whether one of the patients could have transmitted the strain to the other, all K1 and K14 isolates were mapped to a local reference, and pairwise SNP distances between every isolate from K1 and every isolate from K14 were calculated. The mean pairwise SNP distance was 2,737 SNPs, with a minimum distance of 2,680 SNPs. This suggests that although both patients were infected with the same ST, the two patients harbour individual, independent infections. The outgroup-rooted maximum-likelihood phylogeny of the patient K1 and K14 isolates confirm that these patients are infected with separate and distinct populations of isolates from separate infection sources (Figure 3.17).

The second group of closely-related isolates were collected from patients K3 and K6, both of whom are infected with the same novel ST, ST3307. Calculated pairwise SNP distances indicated a mean of 65 SNPs, with a minimum of 24 SNPs. This suggests that the isolates infecting both patients may share a relatively recent common ancestor, which could represent a local circulating strain. The outgroup-rooted maximum-likelihood phylogeny of the K3 and K6 isolates mapped to a local reference, indicate that the K6 isolates form a nested clade within the isolates from patient K3 (Figure 3.18). This is a strong indicator of cross-infection between the two patients (Prosperi *et al.*, 2011).

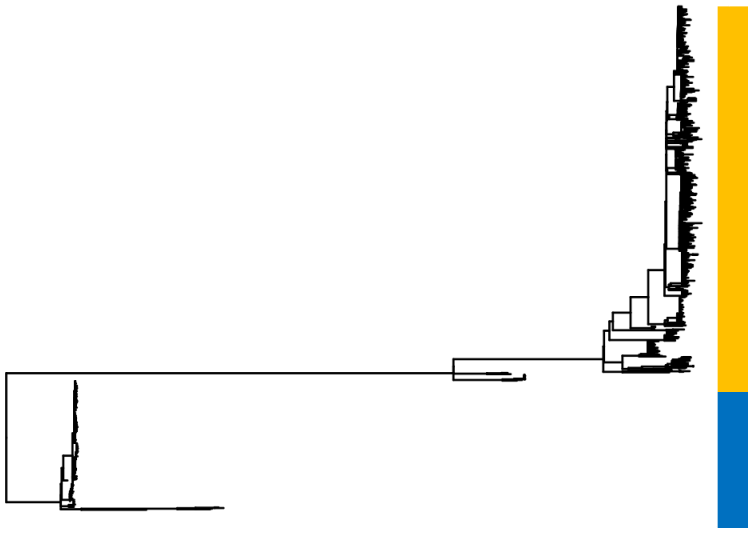


Figure 3.17 An outgroup-rooted maximum-likelihood phylogeny of the K1 (yellow) and K14 (blue) isolates, which indicates two distinct and diverged populations.

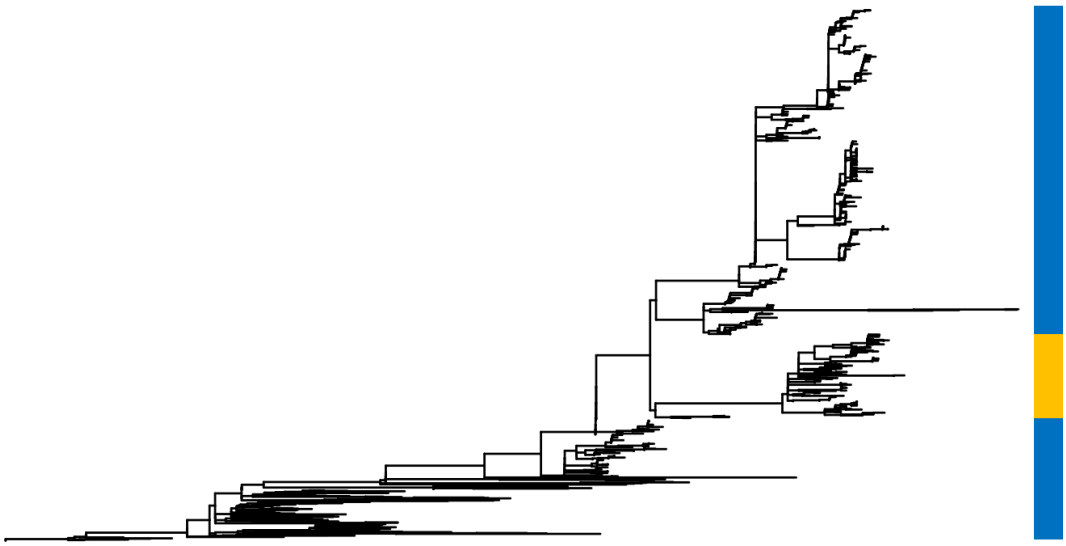


Figure 3.18 Outgroup-rooted phylogeny of patient K3 (blue) and patient K6 (orange) isolates, which indicates that the K6 isolates are nested within the isolates from patient K3. This could be an indication of transmission.

The final group of closely-related isolates are the LES (ST146) isolates, which were collected from patient K7, K9, K11 and K15 sputum samples. The isolates were mapped to the gold-standard reference strain *P. aeruginosa* LESB58, a LES isolate first identified in 1988, and sequenced in 2005 (Salunkhe *et al*., 2005). The pairwise SNP distances between the isolates from these four patients were calculated and are reported in Table 3.9. Patient K7 and K15 share the fewest pairwise SNPs, with a minimum of 4 SNPs, a maximum of 19 SNPs, and a mean of 9 SNPs separating the two groups of isolates. This small number of SNPs may indicate a recent shared common ancestor and potential transmission or common infection source.

Patient ID (age)	Patient ID (age)	Number Pairwise SNPs		
		Minimum	Mean	Maximum
K7 (52)	K9 (36)	48	63	81
K7 (52)	K11 (34)	34	44	56
K7 (52)	K15 (19)	4	9	19
K9 (36)	K11 (34)	37	53	69
K9 (36)	K15 (19)	49	63	78
K11 (34)	K15 (19)	35	44	53

Table 3.9 Pairwise SNP distances between each pair of patients infected with LES.

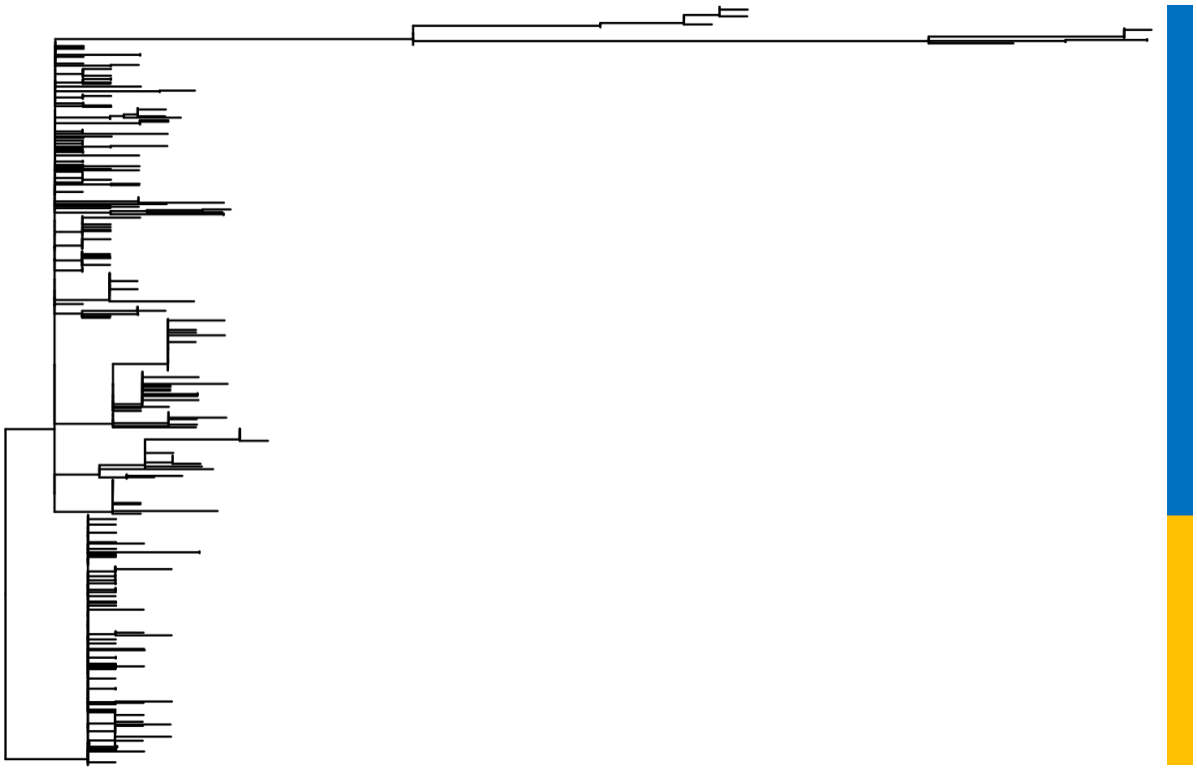


Figure 3.19 Maximum-likelihood tree of patients K7 (blue) and K15 (orange), which shows that each patient's isolates are monophyletic.

An outgroup-rooted maximum-likelihood tree of the K7 and K15 isolates was calculated, to investigate any potential crossover and transmission of isolates between the patients (Figure 3.19). The tree structure is monophyletic for both K7 and K15, with no crossover of isolates. The clinical data (not shown) indicate that both of these patients developed chronic *P. aeruginosa* infection whilst undergoing treatment at Papworth Hospital. Without further epidemiological evidence, it cannot be determined how the patients both obtained this strain; whether one patient transmitted to another directly, they both picked up an infection from a different patient at Papworth Hospital, or whether they picked up the same circulating clone from the environment.

The intrapatient diversity within both the 373 K7 isolates and the 189 K15 isolates, is also small (refer back to Table 3.5). Within the K7 population, the mean pairwise distance is 5 SNPs, with a maximum pairwise distance of 19 SNPs. Twelve of the isolates are identical, and differ by 0 SNPs overall. Within the K15 population, 28% of the isolates are identical, and differ by 0 SNPs. The mean pairwise distance of the K15 isolates is 1.3 SNPs, with a maximum pairwise distance of 8 SNPs. The high proportion of

identical isolates in the patient K15 group suggests that this strain has not had time to diversify within the lungs of the host, which is evidence of a recent infection. This supports the clinical observation that patient K15 had developed this chronic infection recently.

The remaining isolates from the LES patients show no indication of recent transmission, with the minimum pairwise SNP distance ranging from 35 SNPs between the isolates from patients K11 and K15, to 49 SNPs between the isolates from patients K9 and K15.

3.3.8.1 Estimating date of infection

Each patient phylogeny was tested for temporal signal to investigate whether it is possible to date back to the time of infection of the individual patients. Temporal signal was only present for the K1, K3 and K9 isolates. This indicates that the genetic diversity seen within the other patient isolates does not correlate with the phylogeny structure and the date of sampling. Temporal signal could be improved with a sampling period greater than 6 months, to capture more temporal diversity. To investigate any potential transmission, temporal signal was also tested for within the K1 and K14 MES group of isolates, the K7 and K15 LES isolates, and for the K3 and K6 isolates. Temporal signal was present in all of these groups of isolates. Each outgroup-rooted phylogeny was dated three independent times and reported in Figure 3.20.

The MRCA of both the K1 and K14 MES isolates was predicted to be 8.4 to 10.0 years ago. Since the minimum SNP distance between the two groups of isolates is 2,680 SNPs, this time to MRCA would require a mutation rate that is 53x greater than the average *P. aeruginosa* mutation rate of 2.5 SNPs/genome/year (Marvig *et al*., 2013). Mutation rates of 40-60x higher than the average have been previously identified in patients infected with hypermutator strains (Feliziani *et al.*, 2014).

Pairwise SNP distances showed that patients K7 and K15 have both acquired a transmissible strain recently. Dating of the K7 and K15 patient groups estimated a time to MRCA of 0.54 to 1.12 years prior to the study collection period. This recent acquisition is confirmed by the clinical data (not shown), which

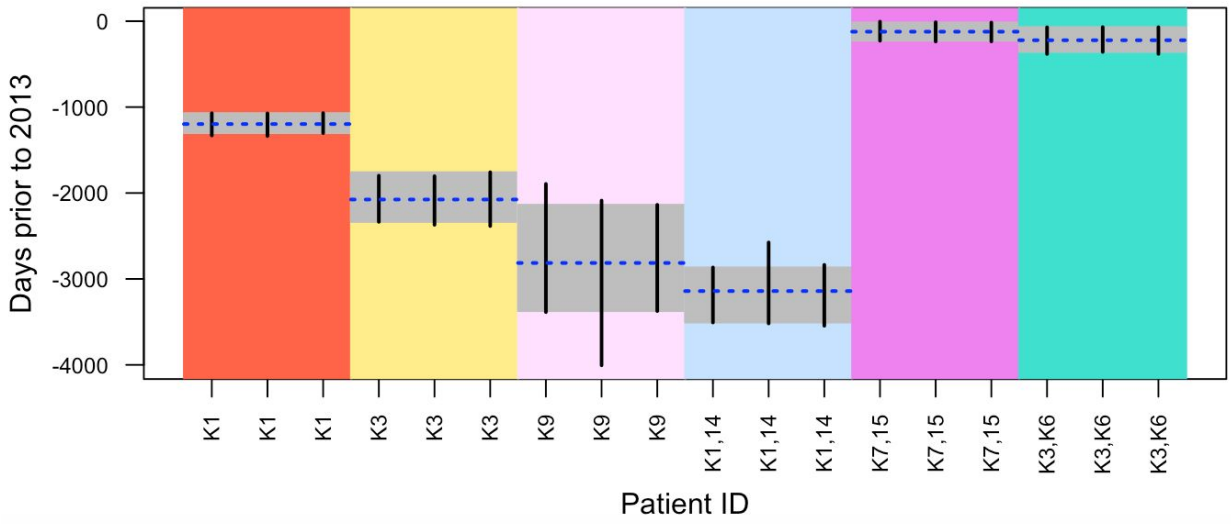


Figure 3.20 The time to MRCA was estimated for the isolates from each of the patient groups which had temporal signal. The consensus confidence intervals shared by all tests, run in triplicate, are highlighted with a grey box, and the mean time to infection is indicated by a blue dotted line.

indicates the patients became chronic at Papworth hospital. However, further epidemiological data would be required to identify whether one patient transmitted directly to the other, or whether both patients picked up a circulating clone separately.

The K6 isolates are nested within the isolates from patient K3, which could also indicate transmission between the two patients (refer back to Figure 3.18). There was a minimum of 24 SNPs between isolates from the two patients, which would assume a time to MRCA of 5-10 years at constant and average mutation rates. The Bayesian dating of the K6 phylogeny, outgroup-rooted by the nearest clade of K3 isolates (Figure 3.21), suggested a time to MRCA of 0.71 - 1.74 years prior to the start of the study. Since the transmission event could have occurred any time between the MRCA of the K3 and K6 isolates, and the MRCA of the K6 isolates only, the lower confidence interval was adjusted to encompass the dating of the MRCA of the K6 isolates only. Therefore, the suggested time to transmission is 0.33 - 1.74 years prior to the start of the study.

Patient K9 contains the greatest number of isolates out of the four patients infected with the LES strain, but the lowest diversity. The Bayesian time to MRCA is predicted to be 6.4 to 9.7 years, suggesting that these isolates capture the majority of the diversity within the patient K9 lung, and there has been no increase in mutation rate compared to the standard average rate.

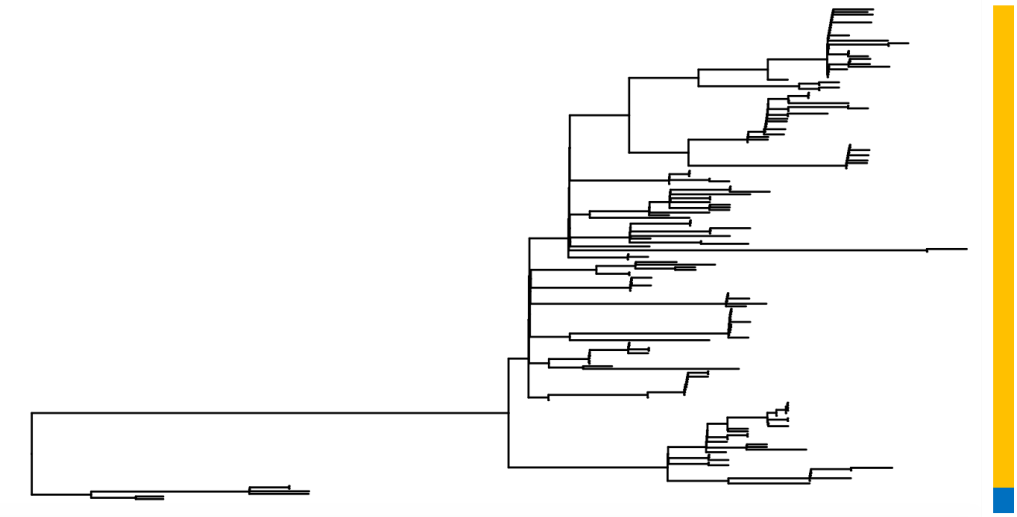


Figure 3.21 The K6 phylogeny (orange) outgroup-rooted with the nearest K3 isolates (blue). Temporal signal was present ($R^2 = 0.26$, $p = 0.0001$).

Bayesian dating of the patient K3 isolates predicted a time to MRCA of 4.8 to 6.4 years prior to the start of the study, which suggests that there is an elevated mutation rate for these isolates, of between 2x and 8x.

An elevated mutation rate is also a plausible explanation for the diversity observed within the isolates from patient K1. Using Bayesian dating, the time to MRCA of the K1 isolates was 3.4 to 4.0 years before the study period. This requires a mutation rate of roughly 4x the standard for *P. aeruginosa*. The MRCA for the K1 and K14 isolates combined was estimated to be 8.4 - 10.0 years ago, which is prior to the time patient K1 is estimated to have obtained the infection. This therefore suggests that patient K1 did not directly transmit the infection to patient K14 and vice versa.

The high mutation rates identified in the isolates from patients K1, K3, and from the patient group K1 and K14, is a further indicator of the hypermutation phenotype having occurred during the evolution of the isolates within the patient lung (see section 3.3.4). Therefore, it's entirely possible that the entire diversity of infection within the lung has been captured after hypermutation has been taken into account. It is unlikely that the isolates from patient K9 contain hypermutators, since the time to MRCA is consistent with a normal mutation rate. For the K7 and K15 isolates, additional clinical evidence suggests that these patients recently became chronic whilst at Papworth hospital, and therefore the dating of a MRCA of 0.54 to 1.12 years ago is likely to be accurate. This timeframe is similar to the time to MRCA of the K6

isolates, which were potentially transmitted from K3 between 0.33 and 1.74 years prior to the sampling period.

The age of each patient at the time to MRCA would be 26-27 for patient K1, 26-27 for patient K3, 26-30 for patient K9, and 23 for patient K6. If both of the K7 and K15 patients developed chronic infection roughly one year before the study period, then they would have been 51 and 18, respectively. A retrospective cohort analysis of 27,000 CF patients in the USA between 2002 and 2012, indicated a median age of chronic infection development of 20.0 years of age, with an interquartile range of 13.0-29.0 (Crull *et al.*, 2018). The majority of infections within this study can be dated to a similar age.

However, it is possible that this study has not captured the entire diversity from the original infection. If there is strong selective pressure in the lungs, which drives replacement of lineages within the lungs, the diversity sampled will have arisen long after the point of original infection, and therefore the time to MRCA observed in this dataset may be later than the original point of infection. The CF lung is a complex environment, with many selection pressures. These include lung inflammation with the corresponding immune response, administration of antibiotics to combat infection, and competition from other microbes found within the lungs (Caballero *et al.*, 2015). However, several studies have identified that the lineages within the CF lung are resilient and resistant to these pressures, and are stable over time (Fodor *et al*., 2012; Tunney *et al*., 2011; Mowat *et al.*, 2011), and therefore the time to MRCAs estimated in this study may capture the original point of infection.

3.3.9 Phylogenetic clustering with exacerbation state

Clustering of exacerbation state timepoints was observed when displayed next to the maximum-likelihood tree from section 2.2 (Figure 3.22). It is therefore possible that, in some cases, some lineages may be being replaced between exacerbation and exacerbation timepoints. Therefore, statistical analysis was carried out to investigate the extent of lineage replacement.

The maximum-likelihood tree of the isolates from patient K1 shows a non-random clustering of exacerbation ID and exacerbation time point (Figure 3.23). Similar patterns are seen within the isolates from all other patients, however missing data from some exacerbation timepoints for some patients (see section 3.3.1) results in a more skewed clustering pattern.

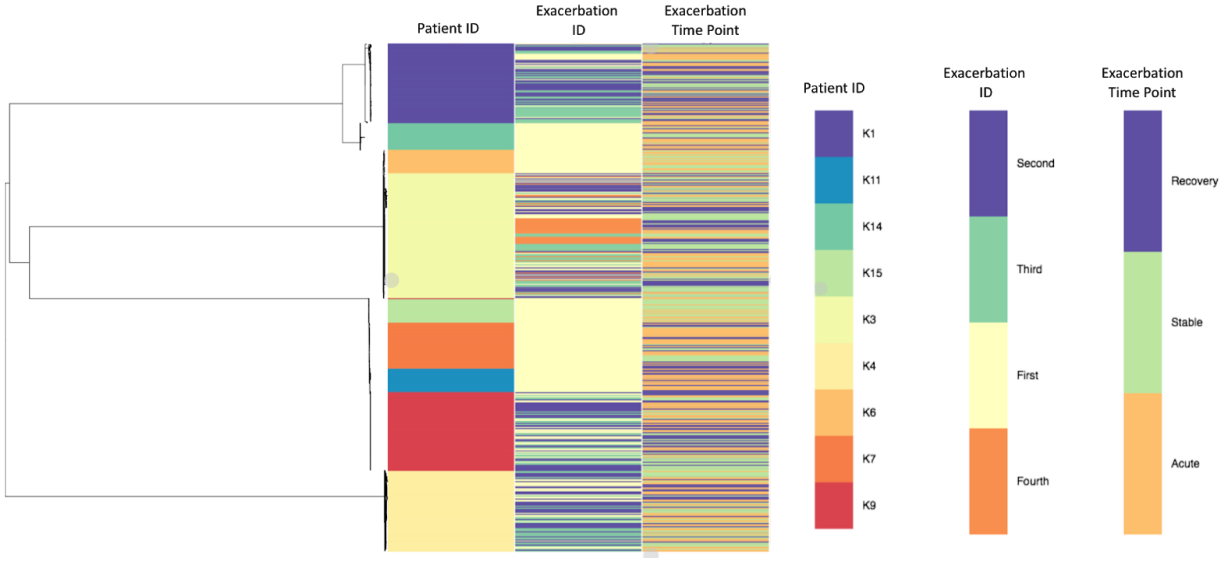


Figure 3.22 A maximum-likelihood tree with exacerbation state timepoints displayed alongside.

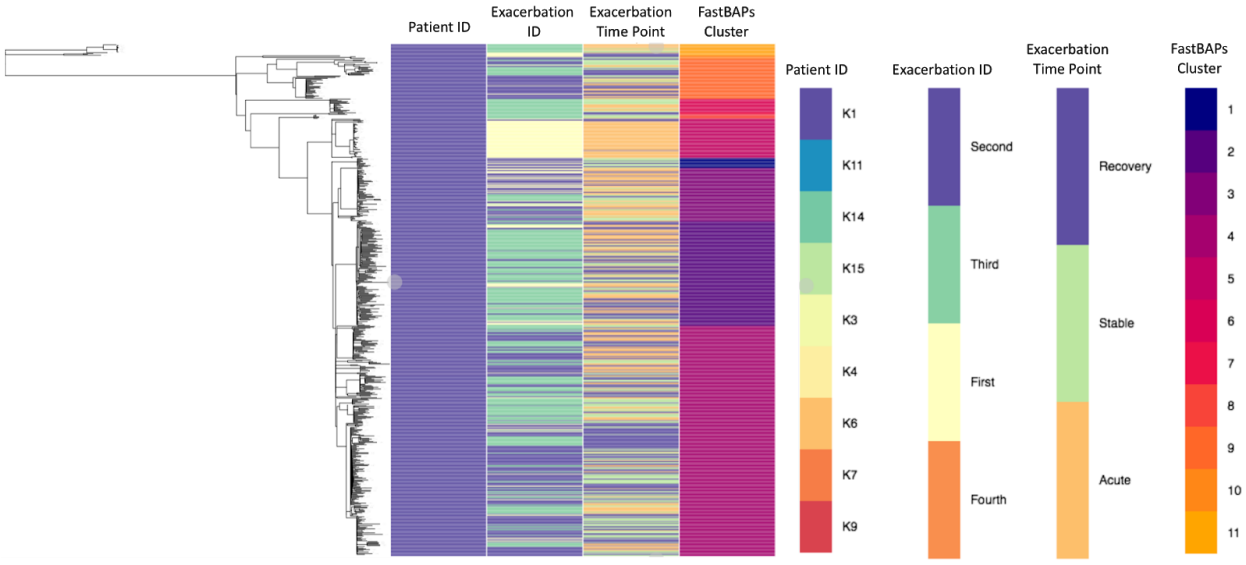


Figure 3.23 A maximum-likelihood tree of isolates from patient K1 with exacerbation state timepoints displayed alongside.

Patient IDs	Exacerbation ID								Exacerbation Timepoint					
	First		Second		Third		Fourth		Stable		Acute		Recovery	
	λ	D	λ	D	λ	D	λ	D	λ	D	λ	D	λ	D
K1	0.779	0.272	0.952	0.162	1.000	-0.199			0.822	0.612	0.909	0.316	0.895	0.627
K3	0.957	0.551	0.944	0.617	0.963	0.464	0.952	0.340	0.937	0.619	0.937	0.423	0.961	0.570
K4	0.967	-0.266	0.936	0.269	0.913	0.284			0.734	0.483	0.487	0.618	0.250	0.644
K6									0.917	0.384	0.917	0.384		
K7									0.787	0.404	0.738	0.427	0.708	0.465
K9	0.828	0.484	0.853	0.240	0.881	0.394			0.850	0.342	0.811	0.464	0.800	0.650
K11											0.790	0.385	0.790	0.381
K14									0.232	0.796	0.405	0.713	0.337	0.804
K15									0.424	0.816	0.424	0.814		

Table 3.10 λ values and D-statistic values for each exacerbation time point. Grey boxes indicate that the value cannot be calculated.

An estimate of the phylogenetic signal was calculated to determine whether the exacerbation state clustering patterns are distributed randomly. Two methods were used to calculate the phylogenetic signal; Pagel's λ (Pagel, 1999), and the Purvis and Fritz D-statistic (Purvis and Fritz, 2010) (Table 3.10).

The Pagel's λ value measures the extent to which correlations in traits reflect their shared evolutionary history. A Pagel's λ value of 0 indicates that the trait is randomly distributed throughout the tree, whereas values close to 1 indicate that the trait has evolved with phylogenetic signal. The majority of the exacerbation state timepoints had Pagel's λ values close to 1, which suggests phylogenetic signal.

The Purvis and Fritz D-statistic builds upon Pagel's λ, and measures the evolution of phylogenetic signal specifically for binary traits. A D-value of 1 indicates that the trait has evolved randomly, whereas a D-value of 0 indicates that the trait has evolved randomly, but with phylogenetic Brownian Motion. A D-value of > 1 indicates that the trait is overdispersed throughout the phylogeny, whereas a D-value of < 0 indicates that the trait is clustered in the phylogeny. Although the D-values are slightly higher than would be expected for the corresponding Pagel's λ value, the phylogenetic signals calculated from both Pagel's λ and the Fritz and Purvis D-statistic are correlated (correlation: -0.637, R^2: 0.389) (Figure 3.24).

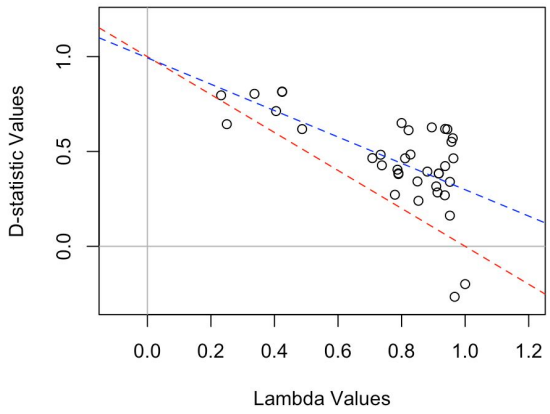


Figure 3.24 The λ values plotted against the D-statistic values for each exacerbation time point. The red line is the theoretical correlation line between the λ value and D-statistic, and the blue line is the observed regression line.

Strong signals of phylogenetic clustering ($D < 0$) for the third exacerbation are present in the isolates from patient K1, and for the first exacerbation in patient K4. This suggests that the lineage architecture of the *P. aeruginosa* population may be affected in the time between exacerbations. To investigate this further, the K1 phylogeny was clustered using FastBaps (Figure 3.25a), and the frequency of isolates in each cluster were plotted and highlighted by APE timepoint (Figure 3.25b). In total, 11 FastBAPs clusters were identified, and vary in size from nine isolates in cluster 7, to 310 isolates in cluster 4. FastBAPs cluster 2 is highly associated with the third exacerbation that patient K1 experienced. Over 70% of the isolates in cluster 2 are associated with that exacerbation. Cluster 4 is the largest cluster, and is associated roughly equally with the first and second exacerbation. Cluster 5 is associated with the first exacerbation only, however, the Purvis and Fritz D-statistic showed that this was not significant, as the isolates are spread across the first five clusters.

This analysis was repeated with the patient K4 isolates, as the Purvis and Fritz-D values indicated clustering within the phylogeny (Figure 3.26). The isolates are separated into fewer clusters than patient K1, but the range of isolates is similar, with nine isolates associated with cluster 2, and 280 isolates associated with cluster 4. The first exacerbation that the patient experienced was associated most with BAPs cluster 5, where nearly 60% of the isolates are associated with the first exacerbation. During the progression from the first exacerbation to the second exacerbation, the BAPS cluster 5 reduces in prevalence, to be replaced by BAPs cluster 4 and 6 as the main clusters for the second exacerbation state. However, the D-value for the second exacerbation indicates a stronger degree of random phylogenetic

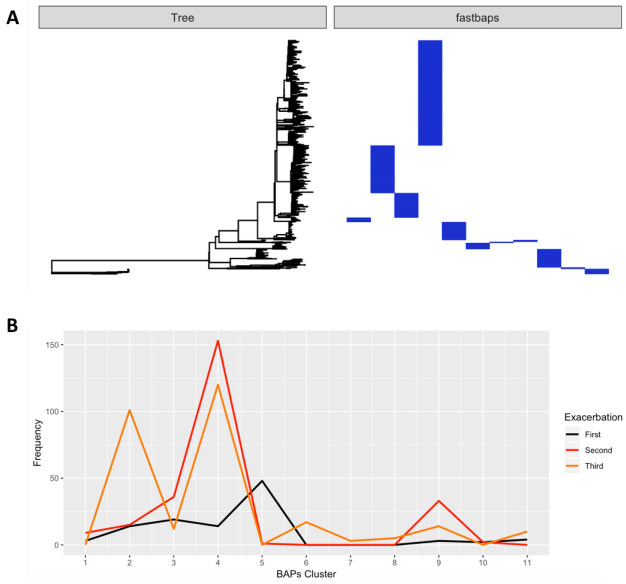


Figure 3.25 a) The K1 phylogeny clustered by FastBAPs, with cluster 1 to cluster 11 from left to right. b) The frequency of isolates plotted against each FastBAPs cluster, and coloured by the exacerbation number.

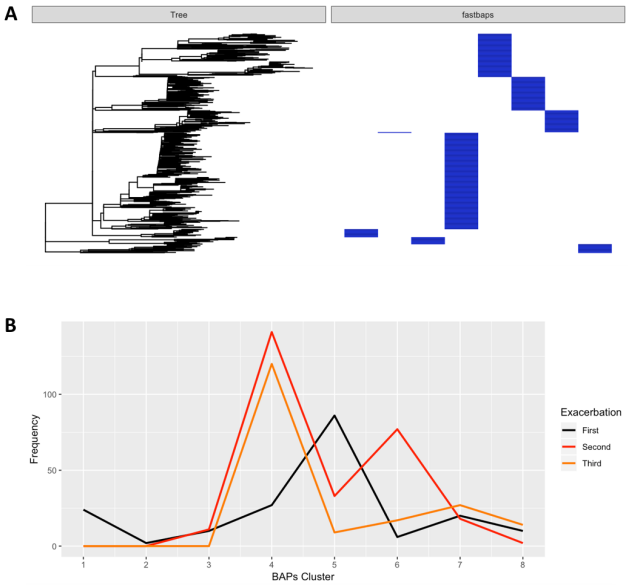


Figure 3.26 a) The K4 phylogeny clustered by FastBAPs, with cluster 1 to cluster 8 from left to right. b) The frequency of isolates plotted against each FastBAPs cluster, and coloured by the exacerbation number.

clustering. The isolates that form the third exacerbation generally follow the lineage architecture of the second exacerbation, with a drop in prevalence of BAPs cluster 6.

These two examples suggest that lineage architecture can be affected between different APEs. No clustering was identified for the patient K1 and K4 individual timepoints within the clustered exacerbation, suggesting that the population of isolates from stable, acute, and recovery time points are unaffected. The majority of timepoints for isolates from patients K3, K6, K7 and K9 suggest a weaker degree of clustering, which is evidence of random distribution throughout the phylogeny. All timepoints from patients K14 and K15 had high signals of random distributions throughout the tree, which suggests there is no evidence of lineages appearing or disappearing. This means that the process of undergoing an APE, and subsequent treatment, do not generally affect the architecture of *P. aeruginosa* lineages within the CF lung. There was also no phylogenetic clustering of the acute, recovery, or stable timepoints across all of the APEs.

3.3.10 AMR prediction

CF lung infections are often resistant to antibiotics. Therefore, in order to inform antimicrobial therapies that might be most effective against CF lung infection, the variation and diversity of AMR genotypes within the 4,094 *P. aeruginosa* isolates was investigated. Although there is no phenotypic AMR data for the isolates in this study, predictions of AMR can be made from the DNA sequence by investigating known resistance genes and gene variants.

3.3.10.1 Mex efflux systems

A strong lineage distinction was observed for the presence of several AMR genes (Figure 3.27). The majority of predicted genes and gene variants that may confer AMR in this dataset belong to the RND-family of efflux pumps, called multidrug efflux pumps (Mex pumps). The Mex systems account for 28 of the 50 (56.0%) predicted AMR genes and gene variants present in the 4,094 isolates. These efflux pumps can make treating *P. aeruginosa* infections more difficult, especially when they are overexpressed, as they confer resistance to several classes of antibiotics. Overexpression of efflux systems is relatively common in *P. aeruginosa*, as very little fitness costs are associated with overexpression (Pacheco *et al*., 2017). In total, nine Mex efflux systems were predicted within the 4,094 isolates (Table 3.11).

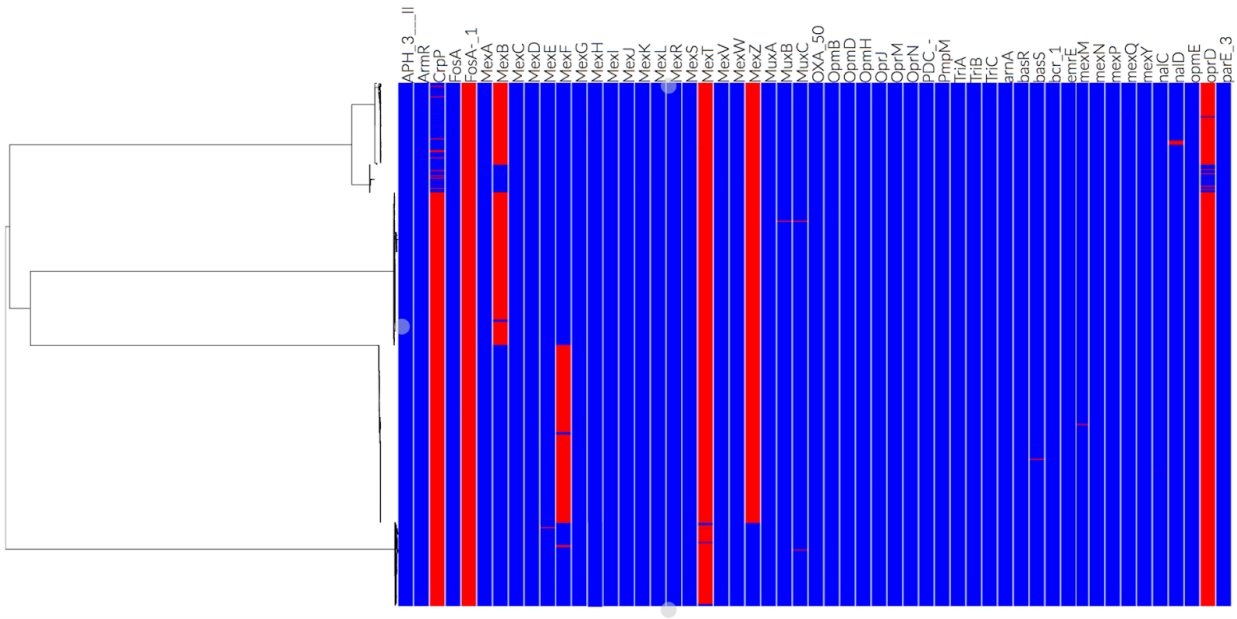


Figure 3.27 Predicted genes and gene variants associated with AMR displayed next to a maximum-likelihood SNP phylogeny of every isolate in this study. Blue = present; red = absent.

Efflux System	Presence in Patient Group (%)									Antibiotic Resistance Conferred
	K1	K3	K4	K6	K7	K9	K11	K14	K15	
MexAB-OprM	0	2	100	0	100	100	100	100	100	Quinolones, macrolides, tetracyclines, chloramphenicol, β-lactams (except carbapenems)
MexAB-OprM (ΔMexB)	100	98	0	100	0	0	0	0	0	--
MexCD-OprJ	100	100	100	100	100	100	100	100	100	Quinolones, macrolides, tetracyclines, chloramphenicol, penicillins, cephems (except ceftazidime), meropenem
MexEF-OprN	100	100	97	100	0	0	9	100	0	Fluoroquinolones, trimethoprim, and chloramphenicol
MexEF-OprN (ΔMexE/ΔMexF)	0	0	3	0	100	100	91	0	100	--
MexGHI-OpmD	100	100	100	100	100	100	100	100	100	Heavy metals
MexGHI-OpmD (ΔMexH)	0	0	0	0	0	0	0	0	0	--
MexJK-OprN	100	100	100	100	100	100	100	100	100	Macrolides, triclosan
MexMN-OprM	100	100	100	100	100	100	100	100	100	Chloramphenicol, heavy metals, β-lactams
MexPQ-OpmE	100	100	100	100	100	100	100	100	100	Macrolides, fluoroquinolones,
MexVW-OprM	100	100	100	100	100	100	100	100	100	Fluoroquinolones, tetracycline, chloramphenicol, macrolides
MexXY-OprM	0	0	0	0	0	0	0	0	0	Quinolones, macrolides, tetracyclines, chloramphenicol, aminoglycosides, penicillins, cephems (except ceftazidime), meropenem
MexXY-OprM (ΔMexX)	100	100	100	100	100	100	100	100	100	--

Table 3.11 Mex-class multidrug efflux pumps are predicted within the genome sequences of the 4,094 CF isolates, which confer resistance to a number of antibiotic classes.

The only efflux pump that is constitutively expressed within *P. aeruginosa* is MexAB-OprM (Verchere *et al*., 2015). This is a non-specific efflux pump that is capable of expelling most classes of antibiotics from the cell. MexB is one of the components of the MexAB-OprM system, and is responsible for substrate recognition. Absence of *mexB* should result in antibiotic susceptibility, as the membrane complex cannot properly form (Lomovskaya *et al.*, 1999). However, *mexB* is absent from nearly all isolates in patients K1, K3 and K6, suggesting that this efflux pump may be ineffective in these patients. All regulators of the MexAB-OprM efflux, *mexR*, *armR*, *nalC, nalD* and *cpxR* are present in all patient groups, suggesting that the efflux pump is not overexpressed in the patients where *mexB* is present, suggesting increased resistance due to overexpression of this efflux pump is not prevalent in this dataset (Rojo-Molinero *et al.*, 2019).

The MexY and OprM proteins, which form part of the MexXY-OprM efflux system responsible for resistance to the aminoglycoside antibiotics (Morita *et al*., 2012), were encoded within all 4,094 genomes. However the MexX membrane-binding protein was completely absent in all cases, suggesting that this efflux pump may also be ineffective. The MexZ regulator protein of the MexXY-OprM efflux system is absent in all patients except patient K4, suggesting that it would be overexpressed in the other patients if the full MexXY-OprM efflux system was present.

In the LES isolates (patients K7, K9, K11 and K15), the MexF substrate recognition protein of the MexEF-OprN complex is absent in most isolates. However, all of the isolates that do not contain *mexF* do encode *emrE*, which can replace MexEF-OprN function and restore aminoglycoside resistance (Li, 2003). The repressor of the MexEF-OprN efflux pump, *mexS*, is present in all of these isolates, and the MexEF-OprN activator, *mexT*, is absent in 99% of isolates (Sobel *et al*., 2005). These gene knockouts have been shown to induce susceptibility to substrate antibiotics (Fargier *et al.*, 2012).

The efflux systems MexCD-OprJ, MexGHI-OpmD, MexJK-OprM, MexMN-OprM, MexPQ-OpmE, and MexVW-OprM were predicted to be present in all 4,094 isolates, potentially providing resistance to a number of classes of antibiotics. However, AMR phenotyping and transcriptomic studies would be required to determine resistance and link that to efflux activity.

3.3.10.2 Non-Mex efflux systems

Efflux System	Resistance conferred
MuxABC-OpmB	Monobactams, macrolides and tetracycline
TriABC-OpmH	Triclosan
PmpM	Fluoroquinolones
OprD	Carbapenems

Table 3.12 Non-Mex multidrug efflux pumps are predicted within the genome sequences of the 4,094 CF isolates, along with the antibiotic classes these efflux pumps confer resistance to.

P. aeruginosa isolates can also encode other non-Mex efflux pumps. In total, three non-Mex efflux pumps are predicted within the 4,094 isolates within this dataset (Table 3.12). Both the MuxABC-OpmB efflux system and the TriABC-OpmH efflux system are present in the majority of isolates. The MuxABC-OpmB system is an efflux pump that is involved in general pathogenesis of *P. aeruginosa*, however, studies have indicated that this efflux system also confers resistance to monobactams, macrolides and tetracycline classes of antibiotics (Yang *et al.*, 2011a). The TriABC-OpmH efflux system confers resistance to triclosan. Additionally, the single-protein exporter protein, PmpM, which confers resistance to the fluoroquinolones is also encoded in all patients.

The absence of the *oprD* porin confers a basal level of resistance to the carbapenem antibiotics (Li *et al.*, 2012). This gene is absent in all patients, except for 2% of isolates in patient K1, and 95% of isolates in patient K14. However, clinical-levels of carbapenem resistance require a further mechanism of resistance on top of *oprD* mutations, which aren't seen on the genetic level in this dataset (Shariati *et al*., 2018).

3.3.10.3 Additional mechanisms of resistance

Nearly 92% of the isolates from patient K1 and 88% from patient K14 carry the plasmid-encoded *crpP* gene, which is a ciprofloxacin modifying enzyme (Chavez-Jacobo *et al.*, 2018). Within the patient K1 and

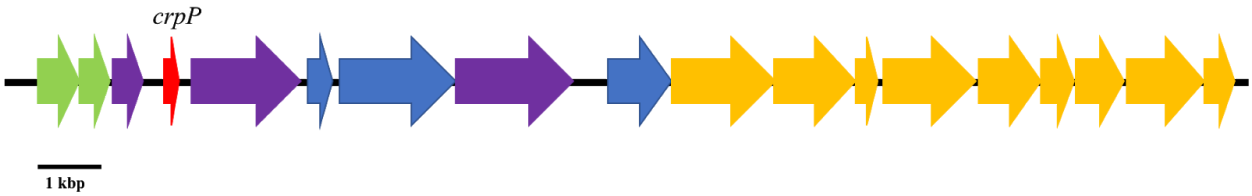


Figure 3.28 The mobile genetic element in which *crpP* is encoded. Green = integrating conjugative element proteins, Purple = DNA binding/helicase proteins, Blue = Hypothetical Proteins, Orange = Type IV secretion machinery, Red = *crpP* resistance gene.

K14 dataset, there is only a 28% prevalence of plasmids assembled from the raw sequencing data, all of which carry the *crpP* gene. The *crpP* gene has been known to integrate into the chromosome, which may explain the difference between the plasmid prevalence and the prevalence of this gene in these isolates (van Belkum *et al.*, 2015; Galetti *et al.*, 2019). In the isolates where the gene is not plasmid encoded, the *crpP* gene is located in a 22 kbp integrative conjugative element (Figure 3.28). These integrative conjugative elements are typically transferred to neighbouring bacteria by conjugation, and encode helicase, and Type IV secretion machinery to facilitate transfer (Johnson & Grossman, 2014).

Ten additional resistance genes and gene variants were predicted within isolates in this study, and are summarised in Table 3.13. All 4,094 isolates are predicted to contain AMR genes which encode; fosfomycin resistance through FosA; aminoglycoside resistance through the aminoglycoside modifying enzyme aph(3')-IIb; oxacillin resistance through the $_{bla}$OXA-50 oxacillinase; cephalosporin resistance through the *Pseudomonas*-derived cephalosporinase PDC; bicyclomycin resistance through Bcr-1; and polymyxin resistance through ArnA and BasR.

Additionally, 97.5% of all isolates encode the chloramphenicol acetyltransferase catB7, which confers resistance to chloramphenicol. Finally, every isolate from patient K1 include mutations in *gyrA* and *parE* genes, which confer additional resistance to the fluoroquinolones.

Gene	Presence in Patient Group (%)									Antibiotic Resistance Conferred
	K1	K3	K4	K6	K7	K9	K11	K14	K15	
GyrA, ParE	100	0	0	0	0	0	0	0	0	Fluoroquinolones
FosA	100	100	100	100	100	100	100	100	100	Fosfomycin
Aph(3')-IIb	100	100	100	100	100	100	100	100	100	Aminoglycosides
blaOXA-50	100	100	100	100	100	100	100	100	100	Oxacillin
PDC	100	100	100	100	100	100	100	100	100	Cephalosporins
catB7	100	90	100	100	100	100	100	100	100	Chloramphenicol
ArnA, basR	100	100	100	100	100	100	100	100	100	Polymyxins
Bcr-1	100	100	100	100	100	100	100	100	100	Bicyclomycin

Table 3.13 Resistance-conferring genes and gene variants are predicted within the genome sequences of the cystic fibrosis isolates, along with the antibiotic classes these genes and gene variants confer resistance to.

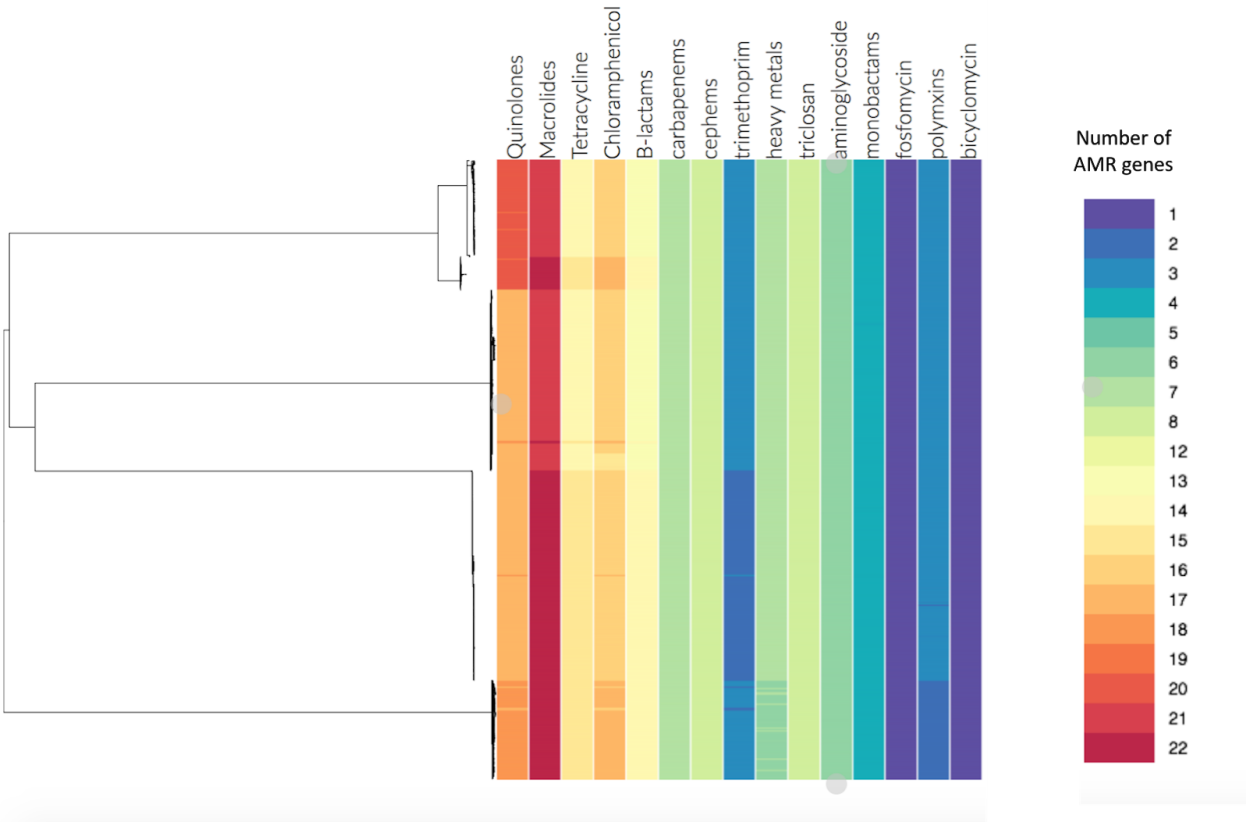


Figure 3.29 The number of AMR genes or gene variants predicted within the genome for each class of antibiotics.

The AMR genes predicted in this dataset confer resistance to a total of 15 different antibiotic classes. Many isolates carry several different genes that may confer resistance to the same antibiotic class (Figure 3.29), for example all isolates within this dataset carry over 20 different genes or gene variants that may confer resistance to macrolide antibiotics.

The variation in resistance-causing gene content within the isolates, especially for patient K4, suggests that some lineages will have a competitive advantage should the antibiotic treatment change for the patient. When there is no variation in predicted resistance, such as for patients K14, K3 and K6, there is no strain that will have a competitive advantage if the antibiotic treatments change.

3.4 Discussion

In this study, 4,094 *P. aeruginosa* isolates from nine patients with CF were whole-genome sequenced to try and develop a greater understanding of within-host diversity and changes in the diversity of isolates across APEs. Initially, the genomes of all 4,094 isolates were clustered and characterised. This showed that each patient was infected with their own, single, distinct strain of *P. aeruginosa*, but that different patients were infected with the same ST. This is in agreement with the current understanding of *P. aeruginosa* lung infection; when a patient picks up a chronic *P. aeruginosa* strain, they remain infected with that strain, and only that strain, for the majority of their life (Fothergill *et al.*, 2012).

MLST typing of the isolates from each patient identified four patients who were infected with the Liverpool Epidemic Strain, which is a highly-virulent and highly-transmittable strain found in clinical CF settings across the UK (Salunkhe *et al*., 2005). Two patients were infected with the Manchester Epidemic Strain, which is another highly-virulent and highly-transmissible UK-wide strain (Ashish *et al*., 2012). No other specially-adapted epidemic strains that have been previously identified from the midlands, Denmark, Australia, USA, or Canada were isolated during this study (Duong *et al*., 2015). Three other patients were infected with previously-undescribed, novel STs. Two of the patients harboured isolates belonging to ST3307, and the other patient harboured isolates belonging to ST3308.

Strains of *P. aeruginosa* across both of the PAO1-like and PA14-like *P. aeruginosa* populations are able to infect patients with CF and adapt to the CF lung (Jeukens *et al.*, 2019). Typically, more PAO1-like than PA14-like infections are found within patients with CF, as the more-toxic ExoU-positive PA14-like isolates are associated with a higher and more-rapid morbidity (Shi & Kan, 2015). The different STs identified in this study are deeply diverged from each other, but are only spread throughout the *P. aeruginosa* PAO1-like population. None of the strains formed part of the PA14-like population, or other *P. aeruginosa* subspecies.

In addition to the between-host population structure, the within-host population structure for each patient was investigated. Evidence of divergent, co-existing lineages of *P. aeruginosa* were identified in patients K1, K4, and K11, which were characterised by the presence of two distinct populations of isolates that diverged at a basal node. This suggests that separate populations of isolates were sampled, most likely

from isolates inhabiting separate niches within the lung. Divergent and co-existing lineages from within a single patient lung have been previously recorded, but sparsely studied (Williams *et al.*, 2015).

Bayesian dating was also used to estimate the date of potential transmission events between patients. Transmission was suspected between patients K7 and K15, as the population of isolates are only diverged by a mean of 9 pairwise SNPs. Bayesian dating suggests that the time to the MRCA between patient K7 and K15 was 0.54 to 1.12 years prior to the study. However, the outgroup-rooted phylogeny was split at a basal node into a distinct clade for each patient. This would suggest that the two infections were picked up at a similar time. However, in order to investigate if this was due to transmission from patient to another, or if both patients picked up a circulating clone, further epidemiological data and historical isolates would be required.

It is likely that transmission of *P. aeruginosa* occurred between patients K3 and K6 however. The combined phylogeny of patient K3 and K6, indicated that the isolates from patient K6 were nested within the isolates collected from patient K3. The two populations of isolates were separated by a minimum of 24 SNPs, which is a strong indicator of recent transmission. Bayesian dating of the MRCA node between the two patients indicated a time to MRCA of 0.71 to 1.74 years prior to the study. This is roughly around the same time that both patient K7 and K15 patients were estimated to have acquired their strains. Transmission is unlikely to have occurred between patient K7 and K15.

Patient K3 provided the most isolates spanning the highest number of APEs. The phylogeny of the patient K3 isolates therefore provides an insight into *P. aeruginosa* evolution over the course of infection. The long branch lengths towards the root of the phylogeny indicate more sparse sampling of those related isolates, whereas the isolates further away from the root were sampled more frequently, and so a more clonal population structure can be observed. Gene presence-absence data displayed alongside this phylogeny indicates large areas of both gene acquisition and gene loss that are associated with the switch from acute to chronic infection. Therefore, the sparsely-sampled isolates at the root of the tree may be the remnants of a population of isolates involved in acute infection, and the frequently-sampled isolates further away from the root of the tree are the more-dominant chronic infection that has subsequently developed.

The acquired T6SS gene block in patient K3 shares over 99% identity with the corresponding region of the *Pseudomonas aeruginosa* LESB58 reference genome. Recombination analysis identified regions of

recombination that may have also originated within the LESB58 genome. Therefore, it is possible that genetic material has been exchanged between the two populations in the lungs of patient K3. No evidence of the LES strain was present in the sputum sample collected from patient K3, suggesting that ST3307 may have outcompeted the LESB ST within the patient K3 lung, to become the dominant clone.

There was very little diversity observed within the isolates from patient K15, with nearly 30% of isolates identical to each other, and a maximum pairwise SNP distance of just eight SNPs. This indicates a newly acquired *P. aeruginosa* infection, which provides a unique insight into the initial adaptation and evolution of *P. aeruginosa* isolates within the CF lung. Many of the non-synonymous mutations were identified in genes related to virulence, the knockdown or knockout of which are indicators of the switch from acute to chronic infection. This offers a rare opportunity to follow the development of a newly-acquired chronic infection, and follow-up isolates could provide an indication of how early *P. aeruginosa* chronic infections evolve within the lungs.

This study also aimed to identify changes in the population of isolates that occurred during three APE timepoints: acute, recovery and stable. The diversity of the isolates was mostly unchanging over the course of the different timepoints, and was also found to be unchanging across different APEs experienced by the same patient. This supports two other studies that investigated *P. aeruginosa* variation over the course of an APE, which concluded that populations remain stable (Fothergill *et al*., 2010; Mowat *et al.*, 2011). In this study, the magnitude of the diversity within each population of isolates was patient-dependent, and suggests that the progression of each infection is unique.

For the patient K1 and K4 isolates, there was evidence of non-random phylogenetic clustering of isolates belonging to a particular APE that each patient underwent. For the patient K4 isolates, BAPS cluster 5 was strongly associated with isolates belonging to the first exacerbation the patient experienced, and for patient K1, BAPS cluster 2 was strongly associated with the third exacerbation the patient experienced. This suggests that the lineage architecture within the lungs of these patients during the time in between exacerbations is affected. However, for all patients, no clustering was observed for isolates belonging to a specific timepoint within each APE, suggesting that the *P. aeruginosa* population structure is not affected by APE-related treatment.

AMR was predicted for all 4,094 isolates in the study based on genetic content. The results showed that the predicted AMR profile of the isolates within each patient did not vary significantly. No genetic cause

for overexpression of efflux pumps was identified, which is a common cause of AMR in patients with CF (Poonsuk *et al.*, 2014). In fact, there was evidence that components of the MexAB-OprM, MexEF-OprN, and MexXY-OprM efflux pumps were missing in some of the isolates in the study, suggesting susceptibility to some antibiotics. Very little variation in the presence/absence of AMR genes was identified, except for *crpP*, which varied in the patient K1 and K14 isolates. This was explained by the presence of the *crpP* gene on a plasmid in some of these isolates, and in the remaining 62% of isolates, the plasmid was integrated within the isolate genome.

This study represents the largest study of comparative genomics in P. aeruginosa clonal isolates from the lungs of patients with CF, investigating 4,094 unique genotypes that were captured from 9 patients with CF. This study has many potential applications, including clinical and evolutionary. By analysing the diversity within individuals with CF, we have shown that the P. aeruginosa population is more varied and more complex than previously identified, and that evolution of the initial P. aeruginosa infection can be rapid and diverse within the enclosed environment of the CF lung. This may help to inform clinical decisions, such as using broad spectrum antibiotics and combination therapies to combat such diverse infections.

This study has provided a unique insight into the evolution of *P. aeruginosa* infection within the CF lung. It has found that the genetic diversity and adaptation of *P. aeruginosa* to the CF lung is relatively stable over the course of APEs, and has probed the population structure and evolution within each individual patient infection.

Chapter 4

Application of genome-wide association studies to identify associations between virulence-related phenotypes and exacerbation-related *Pseudomonas aeruginosa* genotypes from nine patients with cystic fibrosis

Declaration of Contributions

The study was initiated by Andres Floto. Julian Parkhill and Martin Welch supervised this work. All phenotyping assays were carried out by Emem Ukor. I carried out all analysis in this chapter.

4.1 Introduction

P. aeruginosa isolates collected from the lungs of a single patient with cystic fibrosis exhibit significant phenotypic heterogeneity (Lozano *et al.*, 2018; Workentine *et al.*, 2013; Warren *et al.*, 2011). The presence of phenotypic diversity is thought to be caused by environmental diversity, as many different spatial niches exist within the CF lung that are colonised by distinct populations *P. aeruginosa* (Markussen *et al.*, 2014). Intra-species competition between *P. aeruginosa* isolates of the same population has also been shown to lead to distinct phenotypic variation, as each bacteria tries to outcompete the others for resources (Waters & Goldberg, 2019). Variation in both of the phenotypic and antibiotic resistance profiles suggests that the lung environment is complex and no single selection pressure exists within the CF lung (Workentine *et al.*, 2013).

Along with phenotypic heterogeneity, there also exists genotypic heterogeneity, both within the CF lung of a single patient and between different patients (Lozano *et al.*, 2018; Schick & Kassen, 2018). Analysis of different populations of *P. aeruginosa* from different patients with CF revealed common mutations shared between them, which have been termed pathoadaptive mutations (Marvig *et al.*, 2013). The mutations commonly appear in global regulator genes, such as *lasR*, *rpoN*, *mucA*, *mexT*, *retS*, *exsD*, and *ampR*, which control a wide range of virulence-related phenotypes (Marvig *et al.*, 2013). Mutations in the *mucA* regulatory protein result in the overproduction of alginate, which causes mucoid colonies and biofilm formation, and is a common marker that the *P. aeruginosa* infection has switched from acute to chronic (Heltshe *et al.*, 2018). However, the switch from acute to chronic infection is driven through repression of the *rsmA* regulatory protein, and therefore loss-of-function mutations often appear in the *rsmA* gene, and the global regulatory gene, *retS*, which controls *rsmA* expression (Reis *et al.*, 2011). Loss of motility is a common adaptation to the CF lung, which typically occurs through mutations *rpoN* and Type IV pili genes (Cai *et al.*, 2015). Loss of quorum sensing frequently occurs in isolates from the CF lung, which are coupled with mutations in the quorum-sensing regulator proteins, such as *lasR*, *rhlR* and *pqsR* (Winstanley *et al.*, 2016). Other common pathoadaptive traits include loss of Type III Secretion Systems, accumulation of auxotrophic mutations, and mutations in DNA repair genes leading to hypermutation (Winstanley *et al.*, 2016).

However, mutations in global regulator genes are not the only route to achieving a change in a particular phenotype (Jeukens *et al.*, 2014; Khademi *et al.*, 2019). The dataset of 4,094 isolates introduced in

Chapter 3 provides a unique opportunity to investigate the variation in phenotype of a population of isolates collected from a single patient, and to correlate that with the respective genotype. Therefore, as well as whole-genome sequencing, the isolates were screened for 10 virulence-related phenotypes that are known to vary between acute-stage *P. aeruginosa* infection and chronic-stage infection, and also known to vary between isolates of the same patient lung. The phenotypes that were screened for include; biofilm production; rhamnolipid production; protease production (caseinase and gelatinase); quorum sensing (PQS, BHL and OdDHL); siderophore production; and motility (swimming and twitching). These phenotypes are explained in more detail in Chapter 1.

In order to associate phenotype with genotype, this Chapter will use GWAS methodologies. Chapter 1 reviews the packages that have been developed for bacterial GWAS. However, most of these packages are designed for ease-of-use and did not afford the customisability required for this study. Therefore, this Chapter will employ the underlying methodology of several of these packages, GEMMA (Zhou & Stephens, 2012). To date, only one GWAS has tried to associate a *P. aeruginosa* phenotype with a *P. aeruginosa* genotype. This was for the development of the DBGWAS package by Jaillard *et al.*, where the authors associated antimicrobial resistance phenotypes of *P. aeruginosa* with the *P. aeruginosa* genotype, and uncovered both known and novel variants (Jaillard *et al.*, 2018). In this Chapter, complex phenotypes will be associated with the 4,094 genotypes, to try and uncover genetic cause of any phenotypic variation observed within the dataset.

4.1.1 Aims

The events that trigger APEs in patients with CF are not currently well understood, although they may be linked to changes in the virulence behaviour of bacteria that colonise the CF lung. In this study, we aim to use GWAS, an emerging and powerful tool in the field of comparative genomics, to determine whether there is any association between genotypic and phenotypic virulence in the *P. aeruginosa* population of the CF lung that may be causative of APE. By using the genomes of 4,094 isolates and phenotypic data for ten virulence-related phenotypes, this study is the largest *P. aeruginosa* GWAS to date, and will enable the most powerful analysis that will test the limit of current GWAS technologies.

4.2 Methods

4.2.1 Phenotypic assays

In order to identify potential markers of APE, virulence-related phenotypes that may be related to APE were identified (Emem Ukor, Floto Lab, University of Cambridge), and any association of these phenotypes with common genetic changes was investigated. The 4,408 *Pseudomonas* samples that were collected from three time-points before during and after an APE in 9 patients with CF from the TeleCF clinical trial (Chapter 3), 10 virulence-related phenotypes were measured; caseinase production, gelatinase production, siderophore production, rhamnolipid production, swimming motility, twitching motility, biofilm formation, BHL production, PQS production and OdDHL production (Chapter 1) (Emem Ukor, Floto Lab, University of Cambridge).

4.2.1.1 Protease production

In order to identify the presence or absence of caseinase production, each of the 4,408 *Pseudomonas* strains were cultured overnight, and 5 μL aliquots were incubated at 37^{O}C for 48 hours on skim milk plates. Protease activity was identified by the formation of clearing zones around the colony growth due to casein hydrolysis (Emem Ukor, Floto Lab, University of Cambridge).

In order to identify the presence or absence of gelatinase production, 3 μL of each overnight culture was incubated for 24 hours on gelatin agar plates. Plates were flooded with saturated ammonium sulphate solution, and protease activity was identified by the formation of proteolytic clearing zones due to gelatin hydrolysis (Emem Ukor, Floto Lab, University of Cambridge).

4.2.1.2 Siderophore production

In order to detect the presence of siderophores, each of the 4,408 *Pseudomonas* strains were cultured overnight, and 5 μL aliquots were incubated at 37^{O}C for 48 hours on chrome azurol S plates. Siderophore production was identified by the formation of an orange/pink halo around the colony, due to the

Fe^{3+}-chelating action of siderophores excreted from the colony (Emem Ukor, Floto Lab, University of Cambridge).

4.2.1.3 Rhamnolipid production

In order to detect the production of extracellular rhamnolipid, each of the 4,408 *Pseudomonas* strains were cultured overnight, and 5 µL aliquots were incubated at 37^{O}C for 48 hours on proteose peptone-glucose-ammonium salt agar, supplemented with 0.02% cethyltrimethyl ammonium bromide and 0.0005% methylene blue. Plates were incubated at room temperature for at least 24 hours, until extracellular rhamnolipid production was identified by the formation of blue halos around the colony growth, caused to the formation of a complex between the cationic methylene blue dye and partially anionic rhamnolipid molecules (Emem Ukor, Floto Lab, University of Cambridge).

4.2.1.4 Motility

In order to identify the presence or absence of the swimming motility phenotype, each of the 4,408 *Pseudomonas* strains were cultured overnight, and LB agar swim plates were inoculated with 5 µL aliquots. The plates were incubated at 37^{O}C for 8 - 12 hours. Swimming motility was assessed qualitatively by examining for the circular haze of growth/turbid zone formed by bacterial cell migration away from the point of inoculation (Emem Ukor, Floto Lab, University of Cambridge).

In order to identify the presence or absence of the twitching motility phenotype, LB agar swim plates were inoculated with 10 µL of each overnight culture. The plates were incubated at 37^{O}C for 48 - 72 hours. Twitch motility was assessed qualitatively by staining plates with 0.1% crystal violet after removal of agar, in order to assess growth at the interface between the agar and the polystyrene surface due to strong adherence and biofilm formation on the polystyrene surface, consistent with interstitial colony expansion (Emem Ukor, Floto Lab, University of Cambridge).

4.2.1.5 Biofilm

In order to quantify the biofilm phenotype, each of the 4,408 *Pseudomonas* strains were cultured overnight, and stationary-phase cultures were diluted 1:100 in fresh M63 minimal medium supplemented with 20% glucose and 20% casamino acids, then incubated at 37^{O}C for 24 hours. Supernatant was discarded and wells were washed twice in sterile deionized water to remove residual planktonic bacteria. Adherent cells were stained with 0.1% crystal violet for 10 min at room temperature. Wells were washed twice and air-dried overnight. 125 µL of 30% acetic acid was added to each well, and incubated for 10 – 15 min. Absorbance at 595 nm was measured using an EZ Read 400 microplate reader in order to quantify the crystal violet-stained biomass (Emem Ukor, Floto Lab, University of Cambridge).

4.2.1.6 Quorum sensing molecules

In order to identify the presence or absence of OdDHL and BHL quorum sensing molecules, the acylhomoserine lactones in the supernatant were quantified using two lux reporter P. aeruginosa strains; JM109 (pSB1142) for OdDHL and JM109 (pSB536) for BHL. In order to identify the presence or absence of the PQS quorum sensing molecule, the 2-alkyl-4-quinolone PQS was quantified using the lux reporter strain; $P_{pqsA::lux}$ (Emem Ukor, Floto Lab, University of Cambridge).

Each of the 4,408 Pseudomonas strains were cultured in 800 µL of buffered LB at 37^{O}C for 24 hours. The supernatant from planktonic stationary cultures was collected by centrifugation at 4000 rpm for 15 mins at room temperature. 100 µL each supernatant was co-incubated at 37^{O}C for 3 - 5 hours with 100 µL of overnight culture of JM109 pSB536, JM109 pSB1142 or $P_{pqsA::lux}$ reporter strains, in Grenier Cellstar® black, 96 well, polystyrene, flat micro-clear bottom microplates. Bioluminescence was measured with the Fluostar Omega (BMG Labtech) microplate reader. Measurement of AHL and AQ levels were performed by comparison to synthetic standards (Emem Ukor, Floto Lab, University of Cambridge).

4.2.2 Preparing genomes for GWAS

Sequence reads for the 4,094 isolates were generated according to the methods in Chapter 3, and were mapped to the *P. aeruginosa* PAO1 reference genome (BioProject Accession PRJNA331), as in Chapter

3. For analysis of individual patient groups, the highest quality assembled isolate was used as a reference (see Chapter 3, section 3.2.2.1), and all isolates from that patient group were mapped to it.

SNPs were extracted from multiple-sequence alignments using SNP-sites (v2.4.1, Page *et al.*, 2016) with default parameters, and a single VCF file, summarising all SNPs and INDELs across all isolates called against the *P. aeruginosa* PAO1 reference genome, was created. The VCF file was converted into PLINK-readable PED and MAP files using vcftools (v0.1.11, Danecek *et al*., 2011). The corresponding BED and FAM files were created using PLINK (v1.90b3v, Purcell *et al.*, 2007) using default parameters. This was repeated for all phenotypes for all 4,094 isolates, and for each of the phenotypes for each patient subgroup.

In order to aggregate non-synonymous SNPs from each gene into a single variant, an annotation file for the reference sequence was generated, the coding regions extracted, and any non-synonymous SNPs within each coding region aggregated into a single pseudo-SNP. These were then converted into a pseudo-VCF file, and data preparation was as described above.

4.2.3 Running GWAS on SNPs

GEMMA (v0.96, Zhou & Stephens, 2014) was used to perform association tests between SNPs and phenotypes. Linear Models (LM) were run using default settings. Population structure was controlled by GEMMA using a centred-genotype relatedness matrix, which was calculated within the programme. When fitting LMM to the entire dataset, the sequence type for each isolate was included as a covariate.

4.2.4 Running GWAS on gene presence-absence

Gene presence-absence data were generated for each patient subgroup and for the whole dataset, using Roary (v1.7.1, Page *et al*., 2015) with default settings. The gene presence-absence output matrix was converted into a pseudo-VCF file, and subsequent data preparation and association tests were run as previously with GEMMA.

4.2.5 Unitigs

Unitig data (refer to Chapter 1) was generated using the unitig-counter script, as part of the pyseer software suite (Lees *et al.*, 2018) using default settings. Phenotypic data were associated with the unitigs using Fast-LMM within pyseer (v1.2.0).

4.2.6 Analysing GWAS outputs

The output of each association test was visualised as Manhattan and QQ-plots using R (v3.5.0). Each GWAS associated a maximum of 63,008 individual variants to a significance threshold of $\alpha = 0.05$. Bonferroni correction was applied to counteract the problem of multiple testing, and was further rounded down to account for the multiple GWAS run in this chapter.

$$p < \frac{0.05}{63008} = 7.94\text{x}10^{-7} \approx 1\text{x}10^{-8}$$

Therefore, SNPs with reported *p*-values of $< 1\text{x}10^{-8}$ were recorded as significant. SNPs that were present in less than 1% of isolates, regardless of significance, were discarded. SNPs that encoded synonymous mutations were discarded. The gene containing a significant association was annotated within the *P. aeruginosa* PAO1 genome using BLAST (v2.9.0, Camacho *et al.*, 2009), and the gene annotation and function subsequently obtained from www.pseudomonas.com (Winsor *et al.*, 2016).

4.2.6.1 Analysing YfiR variants

Coordinates of the YfiR protein were downloaded from the PDB under accession number 4ZHY. To identify the position of the non-synonymous mutations, the protein structure was investigated using Chimera (v1.13, Petterson *et al.*, 2004).

4.2.7 Estimating phenotypic co-occurrence

In order to measure the pairwise association between each phenotype, Cramer's V was calculated with the Cramer package (v0.93) using R (v3.5.0).

4.3 Results

4.3.1 Phenotypic clustering within patient groups

In order to investigate whether APEs in patients with CF are associated with changes in virulence of the *P. aeruginosa* bacteria that colonise their lungs, a GWAS study was carried out. Initially, the 4,094 isolates from the cystic fibrosis lung of nine patients were tested for ten virulence-related phenotypes (Table 4.1) (Emem Ukor, Floto Lab, University of Cambridge) in order to investigate the phenotypic diversity within each patient compared to the genetic diversity (Chapter 3). Chapter 1 explores these phenotypes in more detail; several phenotypes are known to correlate with the development from acute to chronic *P. aeruginosa* CF infection, and several phenotypes have roles in *P. aeruginosa* virulence and invasive potential.

The BHL production, OdDHL production, and PQS production phenotypes are related to quorum sensing (QS) in *P. aeruginosa* (Lin *et al.*, 2018). QS molecules facilitate intracellular communication, which is cell density dependent (Lee & Zhang, 2015). The systems are all interconnected and regulate many virulence phenotypes in *P. aeruginosa* (Lee & Zhang, 2015).

Biofilm production and rhamnolipid production assays were also carried out on the isolates in this study. A biofilm is a matrix of extracellular polymeric substances which encompass the bacterial colonies and can prevent the incursion of antibiotics to combat the infection (Nickzad & Deziel, 2014). BHL, OdDHL and PQS molecules are known to regulate the production of biofilm and rhamnolipids (Reis *et al*., 2011).

Rhamnolipids are biosurfactants that are involved in increasing the affinity of bacteria to a surface prior to biofilm formation, maintenance of a strong biofilm matrix, and in the dispersion of the biofilm to allow bacteria to invade and colonise other niches. Caseinase and gelatinase production assays were also carried out. These are proteases that allow *P. aeruginosa* to evade the immune system, and are also QS-regulated (Zhang & Li, 2016).

Other phenotypes investigated were the motility mechanisms; twitching motility and swimming motility. These two forms of motility are functionally opposite to each other. When *P. aeruginosa* moves due to

Phenotype	Description
BHL production	An autoinducer that activates the global regulator protein RhlR.
OdDHL production	An autoinducer that activates the global regulator protein LasR.
PQS production	Quinolone-based intercellular signalling which controls biofilm production and virulence factor production.
Biofilm production	A matrix of extracellular polymeric substances that hold microbial cells together to a surface.
Rhamnolipid production	Glycolipids with detergent-like activity and have cytolytic activity against macrophages.
Caseinase production	An exo-enzyme produced by bacteria that breaks down the phosphoprotein casein.
Gelatinase production	An exo-enzyme produced by bacteria that hydrolyses gelatin.
Twitching motility	The movement of *P. aeruginosa* across surfaces using the Type IV pilus.
Swimming motility	A flagellum-dependent form of movement in Gram-negative bacteria.
Siderophore production	Iron-chelating agents that *P. aeruginosa* secrete to uptake iron from the environment.

Table 4.1 Ten phenotypes were experimentally investigated using the 4,094 *P. aeruginosa* isolates in this study.

the twitching mechanism, it has decreased swimming motility, and vice versa (Burrows, 2012). Twitching motility utilises the Type IV pili, whilst swimming motility is flagellum-dependent motion. These forms of motility are also important for biofilm formation and biofilm dispersion (Burrows, 2012).

The siderophore production assay identifies the ability of an isolate to uptake iron from its surroundings, which is essential for bacterial growth. *P. aeruginosa* express two siderophore systems, utilising pyoverdine or pyochelin. These two siderophores can be released into the environment, and then taken back into the cell using specific uptake proteins (Braud *et al.*, 2011).

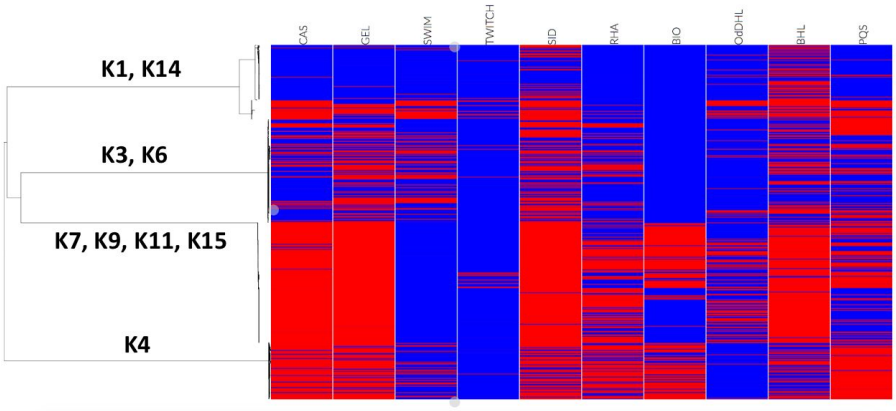


Figure 4.1 A maximum-likelihood phylogeny of the *P. aeruginosa* isolates mapped against PAO1, displayed next to the presence/absence of each phenotype determined by experimental assays. Blue = phenotype absent, Red = phenotype present.

	K1	K14	K3, K6	K4	K7, K9, K11, K15	All
	n = 646	n = 265	n = 1,197	n = 660	n = 1,389	n = 4,094
Biofilm	0.00%	0.00%	0.00%	69.20%	52.80%	28.90%
Caseinase	1.20%	91.80%	25.30%	78.60%	97.70%	58.10%
Gelatinase	4.70%	54.10%	47.30%	81.20%	99.40%	64.10%
Rhamnolipid	1.20%	13.20%	21.20%	75.40%	60.40%	39.60%
Siderophore	23.80%	80.90%	62.40%	68.40%	98.90%	70.70%
Swimming	5.30%	80.90%	30.50%	21.10%	0.20%	17.60%
Twitching	3.00%	31.40%	1.40%	1.80%	7.00%	5.20%
OdDHL	3.40%	60.50%	8.30%	3.50%	39.10%	20.00%
BHL	51.50%	88.20%	39.20%	47.80%	91.60%	62.90%
PQS	4.10%	77.30%	45.40%	91.50%	44.60%	47.80%

Table 4.2 The proportion of isolates where the phenotype is present for the isolates from each BAPS cluster. The colour represents a sliding scale, where red represents 0% presence of the phenotype, yellow represents 50% presence of the phenotype, and green represents 100% presence of the phenotype.

The presence or absence of each phenotype was displayed against the phylogenetic tree that was calculated in Chapter 3, section 3.3.2 (Figure 4.1). Clear clustering of phenotype presence/absence for some of the patient groups was observed.

To further examine the phenotype presence/absence clustering in some patient groups, the proportion of presence and absence of each phenotype in each BAPS cluster (Chapter 3, section 3.3.2) is displayed in Table 4.2.

The isolates from patient K1 were mostly absent for all phenotypes measured; only siderophore production and BHL production were present in more than 10% of the isolates. All of the isolates in this group are Manchester Epidemic Strain (MES) isolates. The other patient that harbours the MES isolates is patient K14. In contrast to the patient K1 isolates, the patient K14 isolates have over 50% presence for 7 out of the 10 phenotypes. This includes the largest proportion of swimming, twitching and OdDHL phenotypes across the whole dataset. This suggests that the K14 isolates may be more virulent than the K1 isolates at the time of isolation, and that although they are both made up of the same epidemic strain, they exhibit vastly different behaviours.

In contrast to the MES isolates, the presence/absence of each phenotype was consistently observed in all patients harbouring the Liverpool Epidemic Strain (LES) isolates (patients K7, K9, K11 and K15). Four

of phenotypes were present in over 90% of these isolates, and only the swimming and twitching phenotypes were present in less than 10% of the isolates. This suggests that the virulence potential that has contributed to the success of the LES isolates to transmit and invade new hosts persists, even after the infection has become chronic (O'Brien *et al.*, 2017).

The phenotype presence/absence ratio in the isolates from patient group K3 and K6 is relatively even in seven of the phenotypes (ranging from 21% - 62% of isolates). Less than 10% of isolates produce biofilm or OdDHL, or have twitching motility. The majority of phenotypes have a high level of presence in the isolates from patient K4, as six of the phenotypes have a presence of over 68%. The number of isolates positive for OdDHL production and twitching motility are very low, as in patient group K3 and K6.

The presence of both motility phenotypes is low across all of the patient groups combined. Twitching motility is only observed in 5.2% of all isolates, and swimming motility is present in just 17.6% of isolates. This suggests that the majority of the isolates are either non-motile, or express a form of motility that was not tested in this study, such as sliding or swarming motility (Murray & Kazmierczak, 2008). Previous studies have suggested that a lack of motility can be a selective advantage when subject to antibiotics, and causes phagocytosis resistance, which explains why a lack of motility is often observed in chronic cystic fibrosis infection (Staudinger *et al*., 2014; Mahenthiralingam *et al.*, 1994). However, in contrast to the overall dataset, over 80% of isolates in the patient K14 group are positive for swimming motility, and over 30% are positive for twitching motility, suggesting that the K14 MES isolates are either more virulent than their K1 counterparts, or the population has not completely switched from acute to chronic infection.

Production of biofilm is an important stage in the development of chronic lung infection, by protecting colonies from antibiotics and the host immune response (Staudinger *et al.*, 2014). However, previous studies have shown that *P. aeruginosa* from long-term chronic cystic fibrosis infections are often impaired in biofilm formation as mutations in the quorum sensing pathways accumulate (Rasamiravaka *et al*., 2015), and therefore the absence of the biofilm phenotype in the majority of isolates is not unexpected. Biofilm presence was only observed in two of the patient groups, the K4 patient group and the K7, K9, K11 and K15 patient group.

The phenotype with the highest prevalence across the dataset was siderophore production (70.7% presence). Iron is essential for bacterial survival (Skaar, 2010), and *P. aeruginosa* contains several

iron-uptake systems that scavenge for this nutrient. This includes two extracellular siderophores, two haem-uptake and utilisation systems, and two ferric-iron uptake systems (Marvig *et al.*, 2014). A study by Harrison *et al*. in 2017 showed that siderophores were produced by wild-type bacteria under standard lab conditions, as well as under artificial CF conditions. However, the CF lung is iron-rich (Tyrrell & Callaghan, 2016), and studies have suggested that siderophore-lacking mutants survive through the uptake of haem and free iron instead (Marvig *et al.*, 2014). As siderophores are secreted into shared environment of a population of *P. aeruginosa*, it is suggested that 'cheaters', which have the ability to uptake siderophores but do not pay the metabolic cost of producing them, are also common within the cystic fibrosis lung (Anderson *et al.*, 2015). This may explain the high, but not complete, prevalence of siderophore production within this dataset.

Cystic fibrosis is a disease of progressive lung function deterioration, with periods of sudden and rapid lung function decline (Ramsey *et al.*, 2017). These declines in lung function are called acute pulmonary exacerbations (APEs) (Bhatt, 2013). The isolates in this study were collected at different stages of an APE for each patient. In this study, the acute-APE timepoint is defined as the day the patients were first treated with antibiotics, the recovery-APE timepoint is defined as seven days after antibiotic treatment has finished, and the stable-APE timepoint is defined as seven days prior to the APE being initially identified. These APEs play an important role in the progression of CF, and therefore, these timepoints have been included as three additional phenotypes in this study.

The proportion of isolates in each BAPS cluster that belong to each APE time point is dependent on how many complete and incomplete exacerbations that have been sequenced, and how many of those isolates passed QC (Table 4.3). In general, more isolates are included in this study from the acute-APE timepoint, which has the highest prevalence in every patient subgroup except for patients K3 and K6 where the most prevalent time point is stable-APEs. In all patient subgroups, the recovery-APE timepoint contains the lowest proportion of isolates.

APE timepoint	Description	Presence in BAPS cluster				
		K1	K14	K3, K6	K4	K7, K9, K11, K15
Stable	Seven days prior to the APE identification and treatment with antibiotics	27.7%	35.9%	35.6%	27.4%	30.5%
Acute	The day the patient was first treated with antibiotics for the APE	38.1%	42.7%	34.9%	45.7%	40.7%
Recovery	Seven days after the course of antibiotics was finished	34.1%	21.3%	29.5%	26.9%	28.7%

Table 4.3 Three time points were investigated surrounding acute pulmonary exacerbations (APEs) using the 4,094 *P. aeruginosa* isolates in this study.

4.3.1.1 Phenotypic co-occurrence

Many phenotypes co-occur in the same isolate, and therefore a relationship may exist between the presence or absence of several phenotypes in this dataset. Cramer's V was calculated for all phenotypes in a pairwise manner (Table 4.4). The interpretation of the Cramer's V statistic was in accordance with Cohen's recommendations (Cohen, 1988). A value of 0 indicates no relationship between the two variables, a value of 0.3 indicates a moderate relationship, and a value of 0.5 indicates a strong relationship between the two variables. A value of 1 would indicate a measurement of complete co-occurrence.

There are five instances of strong relationships (green) between two variables. The strongest association was observed between the two secreted protease phenotypes; caseinase production and gelatinase production. In total, 53.4% of the isolates in the study produced both proteases, and 31.9% of the isolates produced neither. Only 14.7% of the isolates produced either caseinase or gelatinase. This relationship makes biological sense, as several exoenzymes have both caseinolytic and gelatinolytic capabilities, and the individual exoenzymes are often regulated through the same QS network and hence production will be induced together (Yan *et al.*, 2018).

Both caseinase and gelatinase production were also strongly associated with siderophore and rhamnolipid production. Nearly 75% of isolates either both produced or both did not produce caseinase and siderophore, 72% of isolates either produced or both did not produce caseinase and rhamnolipids, 79% of

	CAS	GEL	SWIM	TWITCH	SID	RHA	BIO	OdDHL	BHL	PQS
CAS		0.676	0.02	0.18	0.471	0.481	0.321	0.25	0.307	0.248
GEL			0.126	0.106	0.539	0.483	0.258	0.146	0.205	0.192
SWIM				0.254	0.089	0.004	0.136	0.028	0.06	0.102
TWITCH					0.091	0.035	0.019	0.235	0.125	0.157
SID						0.282	0.273	0.234	0.193	0.174
RHA							0.138	0.141	0.181	0.201
BIO								0.223	0.27	0.073
OdDHL									0.334	0.204
BHL										0.117
PQS										

Table 4.4 A cross-table of the Cramer's V associations between each pair of phenotypes. The scale is a sliding colour scale, where red = 0, yellow = 0.3, and green = 0.5.

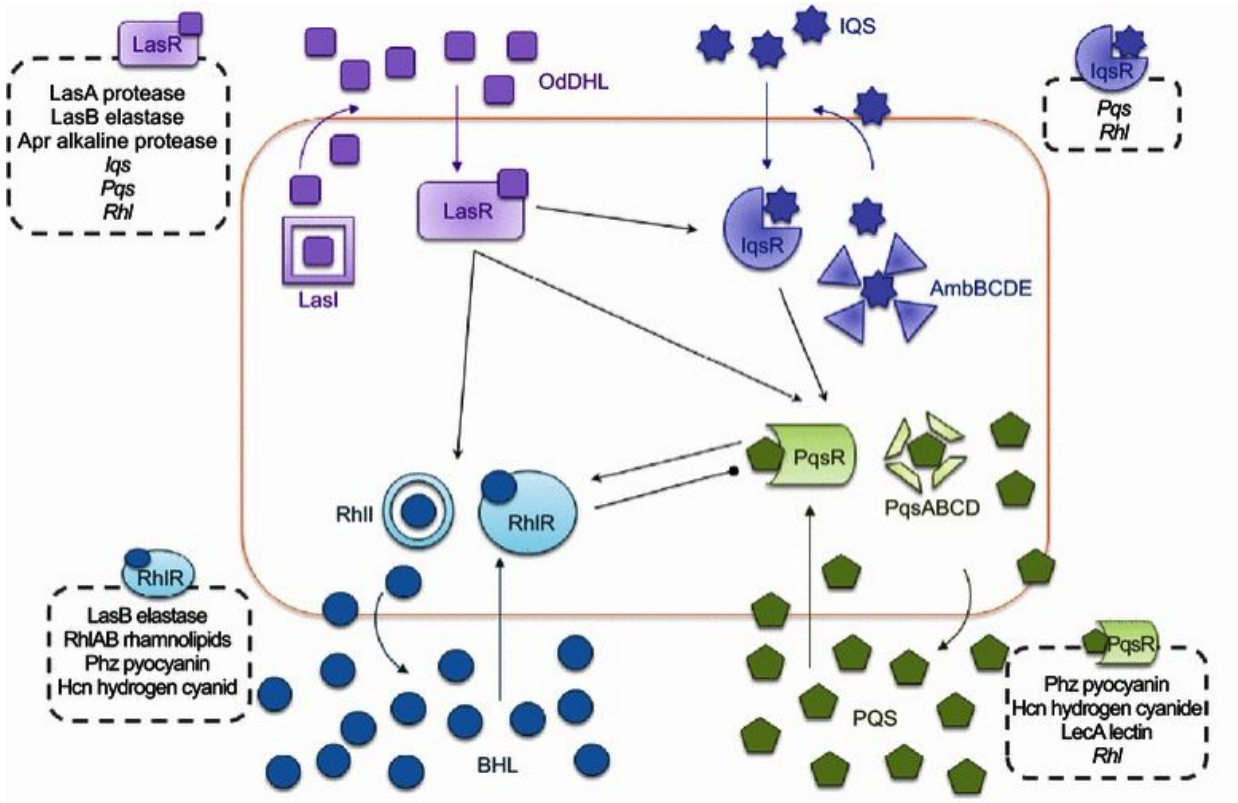


Figure 4.2 A schematic representation of quorum sensing within the *P. aeruginosa* cell. Image has been used with license under creative commons v1.0 from Lee & Zhang, 2015.

isolates either produced or did not produce both gelatinase and siderophores, and 55% of isolates either both produced or both did not produce gelatinase and rhamnolipids. All four of these phenotypes are virulence phenotypes under QS control. The QS systems in *P. aeruginosa* are all hierarchical, with the *lasR* QS system at the top of this hierarchy (Maura *et al.*, 2016). The activation of the *lasR* QS system will activate the other QS systems, which leads to the secretion of protease systems, production of rhamnolipid, and the production of the pyocyanin siderophore system (Maura *et al.*, 2016) (Figure 4.2).

Interestingly, there were only moderate relationships observed between the three QS assays, and moderate relationships observed between the QS assays and the phenotypes under QS control. This suggests that standard lab conditions may be confounding the assay results, where the phenotypes observed under assay conditions may differ to the phenotypes that are present in the CF lung environment, or that random variations have been introduced during the experiments (Wright *et al.*, 2013).

The lowest associations were identified between the two motility phenotypes, swimming and twitching, and all other phenotypes. Nearly 90% of the associations between the swimming phenotype and all other phenotypes scored less than 0.15, which is a very weak association. Just over 50% of the associations between twitching and the other phenotypes were also very weak. This is primarily due to the low presence of both swimming and twitching motilities (79% of the isolates were negative for both swimming and twitching), and the high presence of all of the other phenotypes. Only 3% of isolates tested positive for both motilities, but the two motility mechanisms were found to be moderately associated with one another.

4.3.2 Linear models are unsuitable for GWAS in highly structured datasets

The diversity in the presence of phenotype for each patient suggests that there should be enough power to associate phenotype with the variation of genotype identified in Chapter 3. This could provide genotypic markers that could be indicative of phenotype. Genome-Wide Efficient Mixed Model Association (GEMMA) was used to try and identify any associations with the isolate genotypes and the 13 phenotypes (which include the 10 experimentally-tested phenotypes and the 3 different exacerbation state stages). Initially, a simple Linear Model (LM) association for the caseinase-producing phenotype was fitted against all Single Nucleotide Polymorphisms (SNPs) (n = 63,008) reaching 1% Minor Allele Frequency (MAF) for all 4,094 isolates in the dataset called against the *P. aeruginosa* PAO1 reference genome. The LM association identified several thousand associations at a conservative p-value threshold of $1x10^{-8}$.

The Manhattan plot in Figure 4.3a displays the p-value of each SNP association against its position in the PAO1 genome. Figure 4.3b shows a QQ-plot of the expected vs observed p-values. The majority of the observed p-values deviate significantly from the expected p-values. The horizontal lines at different p-value tiers in Figure 4.3b indicate that the associations reaching the threshold values are reported as positive associations due to strong population structure within the dataset, and hence are false positive results. False positive results due to population structure were observed for all 13 phenotypes using this method of correlation.

The strong population structure is consistent with the population structure observed in Chapter 3, section 3.3.2. The phylogenetic tree (Chapter 3, section 3.3.2) indicated very deep splitting, with long branches

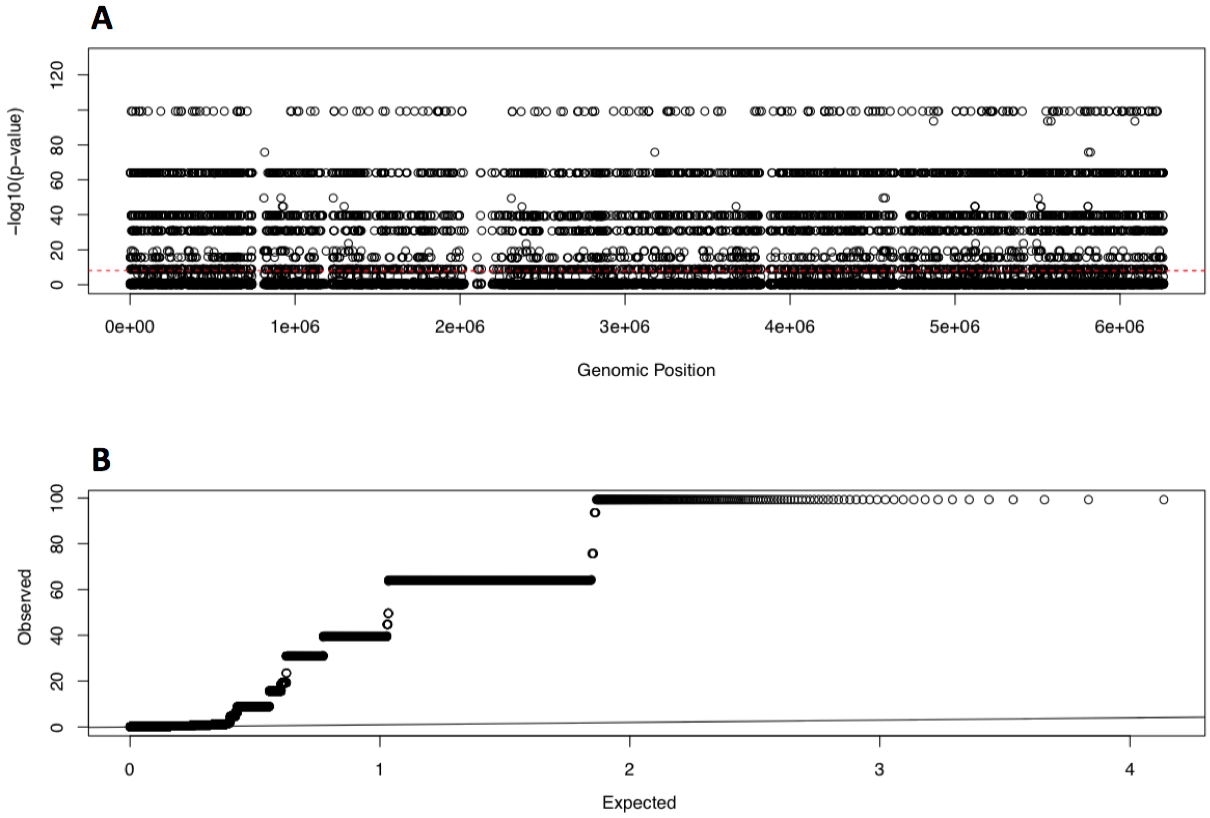


Figure 4.3 All SNPs reaching a 1% MAF cutoff, called against the *P. aeruginosa* PAO1 reference genome for 4,094 isolates, were fit to a LM to test for association with the caseinase-producing phenotype and plotted as a Manhattan plot (a) and QQ-plot (b). The red line in (a) represents the association threshold, and the black line in (b) represent the distribution of expected *p*-values against observed *p*-values.

separating each group of isolates. Principal component analysis identified that 93.4% of genetic variation could be explained by the first three principal components, which confirmed deep divergence and a high degree of clustering of isolates. Clustering of isolates by sequence type showed that each patient is only infected with a single ST, but multiple patients may be infected with the same ST.

Therefore, in order to extract meaningful associations from each GWAS, population structure control must be applied. Several methods of controlling population structure in structured datasets have been suggested, which include, but are not limited to:

- Genomic Control - directly adjust the *p*-values so that the vast majority are non-significant (Devlin *et al.* 1999).

- Subsampling - expect individuals in the subsampled population to have similar phenotypes to the main population (Thomas *et al.* 2009).
- Principal Component Analysis - uses PCA values as covariates in the association study (Abraham *et al*. 2014).
- Linear Mixed Models (LMM) - a method of analysing non-independent, hierarchical, longitudinal or correlated data. It takes random effects into consideration as well as fixed effects (Widmer *et al*. 2014).

4.3.3 Using subsampled groups to control for population structure

The horizontal lines of associated SNPs shown in Figure 4.3b correspond to SNPs that are unique to individual patient groups. Therefore, to extract true significant hits that are associated with a phenotype, population structure control was introduced by separating sequences using BAPs clustering (Chapter 3, section 3.3.2) into five separate datasets; K1 (n = 646) and K14 (n = 265); K3 and K6 (n = 1197); K4 (n = 660); and K7, K9, K11 and K15 (n = 1389). One drawback of subsampling is that the quantity of genotypic and phenotypic data included for each GWAS analysis will be reduced, which will lead to a reduced power to identify correct associations (Peterson *et al.*, 2017).

		Patient Group					
Phenotype		K1 n = 646	K14 n = 265	K3, K6 n = 1,197	K4 n = 660	K7, K9, K11, K15 n = 1,389	All n = 4,094
	Biofilm	0.0%	0.0%	0.0%	69.2%	52.8%	28.9%
	Caseinase	1.2%	91.8%	25.3%	78.6%	97.7%	58.1%
	Gelatinase	4.7%	54.1%	47.3%	81.2%	99.4%	64.1%
	Rhamnolipid	1.2%	13.2%	21.2%	75.4%	60.4%	39.6%
	Siderophore	23.8%	80.9%	62.4%	68.4%	98.9%	70.7%
	Swimming	5.3%	80.9%	30.5%	21.1%	0.2%	17.6%
	Twitching	3.0%	31.4%	1.4%	1.8%	7.0%	5.2%
	OdDHL	3.4%	60.5%	8.3%	3.5%	39.1%	20.0%
	BHL	51.5%	88.2%	39.2%	47.8%	91.6%	62.9%
	PQS	4.1%	77.3%	45.4%	91.5%	44.6%	47.8%
	Stable APE	27.7%	35.9%	35.6%	28.5%	27.4%	30.5%
	Acute APE	38.1%	42.7%	34.9%	43.1%	45.7%	40.7%
	Recovery APE	34.1%	21.3%	29.5%	28.4%	26.9%	28.7%

Table 4.5 The percentage of phenotype-positive isolates in each subsampled group, including presence of each APE timepoint. n = number of isolates in each group.

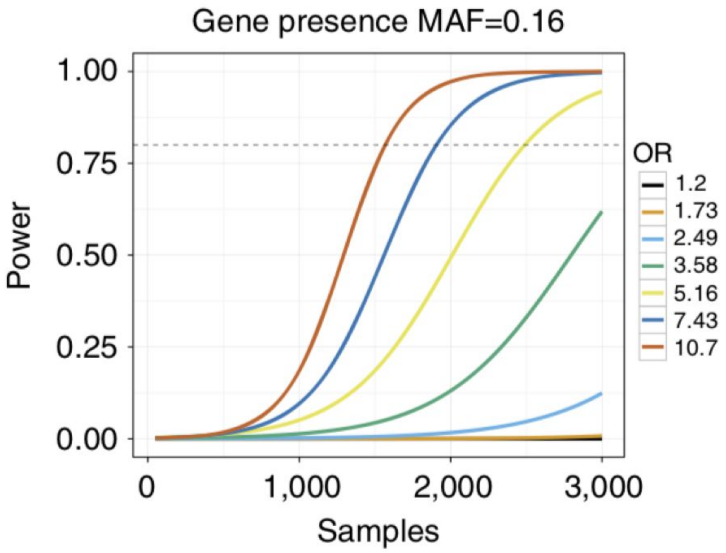


Figure 4.4 Power calculations for different odd ratios (OR) of SNP associations at MAF of 16% can require several thousand isolates in order to have enough power to detect true positive results. By subsampling this dataset, so that each group contains between 265 and 1,389 isolates, it is unlikely that there will be enough power to detect all true positive associations. Image used under license (Creative Commons v4.0) from Lees *et al*. 2016.

The percentage of positive phenotypes for the isolates in each subsampled group are summarised in Table 4.5. The K14 patient group contains the fewest isolates, with only 265 in total. The LES isolates form the largest group, comprising of 1,389 isolates from patients K7, K9, K11, K15. Due to the reduced number of isolates in each subsampled group, even if each SNP has a high odds ratio, even the largest patient group may lack the power to detect all true positive SNPs (Figure 4.4, Lees *et al*., 2016).

Another drawback to using the subsampling method to control for population structure in this dataset is the now-skewed case/control ratio for each phenotype within each group. A similar number of both case phenotypes and control phenotypes is most desirable to result in an effective GWAS (Cook *et al*. 2017). However, within the subsampled groups, case/control ratios range from as low as 0%, up to 99.4%. For example, the LES patient group contains only three isolates that were positive for the swimming motility phenotypes, and just eight of the isolates are negative for the gelatinase phenotype. Associations reported for extreme case/control ratios such as these may not be reliable and cannot necessarily be extrapolated for the rest of the dataset.

Phenotype	Patient group	Phenotype presence in patient group	PA Gene Number	Protein Name	Codon Change	Amino Acid Change	Minor Allele Frequency	Direction of effect on phenotype	*p*-value
Rhamnolipid	K3, K6	21.2%	PA2654	Hypothetical Protein	AGC -> AAC	S50N	14.5%	-ve	2.12×10^{-11}
Rhamnolipid	K3, K6	21.2%	PA4389	SpeA	ACG -> GCG	T77A	15.0%	-ve	3.59×10^{-10}
BHL	K3, K6	39.2%	PA2006	Hypothetical Protein	TTC -> CTC	F101L	8.4%	-ve	1.49×10^{-09}
PQS	K3, K6	45.4%	PA2194	HcnB	CTC -> CCC	L148P	5.30%	+ve	4.34×10^{-14}
PQS	K3, K6	45.4%	PA5344	OxyR	AGG -> AAG	R94K	14.8%	+ve	4.84×10^{-14}
Acute APE	K3, K6	34.9%	PA2194	HcnB	CTC -> CCC	L148P	5.30%	+ve	2.60×10^{-12}
Recovery APE	K3, K6	29.5%	PA2006	Hypothetical Protein	TTC -> CTC	F101L	8.4%	-ve	7.56×10^{-11}

Table 4.6 Results that reached the significance threshold in LM association tests against the 13 phenotypes.

LM association tests were fit for each of the 13 phenotypes to each of the subsampled patient groups, which was successful at reducing the effects of population structure. No associations were reported with any phenotype for the isolates from patient groups K1, K14, and K4. This is probably due to a lack of power to detect true positive results, as there were so few samples within the datasets. No significant associations were identified for the subsampled group containing isolates from patient groups K7, K9, K11 and K15, even though this was the largest of the subsampled groups, suggesting there are multiple different genotypic causes of each phenotype. Five phenotypic associations reached the significance threshold for isolates from patients K3 and K6, which was the second largest subsampled group (Table 4.6).

Two SNP associations reached the significance threshold for the rhamnolipid-producing phenotype; a serine to asparagine mutation in the hypothetical protein, PA2654, and a threonine to alanine mutation in the arginine decarboxylase, SpeA. The PA2654 hypothetical protein is uncharacterised, but predicted to be part of the *P. aeruginosa* motility pathways (Corral-Lugo *et al*., 2018). Rhamnolipids are required for *P. aeruginosa* swarming and sliding motility, and previous studies have shown that the production of rhamnolipid is not altered by the presence or absence of motility mechanisms (Murray & Kazmierczak, 2008). Therefore, if the predicted function of the hypothetical protein is correct, then this association is unlikely to cause a decrease in rhamnolipid production based on our current understanding of *P. aeruginosa* biology.

SpeA is an arginine decarboxylase, responsible for the conversion of *L*-arginine into agmatine (Gilbertson & Williams, 2014). Agmatine is used in the biosynthesis of several polyamines, such as spermidine and putrescine (McCurtain *et al*., 2019). Nitrogen exhaustion in *P. aeruginosa* has previously been shown to inhibit rhamnolipid production (Reis *et al*., 2011), and therefore it is possible that if the T77A mutation in

SpeA inhibits function, the resulting lack of useable nitrogen could reduce rhamnolipid production. However, *L*-arginine is not the only source of nitrogen for the *P. aeruginosa* bacterium, and decarboxylation via SpeA is not the only metabolic pathway for arginine (Lu *et al.*, 2004) (Figure 4.5). Therefore any effect would be small, and this is likely to be an auxotrophic mutation which has been coselected with rhamnolipid production.

Two SNP associations reached the significance threshold for the PQS-producing phenotype; an arginine to lysine mutation in the OxyR global regulator, and a lysine to proline mutation in the HcnB hydrogen cyanide synthase. The *oxyR* gene encodes a global regulator protein that responds to oxidative stress induced by reactive oxygen species (Ochsner *et al.*, 2000). A transcriptional analysis by Wei *et al.* indicated that OxyR knockout mutants increase transcription of several secondary regulators, including the *pqsR* gene (Wei *et al.*, 2012). Previous studies have identified that *pqsR* positively regulates the PQS system (Brouwer *et al.*, 2014), and therefore if R94K inhibits the function of OxyR, an increase in PQS would be observed. However, this mutation is only found within 14.8% of the isolates from patients K3

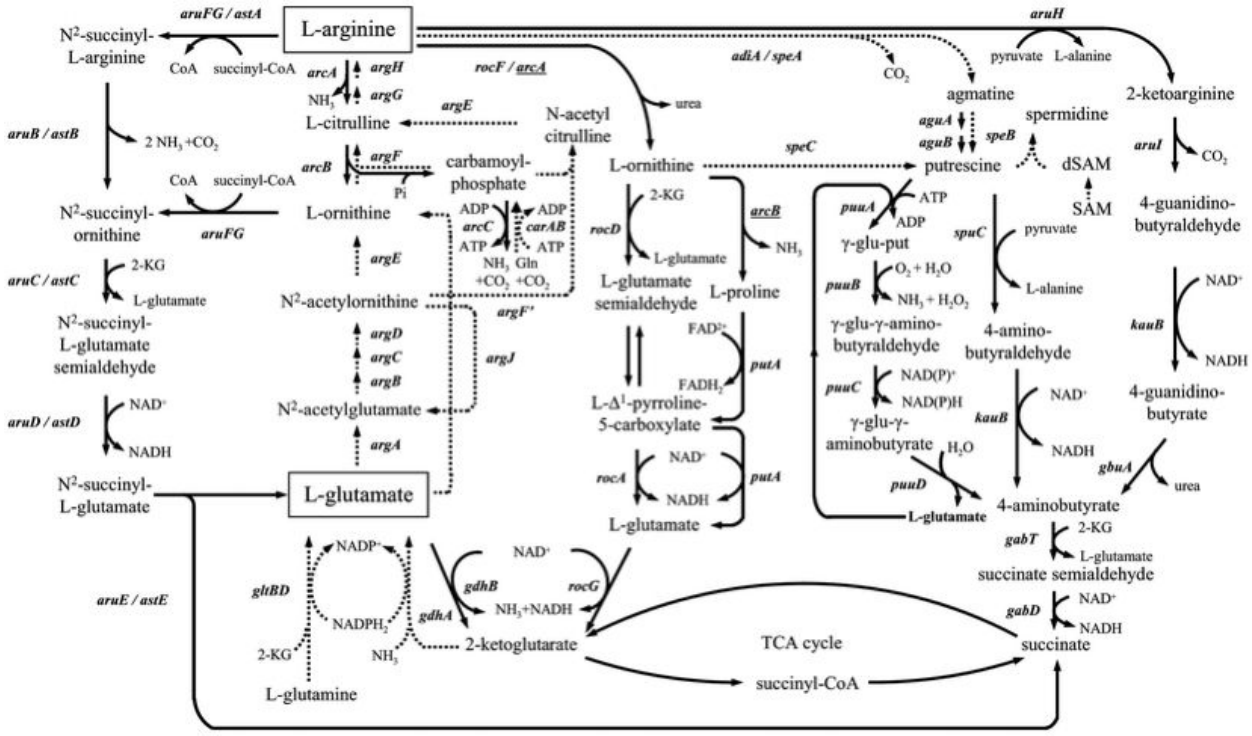


Figure 4.5 A diagram of the arginine metabolic and catabolic pathways. SpeA forms only a small part of this network, decarboxylating *L*-arginine to agmatine. Image used with permission from Lu, 2006.

and K6, whereas the PQS-producing phenotype is present within 45.4% of these isolates. Two other non-synonymous mutations were present in the OxyR protein, but no INDELs were observed. Non-synonymous mutations were present within 35% of the isolates in this group. Although these mutations do not completely explain all of the phenotypic variation observed within this subgroup, there is published experimental support for the biological association.

HcnB is part of the biosynthetic pathway for hydrogen cyanide (HCN), which is a dangerous virulence factor that contributes to high mortality rates during acute infection (Ryall *et al.*, 2008). The HCN biosynthetic cluster is regulated by QS systems, but no feedback mechanism has been identified that suggests that HCN expression causes a change in the expression of QS molecules (Lee & Zhang, 2015). The non-synonymous mutation in *hcnB* that was identified in association with the PQS phenotype, in isolates from patients K3 and K6, was also identified in association with the acute-APE timepoint. As a potent virulence factor, an increase in HCN may lead to a decline in lung function. The non-synonymous mutation in *hcnB* is a reversion to the wild-type *hcnB* allele present in *P. aeruginosa* PAO1 reference strain. If this reversion mutation leads to an induction of HCN production, it may contribute to a decline in lung function and overall health of the patient, similar to what is seen during an acute-APE.

An association for both the BHL-producing phenotype and the recovery-APE timepoint, was identified with a phenylalanine to leucine mutation in the hypothetical protein, PA2006. The *PA2006* gene is hypothesised to encode a transmembrane transport protein, but no function has been definitively assigned. BHL is both a secreted molecule and imported molecule, and if PA2006 was involved in the transport of BHL, a non-synonymous mutation could decrease the detection of BHL in the phenotypic assay. However, PA2006 has not previously been linked to the transport of BHL, and so functional studies would need to be undertaken to understand this association better. No further biological association can be hypothesised for BHL-production or the recovery-APE timepoint until a function is assigned to PA2006.

4.3.4 Aggregating non-synonymous SNPs in subsampled groups

To increase the power of the GWAS to identify less penetrant variants, non-synonymous SNPs from the same gene were aggregated together to form a single aggregated SNP. Aggregating all non-synonymous

Figure 4.6 Aggregating together all non-synonymous SNPs within the same gene may avoid separately testing several mutations that have the same causative effect within the same gene.

mutations found within a single gene may avoid separately testing several mutations that have the same causative effect within the same gene. In this way, all non-synonymous mutations within the same gene can be combined and tested against the phenotype, rather than being tested independently (Figure 4.6).

LM association tests were run for each aggregated SNP with each subsampled patient group for all 13 phenotypes. In total, 13 associations reached the significance threshold for four of the phenotypes (Table 4.7). All associations were identified within the subgroup of isolates from patients K3 and K6, as in the previous section.

A total of eight unique associations were identified with the PQS-producing phenotype. Three of these associations are with hypothetical proteins of no known or predicted function; PA1797, PA1792 and PA4835. No speculation on the validity of these associations can be made until biological function has

Phenotype	Patient group	Phenotype presence in patient group	PA Gene Number	Protein Name	Minor Allele Frequency	Direction of effect on phenotype	*p*-value
OdDHL	K3, K6	8.3%	-	Hypothetical Protein	5.20%	+ve	3.29×10^{-12}
OdDHL	K3, K6	8.3%	PA4193	Hypothetical Protein	19.6%	+ve	5.66×10^{-10}
OdDHL	K3, K6	8.3%	PA0869	PbpG	4.6%	+ve	6.21×10^{-10}
BHL	K3, K6	39.2%	PA2655	Hypothetical Protein	7.40%	-ve	2.41×10^{-09}
PQS	K3, K6	45.4%	PA2008	FahA	19.1%	+ve	9.79×10^{-14}
PQS	K3, K6	45.4%	PA1797	Hypothetical Protein	14.5%	+ve	1.19×10^{-13}
PQS	K3, K6	45.4%	PA3545	AlgG	4.20%	+ve	2.14×10^{-11}
PQS	K3, K6	45.4%	PA4385	GroEL	20.9%	+ve	1.19×10^{-10}
PQS	K3, K6	45.4%	PA1276	CobC	3.90%	-ve	1.62×10^{-09}
PQS	K3, K6	45.4%	PA1792	Hypothetical Protein	35.4%	-ve	1.38×10^{-08}
PQS	K3, K6	45.4%	PA1556	CcoO2	2.90%	+ve	3.50×10^{-08}
PQS	K3, K6	45.4%	PA4835	Hypothetical Protein	5.10%	+ve	9.97×10^{-08}
Recovery-APE	K3, K6	29.5%	PA2655	Hypothetical Protein	7.40%	-ve	1.78×10^{-09}

Table 4.7 Results that reached the significance threshold in LM association tests against the 13 phenotypes.

been assigned. Mutations within the *groEL* gene, which is involved in protein folding (Crouzet *et al.*, 2017), the *cobC* gene, which is involved in cobalamin biosynthesis (Kumari *et al.*, 2014), and the *ccoO2* gene, which is responsible for haeme biosynthesis in anaerobic conditions (Okkotsu *et al*., 2014) were also associated with the PQS-producing phenotype. However, these have no known biological basis for interacting with the PQS system, and therefore are assumed to be false-positive results until such a time as there is a more-detailed understanding of the *P. aeruginosa* biochemical pathways.

Mutations in the *algG* gene were also associated with the PQS-producing phenotype. *AlgG* knockout mutants have been shown to cause the loss of alginate secretion, which is responsible for the mucoid phenotype and biofilm production (Hoffmann *et al.*, 2007). Absence of PQS has been shown to decrease biofilm production (Tettmann *et al.*, 2016). There may therefore be a feedback mechanism between absence of biofilm formation and absence of PQS production, but no such feedback has been identified to date.

Finally, mutations in the *fahA* gene were also associated with the PQS phenotype. The *fahA* gene is responsible for the final step in the production of fumarate. This process has a tenuous link to PQS production, as they share common precursor molecules, such as chorismate (Tielen *et al.*, 2013). A mutation which results in a knockout of the *fahA* gene may affect the enzymatic competition for these precursor molecules, but is unlikely to cause a dramatic increase in PQS production as these precursor molecules are also used in tyrosine and tryptophan biosynthesis. In patient K3, the *fahA* gene (PA2008) is situated 6 kbp downstream of the 48 kbp deletion (PA1957 - PA2002) identified in Chapter 3, section 3.3.7.1. In the isolates containing the large deletion, the *atoBE* genes are absent, preventing the conversion of acetoacetate to 2-acetyl-CoA. The *fahA* gene is involved in the prior step, converting 4-fumarylacetoacetate into acetoacetate and fumarate (Ketelboeter *et al.*, 2014). Therefore, if that pathway is no longer functional, accumulation of mutations in *fahA* could have occurred as a result of spontaneous mutation due to the large deletion.

As no strong causative link could be found for either of the *algG* or *fahA* mutations, it suggests that all eight of the genes that were associated with the PQS phenotype may be false positives.

One association was identified for the BHL-producing phenotype and the same association was also identified for the recovery-APE timepoint. The biological relevancy of this association cannot be

determined, as the SNP association is with the PA2655 hypothetical protein, which has no known or predicted function.

Three associations were identified with the OdDHL-producing phenotype. OdDHL is a quorum-sensing molecule that is responsible for the activation of several secreted virulence-related proteases (e.g *lasA*, *lasB*, *apr* proteases), as well as inducing other quorum sensing systems responsible for further virulence (Lee & Zhang, 2015). Two of the associations are with hypothetical proteins. One of the hypothetical proteins has no similarities with any gene within the *P. aeruginosa* PAO1 reference genome, and has no known function. The PA4193 hypothetical protein is predicted to be a transmembrane transport protein. OdDHL is both a secreted molecule and imported molecule, and if PA4193 was involved in the transport of OdDHL, a non-synonymous mutation in the transport protein could decrease the detection of OdDHL. However no studies have made this biological association, and so functional studies would need to be undertaken to understand this association better.

The final association identified with the OdDHL-producing phenotype was with *pbpG*, which is homologous to the *E. coli* penicillin-binding protein 7 (Song *et al*., 1998). Variation in the penicillin-binding proteins have been generally linked with resistance to β-lactam antibiotics, but *pbpG* modifications do not increase MIC values for the β-lactam antibiotics in *P. aeruginosa* (Ropy *et al.*, 2015). Knock-out of *pbpG* does alter the structure of the outer cell wall (Ropy *et al.*, 2015), but is unlikely to change the secretion or uptake of the OdDHL molecule.

4.3.5 Linear mixed models are effective at controlling for population structure in structured datasets

A second method of population structure control, which addresses some of the problems identified for the subsampling method (Section 4.3.3), is to fit Linear Mixed Models (LMMs). Fitting a LMM requires a correlation matrix of the genetic distances between all of the isolates, equivalent to a PCA matrix. These values are then used to reduce the effects of population structure, by using the correlation matrix as a covariate (Zhou & Stephens, 2012). By fitting a LMM, it is possible to use the genetic information of all 4,094 isolates to uncover associations with the phenotype, substantially increasing the power of the association test compared to the subsampling method.

Initially, the LMM method was applied to each of the subsampled patient groups separately. However, no new associations were uncovered from the LMM results compared to the LM, suggesting that the LMM population structure control had little effect on each patient group (Figure 4.7a). This suggests that each individual patient group population is relatively unstructured. Identical results were reported when LMM were fit to all 13 phenotypes with each subsampled group.

Fitting the LMM to all of the 4,094 isolates effectively controlled for the strong population structure within the dataset (Figure 4.7b), allowing the identification of relevant associations not previously reported due to population structure. Of the 13 phenotypes associated for whole dataset, three yielded results that reached association threshold (Table 4.8); biofilm production, gelatinase production, and BHL production.

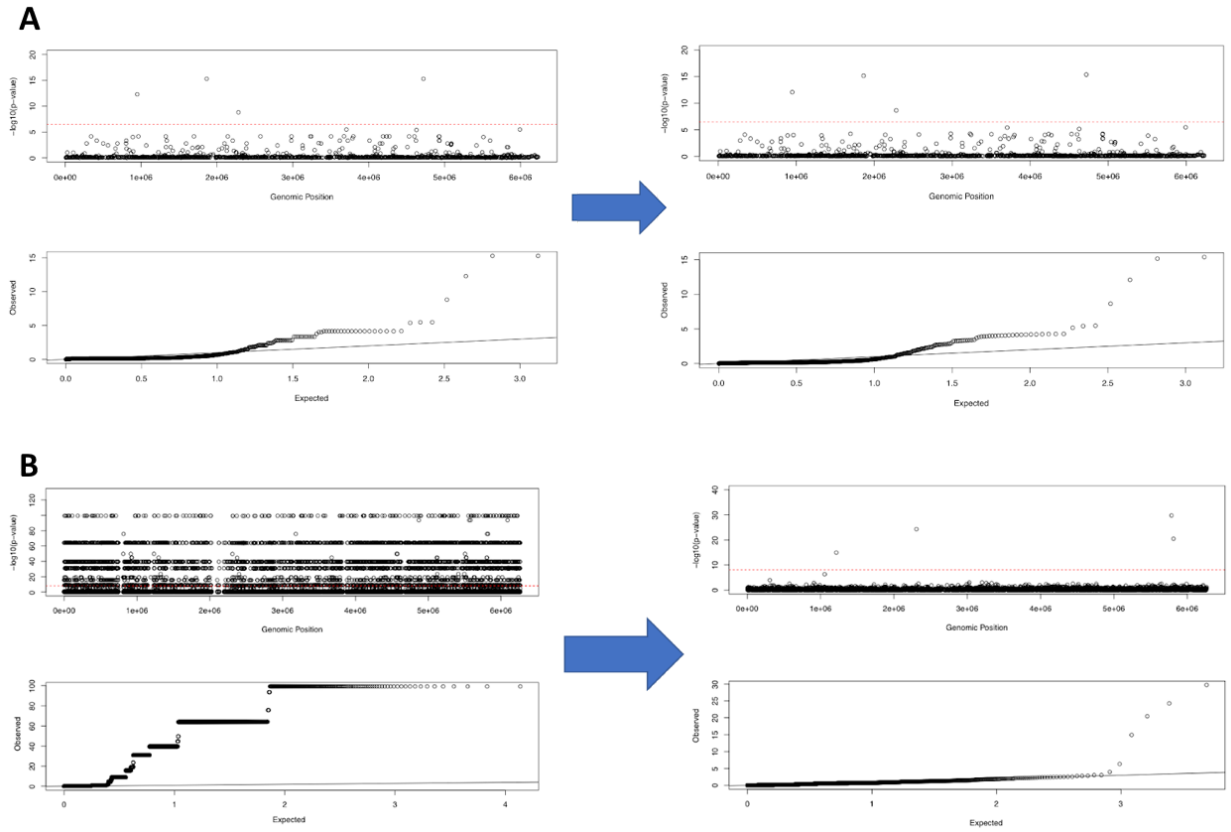


Figure 4.7 a) LM (left) and LMM (right) QQ-plots showed similar results for the subsampled patient groups, and identified the same associated SNPs. b) LMM (right) effectively controlled for population structure in the whole dataset compared to the LM (left) method, and identified several potential associations.

Phenotype	Patient group	Phenotype presence	PA Gene Number	Protein Name	Codon Change	Amino Acid Change	Minor Allele Frequency	Direction of effect on phenotype	*p*-value
Biofilm	All	28.9%	PA5136	Hypothetical Protein	AGT -> GGT	S173G	4.60%	+ve	$2.81x10^{-91}$
Biofilm	All	28.9%	PA5412	Hypothetical Protein	TAC -> GAC	Y192D	4.60%	+ve	$2.81x10^{-91}$
Biofilm	All	28.9%	PA2095	Hypothetical Protein	GAC -> GCC	D248A	4.70%	+ve	$2.48x10^{-89}$
Biofilm	All	28.9%	PA5163	RmlA	CCA -> CAA	P57Q	9.10%	-ve	$9.00x10^{-53}$
Gelatinase	All	64.1%	PA5136	Hypothetical Protein	AGT -> GGT	S173G	4.60%	+ve	$1.74x10^{-09}$
Gelatinase	All	64.1%	PA5412	Hypothetical Protein	TAC -> GAC	Y192D	4.60%	+ve	$1.74x10^{-09}$
Gelatinase	All	64.1%	PA3584	Upstream	C -> G	-	14.3%	-ve	$4.44x10^{-09}$
Gelatinase	All	64.1%	PA2095	Hypothetical Protein	GAC -> GCC	D248A	4.70%	+ve	$1.26x10^{-08}$
Gelatinase	All	64.1%	PA1787 and PA1788	Upstream	T -> G	-	1.00%	-ve	$1.03x10^{-08}$
BHL	All	62.9%	PA5163	RmlA	CCA -> CAA	P57Q	9.10%	-ve	$5.12x10^{-12}$
BHL	All	62.9%	PA5136	Hypothetical Protein	AGT -> GGT	S173G	4.60%	+ve	$1.03x10^{-08}$
BHL	All	62.9%	PA5412	Hypothetical Protein	TAC -> GAC	Y192D	4.60%	+ve	$1.03x10^{-08}$
BHL	All	62.9%	PA2095	Hypothetical Protein	GAC -> GCC	D248A	4.70%	+ve	$1.44x10^{-08}$
BHL	All	62.9%	PA1803	Lon Protease	CCC -> TCC	P32S	4.60%	+ve	$9.46x10^{-08}$

Table 4.8 LMM association tests for each of the 13 phenotypes on the entire dataset mapped against *P. aeruginosa* PAO1. The data which reached significance threshold was filtered for non-synonymous SNPs and for population structure.

Four very strong associations were identified with the biofilm-producing phenotype, with *p*-value results reported from $9.00x10^{-53}$ to $2.81x10^{-91}$. Three of these associations corresponded with mutations in hypothetical proteins; PA5136, PA5412 and PA2095. Mutations in both PA5136 and PA5412 appear with the same MAF and with the same *p*-value. One, or both, of these associations may be a true positive result, and may be causative of the biofilm phenotype. Unfortunately, no function can be predicted for the three significant hypothetical proteins, and therefore functional studies would need to be conducted in order to understand how they may impact biofilm formation.

A negative association was uncovered between biofilm production and a P57Q mutation in RmlA, which has putative biological support. Previous studies by Michael *et al*. have confirmed that transposon knockout of the *rmlA* gene decreases biofilm production by up to 60% (Michael *et al.*, 2016). Therefore, if the P57Q mutation knocks out or knocks down the *rmlA* gene, it would lead to a reduction in biofilm production, which is the predicted direction of the effect. However, this mutation is only present in 9.1% of isolates, and so does not fully explain the 71.1% absence of biofilm production in this study, and would need to be experimentally verified.

Five SNP associations were identified with the gelatinase-producing phenotype. Three of these were identical mutations found in the PA5136, PA5412, and PA2095 hypothetical proteins that were recorded for the biofilm-producing phenotype. This suggests a similar distribution of isolates that produce biofilm also produce gelatinase (Cramer's V = 0.258, suggesting a moderate relationship). The other two results were base-pair changes in intergenic regions. The first is upstream of PA3534 (*glpD*), which is responsible for energy metabolism in *P. aeruginosa* (Shuman *et al.*, 2018). The second is upstream of the divergently transcribed PA1787 (*acnB*) and PA1788 (hypothetical protein), which are essential genes in the TCA cycle of bacteria (Lee *et al.*, 2015). There are no known biological relationships between these genes and the gelatinase-producing phenotype.

Five SNP associations were identified with the BHL-producing phenotype. Three of these were the same hypothetical proteins that were reported for the biofilm and gelatinase-producing phenotype; PA5136, PA5412, and PA2095. The same mutation in *rmlA* that was implicated in biofilm production, was also identified here. Previous studies have shown that *rmlA* is under QS control, but there is no evidence of feedback to QS from the *rml* system (Wagner *et al*., 2003).

A mutation in *lon* protease was also associated with the BHL-producing phenotype. The *lon* protease has been previously implicated in BHL production as a negative regulator of the *rhlR* and *rhlI* genes. Therefore, the P32S mutation may knockout or knockdown the *lon* protease function, leading to an increase in BHL production. However, the P32S mutation is only present in 4.6% of the isolates, and can only explain some of the 62.9% presence of the BHL phenotype, and would need to be verified experimentally.

4.3.6 Aggregating non-synonymous SNPs in the entire dataset

LMM association tests were also run for the SNPs aggregated by gene (see section 4.3.4), in order to address some of the problems identified for the subsampling method (Section 4.3.3) for the associations identified using the aggregated SNPs method. In total, 4,909 genes had one or more non-synonymous SNPs. Using this method, two of the experimental phenotypes yielded results that reached the association threshold (Table 4.9); biofilm production and caseinase production. Additionally, significant associations were identified with the acute-APE timepoint.

Phenotype	Patient group	Phenotype presence	PA Gene Number	Protein Name	Minor Allele Frequency	Direction of effect on phenotype	*p*-value
Biofilm	All	28.9%	PA5136	Hypothetical Protein	45.2%	-ve	$3.94x10^{-24}$
Biofilm	All	28.9%	PA2095	Hypothetical Protein	33.9%	+ve	$9.68x10^{-21}$
Biofilm	All	28.9%	PA5163	RmlA	24.7%	-ve	$2.48x10^{-18}$
Biofilm	All	28.9%	PA1121	YfiR	33.2%	+ve	$5.60x10^{-15}$
Biofilm	All	28.9%	PA0969	TolQ	2.20%	+ve	$2.02x10^{-14}$
Caseinase	All	58.1%	PA5136	Hypothetical Protein	45.2%	-ve	$9.02x10^{-16}$
Caseinase	All	58.1%	PA2095	Hypothetical Protein	33.9%	+ve	$2.84x10^{-13}$
Caseinase	All	58.1%	PA5163	RmlA	24.7%	-ve	$2.55x10^{-12}$
Caseinase	All	58.1%	PA2925	HisM	4.60%	+ve	$1.14x10^{-09}$
Caseinase	All	58.1%	PA1064	Hypothetical Protein	15.6%	-ve	$2.50x10^{-09}$
Caseinase	All	58.1%	PA3189	Hypothetical Protein	31.5%	-ve	$2.70x10^{-09}$
Caseinase	All	58.1%	PA3167	SerC	41.5%	+ve	$1.17x10^{-08}$
Caseinase	All	58.1%	PA3188	Hypothetical Protein	15.5%	-ve	$3.50x10^{-08}$
Acute-APE	All	40.7%	PA5136	Hypothetical Protein	45.2%	-ve	$3.67x10^{-13}$
Acute-APE	All	40.7%	PA2095	Hypothetical Protein	33.9%	+ve	$4.07x10^{-11}$
Acute-APE	All	40.7%	PA5163	RmlA	24.7%	-ve	$6.57x10^{-10}$

Table 4.9 LMM association tests for each of the 13 phenotypes on the entire dataset using aggregated non-synonymous SNPs. Results that reached the significance threshold were filtered for population structure.

Five aggregated SNP associations were identified for the biofilm-producing phenotype. Three of these had previously been reported for the biofilm-producing phenotypes when fitting LMM to all SNPs (section 4.3.5); hypothetical proteins PA5136 and PA2095, and RmlA. These are now reported with increased MAF and a higher *p*-value, indicating the presence of more than one non-synonymous mutation associated with similar causative effects within these genes.

Two additional associations were reported with biofilm production; non-synonymous SNPs in *tolQ* and *yfiR*. The TolQ protein forms part of a complex which maintains outer-membrane integrity (Wei *et al.*, 2009), and expression has been shown to increase under biofilm conditions.

YfiR has previously been implicated in biofilm production. Malone *et al.* showed that a knockout of *yfiR* can have a positive impact on the production of biofilm (Malone *et al*., 2010). YfiR is a repressor of the

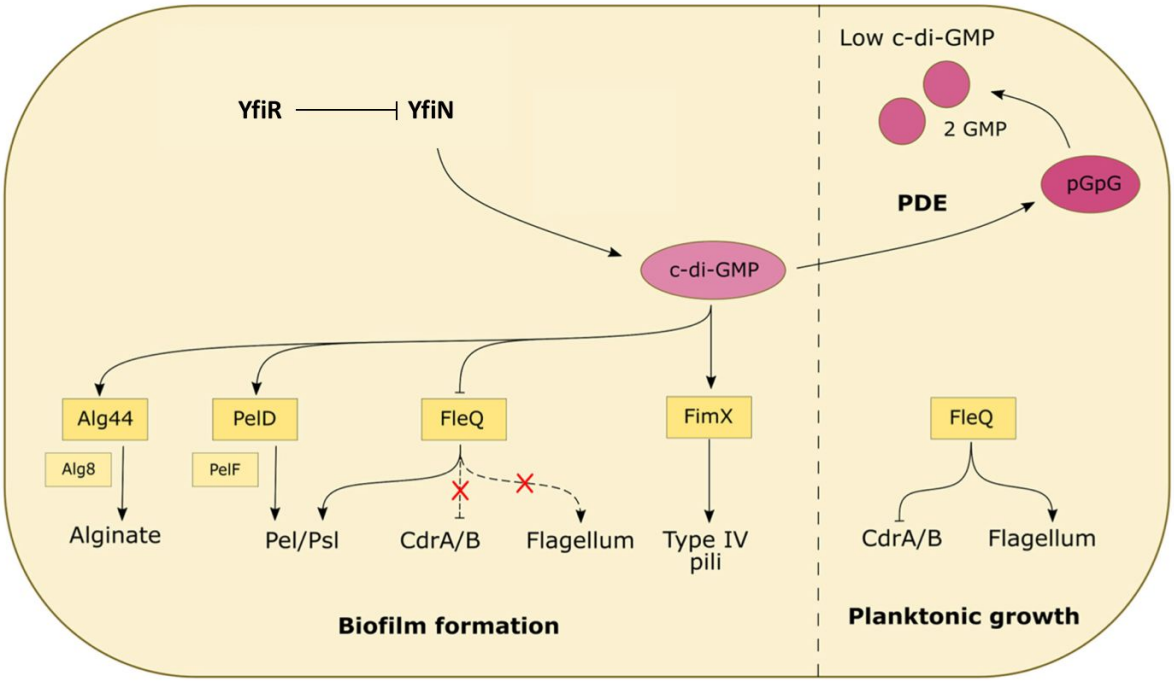


Figure 4.8 A diagram of the cascade that activates biofilm formation. YfiR represses YfiN, preventing biofilm formation. Figure used under licence by creative commons v4.0 from Maunders & Welch, 2017.

YfiN protein, which in turn produces c-di-GMP (Yang *et al.*, 2015). c-di-GMP is an intracellular signalling molecule that activates both the *pel* and *psl* systems, which produce the required secreted molecules to form stable biofilms (Yang *et al.*, 2015). Therefore, a knockout or destabilisation of the YfiR protein will prevent the repression of YfiN, and lead to the constitutive expression of biofilm (Figure 4.8).

YfiR causes repression of YfiN by binding directly to the Per-Arnt-Sim (PAS) domain; a common protein and small molecule binding site found in all kingdoms of life (Moglich *et al*., 2009). Typically, YfiR is released from the YfiN protein, which is located on the inner membrane, by the outer membrane protein YfiB. Crystal studies of YfiB have identified that it is a receptor protein of an as-of-yet undetermined peptidoglycan molecule (Li *et al.*, 2015b). YfiB then binds to YfiR, releasing the YfiN protein to activate the biofilm-producing systems (Yang *et al*., 2015).

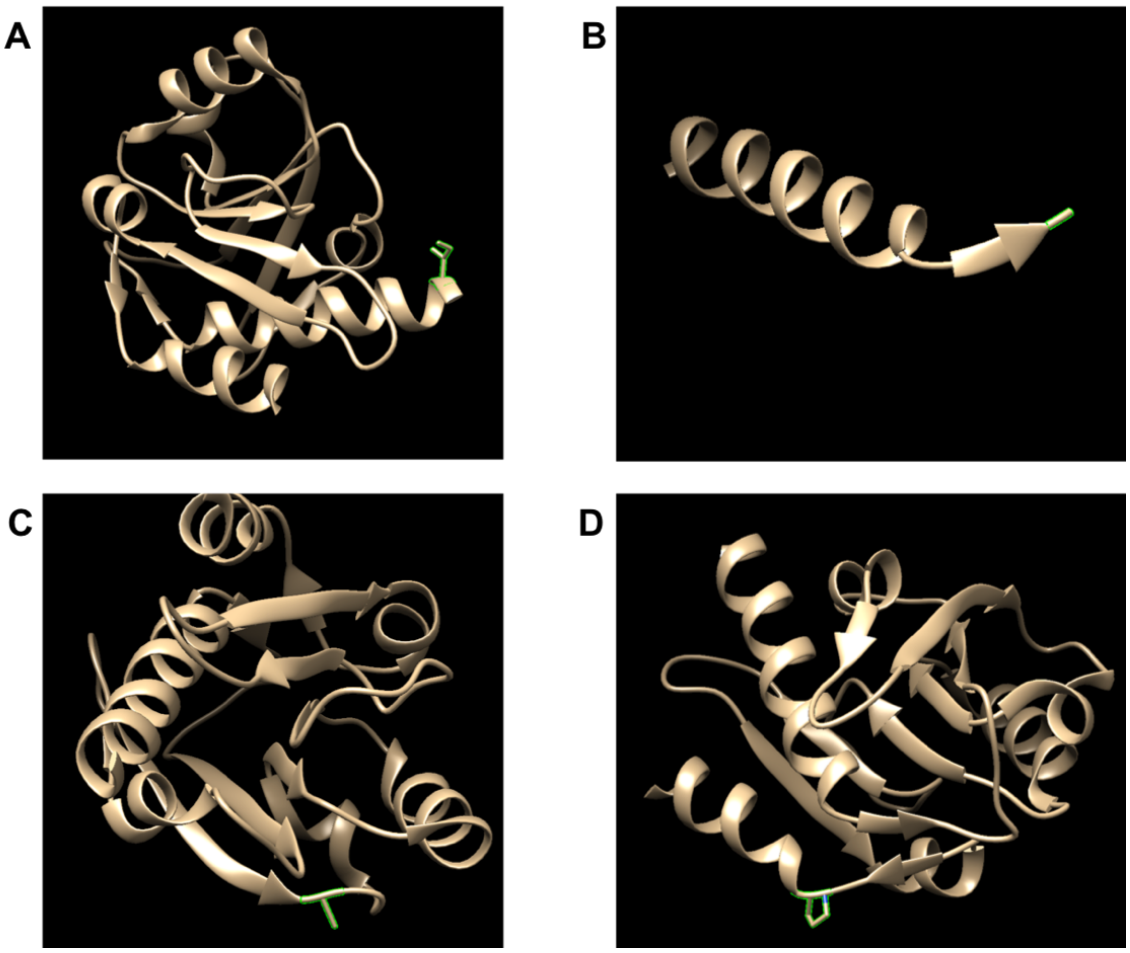


Figure 4.9 The YfiR crystal structure with the four non-synonymous mutations highlighted in green for A) I41V B) W61* C) A100T D) P178L.

Within this dataset, four non-synonymous changes in YfiR were aggregated; I41V, W61*, A100T, and P178L. Figure 4.9 shows the crystal structure (protein data bank: 4ZHY) of YfiR from Yang *et al*., and the mutations identified in this study.

The W61* mutation (Figure 4.9b) introduces an early stop codon, and would result in a natural knockout of the YfiR protein. However, this mutation is only observed in two of the isolates in the dataset. The I41V (Figure 4.9a) and A100T (Figure 4.9c) mutations are present in all of the LES isolates, but no other isolates in this dataset. The structural impact of these mutations is currently unknown. However, neither of the two affected residues interfere with YfiB binding sites, YfiN binding sites, or YfiR dimerisation sites directly, suggesting an indirect impact of the mutations on function. The P178L (Figure 4.9d) mutation is present in twelve isolates. Proline is commonly present at the beginning of alpha-helices, and

is found at the start of the final alpha-helix in YfiR. It does not directly interact with the YfiN binding site, however it is possible that if the final alpha helix was disrupted, the hydrophobic pocket that binds YfiN may also be disrupted.

Eight aggregated SNP associations were identified with the caseinase-producing phenotype. Non-synonymous SNPs in the genes of two hypothetical proteins, PA5136 and PA2095, as well as in RmlA, were reported as significant for both biofilm production and caseinase production. This is not unexpected, as there is a moderately strong association between the caseinase-producing phenotype and biofilm-producing phenotype (Cramer's V = 0.358) (see section 4.3.1.1), but there has been no biological link found between *rmlA* and caseinase production.

Associations were also reported for caseinase production with three other hypothetical proteins; PA1064, PA3189, and PA3188. PA1064 has no known function, and therefore the relevance of this association cannot be determined. PA3188 and PA3189 form two adjoining proteins of an operon responsible for glucose uptake, with no known link to caseinase production (Raneri *et al.*, 2018).

Finally, two further significant associations with caseinase production were reported for HisM and SerC. HisM is a histidine transport protein, and SerC is a protein responsible for the metabolism of several cofactors required for cell survival (Winsor *et al.*, 2016). These are well characterised proteins that have no known links to caseinase production, and therefore they may be false positive results. Without detailed experimental analysis, no biological link with caseinase production can be determined.

Three associations with the acute-APE timepoint were identified for PA5136, PA2095, and RmlA. These were the same associations that were reported for the biofilm-producing and caseinase-producing phenotypes, which suggests that the isolates collected at the acute-APE timepoints may be linked with a change in the biofilm-producing and caseinase-producing phenotypes. An increase in biofilm and caseinase indicates a greater virulence of the strains, however, any link with the acute-APE timepoint would need to be followed up experimentally.

4.3.7 Gene presence-absence in subsampled groups can identify major gene truncations

Phenotype	Patient group	Phenotype presence in patient group	PA Gene Number	Protein Name	Minor Allele Frequency	Direction of effect on phenotype	*p*-value
PQS	K3, K6	45.4%	PA3477	RhlR_2	18.7%	+ve	2.91×10^{-10}
BHL	K7, K9, K11, K15	91.6%	PA4501	OpdD	13.7%	-ve	7.36×10^{-10}
BHL	K7, K9, K11, K15	91.6%	PA4497	DppA2	13.8%	-ve	1.13×10^{-09}

Table 4.10 LM and LMM association tests for each of the 13 phenotypes for the gene presence-absence data for each patient group.

To increase the power of the GWAS to identify structural variants, in contrast with single or aggregate variants, association tests were also run for the presence and absence of genes in the pangenome, which may also be associated with particular phenotypes. LM and LMM were fit for the pangenome of the subsampled patient groups identified in section 4.3.3. Significant associations were reported with the PQS producing and BHL producing phenotypes (Table 4.10).

One significant association was identified with the PQS-producing phenotype and the presence of the PA3477 gene in the isolates from patient group K3 and K6. This gene was annotated as RhlR_2, a second copy of the *rhlR* gene, a QS transcriptional regulator (Medina *et al.*, 2003). Upon closer inspection, RhlR_2 was found to be a truncation of the full RhlR protein (Figure 4.10), where an early stop codon has been introduced in place of glutamine at position 140. A second hypothetical start codon is present at position 158, however, experimental studies would be necessary to determine whether this would be transcribed, and to confirm whether the two independent domains would be functional.

Previous studies have shown that RhlR is linked to PQS production. The BHL QS molecule binds to the RhlR protein, and the RhlR-BHL complex prevents *pqsR* gene expression, inhibiting the PQS quorum sensing system (Brouwer *et al.*, 2014). It is possible that the truncated RhlR_2 may be inactive and cannot successful bind BHL or the *pqsR* regulatory target, leading to an increase in the PQS-producing phenotype. The truncated RhlR_2 gene is only present in 18.7% isolates from patients K3 and K6, and hence cannot fully explain the presence of PQS producing phenotype in this subgroup, which is present in 45.4% of isolates. Follow-up investigations confirmed that only the isolates from patient K3, and not those from patient K6, harbour the truncated RhlR_2. The truncation has been acquired in a distinct sub-lineage of the K3 population (Figure 4.11).

>RhlR COMPLETE SEQUENCE

MRNDGGFLLWWDGLRSEMQPIHDSQGVFAVLEKEVRRLGFDYYAYGVRHTIPFTRPKTEVHGTYPKAWLERY
QMQNYGAVDPAILNGLRSSEMVVWSDSLFDQSRMLWNEARDWGLCVGATLPIRAPNNLLSVLSVARDQQNIS
SFEREEIRLRLRCMIELLTQKLTDLEHPMLMSNPVCLSHREREILQWTADGKSSGEIAIILSISESTVNFHHKNIQK
KFDAPNKTLAAAYAAALGLI*

>RhlR and RhlR_2 COMBINED TRUNCATED SEQUENCE

MRNDGGFLLWWDGLRSEMQPIHDSQGVFAVLEKEVRRLGFDYYAYGVRHTIPFTRPKTEVHGTYPKAWLERY
QMQNYGAVDPAILNGLRSSEMVVWSDSLFDQSRMLWNEARDWGLCVGATLPIRAPNNLLSVLSVARD*QNIS
SFEREEIRLRLRCMIELLTQKLTDLEHPMLMSNPVCLSHREREILQWTADGKSSGEIAIILSISESTVNFHHKNIQK
KFDAPNKTLAAAYAAALGLI*

Figure 4.10 A comparison of the two amino acid sequences for the complete RhlR and the truncated RhlR_2 proteins. The truncated RhlR_2 contains a stop codon in the middle of the sequence (red), prior to a start codon (green), which are separated by 18 amino acids (grey).

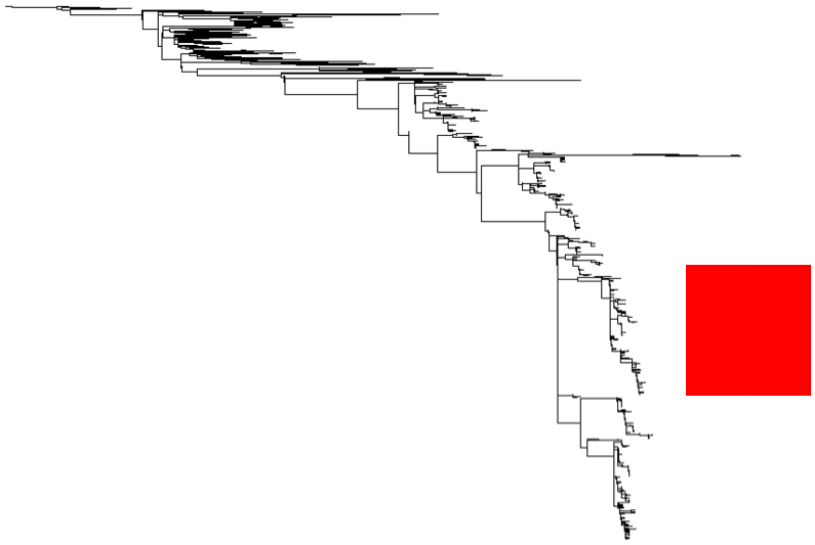


Figure 4.11 The patient K3 phylogeny, with the RhlR truncation indicated next to the phylogeny in red.

This sub-lineage is made up of isolates collected over every time point of each of the 4 exacerbations reported for patient K3. The frequency of the *rhlR* truncation appeared to increase at the acute-APE timepoint for the first three exacerbations, when compared to the stable-APE and recovery-APE timepoints (Figure 4.12). However, the frequency of isolates with the *rhlR* truncation was significantly decreased during the fourth exacerbation, and decreased to zero in the recovery-APE timepoint of the fourth exacerbation. It is possible that this relates to an out-competition of this sub-lineage in subsequent APEs, however additional sequencing data from further APEs would be required to confirm this.

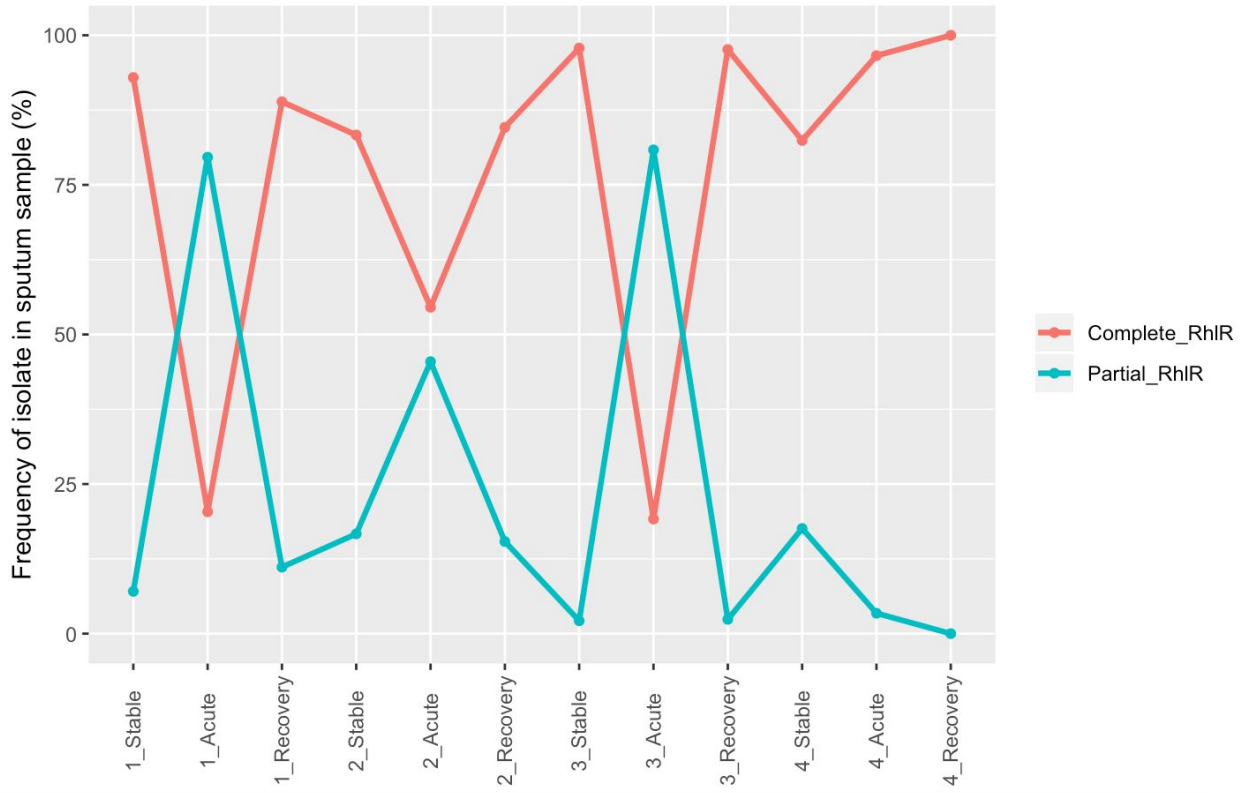


Figure 4.12 A frequency plot showing the presence of the full RhlR protein, compared to the truncated RhlR_2 protein, in isolates collected from patient K3, plotted against the exacerbation timepoints.

Two genes were identified above the significance threshold in isolates from patient group K7, K9, K11 and K15, and were associated with the BHL producing phenotype. The genes were OpdD and DppA2. Both the OpdD porin and the DppA2 protein are involved in dipeptide transport, and knockout mutants of both genes inhibit biofilm formation (Lee *et al.*, 2018). However, no direct associations have been identified between either of these genes and the BHL-producing phenotype.

4.3.7.1 Gene presence-absence for the entire dataset

Association tests were run using LM and LMM for the pangenome of the whole dataset. All results reported for LM were as a result of population structure, and so were uninformative. One LMM association was reported for the PQS-producing phenotype (Table 4.11).

Phenotype	Patient group	Phenotype presence in patient group	PA Gene Number	Protein Name	Minor Allele Frequency	Direction of effect on phenotype	*p*-value
PQS	All	47.8%	PA3477	RhlR_2	5.50%	+ve	3.79×10^{-12}

Table 4.11 LM and LMM association tests for each of the 13 phenotypes on the gene presence-absence data for the whole dataset. Results that reached the significance threshold were filtered for population structure.

The significant result identified using LMM association tests for the gene presence-absence in the whole dataset was the same RhlR_2 protein that was previously reported for the PQS phenotype (section 4.3.7) in isolates from patient K3.

All of the patient groups, apart from patient group K3 and K6, contained isolates that harbour only the complete RhlR protein, and not the truncated version.

4.3.8 Using kmers to identify complex variation

Kmers are lengths of genome that can be used to capture more complex variation than simply SNPs called against a reference, but other variation, such as INDELs, short repeats, and structural rearrangements (Aun *et al*., 2018). To further increase the power of the GWAS to identify these alternative variations that may be associated with particular phenotypes, association tests were also run with kmers.

Due to the large size of the *P. aeruginosa* dataset, counting kmers for all isolates was found to be too computationally intensive for this analysis. However, Jaillard *et al.* identified a way to use compact De Bruijn graphs to compress kmers into unitigs, reducing the computational power required to analyse large datasets (Jaillard *et al*., 2018). In total, 53,000 unitigs were identified in this dataset. LM and LMM were used to identify associations between the unitigs of all 4,094 isolates and each of the 13 phenotypes. Significant unitigs were mapped successfully against the *P. aeruginosa* PAO1 reference genome for annotation.

All LM results were dominated by population structure, as has been seen previously. The LMM association tests only identified two significant associations, one with the biofilm producing phenotype and one with the caseinase producing phenotype. These unitigs each mapped to a single locus within the *P. aeruginosa* genome (Table 4.12).

Phenotype	Patient group	Phenotype presence	PA Gene Number	Protein Name	Minor Allele Frequency	Direction of effect on phenotype	*p*-value
Biofilm	All	28.9%	PA2462	Hypothetical Protein	33.4%	+ve	1.54×10^{-52}
Caseinase	All	58.1%	PA4625	CdrA	32.8%	-ve	6.08×10^{-37}

Table 4.12 Unitig association tests for each of the 13 phenotypes for the whole dataset. Results that reached the significance threshold were filtered for population structure.

An insertion in the region encoding the PA2462 protein, causing a frameshift, was identified as significantly associated with biofilm formation. There is no predicted function associated with this protein, but it is hypothesised to be involved in arginine metabolism and biosynthesis (Winsor *et al*., 2016). It has not previously been linked to biofilm formation.

A single base pair deletion in the *cdrA* gene was identified as significantly associated with caseinase production. This gene is involved in the production of extracellular matrix proteins that bind the *P. aeruginosa* biofilm and prevent proteolysis (Reichhardt *et al.*, 2018). However, this has also never been linked to caseinase production.

4.3.9 Accounting for missing data

In several cases, when determining the experimental phenotypes, some data was not collected. This occurred when the isolates did not grow on a plate, and were subsequently recorded as absent for the phenotype, rather than as no growth. The proportion of missing data is outlined in Table 4.13. Of the ten phenotypes that were investigated experimentally, data was recorded for 100% of isolates for four of the phenotypes; biofilm formation, OdDHL production, BHL production and PQS production. Of the remaining experiments, the highest proportion of missing data was for rhamnolipid production, where no phenotype could be recorded for 1,503 (36.7%) of the isolates. The lowest proportion of missing data was for the twitching motility phenotype, where a phenotype was recorded for 99.4% of isolates.

In order to investigate whether the missing data affected the GWAS results, each LMM association test was re-run for the whole dataset, for the phenotypes with missing data excluded; caseinase production, gelatinase production, rhamnolipid production, siderophore production, swimming motility and twitching

Phenotype	Number of isolates with missing data (%)
Biofilm	0 (0%)
Caseinase	591 (14.4%)
Gelatinase	958 (23.4%)
Rhamnolipid	1,503 (36.7%)
Siderophore	877 (21.4%)
Swimming	1,138 (27.8%)
Twitching	25 (0.6%)
OdDHL	0 (0%)
BHL	0 (0%)
PQS	0 (0%)

Table 4.13 The number of isolates for each phenotype experiment where no phenotype was recorded.

motility. After missing data was removed from these phenotypes, the number of isolates included for each phenotype ranged from 2,591 - 4,069 isolates, which should still have a high enough power to identify any strong associations.

LMM were run against SNPs, aggregated SNPs and gene presence-absence for the whole dataset, to identify any associations with the six phenotypes minus any missing data. No significant results were identified for the gene presence-absence data. This is consistent with section 4.3.7.1, where no significant results were identified for the six phenotypes which were present with missing data still included.

Seven significant results were identified for the caseinase-producing phenotype using LMM against aggregated SNPs (Table 4.14). These genes were the same as those associated with caseinase production in section 4.3.6. Change in *p*-values are due to the altered phenotypic and genotypic proportions in the dataset once missing data was removed. The association between caseinase production and SerC, which was identified in section 4.3.6 as a false positive result, was no longer identified as significant after the missing data was removed.

Finally, LMM were run against SNPs (Table 4.15) for the phenotypes minus missing data. Four significant non-synonymous mutations were identified, and all were associated with both the caseinase

Phenotype	Patient group	Phenotype presence	PA Gene Number	Protein Name	Minor Allele Frequency	Direction of effect on phenotype	*p*-value
Caseinase	All	66.5%	PA5136	Hypothetical Protein	48.7%	-ve	9.02×10^{-24}
Caseinase	All	66.5%	PA2095	Hypothetical Protein	33.0%	+ve	2.84×10^{-21}
Caseinase	All	66.5%	PA5163	RmlA	22.3%	-ve	2.55×10^{-19}
Caseinase	All	66.5%	PA2925	HisM	5.20%	+ve	1.14×10^{-11}
Caseinase	All	66.5%	PA1064	Hypothetical Protein	17.8%	-ve	2.50×10^{-10}
Caseinase	All	66.5%	PA3189	Hypothetical Protein	33.7%	-ve	2.70×10^{-10}
Caseinase	All	66.5%	PA3188	Hypothetical Protein	17.8%	-ve	3.50×10^{-09}

Table 4.14 LMM association tests for each of the 6 phenotypes on the entire dataset using aggregated SNPs non-synonymous SNPs, filtered for the missing data.

producing and siderophore producing phenotypes. All of these were identical to the results of section 4.3.5, with some variation in the *p*-values. A strong association between these two phenotypes was identified in section 4.3.1.1, but after missing data was removed, the Cramer's V statistic increased to 0.581, which suggests a very strong association between these phenotypes.

Overall, no different significant associations were identified when the missing experimental phenotypic data was removed from the analysis.

Phenotype	Patient group	Phenotype presence in patient group	PA Gene Number	Protein Name	Codon Change	Amino Acid Change	Minor Allele Frequency	Direction of effect on phenotype	*p*-value
Caseinase	All	66.5%	PA5163	RmlA	CAA -> CCA	Q57P	9.7%	-ve	1.11×10^{-56}
Caseinase	All	66.5%	PA5136	Hypothetical Protein	AGT -> GGT	S173G	5.3%	+ve	4.34×10^{-50}
Caseinase	All	66.5%	PA5412	Hypothetical Protein	TAC -> GAC	Y193D	5.3%	+ve	4.34×10^{-50}
Caseinase	All	66.5%	PA2095	Hypothetical Protein	GAC -> GCC	D248A	5.3%	+ve	1.12×10^{-48}
Siderophore	All	90.0%	PA5163	RmlA	CAA -> CCA	Q57P	11.5%	-ve	1.11×10^{-14}
Siderophore	All	90.0%	PA5136	Hypothetical Protein	AGT -> GGT	S173G	5.8%	+ve	4.34×10^{-12}
Siderophore	All	90.0%	PA5412	Hypothetical Protein	TAC -> GAC	Y193D	5.8%	+ve	4.34×10^{-12}
Siderophore	All	90.0%	PA2095	Hypothetical Protein	GAC -> GCC	D248A	5.9%	+ve	1.12×10^{-11}

Table 4.15 LMM association tests for each of the 6 phenotypes for SNPs called against a reference, filtered for the missing data. The results that reached the significance threshold were filtered for population structure.

4.3.10 Investigating global transcriptional regulators

Although the GWAS used a wide variety of techniques to identify genotype-phenotype associations, and was able to identify several plausible genotypic associations with the virulence-related phenotypes, there are other methods that may identify further associations that are missed by typical GWAS algorithms. For example, GWAS may not associate multi-variant loci, true significant associations may be hidden by population structure, and GWAS may miss associations that are masked by skewed presence-absence ratios of phenotype. Therefore, some additional known virulence-related genes were manually investigated for their potential association with the ten phenotypes.

A change in pathogenicity commonly occurs with mutations in global transcriptional regulator genes (Marvig *et al*., 2013; Winstanley *et al*., 2016). In this chapter, a change in biofilm has been associated with a premature stop codon in the QS transcriptional regulator *rhlR* (section 4.3.7), and so the sequences of the global transcriptional regulator genes *lasR*, *mucA*, *retS*, *gacS* and *ampR* were extracted from all 4,094 isolates in order to identify any potentially-causative mutations missed by GWAS.

A frameshift-causing insertion within the transcriptional regulator *mucA* was present in 34% of patient K3 isolates and 93% of patient K14 isolates. Only two other mutations were present in *mucA*, both were in patient K4, and both were synonymous mutations. This suggests that all isolates from patients K1, K4, K6, K7, K9, K11, and K15 are not currently overexpressing alginate, which is a frequently-observed phenotype that aids initial chronic adaptation to the CF lung (Feliziani *et al*., 2010). Non-mucoid *P. aeruginosa*, often caused by phenotypic reversion through mutations in alginate biosynthetic pathway, can be isolated alongside mucoid *P. aeruginosa* from CF lungs in late-stage infection (Malhotra *et al*., 2018). The frameshift mutation in patient K3 is only present in isolates close to the root of the phylogeny, and suggests the presence of both mucoid and non-mucoid *P. aeruginosa* in the lungs of patient K3 (Figure 4.13).

The majority of patient K14 isolates that contain the frameshift mutation in *mucA* also harbour a number of mutations in *algU*. Mutating *algU* is a known mechanism by which *P. aeruginosa* reverts back to a non-mucoid phenotype, suggesting that the patient K14 isolates once expressed the mucoid phenotype, but were non-mucoid at the time of sampling (Marvig *et al.*, 2015). However, multiple mutations in

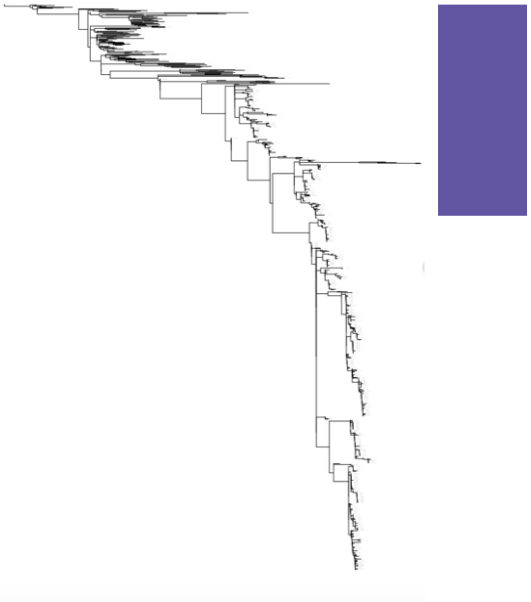


Figure 4.13 A frameshift mutation in *mucA* is only present in the isolates towards the root of the phylogeny (blue), indicating that both mucoid and non-mucoid *P. aeruginosa* exist in patient K3.

hundreds of genes are suspected to change the expression levels of the alginate biosynthetic operon, though mutating *mucA* is by far the most common cause, so a phenotypic assay would be required to confirm the mucoid phenotype of all isolates (Malhotra *et al*., 2018).

One synonymous and one non-synonymous mutation was identified in the QS regulator, *lasR*. The non-synonymous mutation, R217W, was present in 98% of isolates from patient K1, and has been shown to reduce the transcription of *lasR*-regulated genes (Feltner *et al*., 2016). The R217W mutation occurs in the DNA-binding region of *lasR*, suggesting a reduction in transcription is caused by the reduced binding of *lasR* to DNA targets (Feltner *et al*., 2016). A knockdown of *lasR* may be responsible for the low levels of QS and QS-regulated virulence phenotypes observed for the patient K1 isolates.

In total, 26 synonymous mutations were observed in the global regulatory genes *retS* or *gacS*. These are fixed mutations, as they were present in 100% of isolates of the patient the mutations were identified in. Only patient K4 isolates contained non-synonymous mutations in either gene: I339M was found in 99% of patient K4 *retS* sequences; and V440A and A521T was found in 100% and 2.6% of patient K4 *gacS* sequences respectively. Mutations in *retS* complement the activation of the transcriptional switch from acute to chronic infection through the LadS-dependent pathway. However, mutations in *gacS* will

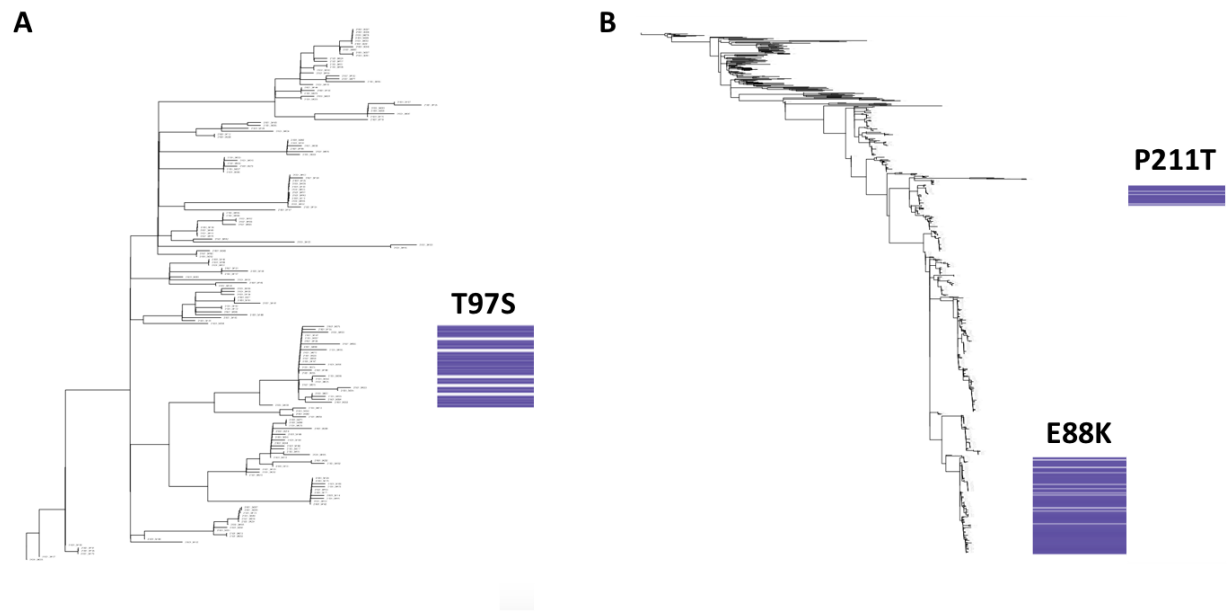


Figure 4.14 Three non-synonymous mutations were identified (blue) in *ampR* in a) patient K6, and b) patient K3.

hypothetically reverse this switch from chronic transcription to the more-virulent acute transcription (Marvig *et al.*, 2015). If the V440A and A512T mutations in *gacS* do revert the isolates from patient K4 to a more virulent state, it may explain the higher prevalence of virulence-related phenotypes observed for these isolates.

Within the global regulator *ampR*, there are two non-synonymous mutations in the isolates from patient K3 (E88K in 16.2% of isolates, and P211T in 3.2% of isolates), and one non-synonymous mutation in the isolates from patient K6 (T97S in 12.6% of isolates). These mutations are associated with specific clonal expansions within each patient phylogeny (Figure 4.16).

The effect of each mutation on the *P. aeruginosa* phenotype depends on the conformational changes they cause to the AmpR protein. For example, the E88K mutations observed in patient K3 has been previously shown to act as a gain-of-function mutation, by locking AmpR in an active conformation (Tueffers *et al*., 2019; Balasubramanian *et al.*, 2015). This would have the main effect of increasing antibiotic resistance through the overexpression of AmpC, which would increase resistance to the β-lactam class of antibiotics, as well as increasing expression of MexXY-OprM efflux pump, which confers resistance to the aminoglycosides (Balasubramanian *et al.*, 2015). AmpR has also been shown to positively regulate the *lasR* QS system, and so locking AmpR in an active conformation would upregulate production of all

QS-regulated virulence factors, such as the T3SS, secreted proteases, and hydrogen cyanide production (Balasubramanian *et al.*, 2015). Finally, an active AmpR has been shown to increase iron uptake, and downregulate chronic phenotypes through direct interaction with RsmZ (Balasubramanian *et al.*, 2015). The T97S and P211T mutations have not yet been characterised, and so could either be a gain-of-function mutation, or a loss-of-function mutation, which would have the opposite effect on the *P. aeruginosa* phenotype. Therefore, any mutation in AmpR can have a significant impact on the *P. aeruginosa* pathogenicity.

4.4 Summary of results and discussion

This study represents the largest study of comparative phenotypic-genotypic correlations of *P. aeruginosa* clonal isolates from the lungs of patients with CF, comparing 4,094 unique genotypes and 13 phenotypes. In this study, a substantial number of GWAS tests were carried out on a large number of isolates, in order to thoroughly examine potential genotype-phenotype correlations. In total, 516 iterations of genome-wide association studies were run to identify any causative genotypic variation for thirteen phenotypes expressed by the 4,094 isolates in this dataset. The results for the five subsampled patient groups are summarised in Table 4.17, and the results for the associations between the whole dataset are summarised in Table 4.18. A list of all associations identified are listed in Table 4.19.

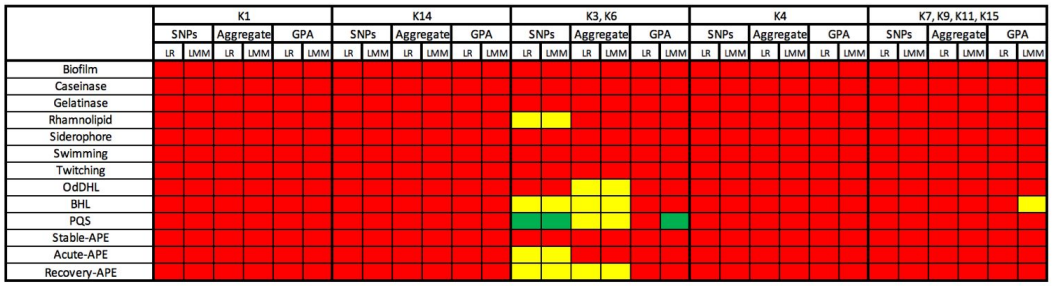


Table 4.16 Summary of all GWAS tests run on the isolates from the five subsampled patient groups. LR = Linear Regression, LMM = Linear Mixed Models, GPA = gene presence absence. Red = No results or results reported due to population structure, yellow = significant results with no known biological link, green = significant results with a biological link.

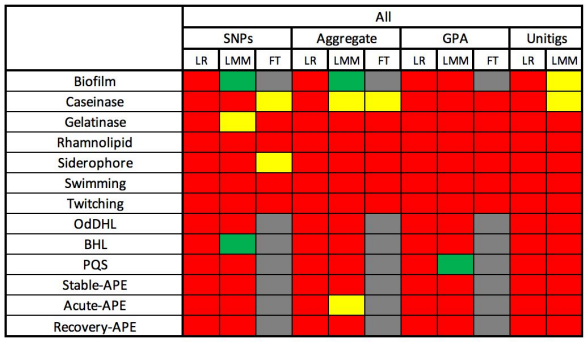


Table 4.17 A table summarising the association tests run for the whole dataset. LR = Linear Regression, LMM = Linear Mixed Models, GPA = gene presence absence. Red = No results or results reported due to population structure, yellow = significant results with no identifiable biological link, green = significant results with identifiable biological link, grey = not run.

Identification Method	Phenotype	Patient Group	PA Gene Number	Protein Name	Direction of Effect	Biologically significant?
LM, Aggregated SNPs, Subsampled Groups	OdDHL	K3, K6	-	Hypothetical Protein	+ve	Unknown
LM, Aggregated SNPs, Subsampled Groups	OdDHL	K3, K6	PA0869	PbpG	+ve	Unlikely
LMM, Aggregated SNPs	Biofilm	All	PA0969	TolQ	+ve	Unlikely
LMM, Aggregated SNPs	Caseinase	All	PA1064	Hypothetical Protein	-ve	Unknown
LMM, Aggregated SNPs	Biofilm	All	PA1121	YfiR	+ve	Likely
LM, Aggregated SNPs, Subsampled Groups	PQS	K3, K6	PA1276	CobC	-ve	Unlikely
LM, Aggregated SNPs, Subsampled Groups	PQS	K3, K6	PA1556	Cco02	+ve	Unlikely
LMM, All SNPs	Gelatinase	All	PA1787/PA1788	Upstream	-ve	Unlikely
LM, Aggregated SNPs, Subsampled Groups	PQS	K3, K6	PA1792	Hypothetical Protein	-ve	Unknown
LM, Aggregated SNPs, Subsampled Groups	PQS	K3, K6	PA1797	Hypothetical Protein	+ve	Unknown
LMM, All SNPs	BHL	All	PA1803	Lon Protease	+ve	Likely
LM, All SNPs, Subsampled Groups	BHL	K3, K6	PA2006	Hypothetical Protein	-ve	Unknown
	Recovery APE					
LM, Aggregated SNPs, Subsampled Groups	PQS	K3, K6	PA2008	FahA	+ve	Unlikely
LMM, All SNPs	Biofilm	All	PA2095	Hypothetical Protein	+ve	Unknown
	Gelatinase					
	BHL					
LMM, Aggregated SNPs	Biofilm					
	Caseinase					
	Acute-APE					
LM, All SNPs, Subsampled Groups	PQS	K3, K6	PA2194	HcnB	+ve	Unlikely
	Acute APE					Possible
Unitigs	Bio	All	PA2462	Hypothetical Protein	+ve	Unknown
LM, All SNPs, Subsampled Groups	Rhamnolipid	K3, K6	PA2654	Hypothetical Protein	-ve	Unknown
LM, Aggregated SNPs, Subsampled Groups	BHL	K3, K6	PA2655	Hypothetical Protein	-ve	Unknown
	Recovery-APE					
LMM, Aggregated SNPs	Caseinase	All	PA2925	HisM	+ve	Unlikely
LMM, Aggregated SNPs	Caseinase	All	PA3167	SerC	+ve	Unlikely
LMM, Aggregated SNPs	Caseinase	All	PA3188	Hypothetical Protein	-ve	Unknown
LMM, Aggregated SNPs	Caseinase	All	PA3189	Hypothetical Protein	-ve	Unknown
Gene Presence-Absence, Subsampled Groups	PQS	K3, K6	PA3477	RhlR_2	+ve	Likely
Gene Presence-Absence		All				
LM, Aggregated SNPs, Subsampled Groups	PQS	K3, K6	PA3545	AlgG	+ve	Unlikely
LMM, All SNPs	Gelatinase	All	PA3584	Upstream	-ve	Unlikely
LM, Aggregated SNPs, Subsampled Groups	OdDHL	K3, K6	PA4139	Hypothetical Protein	+ve	Unknown
LM, Aggregated SNPs, Subsampled Groups	PQS	K3, K6	PA4385	GroEL	+ve	Unlikely
LM, All SNPs, Subsampled Groups	Rhamnolipid	K3, K6	PA4389	SpeA	-ve	Unlikely
Gene Presence-Absence, Subsampled Groups	BHL	K7, K9, K11, K15	PA4497	DppA2	-ve	Unlikely
Gene Presence-Absence, Subsampled Groups	BHL	K7, K9, K11, K15	PA4501	OpdD	-ve	Unlikely
Unitigs	Caseinase	All	PA4625	CdrA	-ve	Unknown
LM, Aggregated SNPs, Subsampled Groups	PQS	K3, K6	PA4835	Hypothetical Protein	+ve	Unknown
LMM, All SNPs	Biofilm	All	PA5136	Hypothetical Protein	+ve	Unknown
	Gelatinase					
	BHL					
LMM, Aggregated SNPs	Biofilm				-ve	
	Caseinase					
	Acute-APE					
LMM, All SNPs	Biofilm	All	PA5163	RmlA	-ve	Possible
	BHL					Unlikely
LMM, Aggregated SNPs	Biofilm					Possible
	Caseinase					Unlikely
	Acute-APE					
LM, All SNPs, Subsampled Groups	PQS	K3, K6	PA5344	OxyR	+ve	Likely
LMM, All SNPs	Biofilm	All	PA5412	Hypothetical Protein	+ve	Unknown
	Gelatinase					
	BHL					

Table 4.18 A list of all of associations identified in this study.

Four different association tests identified biologically plausible genotypic associations; with biofilm production, BHL production, and PQS production. Fitting LMM to aggregated SNPs for the whole dataset identified mutations in *yfiR* that could plausibly cause a change in biofilm production, through an

induction of the c-di-GMP cascade. Fitting LMM to all SNPs for the whole dataset identified a mutation in the *lon* protease that could potentially cause a change in BHL production, through interaction with the *rhl* quorum sensing system. In isolates from patient group K3 and K6, an increase in PQS production was associated with a gene truncation in *rhlR*, which was identified by fitting LMM to gene presence-absence associations for the whole dataset and for the K3 and K6 subsampled group. Additionally, mutations in *oxyR* were plausibly associated with a change in PQS production, through transcriptional interactions with the *pqs* quorum sensing system, for the K3 and K6 subsampled group, which was identified by fitting LM and LMM to all SNPs for this patient subgroup.

As biologically relevant results have been identified using these methods, it shows that there is enough power in this dataset to uncover true positive associations with complex phenotypes. Therefore, associations identified with no clear biological connection, for example the 15 uncharacterised hypothetical proteins, could also be true associations. *P. aeruginosa* molecular biology is far from fully understood, and these uncharacterised mutations may cause changes in these complex phenotypes through mechanisms that are not yet understood. Gene knockout studies would quickly identify whether these uncharacterised genes have the predicted effect on the phenotypes under standard lab conditions.

However, based on current understanding of *P. aeruginosa* biology, 17 of the genes identified are unlikely to be true positive associations. This is because these genes are involved in well-studied pathways that are unrelated to the phenotypes they have been associated with. However, further experimental studies should still be carried out, and as our understanding of *P. aeruginosa* biology improves, these associations may yet yield some biological relevance.

Mutations within six genes were associated with APE timepoints, four for acute-APE and two for recovery-APE. This suggests that there may be genetic associations with APE timepoints. The associations identified with acute-APE timepoints were also significant for many of the virulence related phenotypes; biofilm formation, caseinase production, BHL production, gelatinase production and PQS production. This is evidence that many virulence-related phenotypes may be significantly affected during acute APE, as well as the genotype, and may be one of the triggers or consequences of these exacerbations.

There are several other examples where multiple phenotypes were associated with the same genetic variant. This may occur because there is a correlation between the phenotypes, leading to a similar

distribution of these phenotypes among the isolates. For example, PA2095 was associated with both biofilm and caseinase production. These two phenotypes have a high Cramer's V value, indicating that there is a moderately strong correlation between the phenotypes in the dataset. Other justifications for these multiple-phenotype single-variant associations include similar induction pathways that are related to the genetic variant, the associations may be artifacts of the data due to a small sample size, or both the phenotypes and genetic variants have been co-selected.

The majority of the association studies did not manage to identify any significant genetic variation associated with phenotypic variation. In these cases, it is likely that there are multiple genetic variants which cause the change in phenotype, and there is not enough power in this dataset to identify these variants. Aggregating mutations in genes by known pathway, or by transcriptome response may also increase power to identify variants that influence the expression of the tested phenotype.

This study shows that it is important to use different association techniques to measure the different types of variation within DNA sequences; individual SNPs, aggregated SNPs, kmers (unitigs), and gene presence/absence. In this study, significant biologically-relevant associations were identified using most of the different techniques. Although in this study, using unitigs was not able to uncover any biologically relevant associations, this method can be used to identify short INDELs, as well as repeat regions and structural variation, which may be relevant to other datasets.

It is also noted that there was very little congruence in the associations identified using the different techniques. For example, the stop codon causing the gene truncation in *rhlR* identified in the gene presence-absence study, was not identified in any of the SNP association studies. For the LM and LMM SNP association tests, the SNP that results in the stop codon did not reach the significance threshold, suggesting that other true positive results could be hidden after *p*-value corrections for multiple testing has been applied. Wilson recently identified one method that may be used to overcome this limitation, by applying the harmonic mean *p*-value, which uses a sliding window to combine dependent *p*-values to reduce the burden of significance correction to identify causative mutations (Wilson, 2019).

In all cases, there was congruence between the associations identified using LM and LMM for the unstructured patient groups. In some cases, there was also congruence between the studies using SNPs and those using aggregated SNPs. For example, mutations in *rmlA*, *PA5136,* and *PA2095,* were significant using both SNPs and aggregated SNPs methods. This suggests that there were multiple

mutations within these genes that have the same direction of causative effect on the phenotypes they are associated with. However, in other cases such as the *lon* protease, significant mutations were detected using SNPs techniques, but were not identified using the aggregated SNPs techniques. This suggests that there are multiple non-synonymous mutations in the *lon* protease that do not all cause the same effect, which aren't accounted for when aggregating SNPs together.

The use of simple LM on the structured dataset in this study was uninformative. LM was only able to identify SNPs associated with the strong population structure present within these isolates. This was overcome by using LMM, which allowed the dataset to be analysed as a whole to increase power to identify true positive associations. Both LM and LMM were used to analyse the subsampled patient groups, however the significant *p*-values reported using both models were identical, which suggests that the population structure matrix applied in the LMM did not reduce the ability of the test to identify mutations in unstructured datasets.

One strength of this study was the large sample size, of 4,094 isolates. However, by using subsampling of patient groups into BAPS clusters, associations that are specific for each patient group could be identified. For example, associations with *oxyR*, which were found to have biological relevance, were only significant in the subsampled population group for isolates from patients K3 and K6. Even when identifying associations within the whole dataset, instances where patient specific associations were identified still occurred. The truncation in *rhlR* was identified using LMM and gene presence-absence for the whole dataset, but was only present in the isolates from patient K3 isolates. In order to identify common mutations across different patients that have the same causative effect on the phenotype, we would need a much greater number of patients included in the study. Currently, the effective patient sample size is only nine patients, and therefore the conclusions of this study may not be representative of the whole *P. aeruginosa* CF population. However, the phenotypes investigated in this study are complex, and countless causative mutations in several pathways can impact expression of each phenotype. Therefore, a more diverse dataset may allow the identification of many more mutations in many different pathways that could impact these phenotypes.

In conclusion, we have identified biologically relevant associations for complex phenotypes, and found further associations that require experimental validation. We have shown that LMM can effectively control for population structure in highly structured datasets, and that it is necessary to use multiple different GWAS approaches to capture different types of genetic variation.

Chapter 5

Conclusions

5.1 Restatement of the research aims

P. aeruginosa are metabolically and genetically diverse Gram-negative bacteria that are able to colonise a wide-range of environmental niches. Up to 80% of patients with cystic fibrosis obtain chronic *P. aeruginosa* lung infections that remain within the patient for the rest of their life, and acquisition of chronic *P. aeruginosa* infection is associated with increased lung inflammation and with greater morbidity and mortality. Patients with CF undergo periods of sudden and rapid worsening of symptoms called acute pulmonary exacerbations, which are suspected to be caused by bacteria. Chronic *P. aeruginosa* infections have been associated with an increase in the frequency of APEs and the consequential hospitalisations.

This dissertation forms part of a wider study where 15 patients with CF were asked to undertake home-based measurement of their health over the course of six months, as well as provide daily sputum samples. Sputum samples were selected from nine patients who experienced a total of 18 APEs between them, and 96 *P. aeruginosa* isolates were collected from each sputum sample at the acute timepoint, stable timepoint, and recovery timepoint of each APE that each patient experienced. In total 4,094 *P. aeruginosa* isolates were whole-genome sequenced, and each isolate was screened for 10 virulence-related phenotypes.

In this dissertation, these 4,094 whole-genome sequences were analysed to quantify the inter- and intra-patient diversity of the *P. aeruginosa* communities, and investigate the evolution of each community when facing the pressures of the CF lung. The whole-genome sequences of these 4,094 isolates also offered a unique insight into the changes that occur in the bacterial population over the course of an APE.

In addition, GWAS methods were employed to determine the genetic basis for any phenotypic diversity that was observed within this dataset.

Finally, this dissertation investigated the population structure and genetic diversity of 353 *P. aeruginosa* isolates from two UK-based bloodstream infection surveys. The antimicrobial resistance profiles and whole-genome sequences of all isolates were determined, and phenotypic antimicrobial resistance was associated with genotypic markers.

5.2 Key findings and future directions

5.2.1 Bacteraemia-associated *P. aeruginosa* surveillance

This chapter has provided the first insight into the population structure of bacteraemia-causing *P. aeruginosa* isolates (*n* = 353) from across the UK. These were found to represent the well-characterised diverse *P. aeruginosa* population structure, punctuated with sporadic emergence of highly-successful epidemic clones. Two of these highly-successful epidemic clones were present in the local bacteraemia surveillance dataset; ST395 and ST253. These two international and high-risk clones are typically associated with MDR and XDR bloodstream infections around the globe. However, no evidence of MDR was present within the local ST253 isolates or the majority of the ST395 isolates. However, two subclades of the ST395 isolates were associated with rapid acquisition of multiple AMR phenotypes, most likely caused by overexpression of non-specific efflux pumps. A third epidemic clone was identified within the BSAC surveillance collection; the ST175 clade. The ST175 isolates were found to be distributed across the UK and Ireland, and were associated with MDR and XDR infections. On a global scale, ST175 infection is associated with the presence of MBLs and ESBLs, causing extensive resistance to all classes of β-lactam antibiotics. However, no evidence of MBLs or ESBLs were identified in the whole-genome sequences of the ST175 isolates in this study. Strong temporal signal was present within the ST175 phylogeny, which indicated the emergence of the ST175 epidemic clone in the UK sometime between the mid-1980s and mid-1990s. All three of the high-risk and international STs were found to have a high number of genomic islands compared to the rest of the isolates in the dataset. Most of the acquired genes were related to iron uptake, phage defence, Type VI secretion and motility systems.

Within the local collection, very little antimicrobial resistance was identified for the aminoglycoside antibiotics tobramycin, gentamicin and amikacin, and no resistance was identified for the drug-of-last resort, colistin. Resistance rates were highest for the ticarcillin-clavulanate combination therapy, where almost 50% of isolates were resistant. Widespread MDR was observed for isolates collected as part of the BSAC surveillance programme. Only one isolate showed resistance to colistin, indicating that colistin is still a viable therapeutic option to treat MDR infections. Resistance to the aminoglycoside amikacin was also infrequent, suggesting that resistance determinants to this antibiotic may be more difficult for *P. aeruginosa* to obtain, and should be considered a last-resort treatment for MDR infections alongside colistin.

Given the prevalence of high-risk and international clones within the dataset, continued surveillance of bacteraemia-associated *P. aeruginosa* infection will improve understanding of their distribution, whilst also monitoring the spread of AMR. From an increase in surveillance, possible methods of transmission could be identified, and future transmission of other high-risk clones could potentially be prevented. Future work should try to identify potential reservoirs within hospitals that may contribute to frequent reinfection by these high-risk and international clones.

Further work is also required to elucidate the mechanism of AMR for the isolates which contain no known causative gene or gene variant. Transcriptomic approaches could help to understand AMR-related gene expression, such as overexpression of multi-drug efflux through currently unknown mechanisms. Increasing our understanding of AMR resistance mechanisms will provide new targets for antimicrobial compound development.

5.2.2 Diversity of exacerbation-related *P. aeruginosa*

The second chapter analysed the whole-genome sequences of 4,094 *P. aeruginosa* isolates from nine patients with CF. Each patient was found to be infected with their own individual and distinct strain of *P. aeruginosa*, but different patients were infected with the same ST. This chapter also found evidence of divergent, co-existing lineages of *P. aeruginosa* within patients K1, K4, K11 and K14, which suggests that separate populations of isolates inhabit separate niches within the lung.

This chapter attempted to identify hypermutation within the *P. aeruginosa* isolates, without the need for a phenotypic assay. The presence of large root-to-tip distances and large transition/transversion ratios within the phylogeny indicated that one lineage from patient K14 diverged from the other due to hypermutation. No hypermutators in other patients could be identified through outlier root-to-tip distances. Over 94% of the isolates in patients K1, K3 and K4 had transition/transversion ratios greater than 3:1, which indicates that the isolates are either currently hypermutating, or are descended from a hypermutator strain. Ancestral node-to-node transition/transversion ratios for patient K1, K3, and K4 isolates were also elevated, but no obvious loss-of-function mutations were identified in the mutome, suggesting that these isolates are descended from hypermutators that have reverted to wild-type.

Ancestral hypermutation may also justify the Bayesian time to MRCA estimates for the groups of isolates that showed temporal signal. For example, Bayesian dating suggested that patient K1 isolates were mutating at a rate 5-18x higher than wild-type, and that isolates from patient K3 were mutating at a rate 2-8x higher than wild-type. Bayesian dating suggested that the time to the MRCA isolate of patient K7 and K15 was 0.54 to 1.12 years prior to the study, however, direct transmission between the two patients was ruled out. Transmission of *P. aeruginosa* ST3307 was suspected between patients K3 and K6, as the patient K6 isolates were nested within the patient K3 isolates phylogeny. Bayesian dating indicated a time to MRCA of the patient K6 isolates and the closely-related patient K3 isolates as 0.71 to 1.74 years prior to the study.

The phylogeny of the patient K3 isolates provided an insight into *P. aeruginosa* evolution over the course of infection. Gene presence-absence of the patient K3 isolates indicated large areas of both gene acquisition and gene loss that are associated with the switch from acute to chronic infection. The acquired T6SS genes in the patient K3 isolates shared over 99% identity with the *Pseudomonas aeruginosa* LESB58 reference genome. Regions of recombination were also identified within the patient K3 isolates that may have originated from the LESB58 genome. It is therefore possible that the ST3307 isolates co-existed with a LES clone in the lungs of patient K3, before being outcompeted.

This study found that the diversity of the isolates within each patient was unchanging over the course of the three APE timepoints, and also across different APEs experienced by the same patient. However, for the patient K1 and K4 isolates, there was evidence of non-random phylogenetic clustering of isolates belonging to a particular APE that each patient underwent. This suggests that the lineage architecture within the lungs of these patients is potentially altered during the time in-between exacerbations. However, for all patients, no clustering was observed for isolates belonging to a specific timepoint across each APE, suggesting that the *P. aeruginosa* population structure is not affected over the course of each APE by physiological changes and treatment.

The AMR profiles of all 4,094 isolates in the study were predicted from their genomes. The results showed that the predicted AMR profile of the isolates within each patient did not vary significantly, and no genetic cause for overexpression of efflux pumps could be identified. There was very little variation in the presence/absence of AMR genes.

Further work on this dataset should initially focus on understanding the genetic basis for the non-random phylogenetic clustering of patient K1 and K4 isolates associated with a particular APE, and then characterising any selective advantage these mutations might afford. Secondly, there is evidence that ST3307 has co-existed with, and subsequently outcompeted, a Liverpool Epidemic Strain whilst in the lungs of patient K3. Epidemic strains are highly virulent and adapted to the CF lung, and therefore further work should investigate how ST3307 outcompeted LES. Finally, very little diversity was observed within the isolates from patient K15, providing a unique insight into the initial adaptation and evolution of *P. aeruginosa* isolates within the CF lung. This offers a rare opportunity to follow the development of a newly-acquired chronic infection, and future work should focus on sequencing follow-up isolates to identify how early *P. aeruginosa* chronic infections adapt and evolve within the lungs.

5.2.3 GWAS of complex *P. aeruginosa* phenotypes

This chapter associated genotype with ten virulence related phenotypes, plus three APE timepoints, for the 4,094 *P. aeruginosa* isolates in the CF dataset. This is the first GWAS to associate *P. aeruginosa* genotype with complex phenotypes. In total, 516 iterations of GWAS were run to identify any causative genotypic variation for the thirteen phenotypes exhibited by the 4,094 isolates in this dataset.

Four different association studies identified biologically plausible genotypic associations with; biofilm production, BHL production, and PQS production. Change in biofilm production was associated with mutations in the YfiR protein, which induces the c-di-GMP cascade and results in upregulated biofilm-associated pathways. Change in BHL production was associated with mutations in the *lon* protease, which is known to interact with the *rhl* quorum sensing system. A change in PQS production was associated with a gene truncation in *rhlR*, which interacts with the *pqs* quorum sensing system. Additionally, mutations in *oxyR* were plausibly associated with a change in PQS production, through transcriptional interactions also with the *pqs* quorum sensing system. Mutations within six genes were also associated with APE timepoints, four for acute-APE and two for recovery-APE.

Associations with an additional 15 uncharacterised hypothetical proteins were also uncovered. As biologically relevant results have been identified using these methods, it is possible the 15 uncharacterised hypothetical proteins could also be true associations. *P. aeruginosa* molecular biology is far from fully

understood, and any uncharacterised mutations may cause changes in phenotypes through unknown mechanisms.

This study highlighted the importance of using different association techniques to capture the different types of variation within DNA sequences; individual SNPs, aggregated SNPs, kmers, and gene presence/absence. Additionally, the use of simple LM on the structured dataset in this study was uninformative, due to strong population structure. This limitation was overcome by employing LMM, which successfully controlled for population structure in the highly-structured dataset.

Future work should focus on confirming the associations identified in this chapter. Lab experiments would be required to prove whether the associations identified are true positive results. Secondly, lab experiments should characterise whether the hypothetical proteins are associated with a change in phenotype and through which mechanisms.

Finally, using GWAS to understand genotypic and phenotypic changes that occur frequently in the *P. aeruginosa* CF population would require a more diverse sample of isolates. Therefore, increasing the number of patients recruited in future studies could identify more frequently-occurring *P. aeruginosa* mutations in the CF lung. An improved understanding of common mutations within the CF lung could provide new targets for therapeutics to prevent chronic infection.

5.3 Concluding remarks

This dissertation has provided evidence that high-risk epidemic *P. aeruginosa* strains are present in the UK and Ireland bacteraemia surveillance programmes on both a local and national level. However, in this dataset, these high-risk and epidemic strains are not associated with the presence of MBLs and EMBLs that are widely reported in the literature. Phenotypic and genotypic antimicrobial susceptibility testing demonstrated that colistin and the aminoglycoside antibiotics are still viable therapeutic options to treat MDR bacteraemia infections.

This dissertation has also applied whole-genome sequencing to improve current understanding of the diversity of *P. aeruginosa* infection in the CF lung. Changes in lineage architecture associated with some APEs have been newly identified.

Finally, this dissertation has identified biologically relevant associations for complex phenotypes in *P. aeruginosa*, which will be validated experimentally. LMM have been shown to effectively control for population structure in highly structured datasets, and multiple different GWAS approaches are required to capture different types of genetic variation that may be associated with a phenotype.

6 References

Abdel-Mawgoud, A.M., Lépine, F., Déziel, E., 2010. Rhamnolipids: diversity of structures, microbial origins and roles. Applied Microbiology and Biotechnology 86, 1323–1336. https://doi.org/10.1007/s00253-010-2498-2

Aendekerk, S., Ghysels, B., Cornelis, P., Baysse, C., 2002. Characterization of a new efflux pump, MexGHI-OpmD, from Pseudomonas aeruginosa that confers resistance to vanadium. Microbiology 148, 2371–2381. https://doi.org/10.1099/00221287-148-8-2371

Al-Aloul, M., 2004. Increased morbidity associated with chronic infection by an epidemic Pseudomonas aeruginosa strain in CF patients. Thorax 59, 334–336. https://doi.org/10.1136/thx.2003.014258

Aldred, K.J., Kerns, R.J., Osheroff, N., 2014. Mechanism of Quinolone Action and Resistance. Biochemistry 53, 1565–1574. https://doi.org/10.1021/bi5000564

Andersen, S.B., Marvig, R.L., Molin, S., Krogh Johansen, H., Griffin, A.S., 2015. Long-term social dynamics drive loss of function in pathogenic bacteria. Proceedings of the National Academy of Sciences 112, 10756–10761. https://doi.org/10.1073/pnas.1508324112

Antunes, L.C.M., Ferreira, R.B.R., Buckner, M.M.C., Finlay, B.B., 2010. Quorum sensing in bacterial virulence. Microbiology 156, 2271–2282. https://doi.org/10.1099/mic.0.038794-0

Armstrong, E.S., Miller, G.H., 2010. Combating evolution with intelligent design: the neoglycoside ACHN-490. Current Opinion in Microbiology 13, 565–573. https://doi.org/10.1016/j.mib.2010.09.004

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., Wishart, D.S., 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Research 44, W16–W21. https://doi.org/10.1093/nar/gkw387

Ashish, A., Shaw, M., Winstanley, C., Ledson, M.J., Walshaw, M.J., 2012. Increasing resistance of the Liverpool Epidemic Strain (LES) of Pseudomonas aeruginosa (Psa) to antibiotics in cystic fibrosis (CF)—A cause for concern? Journal of Cystic Fibrosis 11, 173–179. https://doi.org/10.1016/j.jcf.2011.11.004

Aun, E., Brauer, A., Kisand, V., Tenson, T., Remm, M., 2018. A k-mer-based method for the identification of phenotype-associated genomic biomarkers and predicting phenotypes of sequenced bacteria. PLOS Computational Biology 14, e1006434. https://doi.org/10.1371/journal.pcbi.1006434

Bahemia, I.A., Muganza, A., Moore, R., Sahid, F., Menezes, C.N., 2015. Microbiology and antibiotic resistance in severe burns patients: A 5 year review in an adult burns unit. Burns 41, 1536–1542. https://doi.org/10.1016/j.burns.2015.05.007

Balasubramanian, D., Kumari, H., Mathee, K., 2014. Pseudomonas aeruginosa AmpR: an acute-chronic switch regulator. Pathogens and Disease. https://doi.org/10.1111/2049-632x.12208

Balasubramanian, D., Schneper, L., Kumari, H., Mathee, K., 2012. A dynamic and intricate regulatory network determines Pseudomonas aeruginosa virulence. Nucleic Acids Research 41, 1–20. https://doi.org/10.1093/nar/gks1039

Bals, R., Hiemstra, P.S., 2004. Innate immunity in the lung: how epithelial cells fight against respiratory pathogens. European Respiratory Journal 23, 327–333. https://doi.org/10.1183/09031936.03.00098803

Barr, I., Latham, J.A., Iavarone, A.T., Chantarojsiri, T., Hwang, J.D., Klinman, J.P., 2016. Demonstration That the RadicalS-Adenosylmethionine (SAM) Enzyme PqqE Catalyzes de Novo Carbon-Carbon Cross-linking within a Peptide Substrate PqqA in the Presence of the Peptide Chaperone PqqD. Journal of Biological Chemistry 291, 8877–8884. https://doi.org/10.1074/jbc.c115.699918

Barry, P.J., Donaldson, A.L., Jones, A.M., 2018. Ivacaftor for cystic fibrosis. BMJ k1783. https://doi.org/10.1136/bmj.k1783

Bassetti, M., Vena, A., Croxatto, A., Righi, E., Guery, B., 2018. How to manage Pseudomonas aeruginosa infections. Drugs in Context 7, 1–18. https://doi.org/10.7573/dic.212527

Battle, S.E., Rello, J., Hauser, A.R., 2009. Genomic islands ofPseudomonas aeruginosa. FEMS Microbiology Letters 290, 70–78. https://doi.org/10.1111/j.1574-6968.2008.01406.x

Beaman, T.W., Sugantino, M., Roderick, S.L., 1998. Structure of the Hexapeptide Xenobiotic Acetyltransferase from Pseudomonas aeruginosa. Biochemistry 37, 6689–6696. https://doi.org/10.1021/bi980106v

Beare, P.A., For, R.J., Martin, L.W., Lamont, I.L., 2002. Siderophore-mediated cell signalling in Pseudomonas aeruginosa: divergent pathways regulate virulence factor production and siderophore receptor synthesis. Molecular Microbiology 47, 195–207. https://doi.org/10.1046/j.1365-2958.2003.03288.x

Bédard, M., McClure, C.D., Schiller, N.L., Francoeur, C., Cantin, A., Denis, M., 1993. Release of Interleukin-8, Interleukin-6, and Colony-stimulating Factors by Upper Airway Epithelial Cells: Implications for Cystic Fibrosis. American Journal of Respiratory Cell and Molecular Biology 9, 455–462. https://doi.org/10.1165/ajrcmb/9.4.455

Bennett, P.M., 2009. Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. British Journal of Pharmacology 153, S347–S357. https://doi.org/10.1038/sj.bjp.0707607

Benz, J., Sendlmeier, C., Barends, T.R.M., Meinhart, A., 2012. Structural Insights into the Effector – Immunity System Tse1/Tsi1 from Pseudomonas aeruginosa. PLoS ONE 7, e40453. https://doi.org/10.1371/journal.pone.0040453

Berrazeg, M., Jeannot, K., Ntsogo Enguéné, V.Y., Broutin, I., Loeffert, S., Fournier, D., Plésiat, P., 2015. Mutations in β-Lactamase AmpC Increase Resistance of Pseudomonas aeruginosa Isolates to Antipseudomonal Cephalosporins. Antimicrobial Agents and Chemotherapy 59, 6248–6255. https://doi.org/10.1128/aac.00825-15

Bhagirath, A.Y., Li, Y., Somayajula, D., Dadashi, M., Badr, S., Duan, K., 2016. Cystic fibrosis lung environment and Pseudomonas aeruginosa infection. BMC Pulmonary Medicine 16. https://doi.org/10.1186/s12890-016-0339-5

Bhatt, J.M., 2013. Treatment of pulmonary exacerbations in cystic fibrosis. European Respiratory Review 22, 205–216. https://doi.org/10.1183/09059180.00006512

Bi, D., Xie, Y., Tai, C., Jiang, X., Zhang, J., Harrison, E.M., Jia, S., Deng, Z., Rajakumar, K., Ou, H.-Y., 2016. A Site-Specific Integrative Plasmid Found in Pseudomonas aeruginosa Clinical Isolate HS87 along with A Plasmid Carrying an Aminoglycoside-Resistant Gene. PLOS ONE 11, e0148367. https://doi.org/10.1371/journal.pone.0148367

Bianconi, I., D'Arcangelo, S., Esposito, A., Benedet, M., Piffer, E., Dinnella, G., Gualdi, P., Schinella, M., Baldo, E., Donati, C., Jousson, O., 2019. Persistence and Microevolution of Pseudomonas aeruginosa in the Cystic Fibrosis Lung: A Single-Patient Longitudinal Genomic Study. Frontiers in Microbiology 9. https://doi.org/10.3389/fmicb.2018.03242

Bleriot, C., Effantin, G., Lagarde, F., Mandrand-Berthelot, M.-A., Rodrigue, A., 2011. RcnB Is a Periplasmic Protein Essential for Maintaining Intracellular Ni and Co Concentrations in Escherichia coli. Journal of Bacteriology 193, 3785–3793. https://doi.org/10.1128/jb.05032-11

Boles, B.R., Thoendel, M., Singh, P.K., 2005. Rhamnolipids mediate detachment ofPseudomonas aeruginosafrom biofilms. Molecular Microbiology 57, 1210–1223. https://doi.org/10.1111/j.1365-2958.2005.04743.x

Bondy-Denomy, J., Pawluk, A., Maxwell, K.L., Davidson, A.R., 2012. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature 493, 429–432. https://doi.org/10.1038/nature11723

Boursier, M.E., Moore, J.D., Heitman, K.M., Shepardson-Fungairino, S.P., Combs, J.B., Koenig, L.C., Shin, D., Brown, E.C., Nagarajan, R., Blackwell, H.E., 2018. Structure–Function Analyses of the N-Butanoyl l-Homoserine Lactone Quorum-Sensing Signal Define Features Critical to Activity in RhlR. ACS Chemical Biology 13, 2655–2662. https://doi.org/10.1021/acschembio.8b00577

Braud, A., Geoffroy, V., Hoegy, F., Mislin, G.L.A., Schalk, I.J., 2010. Presence of the siderophores pyoverdine and pyochelin in the extracellular medium reduces toxic metal accumulation in Pseudomonas aeruginosa and increases bacterial metal tolerance. Environmental Microbiology Reports 2, 419–425. https://doi.org/10.1111/j.1758-2229.2009.00126.x

Breidenstein, E.B.M., de la Fuente-Núñez, C., Hancock, R.E.W., 2011. Pseudomonas aeruginosa: all roads lead to resistance. Trends in Microbiology 19, 419–426. https://doi.org/10.1016/j.tim.2011.04.005

Brint, J.M., Ohman, D.E., 1995. Synthesis of multiple exoproducts in Pseudomonas aeruginosa is under the control of RhlR-RhlI, another set of regulators in strain PAO1 with homology to the autoinducer-responsive LuxR-LuxI family. Journal of Bacteriology 177, 7155–7163. https://doi.org/10.1128/jb.177.24.7155-7163.1995

Broder, U.N., Jaeger, T., Jenal, U., 2016. LadS is a calcium-responsive kinase that induces acute-to-chronic virulence switch in Pseudomonas aeruginosa. Nature Microbiology 2. https://doi.org/10.1038/nmicrobiol.2016.184

Brodlie, M., Haq, I.J., Roberts, K., Elborn, J.S., 2015. Targeted therapies to improve CFTR function in cystic fibrosis. Genome Medicine 7. https://doi.org/10.1186/s13073-015-0223-6

Bröms, J.E., Meyer, L., Lavander, M., Larsson, P., Sjöstedt, A., 2012. DotU and VgrG, Core Components of Type VI Secretion Systems, Are Essential for Francisella LVS Pathogenicity. PLoS ONE 7, e34639. https://doi.org/10.1371/journal.pone.0034639

Brouwer, S., Pustelny, C., Ritter, C., Klinkert, B., Narberhaus, F., Haussler, S., 2014. The PqsR and RhlR Transcriptional Regulators Determine the Level of Pseudomonas Quinolone Signal Synthesis in Pseudomonas aeruginosa by Producing Two Different pqsABCDE mRNA Isoforms. Journal of Bacteriology 196, 4163–4171. https://doi.org/10.1128/jb.02000-14

Brown, M.R.W., Foster, J.H.S., 1970. A simple diagnostic milk medium for Pseudomonas aeruginosa. Journal of Clinical Pathology 23, 172–177. https://doi.org/10.1136/jcp.23.2.172

Bruchmann, S., Dötsch, A., Nouri, B., Chaberny, I.F., Häussler, S., 2012. Quantitative Contributions of Target Alteration and Decreased Drug Accumulation to Pseudomonas aeruginosa Fluoroquinolone Resistance. Antimicrobial Agents and Chemotherapy 57, 1361–1368. https://doi.org/10.1128/aac.01581-12

Brünker, P., Altenbuchner, J., Mattes, R., 1998. Structure and function of the genes involved in mannitol, arabitol and glucitol utilization from Pseudomonas fluorescens DSM50106. Gene 206, 117–126. https://doi.org/10.1016/s0378-1119(97)00574-x

Bruscia, E.M., Bonfield, T.L., 2016. Cystic Fibrosis Lung Immunity: The Role of the Macrophage. Journal of Innate Immunity 8, 550–563. https://doi.org/10.1159/000446825

Brynildsrud, O., Bohlin, J., Scheffer, L., Eldholm, V., 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biology 17. https://doi.org/10.1186/s13059-016-1108-8

Buehrle, D.J., Shields, R.K., Clarke, L.G., Potoski, B.A., Clancy, C.J., Nguyen, M.H., 2016. Carbapenem-Resistant Pseudomonas aeruginosa Bacteremia: Risk Factors for Mortality and

Microbiologic Treatment Failure. Antimicrobial Agents and Chemotherapy 61. https://doi.org/10.1128/aac.01243-16

Burrowes, E., 2006. Influence of the regulatory protein RsmA on cellular functions in Pseudomonas aeruginosa PAO1, as revealed by transcriptome analysis. Microbiology 152, 405–418. https://doi.org/10.1099/mic.0.28324-0

Burrows, L.L., 2012. Pseudomonas aeruginosaTwitching Motility: Type IV Pili in Action. Annual Review of Microbiology 66, 493–520. https://doi.org/10.1146/annurev-micro-092611-150055

Bush, K., 2008. Extended-spectrum β-lactamases in North America, 1987–2006. Clinical Microbiology and Infection 14, 134–143. https://doi.org/10.1111/j.1469-0691.2007.01848.x

Caballero, A.R., Moreau, J.M., Engel, L.S., Marquart, M.E., Hill, J.M., O'Callaghan, R.J., 2001. Pseudomonas aeruginosa Protease IV Enzyme Assays and Comparison to Other Pseudomonas Proteases. Analytical Biochemistry 290, 330–337. https://doi.org/10.1006/abio.2001.4999

Cabot, G., López-Causapé, C., Ocampo-Sosa, A.A., Sommer, L.M., Domínguez, M.Á., Zamorano, L., Juan, C., Tubau, F., Rodríguez, C., Moyà, B., Peña, C., Martínez-Martínez, L., Plesiat, P., Oliver, A., 2016. Deciphering the resistome of the widespread P. aeruginosa ST175 international high-risk clone through whole genome sequencing. Antimicrobial Agents and Chemotherapy AAC.01720-16. https://doi.org/10.1128/aac.01720-16

Cadoret, F., Ball, G., Douzi, B., Voulhoux, R., 2014. Txc, a New Type II Secretion System of Pseudomonas aeruginosa Strain PA7, Is Regulated by the TtsS/TtsR Two-Component System and Directs Specific Secretion of the CbpE Chitin-Binding Protein. Journal of Bacteriology 196, 2376–2386. https://doi.org/10.1128/jb.01563-14

Cai, Z., Liu, Y., Chen, Y., Yam, J., Chew, S., Chua, S., Wang, K., Givskov, M., Yang, L., 2015. RpoN Regulates Virulence Factors of Pseudomonas aeruginosa via Modulating the PqsR Quorum Sensing Regulator. International Journal of Molecular Sciences 16, 28311–28319. https://doi.org/10.3390/ijms161226103

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421. https://doi.org/10.1186/1471-2105-10-421

Campodonico, V.L., Llosa, N.J., Grout, M., Doring, G., Maira-Litran, T., Pier, G.B., 2009. Evaluation of Flagella and Flagellin of Pseudomonas aeruginosa as Vaccines. Infection and Immunity 78, 746–755. https://doi.org/10.1128/iai.00806-09

Cao, H., Krishnan, G., Goumnerov, B., Tsongalis, J., Tompkins, R., Rahme, L.G., 2001. A quorum sensing-associated virulence gene of Pseudomonas aeruginosa encodes a LysR-like transcription regulator with a unique self-regulatory mechanism. Proceedings of the National Academy of Sciences 98, 14613–14618. https://doi.org/10.1073/pnas.251465298

Cazares, A., Moore, M. P., Grimes, M., Emond-Rheault, J-G., Wright, L. L., Pongchaikul, P., Santanirand, P., Levesque, R. C., Fothergill, J. L., Winstanley, C., 2019. A megaplasmid family responsible for dissemination of multidrug resistance in Pseudomonas bioRxiv 10.1101/630780

Chaudary, N., 2018. Triplet CFTR modulators: future prospects for treatment of cystic fibrosis. Therapeutics and Clinical Risk Management Volume 14, 2375–2383. https://doi.org/10.2147/tcrm.s147164

Chávez-Jacobo, V.M., Hernández-Ramírez, K.C., Romo-Rodríguez, P., Pérez-Gallardo, R.V., Campos-García, J., Gutiérrez-Corona, J.F., García-Merinos, J.P., Meza-Carmen, V., Silva-Sánchez, J., Ramírez-Díaz, M.I., 2018. CrpP Is a Novel Ciprofloxacin-Modifying Enzyme Encoded by the Pseudomonas aeruginosa pUM505 Plasmid. Antimicrobial Agents and Chemotherapy 62, e02629-17. https://doi.org/10.1128/aac.02629-17

Che, S., Chen, Y., Liang, Y., Zhang, Q., Bartlam, M., 2018. Crystal structure of RecR, a member of the RecFOR DNA-repair pathway, from Pseudomonas aeruginosa PAO1. Acta Crystallographica Section F Structural Biology Communications 74, 222–230. https://doi.org/10.1107/s2053230x18003503

Chen, L., Zou, Y., She, P., Wu, Y., 2015. Composition, function, and regulation of T6SS in Pseudomonas aeruginosa. Microbiological Research 172, 19–25. https://doi.org/10.1016/j.micres.2015.01.004

Chevalier, S., Bouffartigues, E., Bodilis, J., Maillot, O., Lesouhaitier, O., Feuilloley, M.G.J., Orange, N., Dufour, A., Cornelis, P., 2017. Structure, function and regulation of Pseudomonas aeruginosa porins. FEMS Microbiology Reviews 41, 698–722. https://doi.org/10.1093/femsre/fux020

Chuanchuen, R., Narasaki, C.T., Schweizer, H.P., 2002. The MexJK Efflux Pump of Pseudomonas aeruginosa Requires OprM for Antibiotic Efflux but Not for Efflux of Triclosan. Journal of Bacteriology 184, 5036–5044. https://doi.org/10.1128/jb.184.18.5036-5044.2002

Ciszek-Lenda, M., Strus, M., Walczewska, M., Majka, G., Machul-Żwirbla, A., Mikołajczyk, D., Górska, S., Gamian, A., Chain, B., Marcinkiewicz, J., 2019. Pseudomonas aeruginosa biofilm is a potent inducer of phagocyte hyperinflammation. Inflammation Research 68, 397–413. https://doi.org/10.1007/s00011-019-01227-x

Co, J.Y., Cárcamo-Oyarce, G., Billings, N., Wheeler, K.M., Grindy, S.C., Holten-Andersen, N., Ribbeck, K., 2018. Mucins trigger dispersal of Pseudomonas aeruginosa biofilms. npj Biofilms and Microbiomes 4. https://doi.org/10.1038/s41522-018-0067-0

Codjoe, F., Donkor, E., 2017. Carbapenem Resistance: A Review. Medical Sciences 6, 1. https://doi.org/10.3390/medsci6010001

Cogen, J.D., Oron, A.P., Gibson, R.L., Hoffman, L.R., Kronman, M.P., Ong, T., Rosenfeld, M., 2017. Characterization of Inpatient Cystic Fibrosis Pulmonary Exacerbations. Pediatrics 139, e20162642. https://doi.org/10.1542/peds.2016-2642

Cogen, J.D., Rosenfeld, M., 2018. Treating Cystic Fibrosis Pulmonary Exacerbations: In the Hospital with a Physician or at Home under Your Own Supervision? Annals of the American Thoracic Society 15, 169–170. https://doi.org/10.1513/annalsats.201711-891ed

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, New Jersey: Lawrence Erlbaum Associates. ISBN-13 978-0805802832

Cohen-Cymberknoh, M., Shoseyov, D., Kerem, E., 2011. Managing Cystic Fibrosis. American Journal of Respiratory and Critical Care Medicine 183, 1463–1471. https://doi.org/10.1164/rccm.201009-1478ci

Coll, F., Phelan, J., Hill-Cawthorne, G.A., Nair, M.B., Mallard, K., Ali, S., Abdallah, A.M., Alghamdi, S., Alsomali, M., Ahmed, A.O., Portelli, S., Oppong, Y., Alves, A., Bessa, T.B., Campino, S., Caws, M., Chatterjee, A., Crampin, A.C., Dheda, K., Furnham, N., Glynn, J.R., Grandjean, L., Minh Ha, D., Hasan, R., Hasan, Z., Hibberd, M.L., Joloba, M., Jones-López, E.C., Matsumoto, T., Miranda, A., Moore, D.J., Mocillo, N., Panaiotov, S., Parkhill, J., Penha, C., Perdigão, J., Portugal, I., Rchiad, Z., Robledo, J., Sheen, P., Shesha, N.T., Sirgel, F.A., Sola, C., Oliveira Sousa, E., Streicher, E.M., Helden, P.V., Viveiros, M., Warren, R.M., McNerney, R., Pain, A., Clark, T.G., 2018. Genome-wide analysis of multi- and extensively drug-resistant Mycobacterium tuberculosis. Nature Genetics 50, 307–316. https://doi.org/10.1038/s41588-017-0029-0

Collawn, J.F., Matalon, S., 2014. CFTR and lung homeostasis. American Journal of Physiology-Lung Cellular and Molecular Physiology 307, L917–L923. https://doi.org/10.1152/ajplung.00326.2014

Collins, C., Didelot, X., 2018. A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. PLOS Computational Biology 14, e1005958. https://doi.org/10.1371/journal.pcbi.1005958

Comolli, J. C., Hauser, A. R, Waite, L., Whitchurc,h C. B., Mattick, J. S., Engel, J. N. 1999. Pseudomonas aeruginosa gene products PilT and PilU are required for cytotoxicity in vitro and virulence in a mouse model of acute pneumonia. Infect Immun. 67(7), 3625-30.

Cornelis, P., Dingemans, J., 2013. Pseudomonas aeruginosa adapts its iron uptake strategies in function of the type of infections. Frontiers in Cellular and Infection Microbiology 3. https://doi.org/10.3389/fcimb.2013.00075

Corral-Lugo, A., Matilla, M.A., Martín-Mora, D., Silva Jiménez, H., Mesa Torres, N., Kato, J., Hida, A., Oku, S., Conejero-Muriel, M., Gavira, J.A., Krell, T., 2018. High-Affinity Chemotaxis to Histamine Mediated by the TlpQ Chemoreceptor of the Human Pathogen Pseudomonas aeruginosa. mBio 9. https://doi.org/10.1128/mbio.01894-18

Cramer, N., Wiehlmann, L., Tümmler, B., 2010. Clonal epidemiology of Pseudomonas aeruginosa in cystic fibrosis. International Journal of Medical Microbiology 300, 526–533. https://doi.org/10.1016/j.ijmm.2010.08.004

Crocker, A. W., Harty, C. E., Hammond, J. H., Willger, S. D., Salazar, P., Jacobs, N. J., Hogan, D. A., 2019, Pseudomonas aeruginosa ethanol oxidation by AdhA in low oxygen environments. BioRXiv.

Croucher, N.J., Page, A.J., Connor, T.R., Delaney, A.J., Keane, J.A., Bentley, S.D., Parkhill, J., Harris, S.R., 2014. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Research 43, e15–e15. https://doi.org/10.1093/nar/gku1196

Crouzet, M., Claverol, S., Lomenech, A.-M., Le Sénéchal, C., Costaglioli, P., Barthe, C., Garbay, B., Bonneu, M., Vilain, S., 2017. Pseudomonas aeruginosa cells attached to a surface display a typical proteome early as 20 minutes of incubation. PLOS ONE 12, e0180341. https://doi.org/10.1371/journal.pone.0180341

Crull, M.R., Ramos, K.J., Caldwell, E., Mayer-Hamblett, N., Aitken, M.L., Goss, C.H., 2016. Change in Pseudomonas aeruginosa prevalence in cystic fibrosis adults over time. BMC Pulmonary Medicine 16. https://doi.org/10.1186/s12890-016-0333-y

Crull, M.R., Somayaji, R., Ramos, K.J., Caldwell, E., Mayer-Hamblett, N., Aitken, M.L., Nichols, D.P., Rowhani-Rahbar, A., Goss, C.H., 2018. Changing Rates of Chronic Pseudomonas aeruginosa Infections in Cystic Fibrosis: A Population-Based Cohort Study. Clinical Infectious Diseases 67, 1089–1095. https://doi.org/10.1093/cid/ciy215

Culhane, S., George, C., Pearo, B., Spoede, E., 2013. Malnutrition in Cystic Fibrosis. Nutrition in Clinical Practice 28, 676–683. https://doi.org/10.1177/0884533613507086

Culp, E., Wright, G.D., 2016. Bacterial proteases, untapped antimicrobial drug targets. The Journal of Antibiotics 70, 366–377. https://doi.org/10.1038/ja.2016.138

Cuthbertson, L., Rogers, G.B., Walker, A.W., Oliver, A., Green, L.E., Daniels, T.W.V., Carroll, M.P., Parkhill, J., Bruce, K.D., van der Gast, C.J., 2015. Respiratory microbiota resistance and resilience to pulmonary exacerbation and subsequent antimicrobial intervention. The ISME Journal 10, 1081–1091. https://doi.org/10.1038/ismej.2015.198

Cutting, G.R., 2014. Cystic fibrosis genetics: from molecular understanding to clinical application. Nature Reviews Genetics 16, 45–56. https://doi.org/10.1038/nrg3849

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Darch, S.E., McNally, A., Harrison, F., Corander, J., Barr, H.L., Paszkiewicz, K., Holden, S., Fogarty, A., Crusz, S.A., Diggle, S.P., 2015. Recombination is a key driver of genomic and phenotypic diversity in a Pseudomonas aeruginosa population during cystic fibrosis infection. Scientific Reports 5. https://doi.org/10.1038/srep07649

Davey, M.E., Caiazza, N.C., O'Toole, G.A., 2003. Rhamnolipid Surfactant Production Affects Biofilm Architecture in Pseudomonas aeruginosa PAO1. Journal of Bacteriology 185, 1027–1036. https://doi.org/10.1128/jb.185.3.1027-1036.2003

De Boeck, K., Amaral, M.D., 2016. Progress in therapies for cystic fibrosis. The Lancet Respiratory Medicine 4, 662–674. https://doi.org/10.1016/s2213-2600(16)00023-0

de Brito, D.M., Maracaja-Coutinho, V., de Farias, S.T., Batista, L.V., do Rêgo, T.G., 2016. A Novel Method to Predict Genomic Islands Based on Mean Shift Clustering Algorithm. PLOS ONE 11, e0146352. https://doi.org/10.1371/journal.pone.0146352

de Kievit, T., 2002. Role of the Pseudomonas aeruginosa las and rhl quorum-sensing systems in rhlI regulation. FEMS Microbiology Letters 212, 101–106. https://doi.org/10.1016/s0378-1097(02)00735-8

Deeks, E.D., 2016. Lumacaftor/Ivacaftor: A Review in Cystic Fibrosis. Drugs 76, 1191–1201. https://doi.org/10.1007/s40265-016-0611-2

del Barrio-Tofiño, E., López-Causapé, C., Cabot, G., Rivera, A., Benito, N., Segura, C., Montero, M.M., Sorlí, L., Tubau, F., Gómez-Zorrilla, S., Tormo, N., Durá-Navarro, R., Viedma, E., Resino-Foz, E., Fernández-Martínez, M., González-Rico, C., Alejo-Cancho, I., Martínez, J.A., Labayru-Echverria, C., Dueñas, C., Ayestarán, I., Zamorano, L., Martinez-Martinez, L., Horcajada, J.P., Oliver, A., 2017. Genomics and Susceptibility Profiles of Extensively Drug-Resistant Pseudomonas aeruginosa Isolates from Spain. Antimicrobial Agents and Chemotherapy 61. https://doi.org/10.1128/aac.01589-17

Dettman, J.R., Sztepanacz, J.L., Kassen, R., 2016. The properties of spontaneous mutations in the opportunistic pathogen Pseudomonas aeruginosa. BMC Genomics 17. https://doi.org/10.1186/s12864-015-2244-3

Deziel, E., Lepine, F., Milot, S., He, J., Mindrinos, M.N., Tompkins, R.G., Rahme, L.G., 2004. Analysis of Pseudomonas aeruginosa 4-hydroxy-2-alkylquinolines (HAQs) reveals a role for 4-hydroxy-2-heptylquinoline in cell-to-cell communication. Proceedings of the National Academy of Sciences 101, 1339–1344. https://doi.org/10.1073/pnas.0307694100

Diaz Caballero, J., Clark, S.T., Coburn, B., Zhang, Y., Wang, P.W., Donaldson, S.L., Tullis, D.E., Yau, Y.C.W., Waters, V.J., Hwang, D.M., Guttman, D.S., 2015. Selective Sweeps and Parallel Pathoadaptation Drive Pseudomonas aeruginosa Evolution in the Cystic Fibrosis Lung. mBio 6. https://doi.org/10.1128/mbio.00981-15

Didelot, X., Croucher, N.J., Bentley, S.D., Harris, S.R., Wilson, D.J., 2018. Bayesian inference of ancestral dates on bacterial phylogenetic trees. Nucleic Acids Research 46, e134–e134. https://doi.org/10.1093/nar/gky783

Downey, D.G., Bell, S.C., Elborn, J.S., 2008. Neutrophils in cystic fibrosis. Thorax 64, 81–88. https://doi.org/10.1136/thx.2007.082388

Duong, J., Booth, S.C., McCartney, N.K., Rabin, H.R., Parkins, M.D., Storey, D.G., 2015. Phenotypic and Genotypic Comparison of Epidemic and Non-Epidemic Strains of Pseudomonas aeruginosa from Individuals with Cystic Fibrosis. PLOS ONE 10, e0143466. https://doi.org/10.1371/journal.pone.0143466

Earle, S.G., Wu, C.-H., Charlesworth, J., Stoesser, N., Gordon, N.C., Walker, T.M., Spencer, C.C.A., Iqbal, Z., Clifton, D.A., Hopkins, K.L., Woodford, N., Smith, E.G., Ismail, N., Llewelyn, M.J., Peto, T.E., Crook, D.W., McVean, G., Walker, A.S., Wilson, D.J., 2016. Identifying lineage effects when controlling for population structure improves power in bacterial association studies. Nature Microbiology 1. https://doi.org/10.1038/nmicrobiol.2016.41

Elborn, S., Vallieres, E., 2014. Cystic fibrosis gene mutations: evaluation and assessment of disease severity. Advances in Genomics and Genetics 161. https://doi.org/10.2147/agg.s53768

Espel, J.C., Palac, H.L., Cullina, J.F., Clarke, A.P., McColley, S.A., Prickett, M.H., Jain, M., 2017. Antibiotic duration and changes in FEV1 are not associated with time until next exacerbation in adult cystic fibrosis: a single center study. BMC Pulmonary Medicine 17. https://doi.org/10.1186/s12890-017-0503-6

Etherington, C., Naseer, R., Conway, S.P., Whitaker, P., Denton, M., Peckham, D.G., 2014. The role of respiratory viruses in adult patients with cystic fibrosis receiving intravenous antibiotics for a pulmonary exacerbation. Journal of Cystic Fibrosis 13, 49–55. https://doi.org/10.1016/j.jcf.2013.06.004

Evans, K., Poole, K., 1999. The MexA-MexB-OprM multidrug efflux system of Pseudomonas aeruginosa is growth-phase regulated. FEMS Microbiology Letters 173, 35–39. https://doi.org/10.1111/j.1574-6968.1999.tb13481.x

Fàbrega, A., Madurga, S., Giralt, E., Vila, J., 2009. Mechanism of action of and resistance to quinolones. Microbial Biotechnology 2, 40–61. https://doi.org/10.1111/j.1751-7915.2008.00063.x

Fahy, J.V., Dickey, B.F., 2010. Airway Mucus Function and Dysfunction. New England Journal of Medicine 363, 2233–2247. https://doi.org/10.1056/nejmra0910061

Falagas, M.E., Kasiakou, S.K., Saravolatz, L.D., 2005. Colistin: The Revival of Polymyxins for the Management of Multidrug-Resistant Gram-Negative Bacterial Infections. Clinical Infectious Diseases 40, 1333–1341. https://doi.org/10.1086/429323

Fan, X., Wu, Y., Xiao, M., Xu, Z.-P., Kudinha, T., Bazaj, A., Kong, F., Xu, Y.-C., 2016. Diverse Genetic Background of Multidrug-Resistant Pseudomonas aeruginosa from Mainland China and Emergence of an Extensively Drug-Resistant ST292 Clone in Kunming. Scientific Reports 6. https://doi.org/10.1038/srep26522

Fargier, E., Mac Aogain, M., Mooij, M.J., Woods, D.F., Morrissey, J.P., Dobson, A.D.W., Adams, C., O'Gara, F., 2012. MexT Functions as a Redox-Responsive Regulator Modulating Disulfide Stress Resistance in Pseudomonas aeruginosa. Journal of Bacteriology 194, 3502–3511. https://doi.org/10.1128/jb.06632-11

Farnia, P., Mirtajani, S., Hassanzad, M., Ghanavi, J., Farnia, P., Velayati, A., 2017. Geographical distribution of cystic fibrosis; The past 70 years of data analysis. Biomedical and Biotechnology Research Journal (BBRJ) 1, 105. https://doi.org/10.4103/bbrj.bbrj_81_17

Feil, H., Feil, W.S., Chain, P., Larimer, F., DiBartolo, G., Copeland, A., Lykidis, A., Trong, S., Nolan, M., Goltsman, E., Thiel, J., Malfatti, S., Loper, J.E., Lapidus, A., Detter, J.C., Land, M., Richardson, P.M., Kyrpides, N.C., Ivanova, N., Lindow, S.E., 2005. Comparison of the complete genome sequences of Pseudomonas syringae pv. syringae B728a and pv. tomato DC3000. Proceedings of the National Academy of Sciences 102, 11064–11069. https://doi.org/10.1073/pnas.0504930102

Feliziani, S., Luján, A.M., Moyano, A.J., Sola, C., Bocco, J.L., Montanaro, P., Canigia, L.F., Argaraña, C.E., Smania, A.M., 2010. Mucoidy, Quorum Sensing, Mismatch Repair and Antibiotic Resistance in Pseudomonas aeruginosa from Cystic Fibrosis Chronic Airways Infections. PLoS ONE 5, e12669. https://doi.org/10.1371/journal.pone.0012669

Feliziani, S., Marvig, R.L., Luján, A.M., Moyano, A.J., Di Rienzo, J.A., Krogh Johansen, H., Molin, S., Smania, A.M., 2014. Coexistence and Within-Host Evolution of Diversified Lineages of Hypermutable Pseudomonas aeruginosa in Long-term Cystic Fibrosis Infections. PLoS Genetics 10, e1004651. https://doi.org/10.1371/journal.pgen.1004651

Feltman, H., Jain, M., Peterson, L., Schulert, G., Khan, S., Hauser, A.R., 2001. Prevalence of type III secretion genes in clinical and environmental isolates of Pseudomonas aeruginosa. Microbiology 147, 2659–2669. https://doi.org/10.1099/00221287-147-10-2659

Feltner, J.B., Wolter, D.J., Pope, C.E., Groleau, M.-C., Smalley, N.E., Greenberg, E.P., Mayer-Hamblett, N., Burns, J., Déziel, E., Hoffman, L.R., Dandekar, A.A., 2016. LasR Variant Cystic Fibrosis Isolates Reveal an Adaptable Quorum-Sensing Hierarchy in Pseudomonas aeruginosa. mBio 7. https://doi.org/10.1128/mbio.01513-16

Filkins, L.M., O'Toole, G.A., 2015. Cystic Fibrosis Lung Infections: Polymicrobial, Complex, and Hard to Treat. PLOS Pathogens 11, e1005258. https://doi.org/10.1371/journal.ppat.1005258

Filloux, A., 2011. Protein Secretion Systems in Pseudomonas aeruginosa: An Essay on Diversity, Evolution, and Function. Frontiers in Microbiology 2. https://doi.org/10.3389/fmicb.2011.00155

Filloux, A., Hachani, A., Bleves, S., 2008. The bacterial type VI secretion machine: yet another player for protein transport across membranes. Microbiology 154, 1570–1583. https://doi.org/10.1099/mic.0.2008/016840-0

Fluge, G., Ojeniyi, B., Høiby, N., Digranes, A., Ciofu, O., Hunstad, E., Haanaes, O.C., Storrøsten, O.-T., 2001. Typing of Pseudomonas aeruginosa strains in Norwegian cystic fibrosis patients. Clinical Microbiology and Infection 7, 238–243. https://doi.org/10.1046/j.1469-0691.2001.00247.x

Flynn, J.M., Niccum, D., Dunitz, J.M., Hunter, R.C., 2016. Evidence and Role for Bacterial Mucin Degradation in Cystic Fibrosis Airway Disease. PLOS Pathogens 12, e1005846. https://doi.org/10.1371/journal.ppat.1005846

Fodor, A.A., Klem, E.R., Gilpin, D.F., Elborn, J.S., Boucher, R.C., Tunney, M.M., Wolfgang, M.C., 2012. The Adult Cystic Fibrosis Airway Microbiota Is Stable over Time and Infection Type, and Highly Resilient to Antibiotic Treatment of Exacerbations. PLoS ONE 7, e45001. https://doi.org/10.1371/journal.pone.0045001

Fohner, A.E., McDonagh, E.M., Clancy, J.P., Whirl Carrillo, M., Altman, R.B., Klein, T.E., 2017. PharmGKB summary. Pharmacogenetics and Genomics 27, 39–42. https://doi.org/10.1097/fpc.0000000000000246

Folkesson, A., Jelsbak, L., Yang, L., Johansen, H.K., Ciofu, O., Høiby, N., Molin, S., 2012. Adaptation of Pseudomonas aeruginosa to the cystic fibrosis airway: an evolutionary perspective. Nature Reviews Microbiology 10, 841–851. https://doi.org/10.1038/nrmicro2907

Fonseca, E.L., Marin, M.A., Encinas, F., Vicente, A.C.P., 2015. Full characterization of the integrative and conjugative element carrying the metallo-β-lactamase blaSPM-1 and bicyclomycin bcr1 resistance genes found in the pandemic Pseudomonas aeruginosa clone SP/ST277 : Figure 1. Journal of Antimicrobial Chemotherapy 70, 2547–2550. https://doi.org/10.1093/jac/dkv152

Fothergill, J.L., Mowat, E., Ledson, M.J., Walshaw, M.J., Winstanley, C., 2009. Fluctuations in phenotypes and genotypes within populations of Pseudomonas aeruginosa in the cystic fibrosis lung during pulmonary exacerbations. Journal of Medical Microbiology 59, 472–481. https://doi.org/10.1099/jmm.0.015875-0

Fothergill, J.L., Walshaw, M.J., Winstanley, C., 2012. Transmissible strains of Pseudomonas aeruginosa in cystic fibrosis lung infections. European Respiratory Journal 40, 227–238. https://doi.org/10.1183/09031936.00204411

Fowler, R.C., Hanson, N.D., 2014. Emergence of Carbapenem Resistance Due to the Novel Insertion Sequence ISPa8 in Pseudomonas aeruginosa. PLoS ONE 9, e91299. https://doi.org/10.1371/journal.pone.0091299

Francis, V.I., Waters, E.M., Finton-James, S.E., Gori, A., Kadioglu, A., Brown, A.R., Porter, S.L., 2018. Multiple communication mechanisms between sensor kinases are crucial for virulence in Pseudomonas aeruginosa. Nature Communications 9. https://doi.org/10.1038/s41467-018-04640-8

Freschi, L., Bertelli, C., Jeukens, J., Moore, M.P., Kukavica-Ibrulj, I., Emond-Rheault, J.-G., Hamel, J., Fothergill, J.L., Tucker, N.P., McClean, S., Klockgether, J., de Soyza, A., Brinkman, F.S.L., Levesque, R.C., Winstanley, C., 2018. Genomic characterisation of an international Pseudomonas aeruginosa reference panel indicates that the two major groups draw upon distinct mobile gene pools. FEMS Microbiology Letters 365. https://doi.org/10.1093/femsle/fny120

Freschi, L., Vincent, A.T., Jeukens, J., Emond-Rheault, J.-G., Kukavica-Ibrulj, I., Dupont, M.-J., Charette, S.J., Boyle, B., Levesque, R.C., 2019. The Pseudomonas aeruginosa Pan-Genome Provides New Insights on Its Population Structure, Horizontal Gene Transfer, and Pathogenicity. Genome Biology and Evolution 11, 109–120. https://doi.org/10.1093/gbe/evy259

Fritz, S.A., Purvis, A., 2010. Phylogenetic diversity does not capture body size variation at risk in the world's mammals. Proceedings of the Royal Society B: Biological Sciences 277, 2435–2441. https://doi.org/10.1098/rspb.2010.0030

Galán-Vásquez, E., Luna, B., Martínez-Antonio, A., 2011. The Regulatory Network of Pseudomonas aeruginosa. Microbial Informatics and Experimentation 1, 3. https://doi.org/10.1186/2042-5783-1-3

Galdino, A.C.M., Branquinha, M.H., Santos, A.L.S., Viganor, L., 2017. Pseudomonas aeruginosa and Its Arsenal of Proteases: Weapons to Battle the Host, in: Pathophysiological Aspects of Proteases. Springer Singapore, pp. 381–397. https://doi.org/10.1007/978-981-10-6141-7_16

Galetti, R., Andrade, L.N., Varani, A.M., Darini, A.L.C., 2019. SPM-1-producing Pseudomonas aeruginosa ST277 carries a chromosomal pack of acquired resistance genes: An example of high-risk clone associated with 'intrinsic resistome'. Journal of Global Antimicrobial Resistance 16, 183–186. https://doi.org/10.1016/j.jgar.2018.12.009

Galle, M., Carpentier, I., Beyaert, R., 2012. Structure and Function of the Type III Secretion System of Pseudomonas aeruginosa. Current Protein and Peptide Science 13, 831–842. https://doi.org/10.2174/138920312804871210

Geer, L.Y., Marchler-Bauer, A., Geer, R.C., Han, L., He, J., He, S., Liu, C., Shi, W., Bryant, S.H., 2009. The NCBI BioSystems database. Nucleic Acids Research 38, D492–D496. https://doi.org/10.1093/nar/gkp858

Gellatly, S.L., Hancock, R.E.W., 2013. Pseudomonas aeruginosa: new insights into pathogenesis and host defenses. Pathogens and Disease 67, 159–173. https://doi.org/10.1111/2049-632x.12033

Gersch, M., List, A., Groll, M., Sieber, S.A., 2012. Insights into Structural Network Responsible for Oligomerization and Activity of Bacterial Virulence Regulator Caseinolytic Protease P (ClpP) Protein. Journal of Biological Chemistry 287, 9484–9494. https://doi.org/10.1074/jbc.m111.336222

Gilbert, K.B., Kim, T.H., Gupta, R., Greenberg, E.P., Schuster, M., 2009. Global position analysis of thePseudomonas aeruginosa quorum-sensing transcription factor LasR. Molecular Microbiology 73, 1072–1085. https://doi.org/10.1111/j.1365-2958.2009.06832.x

Gilbertsen, A., Williams, B., 2014. Development of a Pseudomonas aeruginosa Agmatine Biosensor. Biosensors 4, 387–402. https://doi.org/10.3390/bios4040387

Golle, A., Janezic, S., Rupnik, M., 2017. Low overlap between carbapenem resistant Pseudomonas aeruginosa genotypes isolated from hospitalized patients and wastewater treatment plants. PLOS ONE 12, e0186736. https://doi.org/10.1371/journal.pone.0186736

Golovkine, G., Reboud, E., Huber, P., 2018. Pseudomonas aeruginosa Takes a Multi-Target Approach to Achieve Junction Breach. Frontiers in Cellular and Infection Microbiology 7. https://doi.org/10.3389/fcimb.2017.00532

Gomila, M., del Carmen Gallegos, M., Fernández-Baca, V., Pareja, A., Pascual, M., Díaz-Antolín, P., García-Valdés, E., Lalucat, J., 2013. Genetic diversity of clinical Pseudomonas aeruginosa isolates in a public hospital in Spain. BMC Microbiology 13, 138. https://doi.org/10.1186/1471-2180-13-138

Haardt, M., Benharouga, M., Lechardeur, D., Kartner, N., Lukacs, G.L., 1999. C-terminal Truncations Destabilize the Cystic Fibrosis Transmembrane Conductance Regulator without Impairing Its Biogenesis. Journal of Biological Chemistry 274, 21873–21877. https://doi.org/10.1074/jbc.274.31.21873

Hadfield, J., Croucher, N.J., Goater, R.J., Abudahab, K., Aanensen, D.M., Harris, S.R., 2017. Phandango: an interactive viewer for bacterial population genomics. Bioinformatics 34, 292–293. https://doi.org/10.1093/bioinformatics/btx610

Harrison, F., 2007. Microbial ecology of the cystic fibrosis lung. Microbiology 153, 917–923. https://doi.org/10.1099/mic.0.2006/004077-0

Harrison, F., McNally, A., da Silva, A.C., Heeb, S., Diggle, S.P., 2017. Optimised chronic infection models demonstrate that siderophore 'cheating' in Pseudomonas aeruginosa is context specific. The ISME Journal 11, 2492–2509. https://doi.org/10.1038/ismej.2017.103

Hauser, A.R., 2009. The type III secretion system of Pseudomonas aeruginosa: infection by injection. Nature Reviews Microbiology 7, 654–665. https://doi.org/10.1038/nrmicro2199

Häussler, S., Becker, T., 2008. The Pseudomonas Quinolone Signal (PQS) Balances Life and Death in Pseudomonas aeruginosa Populations. PLoS Pathogens 4, e1000166. https://doi.org/10.1371/journal.ppat.1000166

Hawver, L.A., Jung, S.A., Ng, W.-L., 2016. Specificity and complexity in bacterial quorum-sensing systems. FEMS Microbiology Reviews 40, 738–752. https://doi.org/10.1093/femsre/fuw014

He, G.-X., Kuroda, T., Mima, T., Morita, Y., Mizushima, T., Tsuchiya, T., 2003. An H+-Coupled Multidrug Efflux Pump, PmpM, a Member of the MATE Family of Transporters, from Pseudomonas aeruginosa. Journal of Bacteriology 186, 262–265. https://doi.org/10.1128/jb.186.1.262-265.2004

Hector, A., Griese, M., Hartl, D., 2014. Oxidative stress in cystic fibrosis lung disease: an early event, but worth targeting? European Respiratory Journal 44, 17–19. https://doi.org/10.1183/09031936.00038114

Heltshe, S.L., Khan, U., Beckett, V., Baines, A., Emerson, J., Sanders, D.B., Gibson, R.L., Morgan, W., Rosenfeld, M., 2018. Longitudinal development of initial, chronic and mucoid Pseudomonas aeruginosa

infection in young children with cystic fibrosis. Journal of Cystic Fibrosis 17, 341–347. https://doi.org/10.1016/j.jcf.2017.10.008

Henrichfreise, B., Wiegand, I., Pfister, W., Wiedemann, B., 2007. Resistance Mechanisms of Multiresistant Pseudomonas aeruginosa Strains from Germany and Correlation with Hypermutation. Antimicrobial Agents and Chemotherapy 51, 4062–4070. https://doi.org/10.1128/aac.00148-07

Héry-Arnaud, G., Boutin, S., Cuthbertson, L., Elborn, S.J., Tunney, M.M., 2019. The lung and gut microbiome: what has to be taken into consideration for cystic fibrosis? Journal of Cystic Fibrosis 18, 13–21. https://doi.org/10.1016/j.jcf.2018.11.003

Hickey, C., Schaible, B., Nguyen, S., Hurley, D., Srikumar, S., Fanning, S., Brown, E., Crifo, B., Matallanas, D., McClean, S., Taylor, C.T., Schaffer, K., 2018. Increased Virulence of Bloodstream Over Peripheral Isolates of P. aeruginosa Identified Through Post-transcriptional Regulation of Virulence Factors. Frontiers in Cellular and Infection Microbiology 8. https://doi.org/10.3389/fcimb.2018.00357

Highsmith, W.E., Burch, L.H., Zhou, Z., Olsen, J.C., Strong, T.V., Smith, T., Friedman, K.J., Silverman, L.M., Boucher, R.C., Collins, F.S., Knowles, M.R., 1997. Identification of a splice site mutation (2789+5 G>A) associated with small amounts of normal CFTRmRNA and mild cystic fibrosis. Human Mutation 9, 332–338. https://doi.org/10.1002/(sici)1098-1004(1997)9:4<332::aid-humu5>3.0.co;2-7

Hill, D.B., Long, R.F., Kissner, W.J., Atieh, E., Garbarine, I.C., Markovetz, M.R., Fontana, N.C., Christy, M., Habibpour, M., Tarran, R., Forest, M.G., Boucher, R.C., Button, B., 2018. Pathological mucus and impaired mucus clearance in cystic fibrosis patients result from increased concentration, not altered pH. European Respiratory Journal 52, 1801297. https://doi.org/10.1183/13993003.01297-2018

Hilliam, Y., Moore, M.P., Lamont, I.L., Bilton, D., Haworth, C.S., Foweraker, J., Walshaw, M.J., Williams, D., Fothergill, J.L., De Soyza, A., Winstanley, C., 2017. Pseudomonas aeruginosa adaptation and diversification in the non-cystic fibrosis bronchiectasis lung. European Respiratory Journal 49, 1602108. https://doi.org/10.1183/13993003.02108-2016

Hirsch, E.B., Tam, V.H., 2010. Impact of multidrug-resistant Pseudomonas aeruginosa infection on patient outcomes. Expert Review of Pharmacoeconomics & Outcomes Research 10, 441–451. https://doi.org/10.1586/erp.10.49

Hoffman, L., Surette, M., 2013. What to Expect When You're Expectorating: Cystic Fibrosis Exacerbations and Microbiota. Annals of the American Thoracic Society 10, 249–250. https://doi.org/10.1513/annalsats.201304-086ed

Hoffmann, N., Lee, B., Hentzer, M., Rasmussen, T.B., Song, Z., Johansen, H.K., Givskov, M., Hoiby, N., 2007. Azithromycin Blocks Quorum Sensing and Alginate Polymer Formation and Increases the Sensitivity to Serum and Stationary-Growth-Phase Killing of Pseudomonas aeruginosa and Attenuates Chronic P. aeruginosa Lung Infection in Cftr / Mice. Antimicrobial Agents and Chemotherapy 51, 3677–3687. https://doi.org/10.1128/aac.01011-06

Hogardt, M., Heesemann, J., 2011. Microevolution of Pseudomonas aeruginosa to a Chronic Pathogen of the Cystic Fibrosis Lung, in: Between Pathogenicity and Commensalism. Springer Berlin Heidelberg, pp. 91–118. https://doi.org/10.1007/82_2011_199

Høiby, N., Ciofu, O., Bjarnsholt, T., 2010. Pseudomonas aeruginosa biofilms in cystic fibrosis. Future Microbiology 5, 1663–1674. https://doi.org/10.2217/fmb.10.125

Høiby, N., Frederiksen, B., Pressler, T., 2005. Eradication of early Pseudomonas aeruginosa infection. Journal of Cystic Fibrosis 4, 49–54. https://doi.org/10.1016/j.jcf.2005.05.018

Housseini B Issa, K., Phan, G., Broutin, I., 2018. Functional Mechanism of the Efflux Pumps Transcription Regulators From Pseudomonas aeruginosa Based on 3D Structures. Frontiers in Molecular Biosciences 5. https://doi.org/10.3389/fmolb.2018.00057

Hunt, M., Mather, A.E., Sánchez-Busó, L., Page, A.J., Parkhill, J., Keane, J.A., Harris, S.R., 2017. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. Microbial Genomics 3. https://doi.org/10.1099/mgen.0.000131

Hurley, M.N., Cámara, M., Smyth, A.R., 2012. Novel approaches to the treatment of Pseudomonas aeruginosa infections in cystic fibrosis. European Respiratory Journal 40, 1014–1023. https://doi.org/10.1183/09031936.00042012

Huson, D.H., Scornavacca, C., 2012. Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks. Systematic Biology 61, 1061–1067. https://doi.org/10.1093/sysbio/sys062

International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921. https://doi.org/10.1038/35057062

Ito, R., Mustapha, M.M., Tomich, A.D., Callaghan, J.D., McElheny, C.L., Mettus, R.T., Shanks, R.M.Q., Sluis-Cremer, N., Doi, Y., 2017. Widespread Fosfomycin Resistance in Gram-Negative Bacteria Attributable to the Chromosomal fosA Gene. mBio 8. https://doi.org/10.1128/mbio.00749-17

Jaillard, M., Lima, L., Tournoud, M., Mahé, P., van Belkum, A., Lacroix, V., Jacob, L., 2018. A fast and agnostic method for bacterial genome-wide association studies: Bridging the gap between k-mers and genetic events. PLOS Genetics 14, e1007758. https://doi.org/10.1371/journal.pgen.1007758

Jain, M., Ramirez, D., Seshadri, R., Cullina, J.F., Powers, C.A., Schulert, G.S., Bar-Meir, M., Sullivan, C.L., McColley, S.A., Hauser, A.R., 2004. Type III Secretion Phenotypes of Pseudomonas aeruginosa Strains Change during Infection of Individuals with Cystic Fibrosis. Journal of Clinical Microbiology 42, 5229–5237. https://doi.org/10.1128/jcm.42.11.5229-5237.2004

Jani, M., Mathee, K., Azad, R.K., 2016. Identification of Novel Genomic Islands in Liverpool Epidemic Strain of Pseudomonas aeruginosa Using Segmentation and Clustering. Frontiers in Microbiology 7. https://doi.org/10.3389/fmicb.2016.01210

Javed, M., Ueltzhoeffer, V., Heinrich, M., Siegrist, H.J., Wildermuth, R., Lorenz, F.-R., Neher, R.A., Willmann, M., 2018. Colistin susceptibility test evaluation of multiple-resistance-level Pseudomonas aeruginosa isolates generated in a morbidostat device. Journal of Antimicrobial Chemotherapy. https://doi.org/10.1093/jac/dky337

Jeukens, J., Boyle, B., Kukavica-Ibrulj, I., Ouellet, M.M., Aaron, S.D., Charette, S.J., Fothergill, J.L., Tucker, N.P., Winstanley, C., Levesque, R.C., 2014. Comparative Genomics of Isolates of a Pseudomonas aeruginosa Epidemic Strain Associated with Chronic Lung Infections of Cystic Fibrosis Patients. PLoS ONE 9, e87611. https://doi.org/10.1371/journal.pone.0087611

Jeukens, J., Freschi, L., Kukavica-Ibrulj, I., Emond-Rheault, J.-G., Allard, C., Barbeau, J., Cantin, A., Charette, S.J., Déziel, E., Malouin, F., Milot, J., Nguyen, D., Popa, C., Boyle, B., Levesque, R.C., 2019. The Pseudomonas aeruginosa Population among Cystic Fibrosis Patients in Quebec, Canada: a Disease Hot Spot without Known Epidemic Isolates. Journal of Clinical Microbiology 57. https://doi.org/10.1128/jcm.02019-18

Jia, B., Raphenya, A.R., Alcock, B., Waglechner, N., Guo, P., Tsang, K.K., Lago, B.A., Dave, B.M., Pereira, S., Sharma, A.N., Doshi, S., Courtot, M., Lo, R., Williams, L.E., Frye, J.G., Elsayegh, T., Sardar, D., Westman, E.L., Pawlowski, A.C., Johnson, T.A., Brinkman, F.S.L., Wright, G.D., McArthur, A.G., 2017. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Research 45, D566–D573. https://doi.org/10.1093/nar/gkw1004

Jochumsen, N., Marvig, R.L., Damkiær, S., Jensen, R.L., Paulander, W., Molin, S., Jelsbak, L., Folkesson, A., 2016. The evolution of antimicrobial peptide resistance in Pseudomonas aeruginosa is shaped by strong epistatic interactions. Nature Communications 7. https://doi.org/10.1038/ncomms13002

Johnson, C.M., Grossman, A.D., 2015. Integrative and Conjugative Elements (ICEs): What They Do and How They Work. Annual Review of Genetics 49, 577–601. https://doi.org/10.1146/annurev-genet-112414-055018

Jolley, K.A., Maiden, M.C., 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11. https://doi.org/10.1186/1471-2105-11-595

Jombart, T., Balloux, F., Dray, S., 2010. adephylo: new tools for investigating the phylogenetic signal in biological traits. Bioinformatics 26, 1907–1909. https://doi.org/10.1093/bioinformatics/btq292

Jorth, P., McLean, K., Ratjen, A., Secor, P.R., Bautista, G.E., Ravishankar, S., Rezayat, A., Garudathri, J., Harrison, J.J., Harwood, R.A., Penewit, K., Waalkes, A., Singh, P.K., Salipante, S.J., 2017. Evolved Aztreonam Resistance Is Multifactorial and Can Produce Hypervirulence in Pseudomonas aeruginosa. mBio 8. https://doi.org/10.1128/mbio.00517-17

Kainuma, A., Momiyama, K., Kimura, T., Akiyama, K., Inoue, K., Naito, Y., Kinoshita, M., Shimizu, M., Kato, H., Shime, N., Fujita, N., Sawa, T., 2018. An outbreak of fluoroquinolone-resistant Pseudomonas

aeruginosa ST357 harboring the exoU gene. Journal of Infection and Chemotherapy 24, 615–622. https://doi.org/10.1016/j.jiac.2018.03.008

Kang, D., Kirienko, N.V., 2018. Interdependence between iron acquisition and biofilm formation in Pseudomonas aeruginosa. Journal of Microbiology 56, 449–457. https://doi.org/10.1007/s12275-018-8114-3

Kapoor, P., Murphy, P., 2018. Combination antibiotics against Pseudomonas aeruginosa, representing common and rare cystic fibrosis strains from different Irish clinics. Heliyon 4, e00562. https://doi.org/10.1016/j.heliyon.2018.e00562

Kaur, A.P., Lansky, I.B., Wilks, A., 2008. The Role of the Cytoplasmic Heme-binding Protein (PhuS) of Pseudomonas aeruginosa in Intracellular Heme Trafficking and Iron Homeostasis. Journal of Biological Chemistry 284, 56–66. https://doi.org/10.1074/jbc.m806068200

Kaye, K., Pragasam, A., Veeraraghavan, B., Nalini, E., Anandan, S., 2018. An update on antimicrobial resistance and the role of newer antimicrobial agents for Pseudomonas aeruginosa. Indian Journal of Medical Microbiology 36, 303. https://doi.org/10.4103/ijmm.ijmm_18_334

Kazmierczak, K.M., Rabine, S., Hackel, M., McLaughlin, R.E., Biedenbach, D.J., Bouchillon, S.K., Sahm, D.F., Bradford, P.A., 2015. Multiyear, Multinational Survey of the Incidence and Global Distribution of Metallo-β-Lactamase-Producing Enterobacteriaceae and Pseudomonas aeruginosa. Antimicrobial Agents and Chemotherapy 60, 1067–1078. https://doi.org/10.1128/aac.02379-15

Keogh, R.H., Szczesniak, R., Taylor-Robinson, D., Bilton, D., 2018. Up-to-date and projected estimates of survival for people with cystic fibrosis using baseline characteristics: A longitudinal study using UK patient registry data. Journal of Cystic Fibrosis 17, 218–227. https://doi.org/10.1016/j.jcf.2017.11.019

Khademi, S.M.H., Sazinas, P., Jelsbak, L., 2019. Within-Host Adaptation Mediated by Intergenic Evolution in Pseudomonas aeruginosa. Genome Biology and Evolution 11, 1385–1397. https://doi.org/10.1093/gbe/evz083

Khan, A.U., Maryam, L., Zarrilli, R., 2017. Structure, Genetics and Worldwide Spread of New Delhi Metallo-β-lactamase (NDM): a threat to public health. BMC Microbiology 17. https://doi.org/10.1186/s12866-017-1012-8

Kidd, T.J., Ritchie, S.R., Ramsay, K.A., Grimwood, K., Bell, S.C., Rainey, P.B., 2012. Pseudomonas aeruginosa Exhibits Frequent Recombination, but Only a Limited Association between Genotype and Ecological Setting. PLoS ONE 7, e44199. https://doi.org/10.1371/journal.pone.0044199

Kiser, T.H., Obritsch, M.D., Jung, R., MacLaren, R., Fish, D.N., 2010. Efflux Pump Contribution to Multidrug Resistance in Clinical Isolates of Pseudomonas aeruginosa. Pharmacotherapy 30, 632–638. https://doi.org/10.1592/phco.30.7.632

Klockgether, J., Cramer, N., Wiehlmann, L., Davenport, C.F., Tümmler, B., 2011. Pseudomonas aeruginosa Genomic Structure and Diversity. Frontiers in Microbiology 2. https://doi.org/10.3389/fmicb.2011.00150

Klontz, E.H., Tomich, A.D., Günther, S., Lemkul, J.A., Deredge, D., Silverstein, Z., Shaw, J.F., McElheny, C., Doi, Y., Wintrode, P.L., MacKerell, A.D., Jr., Sluis-Cremer, N., Sundberg, E.J., 2017. Structure and Dynamics of FosA-Mediated Fosfomycin Resistance in Klebsiella pneumoniae and Escherichia coli. Antimicrobial Agents and Chemotherapy 61. https://doi.org/10.1128/aac.01572-17

Kos, V.N., Déraspe, M., McLaughlin, R.E., Whiteaker, J.D., Roy, P.H., Alm, R.A., Corbeil, J., Gardner, H., 2014. The Resistome of Pseudomonas aeruginosa in Relationship to Phenotypic Susceptibility. Antimicrobial Agents and Chemotherapy 59, 427–436. https://doi.org/10.1128/aac.03954-14

Kos, V.N., McLaughlin, R.E., Gardner, H.A., 2016. Elucidation of Mechanisms of Ceftazidime Resistance among Clinical Isolates of Pseudomonas aeruginosa by Using Genomic Data. Antimicrobial Agents and Chemotherapy 60, 3856–3861. https://doi.org/10.1128/aac.03113-15

Kumari, H., Murugapiran, S.K., Balasubramanian, D., Schneper, L., Merighi, M., Sarracino, D., Lory, S., Mathee, K., 2014. LTQ-XL mass spectrometry proteome analysis expands the Pseudomonas aeruginosa AmpR regulon to include cyclic di-GMP phosphodiesterases and phosphoproteins, and identifies novel open reading frames. Journal of Proteomics 96, 328–342. https://doi.org/10.1016/j.jprot.2013.11.018

Kung, V.L., Ozer, E.A., Hauser, A.R., 2010. The Accessory Genome of Pseudomonas aeruginosa. Microbiology and Molecular Biology Reviews 74, 621–641. https://doi.org/10.1128/mmbr.00027-10

Kwan, T., Liu, J., DuBow, M., Gros, P., Pelletier, J., Comparative Genomic Analysis of 18 Pseudomonas aeruginosa Bacteriophages. Journal of Bacteriology 3, 1184 - 1187.

La Rosa, R., Johansen, H.K., Molin, S., 2018. Convergent Metabolic Specialization through Distinct Evolutionary Paths in Pseudomonas aeruginosa. mBio 9. https://doi.org/10.1128/mbio.00269-18

Laarman, A.J., Bardoel, B.W., Ruyken, M., Fernie, J., Milder, F.J., van Strijp, J.A.G., Rooijakkers, S.H.M., 2011. Pseudomonas aeruginosa Alkaline Protease Blocks Complement Activation via the Classical and Lectin Pathways. The Journal of Immunology 188, 386–393. https://doi.org/10.4049/jimmunol.1102162

LaFayette, S.L., Houle, D., Beaudoin, T., Wojewodka, G., Radzioch, D., Hoffman, L.R., Burns, J.L., Dandekar, A.A., Smalley, N.E., Chandler, J.R., Zlosnik, J.E., Speert, D.P., Bernier, J., Matouk, E., Brochiero, E., Rousseau, S., Nguyen, D., 2015. Cystic fibrosis–adapted Pseudomonas aeruginosa quorum sensing lasR mutants cause hyperinflammatory responses. Science Advances 1, e1500199. https://doi.org/10.1126/sciadv.1500199

Lappalainen, T., Scott, A.J., Brandt, M., Hall, I.M., 2019. Genomic Analysis in the Age of Human Genome Sequencing. Cell 177, 70–84. https://doi.org/10.1016/j.cell.2019.02.032

Latifi, A., Foglino, M., Tanaka, K., Williams, P., Lazdunski, A., 1996. A hierarchical quorum-sensing cascade in Pseudomonas aeruginosa links the transcriptional activators LasR and RhlR (VsmR) to expression of the stationary-phase sigma factor RpoS. Molecular Microbiology 21, 1137–1146. https://doi.org/10.1046/j.1365-2958.1996.00063.x

Laval, J., Ralhan, A., Hartl, D., 2016. Neutrophils in cystic fibrosis. Biological Chemistry 397. https://doi.org/10.1515/hsz-2015-0271

Lee, S.-W., Glickmann, E., Cooksey, D.A., 2001. Chromosomal Locus for Cadmium Resistance in Pseudomonas putida Consisting of a Cadmium-Transporting ATPase and a MerR Family Response Regulator. Applied and Environmental Microbiology 67, 1437–1444. https://doi.org/10.1128/aem.67.4.1437-1444.2001

Lee, J.-Y., Park, Y.K., Chung, E.S., Na, I.Y., Ko, K.S., 2016. Evolved resistance to colistin and its loss due to genetic reversion in Pseudomonas aeruginosa. Scientific Reports 6. https://doi.org/10.1038/srep25543

Lee, J., Zhang, L., 2015. The hierarchy quorum sensing network in Pseudomonas aeruginosa. Protein & Cell 6, 26–41. https://doi.org/10.1007/s13238-014-0100-x

Lee, S.A., Gallagher, L.A., Thongdee, M., Staudinger, B.J., Lippman, S., Singh, P.K., Manoil, C., 2015. General and condition-specific essential functions of Pseudomonas aeruginosa. Proceedings of the National Academy of Sciences 112, 5189–5194. https://doi.org/10.1073/pnas.1422186112

Lee, Y., Song, S., Sheng, L., Zhu, L., Kim, J.-S., Wood, T.K., 2018. Substrate Binding Protein DppA1 of ABC Transporter DppBCDF Increases Biofilm Formation in Pseudomonas aeruginosa by Inhibiting Pf5 Prophage Lysis. Frontiers in Microbiology 9. https://doi.org/10.3389/fmicb.2018.00030

Lees, J. A., Croucher, N. J., Goldblatt, D., Nosten, F., Parkhill, J., Turner, C., Turner, P., Bentley, S. D., 2017. Genome-wide identification of lineage and locus specific variation associated with pneumococcal carriage duration. Elife 6;e26255 10.7554/eLife.26255.

Lees, J.A., Bentley, S.D., 2016. Bacterial GWAS: not just gilding the lily. Nature Reviews Microbiology 14, 406–406. https://doi.org/10.1038/nrmicro.2016.82

Lees, J.A., Ferwerda, B., Kremer, P.H.C., Wheeler, N.E., Serón, M.V., Croucher, N.J., Gladstone, R.A., Bootsma, H.J., Rots, N.Y., Wijmega-Monsuur, A.J., Sanders, E.A.M., Trzciński, K., Wyllie, A.L., Zwinderman, A.H., van den Berg, L.H., van Rheenen, W., Veldink, J.H., Harboe, Z.B., Lundbo, L.F., de Groot, L.C.P.G.M., van Schoor, N.M., van der Velde, N., Ängquist, L.H., Sørensen, T.I.A., Nohr, E.A., Mentzer, A.J., Mills, T.C., Knight, J.C., du Plessis, M., Nzenze, S., Weiser, J.N., Parkhill, J., Madhi, S., Benfield, T., von Gottberg, A., van der Ende, A., Brouwer, M.C., Barrett, J.C., Bentley, S.D., van de Beek, D., 2019. Joint sequencing of human and pathogen genomes reveals the genetics of pneumococcal meningitis. Nature Communications 10. https://doi.org/10.1038/s41467-019-09976-3

Lees, J.A., Galardini, M., Bentley, S.D., Weiser, J.N., Corander, J., 2018. pyseer: a comprehensive tool for microbial pangenome-wide association studies. Bioinformatics 34, 4310–4312. https://doi.org/10.1093/bioinformatics/bty539

Leid, J.G., Kerr, M., Selgado, C., Johnson, C., Moreno, G., Smith, A., Shirtliff, M.E., O'Toole, G.A., Cope, E.K., 2009. Flagellum-Mediated Biofilm Defense Mechanisms of Pseudomonas aeruginosa against Host-Derived Lactoferrin. Infection and Immunity 77, 4559–4566. https://doi.org/10.1128/iai.00075-09

Letunic, I., Bork, P., 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Research. https://doi.org/10.1093/nar/gkz239

Levitt, D., Levitt, M., 2016. Human serum albumin homeostasis: a new look at the roles of synthesis, catabolism, renal and gastrointestinal excretion, and the clinical value of serum albumin measurements. International Journal of General Medicine Volume 9, 229–255. https://doi.org/10.2147/ijgm.s102819

Li, X.-Z., Poole, K., Nikaido, H., 2003. Contributions of MexAB-OprM and an EmrE Homolog to Intrinsic Resistance of Pseudomonas aeruginosa to Aminoglycosides and Dyes. Antimicrobial Agents and Chemotherapy 47, 27–33. https://doi.org/10.1128/aac.47.1.27-33.2003

Li, Y., 2003. A new member of the tripartite multidrug efflux pumps,MexVW-OprM, in Pseudomonas aeruginosa. Journal of Antimicrobial Chemotherapy 52, 572–575. https://doi.org/10.1093/jac/dkg390

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, H., Luo, Y.-F., Williams, B.J., Blackwell, T.S., Xie, C.-M., 2012. Structure and function of OprD protein in Pseudomonas aeruginosa: From antibiotic resistance to novel therapies. International Journal of Medical Microbiology 302, 63–68. https://doi.org/10.1016/j.ijmm.2011.10.001

Li, X.-Z., Plésiat, P., Nikaido, H., 2014. The Challenge of Efflux-Mediated Antibiotic Resistance in Gram-Negative Bacteria. Clinical Microbiology Reviews 28, 337–418. https://doi.org/10.1128/cmr.00117-14

Li, Y., Bai, F., Xia, H., Zhuang, L., Xu, H., Jin, Y., Zhang, X., Bai, Y., Qiao, M., 2015a. A novel regulator PA5022 (aefA) is involved in swimming motility, biofilm formation and elastase activity of Pseudomonas aeruginosa. Microbiological Research 176, 14–20. https://doi.org/10.1016/j.micres.2015.04.001

Li, S., Li, T., Xu, Y., Zhang, Q., Zhang, W., Che, S., Liu, R., Wang, Y., Bartlam, M., 2015b. Structural insights into YfiR sequestering by YfiB in Pseudomonas aeruginosa PAO1. Scientific Reports 5. https://doi.org/10.1038/srep16915

Li, X.-Z., Plésiat, P., 2016. Antimicrobial Drug Efflux Pumps in Pseudomonas aeruginosa, in: Efflux-Mediated Antimicrobial Resistance in Bacteria. Springer International Publishing, pp. 359–400. https://doi.org/10.1007/978-3-319-39658-3_14

Lim, L.M., Ly, N., Anderson, D., Yang, J.C., Macander, L., Jarkowski, A., III, Forrest, A., Bulitta, J.B., Tsuji, B.T., 2010. Resurgence of Colistin: A Review of Resistance, Toxicity, Pharmacodynamics, and Dosing. Pharmacotherapy 30, 1279–1291. https://doi.org/10.1592/phco.30.12.1279

Lim, M., Wallis, C., Price, J.F., Carr, S.B., Chavasse, R.J., Shankar, A., Seddon, P., Balfour-Lynn, I.M., 2013. Diagnosis of cystic fibrosis in London and South East England before and after the introduction of newborn screening. Archives of Disease in Childhood 99, 197–202. https://doi.org/10.1136/archdischild-2013-304766

Lin, C.K., Kazmierczak, B.I., 2017. Inflammation: A Double-Edged Sword in the Response to Pseudomonas aeruginosa Infection. Journal of Innate Immunity 9, 250–261. https://doi.org/10.1159/000455857

Lin, J., Cheng, J., Wang, Y., Shen, X., 2018. The Pseudomonas Quinolone Signal (PQS): Not Just for Quorum Sensing Anymore. Frontiers in Cellular and Infection Microbiology 8. https://doi.org/10.3389/fcimb.2018.00230

Lister, P.D., Gardner, V.M., Sanders, C.C., 1999. Clavulanate Induces Expression of the Pseudomonas aeruginosa AmpC Cephalosporinase at Physiologically Relevant Concentrations and Antagonizes the Antibacterial Activity of Ticarcillin. Antimicrobial Agents and Chemotherapy 43, 882–889. https://doi.org/10.1128/aac.43.4.882

Liu, B., Zheng, D., Jin, Q., Chen, L., Yang, J., 2019. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucleic Acids Research 47, D687–D692. https://doi.org/10.1093/nar/gky1080

Liu, X., Long, D., You, H., Yang, D., Zhou, S., Zhang, S., Li, M., He, M., Xiong, M., Wang, X., 2016. Phosphatidylcholine affects the secretion of the alkaline phosphatase PhoA in Pseudomonas strains. Microbiological Research 192, 21–29. https://doi.org/10.1016/j.micres.2016.02.001

Llanes, C., Köhler, T., Patry, I., Dehecq, B., van Delden, C., Plésiat, P., 2011. Role of the MexEF-OprN Efflux System in Low-Level Resistance of Pseudomonas aeruginosa to Ciprofloxacin. Antimicrobial Agents and Chemotherapy 55, 5676–5684. https://doi.org/10.1128/aac.00101-11

Lomovskaya, O., Lee, A., Hoshino, K., Ishida, H., Mistry, A., Warren, M.S., Boyer, E., Chamberland, S., Lee, V.J., 1999. Use of a Genetic Approach To Evaluate the Consequences of Inhibition of Efflux Pumps in Pseudomonas aeruginosa. Antimicrobial Agents and Chemotherapy 43, 1340–1346. https://doi.org/10.1128/aac.43.6.1340

López-Causapé, C., Sommer, L.M., Cabot, G., Rubio, R., Ocampo-Sosa, A.A., Johansen, H.K., Figuerola, J., Cantón, R., Kidd, T.J., Molin, S., Oliver, A., 2017. Evolution of the Pseudomonas aeruginosa

mutational resistome in an international Cystic Fibrosis clone. Scientific Reports 7. https://doi.org/10.1038/s41598-017-05621-5

Lozano, C., Azcona-Gutiérrez, J.M., Van Bambeke, F., Sáenz, Y., 2018. Great phenotypic and genetic variation among successive chronic Pseudomonas aeruginosa from a cystic fibrosis patient. PLOS ONE 13, e0204167. https://doi.org/10.1371/journal.pone.0204167

Lu, C.-D., 2006. Pathways and regulation of bacterial arginine metabolism and perspectives for obtaining arginine overproducing strains. Applied Microbiology and Biotechnology 70, 261–272. https://doi.org/10.1007/s00253-005-0308-z

Lu, C.-D., Itoh, Y., Nakada, Y., Jiang, Y., 2002. Functional Analysis and Regulation of the Divergent spuABCDEFGH-spuI Operons for Polyamine Uptake and Utilization in Pseudomonas aeruginosa PAO1. Journal of Bacteriology 184, 3765–3773. https://doi.org/10.1128/jb.184.14.3765-3773.2002

Lu, C.-D., Yang, Z., Li, W., 2004. Transcriptome Analysis of the ArgR Regulon in Pseudomonas aeruginosa. Journal of Bacteriology 186, 3855–3861. https://doi.org/10.1128/jb.186.12.3855-3861.2004

Lu, D., Shang, G., Zhang, H., Yu, Q., Cong, X., Yuan, J., He, F., Zhu, C., Zhao, Y., Yin, K., Chen, Y., Hu, J., Zhang, X., Yuan, Z., Xu, S., Hu, W., Cang, H., Gu, L., 2014. Structural insights into the T6SS effector protein Tse3 and the Tse3-Tsi3 complex from Pseudomonas aeruginosa reveal a calcium-dependent membrane-binding mechanism. Molecular Microbiology 92, 1092–1112. https://doi.org/10.1111/mmi.12616

Lukacs, G.L., Verkman, A.S., 2012. CFTR: folding, misfolding and correcting the ΔF508 conformational defect. Trends in Molecular Medicine 18, 81–91. https://doi.org/10.1016/j.molmed.2011.10.003

Ly, N.S., Yang, J., Bulitta, J.B., Tsuji, B.T., 2012. Impact of Two-Component Regulatory Systems PhoP-PhoQ and PmrA-PmrB on Colistin Pharmacodynamics in Pseudomonas aeruginosa. Antimicrobial Agents and Chemotherapy 56, 3453–3456. https://doi.org/10.1128/aac.06380-11

Lyczak, J.B., Cannon, C.L., Pier, G.B., 2002. Lung Infections Associated with Cystic Fibrosis. Clinical Microbiology Reviews 15, 194–222. https://doi.org/10.1128/cmr.15.2.194-222.2002

M. Ketelboeter, L., Y. Potharla, V., L. Bardy, S., 2014. NTBC Treatment of the Pyomelanogenic Pseudomonas aeruginosa Clinical Isolate PA1111 Inhibits Pigment Production and Increases Sensitivity to Oxidative Stress. Current Microbiology 69, 343–348. https://doi.org/10.1007/s00284-014-0593-9

Macia, M.D., Blanquer, D., Togores, B., Sauleda, J., Perez, J.L., Oliver, A., 2005. Hypermutation Is a Key Factor in Development of Multiple-Antimicrobial Resistance in Pseudomonas aeruginosa Strains Causing Chronic Lung Infections. Antimicrobial Agents and Chemotherapy 49, 3382–3386. https://doi.org/10.1128/aac.49.8.3382-3386.2005

Maeda, Y., Murayama, M., Goldsmith, C.E., Coulter, W.A., Mason, C., Millar, B.C., Dooley, J.S.G., Lowery, C.J., Matsuda, M., Rendall, J.C., Elborn, J.S., Moore, J.E., 2010. Molecular characterization and phylogenetic analysis of quinolone resistance-determining regions (QRDRs) of gyrA, gyrB, parC and

parE gene loci in viridans group streptococci isolated from adult patients with cystic fibrosis. Journal of Antimicrobial Chemotherapy 66, 476–486. https://doi.org/10.1093/jac/dkq485

Mahenthiralingam, E., Campbell, M. E., Speert, D. P., 1994. Nonmotility and phagocytic resistance of Pseudomonas-aeruginosa isolates from chronically colonized patients with cystic-fibrosis. Infection and Immunity 62 (2), 596-605.

Maiuri, L., Raia, V., Kroemer, G., 2017. Strategies for the etiological therapy of cystic fibrosis. Cell Death & Differentiation 24, 1825–1844. https://doi.org/10.1038/cdd.2017.126

Malhotra, S., Limoli, D.H., English, A.E., Parsek, M.R., Wozniak, D.J., 2018. Mixed Communities of Mucoid and Nonmucoid Pseudomonas aeruginosa Exhibit Enhanced Resistance to Host Antimicrobials. mBio 9. https://doi.org/10.1128/mbio.00275-18

Mall, M.A., Hartl, D., 2014. CFTR: cystic fibrosis and beyond. European Respiratory Journal 44, 1042–1054. https://doi.org/10.1183/09031936.00228013

Malone, J.G., Jaeger, T., Spangler, C., Ritz, D., Spang, A., Arrieumerlou, C., Kaever, V., Landmann, R., Jenal, U., 2010. YfiBNR Mediates Cyclic di-GMP Dependent Small Colony Variant Formation and Persistence in Pseudomonas aeruginosa. PLoS Pathogens 6, e1000804. https://doi.org/10.1371/journal.ppat.1000804

Mann, E.E., Wozniak, D.J., 2012. Pseudomonas biofilm matrix composition and niche biology. FEMS Microbiology Reviews 36, 893–916. https://doi.org/10.1111/j.1574-6976.2011.00322.x

Mantere, T., Kersten, S., Hoischen, A., 2019. Long-Read Sequencing Emerging in Medical Genetics. Frontiers in Genetics 10. https://doi.org/10.3389/fgene.2019.00426

Markussen, T., Marvig, R.L., Gómez-Lozano, M., Aanæs, K., Burleigh, A.E., Høiby, N., Johansen, H.K., Molin, S., Jelsbak, L., 2014. Environmental Heterogeneity Drives Within-Host Diversification and Evolution of Pseudomonas aeruginosa. mBio 5. https://doi.org/10.1128/mbio.01592-14

Marson, F.A.L., Bertuzzo, C.S., Ribeiro, J.D., 2016. Classification of CFTR mutation classes. The Lancet Respiratory Medicine 4, e37–e38. https://doi.org/10.1016/s2213-2600(16)30188-6

Martha, B., Croisier, D., Durand, D., Hocquet, D., Plesiat, P., Piroth, L., Portier, H., Chavanet, P., 2006. In-vivo impact of the MexXY efflux system on aminoglycoside efficacy in an experimental model of Pseudomonas aeruginosa pneumonia treated with tobramycin. Clinical Microbiology and Infection 12, 426–432. https://doi.org/10.1111/j.1469-0691.2006.01371.x

Martin, C., Hamard, C., Kanaan, R., Boussaud, V., Grenet, D., Abély, M., Hubert, D., Munck, A., Lemonnier, L., Burgel, P.-R., 2016. Causes of death in French cystic fibrosis patients: The need for improvement in transplantation referral strategies! Journal of Cystic Fibrosis 15, 204–212. https://doi.org/10.1016/j.jcf.2015.09.002

Martin, D.W., Schurr, M.J., Yu, H., Deretic, V., 1994. Analysis of promoters controlled by the putative sigma factor AlgU regulating conversion to mucoidy in Pseudomonas aeruginosa: relationship to sigma E and stress response. Journal of Bacteriology 176, 6688–6696. https://doi.org/10.1128/jb.176.21.6688-6696.1994

Martinez, J.L., Sánchez, M.B., Martínez-Solano, L., Hernandez, A., Garmendia, L., Fajardo, A., Alvarez-Ortega, C., 2009. Functional role of bacterial multidrug efflux pumps in microbial natural ecosystems. FEMS Microbiology Reviews 33, 430–449. https://doi.org/10.1111/j.1574-6976.2008.00157.x

Marvig, R.L., Johansen, H.K., Molin, S., Jelsbak, L., 2013. Genome Analysis of a Transmissible Lineage of Pseudomonas aeruginosa Reveals Pathoadaptive Mutations and Distinct Evolutionary Paths of Hypermutators. PLoS Genetics 9, e1003741. https://doi.org/10.1371/journal.pgen.1003741

Marvig, R.L., Damkiær, S., Khademi, S.M.H., Markussen, T.M., Molin, S., Jelsbak, L., 2014. Within-Host Evolution of Pseudomonas aeruginosa Reveals Adaptation toward Iron Acquisition from Hemoglobin. mBio 5. https://doi.org/10.1128/mbio.00966-14

Marvig, R.L., Sommer, L.M., Molin, S., Johansen, H.K., 2015. Convergent evolution and adaptation of Pseudomonas aeruginosa within patients with cystic fibrosis. Nature Genetics 47, 57–64. https://doi.org/10.1038/ng.3148

Masuda, N., Sakagawa, E., Ohya, S., Gotoh, N., Tsujimoto, H., Nishino, T., 2000. Substrate Specificities of MexAB-OprM, MexCD-OprJ, and MexXY-OprM Efflux Pumps in Pseudomonas aeruginosa. Antimicrobial Agents and Chemotherapy 44, 3322–3327. https://doi.org/10.1128/aac.44.12.3322-3327.2000

Mathee, K., Narasimhan, G., Valdes, C., Qiu, X., Matewish, J.M., Koehrsen, M., Rokas, A., Yandava, C.N., Engels, R., Zeng, E., Olavarietta, R., Doud, M., Smith, R.S., Montgomery, P., White, J.R., Godfrey, P.A., Kodira, C., Birren, B., Galagan, J.E., Lory, S., 2008. Dynamics of Pseudomonas aeruginosa genome evolution. Proceedings of the National Academy of Sciences 105, 3100–3105. https://doi.org/10.1073/pnas.0711982105

Maunders, E., Welch, M., 2017. Matrix exopolysaccharides; the sticky side of biofilm formation. FEMS Microbiology Letters 364. https://doi.org/10.1093/femsle/fnx120

Maura, D., Hazan, R., Kitao, T., Ballok, A.E., Rahme, L.G., 2016. Evidence for Direct Control of Virulence and Defense Gene Circuits by the Pseudomonas aeruginosa Quorum Sensing Regulator, MvfR. Scientific Reports 6. https://doi.org/10.1038/srep34083

Mayer, M., 2015. Lumacaftor-ivacaftor (Orkambi) for cystic fibrosis: behind the 'breakthrough'. Evidence Based Medicine 21, 83–86. https://doi.org/10.1136/ebmed-2015-110325

McCallum, S.J., Corkill, J., Gallagher, M., Ledson, M.J., Anthony Hart, C., Walshaw, M.J., 2001. Superinfection with a transmissible strain of Pseudomonas aeruginosa in adults with cystic fibrosis

chronically colonised by P aeruginosa. The Lancet 358, 558–560. https://doi.org/10.1016/s0140-6736(01)05715-4

McCarthy, K.L., Paterson, D.L., 2017. Long-term mortality following Pseudomonas aeruginosa bloodstream infection. Journal of Hospital Infection 95, 292–299. https://doi.org/10.1016/j.jhin.2016.11.014

McCormick, J., Green, M.W., Mehta, G., Culross, F., Mehta, A., 2002. Demographics of the UK cystic fibrosis population: implications for neonatal screening. European Journal of Human Genetics 10, 583–590. https://doi.org/10.1038/sj.ejhg.5200850

McCurtain, J.L., Gilbertsen, A.J., Evert, C., Williams, B.J., Hunter, R.C., 2019. Agmatine accumulation by Pseudomonas aeruginosa clinical isolates confers antibiotic tolerance and dampens host inflammation. Journal of Medical Microbiology 68, 446–455. https://doi.org/10.1099/jmm.0.000928

McInerney, J.O., McNally, A., O'Connell, M.J., 2017. Why prokaryotes have pangenomes. Nature Microbiology 2. https://doi.org/10.1038/nmicrobiol.2017.40

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20, 1297–1303. https://doi.org/10.1101/gr.107524.110

Medina, G., Juarez, K., Valderrama, B., Soberon-Chavez, G., 2003. Mechanism of Pseudomonas aeruginosa RhlR Transcriptional Regulation of the rhlAB Promoter. Journal of Bacteriology 185, 5976–5983. https://doi.org/10.1128/jb.185.20.5976-5983.2003

Meletis, G., Exindari, M., Vavatsi, N., Sofianou, D., Diza, E., 2012. Mechanisms responsible for the emergence of carbapenem resistance in Pseudomonas aeruginosa. Hippokratia 16 (4), 303-307.

Mena, A., Smith, E.E., Burns, J.L., Speert, D.P., Moskowitz, S.M., Perez, J.L., Oliver, A., 2008. Genetic Adaptation of Pseudomonas aeruginosa to the Airways of Cystic Fibrosis Patients Is Catalyzed by Hypermutation. Journal of Bacteriology 190, 7910–7917. https://doi.org/10.1128/jb.01147-08

Meneely, K.M., Luo, Q., Dhar, P., Lamb, A.L., 2013. Lysine221 is the general base residue of the isochorismate synthase from Pseudomonas aeruginosa (PchA) in a reaction that is diffusion limited. Archives of Biochemistry and Biophysics 538, 49–56. https://doi.org/10.1016/j.abb.2013.07.026

Mern, D.S., Ha, S.W., Khodaverdi, V., Gliese, N., Gorisch, H., 2010. A complex regulatory network controls aerobic ethanol oxidation in Pseudomonas aeruginosa: indication of four levels of sensor kinases and response regulators. Microbiology 156, 1505–1516. https://doi.org/10.1099/mic.0.032847-0

Mesaros, N., Nordmann, P., Plésiat, P., Roussel-Delvallez, M., Van Eldere, J., Glupczynski, Y., Van Laethem, Y., Jacobs, F., Lebecque, P., Malfroot, A., Tulkens, P.M., Van Bambeke, F., 2007.

Pseudomonas aeruginosa: resistance and therapeutic options at the turn of the new millennium. Clinical Microbiology and Infection 13, 560–578. https://doi.org/10.1111/j.1469-0691.2007.01681.x

Micek, S.T., Lloyd, A.E., Ritchie, D.J., Reichley, R.M., Fraser, V.J., Kollef, M.H., 2005. Pseudomonas aeruginosa Bloodstream Infection: Importance of Appropriate Initial Antimicrobial Treatment. Antimicrobial Agents and Chemotherapy 49, 1306–1311. https://doi.org/10.1128/aac.49.4.1306-1311.2005

Michael, V., Frank, O., Bartling, P., Scheuner, C., Göker, M., Brinkmann, H., Petersen, J., 2016. Biofilm plasmids with a rhamnose operon are widely distributed determinants of the 'swim-or-stick' lifestyle in roseobacters. The ISME Journal 10, 2498–2513. https://doi.org/10.1038/ismej.2016.30

Mima, T., Sekiya, H., Mizushima, T., Kuroda, T., Tsuchiya, T., 2005. Gene Cloning and Properties of the RND-Type Multidrug Efflux Pumps MexPQ-OpmE and MexMN-OprM from Pseudomonas aeruginosa. Microbiology and Immunology 49, 999–1002. https://doi.org/10.1111/j.1348-0421.2005.tb03696.x

Möglich, A., Ayers, R.A., Moffat, K., 2009. Structure and Signaling Mechanism of Per-ARNT-Sim Domains. Structure 17, 1282–1294. https://doi.org/10.1016/j.str.2009.08.011

Molinski, S.V., Ahmadi, S., Ip, W., Ouyang, H., Villella, A., Miller, J.P., Lee, P., Kulleperuma, K., Du, K., Di Paola, M., Eckford, P.D., Laselva, O., Huan, L.J., Wellhauser, L., Li, E., Ray, P.N., Pomès, R., Moraes, T.J., Gonska, T., Ratjen, F., Bear, C.E., 2017. Orkambi® and amplifier co‐therapy improves function from a rare CFTR mutation in gene‐edited cells and patient tissue. EMBO Molecular Medicine 9, 1224–1243. https://doi.org/10.15252/emmm.201607137

Moradali, M.F., Ghods, S., Rehm, B.H.A., 2017. Pseudomonas aeruginosa Lifestyle: A Paradigm for Adaptation, Survival, and Persistence. Frontiers in Cellular and Infection Microbiology 7. https://doi.org/10.3389/fcimb.2017.00039

Morales-Espinosa, R., Delgado, G., Espinosa, L.F., Isselo, D., Méndez, J.L., Rodriguez, C., Miranda, G., Cravioto, A., 2017. Fingerprint Analysis and Identification of Strains ST309 as a Potential High Risk Clone in a Pseudomonas aeruginosa Population Isolated from Children with Bacteremia in Mexico City. Frontiers in Microbiology 8. https://doi.org/10.3389/fmicb.2017.00313

Morita, Y., Tomida, J., Kawamura, Y., 2012. MexXY multidrug efflux system of Pseudomonas aeruginosa. Frontiers in Microbiology 3. https://doi.org/10.3389/fmicb.2012.00408

Morita, Y., Tomida, J., Kawamura, Y., 2014. Responses of Pseudomonas aeruginosa to antimicrobials. Frontiers in Microbiology 4. https://doi.org/10.3389/fmicb.2013.00422

Moscoso, J.A., Mikkelsen, H., Heeb, S., Williams, P., Filloux, A., 2011. The Pseudomonas aeruginosa sensor RetS switches Type III and Type VI secretion via c-di-GMP signalling. Environmental Microbiology 13, 3128–3138. https://doi.org/10.1111/j.1462-2920.2011.02595.x

Mosquera-Rendón, J., Rada-Bravo, A.M., Cárdenas-Brito, S., Corredor, M., Restrepo-Pineda, E., Benítez-Páez, A., 2016. Pangenome-wide and molecular evolution analyses of the Pseudomonas aeruginosa species. BMC Genomics 17. https://doi.org/10.1186/s12864-016-2364-4

Mougous, J.D., 2006. A Virulence Locus of Pseudomonas aeruginosa Encodes a Protein Secretion Apparatus. Science 312, 1526–1530. https://doi.org/10.1126/science.1128393

Mowat, E., Paterson, S., Fothergill, J.L., Wright, E.A., Ledson, M.J., Walshaw, M.J., Brockhurst, M.A., Winstanley, C., 2011. Pseudomonas aeruginosa Population Diversity and Turnover in Cystic Fibrosis Chronic Infections. American Journal of Respiratory and Critical Care Medicine 183, 1674–1679. https://doi.org/10.1164/rccm.201009-1430oc

Munck, A., Bonacorsi, S., Mariani‐Kurkdjian, P., Lebourgeois, M., Gérardin, M., Brahimi, N., Navarro, J., Bingen, E., 2001. Genotypic characterization of Pseudomonas aeruginosa strains recovered from patients with cystic fibrosis after initial and subsequent colonization. Pediatric Pulmonology 32, 288–292. https://doi.org/10.1002/ppul.1121

Murray, T.S., Kazmierczak, B.I., 2007. Pseudomonas aeruginosa Exhibits Sliding Motility in the Absence of Type IV Pili and Flagella. Journal of Bacteriology 190, 2700–2708. https://doi.org/10.1128/jb.01620-07

Murugan, N., Malathi, J., Umashankar, V., Madhavan, H.N.R., 2014. Comparative Genomic Analysis of Multidrug-Resistant Pseudomonas aeruginosa Clinical Isolates VRFPA06 and VRFPA08 with VRFPA07. Genome Announcements 2. https://doi.org/10.1128/genomea.00140-14

Nakada, Y., 2003. Identification of the putrescine biosynthetic genes in Pseudomonas aeruginosa and characterization of agmatine deiminase and N-carbamoylputrescine amidohydrolase of the arginine decarboxylase pathway. Microbiology 149, 707–714. https://doi.org/10.1099/mic.0.26009-0

Nakada, Y., 2005. Pseudomonas aeruginosa PAO1 genes for 3-guanidinopropionate and 4-guanidinobutyrate utilization may be derived from a common ancestor. Microbiology 151, 4055–4062. https://doi.org/10.1099/mic.0.28258-0

Nickzad, A., Déziel, E., 2014. The involvement of rhamnolipids in microbial cell adhesion and biofilm development - an approach for control? Letters in Applied Microbiology 58, 447–453. https://doi.org/10.1111/lam.12211

O'Brien, S., Williams, D., Fothergill, J.L., Paterson, S., Winstanley, C., Brockhurst, M.A., 2017. High virulence sub-populations in Pseudomonas aeruginosa long-term cystic fibrosis airway infections. BMC Microbiology 17. https://doi.org/10.1186/s12866-017-0941-6

Ochsner, U.A., Vasil, M.L., Alsabbagh, E., Parvatiyar, K., Hassett, D.J., 2000. Role of the Pseudomonas aeruginosa oxyR-recG Operon in Oxidative Stress Defense and DNA Repair: OxyR-Dependent Regulation of katB-ankB, ahpB, and ahpC-ahpF. Journal of Bacteriology 182, 4533–4544. https://doi.org/10.1128/jb.182.16.4533-4544.2000

Okkotsu, Y., Little, A.S., Schurr, M.J., 2014. The Pseudomonas aeruginosa AlgZR two-component system coordinates multiple phenotypes. Frontiers in Cellular and Infection Microbiology 4. https://doi.org/10.3389/fcimb.2014.00082

Olivares Pacheco, J., Alvarez-Ortega, C., Alcalde Rico, M., Martínez, J.L., 2017. Metabolic Compensation of Fitness Costs Is a General Outcome for Antibiotic-Resistant Pseudomonas aeruginosa Mutants Overexpressing Efflux Pumps. mBio 8. https://doi.org/10.1128/mbio.00500-17

Olivares, E., Badel-Berchoux, S., Provot, C., Jaulhac, B., Prévost, G., Bernardi, T., Jehl, F., 2017. Tobramycin and Amikacin Delay Adhesion and Microcolony Formation in Pseudomonas aeruginosa Cystic Fibrosis Isolates. Frontiers in Microbiology 8. https://doi.org/10.3389/fmicb.2017.01289

Oliver, A., 2000. High Frequency of Hypermutable Pseudomonas aeruginosa in Cystic Fibrosis Lung Infection. Science 288, 1251–1253. https://doi.org/10.1126/science.288.5469.1251

Oliver, A., Mena, A., 2010. Bacterial hypermutation in cystic fibrosis, not only for antibiotic resistance. Clinical Microbiology and Infection 16, 798–808. https://doi.org/10.1111/j.1469-0691.2010.03250.x

Oliver, A., 2015. Clinical relevance of Pseudomonas aeruginosa hypermutation in cystic fibrosis chronic respiratory infection. Journal of Cystic Fibrosis 14, e1–e2. https://doi.org/10.1016/j.jcf.2014.12.009

Oliver, A., Mulet, X., López-Causapé, C., Juan, C., 2015. The increasing threat of Pseudomonas aeruginosa high-risk clones. Drug Resistance Updates 21–22, 41–59. https://doi.org/10.1016/j.drup.2015.08.002

Ozer, E.A., Nnah, E., Didelot, X., Whitaker, R.J., Hauser, A.R., 2019. The Population Structure of Pseudomonas aeruginosa Is Characterized by Genetic Isolation of exoU+ and exoS+ Lineages. Genome Biology and Evolution 11, 1780–1796. https://doi.org/10.1093/gbe/evz119

Padoa, C., Goldman, A., Jenkins, T., Ramsay, M., 1999. Cystic fibrosis carrier frequencies in populations of African origin. Journal of Medical Genetics (J Med Genet), 36. 10.1136/jmg.36.1.41

Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., Fookes, M., Falush, D., Keane, J.A., Parkhill, J., 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31, 3691–3693. https://doi.org/10.1093/bioinformatics/btv421

Page, A.J., Taylor, B., Delaney, A.J., Soares, J., Seemann, T., Keane, J.A., Harris, S.R., 2016. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microbial Genomics 2. https://doi.org/10.1099/mgen.0.000056

Pagel, M., 1999. Inferring the historical patterns of biological evolution. Nature 401, 877–884. https://doi.org/10.1038/44766

Palmer, G.C., Palmer, K.L., Jorth, P.A., Whiteley, M., 2010. Characterization of the Pseudomonas aeruginosa Transcriptional Response to Phenylalanine and Tyrosine. Journal of Bacteriology 192, 2722–2728. https://doi.org/10.1128/jb.00112-10

Palmer, K.L., Aye, L.M., Whiteley, M., 2007. Nutritional Cues Control Pseudomonas aeruginosa Multicellular Behavior in Cystic Fibrosis Sputum. Journal of Bacteriology 189, 8079–8087. https://doi.org/10.1128/jb.01138-07

Pamp, S.J., Tolker-Nielsen, T., 2007. Multiple Roles of Biosurfactants in Structural Biofilm Development by Pseudomonas aeruginosa. Journal of Bacteriology 189, 2531–2539. https://doi.org/10.1128/jb.01515-06

Papagiannitsis, C.C., Medvecky, M., Chudejova, K., Skalova, A., Rotova, V., Spanelova, P., Jakubu, V., Zemlickova, H., Hrabak, J., 2017. Molecular Characterization of Carbapenemase-Producing Pseudomonas aeruginosa of Czech Origin and Evidence for Clonal Spread of Extensively Resistant Sequence Type 357 Expressing IMP-7 Metallo-β-Lactamase. Antimicrobial Agents and Chemotherapy 61. https://doi.org/10.1128/aac.01811-17

Paradis, E., Claude, J., Strimmer, K., 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20, 289–290. https://doi.org/10.1093/bioinformatics/btg412

Passos da Silva, D., Matwichuk, M.L., Townsend, D.O., Reichhardt, C., Lamba, D., Wozniak, D.J., Parsek, M.R., 2019. The Pseudomonas aeruginosa lectin LecB binds to the exopolysaccharide Psl and stabilizes the biofilm matrix. Nature Communications 10. https://doi.org/10.1038/s41467-019-10201-4

Patterson, N., Price, A.L., Reich, D., 2006. Population Structure and Eigenanalysis. PLoS Genetics 2, e190. https://doi.org/10.1371/journal.pgen.0020190

Pazos, M.A., Lanter, B.B., Yonker, L.M., Eaton, A.D., Pirzai, W., Gronert, K., Bonventre, J.V., Hurley, B.P., 2017. Pseudomonas aeruginosa ExoU augments neutrophil transepithelial migration. PLOS Pathogens 13, e1006548. https://doi.org/10.1371/journal.ppat.1006548

Pena, R.T., Blasco, L., Ambroa, A., González-Pedrajo, B., Fernández-García, L., López, M., Bleriot, I., Bou, G., García-Contreras, R., Wood, T.K., Tomás, M., 2019. Relationship Between Quorum Sensing and Secretion Systems. Frontiers in Microbiology 10. https://doi.org/10.3389/fmicb.2019.01100

Perez, F., Bonomo, R.A., 2018. Evidence to improve the treatment of infections caused by carbapenem-resistant Gram-negative bacteria. The Lancet Infectious Diseases 18, 358–360. https://doi.org/10.1016/s1473-3099(18)30112-9

Peterson, R.E., Edwards, A.C., Bacanu, S.-A., Dick, D.M., Kendler, K.S., Webb, B.T., 2017. The utility of empirically assigning ancestry groups in cross-population genetic studies of addiction. The American Journal on Addictions 26, 494–501. https://doi.org/10.1111/ajad.12586

Petitjean, M., Martak, D., Silvant, A., Bertrand, X., Valot, B., Hocquet, D., 2017. Genomic characterization of a local epidemic Pseudomonas aeruginosa reveals specific features of the widespread clone ST395. Microbial Genomics 3. https://doi.org/10.1099/mgen.0.000129

Pirnay, J.-P., Bilocq, F., Pot, B., Cornelis, P., Zizi, M., Van Eldere, J., Deschaght, P., Vaneechoutte, M., Jennes, S., Pitt, T., De Vos, D., 2009. Pseudomonas aeruginosa Population Structure Revisited. PLoS ONE 4, e7740. https://doi.org/10.1371/journal.pone.0007740

Pissaridou, P., Allsopp, L.P., Wettstadt, S., Howard, S.A., Mavridou, D.A.I., Filloux, A., 2018. The Pseudomonas aeruginosa T6SS-VgrG1b spike is topped by a PAAR protein eliciting DNA damage to bacterial competitors. Proceedings of the National Academy of Sciences 115, 12519–12524. https://doi.org/10.1073/pnas.1814181115

Pittman, J.E., Ferkol, T.W., 2015. The Evolution of Cystic Fibrosis Care. Chest 148, 533–542. https://doi.org/10.1378/chest.14-1997

Poirel, L., Lambert, T., Turkoglu, S., Ronco, E., Gaillard, J.-L., Nordmann, P., 2001. Characterization of Class 1 Integrons from Pseudomonas aeruginosa That Contain the blaVIM-2 Carbapenem-Hydrolyzing -Lactamase Gene and of Two Novel Aminoglycoside Resistance Gene Cassettes. Antimicrobial Agents and Chemotherapy 45, 546–552. https://doi.org/10.1128/aac.45.2.546-552.2001

Poole, K. 2010. Multidrug Efflux Pumps and Antimicrobial Resistance in Pseudomonas aeruginosa and Related Organisms. Journal of molecular microbiology and biotechnology. 3. 255-64.

Poole, K., 2004. Resistance to beta-lactam antibiotics. Cellular and Molecular Life Sciences 61. https://doi.org/10.1007/s00018-004-4060-9

Poole, K., 2011. Pseudomonas Aeruginosa: Resistance to the Max. Frontiers in Microbiology 2. https://doi.org/10.3389/fmicb.2011.00065

Poon, K.K.H., Westman, E.L., Vinogradov, E., Jin, S., Lam, J.S., 2008. Functional Characterization of MigA and WapR: Putative Rhamnosyltransferases Involved in Outer Core Oligosaccharide Biosynthesis of Pseudomonas aeruginosa. Journal of Bacteriology 190, 1857–1865. https://doi.org/10.1128/jb.01546-07

Poonsuk, K., Tribuddharat, C., Chuanchuen, R., 2014. Simultaneous overexpression of multidrug efflux pumps in Pseudomonas aeruginosa non-cystic fibrosis clinical isolates. Canadian Journal of Microbiology 60, 437–443. https://doi.org/10.1139/cjm-2014-0239

Poulsen, B.E., Yang, R., Clatworthy, A.E., White, T., Osmulski, S.J., Li, L., Penaranda, C., Lander, E.S., Shoresh, N., Hung, D.T., 2019. Defining the core essential genome of Pseudomonas aeruginosa. Proceedings of the National Academy of Sciences 201900570. https://doi.org/10.1073/pnas.1900570116

Power, R.A., Parkhill, J., de Oliveira, T., 2016. Microbial genome-wide association studies: lessons from human GWAS. Nature Reviews Genetics 18, 41–50. https://doi.org/10.1038/nrg.2016.132

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE 5, e9490. https://doi.org/10.1371/journal.pone.0009490

Pritt, B., O'Brien, L., Winn, W., 2007. Mucoid Pseudomonas in Cystic Fibrosis. American Journal of Clinical Pathology 128, 32–34. https://doi.org/10.1309/kjrpc7dd5tr9ntdm

Prosperi, M.C.F., Ciccozzi, M., Fanti, I., Saladini, F., Pecorari, M., Borghi, V., Di Giambenedetto, S., Bruzzone, B., Capetti, A., Vivarelli, A., Rusconi, S., Re, M.C., Gismondo, M.R., Sighinolfi, L., Gray, R.R., Salemi, M., Zazzi, M., De Luca, A., 2011. A novel methodology for large-scale phylogeny partition. Nature Communications 2. https://doi.org/10.1038/ncomms1325

Proteostasis Therapeutics, 2019. (http://ir.proteostasis.com/news-releases/news-release-details/proteostasis-therapeutics-announces-broad-new-dataset).

Psoter, K.J., Rosenfeld, M., De Roos, A.J., Mayer, J.D., Wakefield, J., 2014. Differential Geographical Risk of Initial Pseudomonas aeruginosa Acquisition in Young US Children With Cystic Fibrosis. American Journal of Epidemiology 179, 1503–1513. https://doi.org/10.1093/aje/kwu077

Puchelle, E., Bajolet, O., Abély, M., 2002. Airway mucus in cystic fibrosis. Paediatric Respiratory Reviews 3, 115–119. https://doi.org/10.1016/s1526-0550(02)00005-7

Puehringer, S., Metlitzky, M., Schwarzenbacher, R., 2008. The pyrroloquinoline quinone biosynthesis pathway revisited: A structural approach. BMC Biochemistry 9, 8. https://doi.org/10.1186/1471-2091-9-8

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics 81, 559–575. https://doi.org/10.1086/519795

Purssell, A., Poole, K., 2013. Functional characterization of the NfxB repressor of the mexCD-oprJ multidrug efflux operon of Pseudomonas aeruginosa. Microbiology 159, 2058–2073. https://doi.org/10.1099/mic.0.069286-0

Quesada, J.M., Otero-Asman, J.R., Bastiaansen, K.C., Civantos, C., Llamas, M.A., 2016. The Activity of the Pseudomonas aeruginosa Virulence Regulator σVreI Is Modulated by the Anti-σ Factor VreR and the Transcription Factor PhoB. Frontiers in Microbiology 7. https://doi.org/10.3389/fmicb.2016.01159

Quick, J., Cumley, N., Wearn, C.M., Niebel, M., Constantinidou, C., Thomas, C.M., Pallen, M.J., Moiemen, N.S., Bamford, A., Oppenheim, B., Loman, N.J., 2014. Seeking the source of Pseudomonas aeruginosa infections in a recently opened hospital: an observational study using whole-genome sequencing. BMJ Open 4, e006278. https://doi.org/10.1136/bmjopen-2014-006278

Quinton, P.M., 1999. Physiological Basis of Cystic Fibrosis: A Historical Perspective. Physiological Reviews 79, S3–S22. https://doi.org/10.1152/physrev.1999.79.1.s3

Rambaut, A., Lam, T.T., Max Carvalho, L., Pybus, O.G., 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evolution 2, vew007. https://doi.org/10.1093/ve/vew007

Ramirez, M.S., Tolmasky, M.E., 2010. Aminoglycoside modifying enzymes. Drug Resistance Updates 13, 151–171. https://doi.org/10.1016/j.drup.2010.08.003

Ramiro, R. S., Durao, P., Bank, C., Gordo, I., 2019. Low Mutational Load Allows for High Mutation Rate Variation in Gut Commensal Bacteria bioRxiv 10.1101/568709

Rampioni, G., Pillai, C.R., Longo, F., Bondì, R., Baldelli, V., Messina, M., Imperi, F., Visca, P., Leoni, L., 2017. Effect of efflux pump inhibition on Pseudomonas aeruginosa transcriptome and virulence. Scientific Reports 7. https://doi.org/10.1038/s41598-017-11892-9

Ramsey, B.W., Davies, J., McElvaney, N.G., Tullis, E., Bell, S.C., Dřevínek, P., Griese, M., McKone, E.F., Wainwright, C.E., Konstan, M.W., Moss, R., Ratjen, F., Sermet-Gaudelus, I., Rowe, S.M., Dong, Q., Rodriguez, S., Yen, K., Ordoñez, C., Elborn, J.S., 2011. A CFTR Potentiator in Patients with Cystic Fibrosis and the G551D Mutation. New England Journal of Medicine 365, 1663–1672. https://doi.org/10.1056/nejmoa1105185

Ramsey, K., Ratjen, F., Latzin, P., 2017. Elucidating progression of early cystic fibrosis lung disease. European Respiratory Journal 50, 1701916. https://doi.org/10.1183/13993003.01916-2017

Raneri, M., Pinatel, E., Peano, C., Rampioni, G., Leoni, L., Bianconi, I., Jousson, O., Dalmasio, C., Ferrante, P., Briani, F., 2018. Pseudomonas aeruginosa mutants defective in glucose uptake have pleiotropic phenotype and altered virulence in non-mammal infection models. Scientific Reports 8. https://doi.org/10.1038/s41598-018-35087-y

Ranjan, A., Scholz, J., Semmler, T., Wieler, L.H., Ewers, C., Müller, S., Pickard, D.J., Schierack, P., Tedin, K., Ahmed, N., Schaufler, K., Guenther, S., 2018. ESBL-plasmid carriage in E. coli enhances in vitro bacterial competition fitness and serum resistance in some strains of pandemic sequence types without overall fitness cost. Gut Pathogens 10. https://doi.org/10.1186/s13099-018-0243-z

Rasamiravaka, T., Labtani, Q., Duez, P., El Jaziri, M., 2015. The Formation of Biofilms byPseudomonas aeruginosa: A Review of the Natural and Synthetic Compounds Interfering with Control Mechanisms. BioMed Research International 2015, 1–17. https://doi.org/10.1155/2015/759348

Ratjen, F., Brockhaus, F., Angyalosi, G., 2009. Aminoglycoside therapy against Pseudomonas aeruginosa in cystic fibrosis: A review. Journal of Cystic Fibrosis 8, 361–369. https://doi.org/10.1016/j.jcf.2009.08.004

Rau, M.H., Marvig, R.L., Ehrlich, G.D., Molin, S., Jelsbak, L., 2012. Deletion and acquisition of genomic content during early stage adaptation of Pseudomonas aeruginosa to a human host environment. Environmental Microbiology 14, 2200–2211. https://doi.org/10.1111/j.1462-2920.2012.02795.x

Redgrave, L.S., Sutton, S.B., Webber, M.A., Piddock, L.J.V., 2014. Fluoroquinolone resistance: mechanisms, impact on bacteria, and role in evolutionary success. Trends in Microbiology 22, 438–445. https://doi.org/10.1016/j.tim.2014.04.007

Rees, V.E., Deveson Lucas, D.S., López-Causapé, C., Huang, Y., Kotsimbos, T., Bulitta, J.B., Rees, M.C., Barugahare, A., Peleg, A.Y., Nation, R.L., Oliver, A., Boyce, J.D., Landersdorfer, C.B., 2019. Characterization of Hypermutator Pseudomonas aeruginosa Isolates from Patients with Cystic Fibrosis in Australia. Antimicrobial Agents and Chemotherapy 63. https://doi.org/10.1128/aac.02538-18

Reichhardt, C., Wong, C., Passos da Silva, D., Wozniak, D.J., Parsek, M.R., 2018. CdrA Interactions within the Pseudomonas aeruginosa Biofilm Matrix Safeguard It from Proteolysis and Promote Cellular Packing. mBio 9. https://doi.org/10.1128/mbio.01376-18

Reis, R.S., Pereira, A.G., Neves, B.C., Freire, D.M.G., 2011. Gene regulation of rhamnolipid production in Pseudomonas aeruginosa – A review. Bioresource Technology 102, 6377–6384. https://doi.org/10.1016/j.biortech.2011.03.074

Revell, L.J., 2011. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3, 217–223. https://doi.org/10.1111/j.2041-210x.2011.00169.x

Reyes-Darias, J.A., Yang, Y., Sourjik, V., Krell, T., 2015. Correlation between signal input and output in PctA and PctB amino acid chemoreceptor ofPseudomonas aeruginosa. Molecular Microbiology 96, 513–525. https://doi.org/10.1111/mmi.12953

Reynolds, R., Hope, R., Williams, L., 2008. Survey, laboratory and statistical methods for the BSAC Resistance Surveillance Programmes. Journal of Antimicrobial Chemotherapy 62, ii15-ii28. https://doi.org/10.1093/jac/dkn349

Riordan, J.R., 2008. CFTR Function and Prospects for Therapy. Annual Review of Biochemistry 77, 701–726. https://doi.org/10.1146/annurev.biochem.75.103004.142532

Riordan, Rommens, J., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J., et, al., 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 245, 1066–1073. https://doi.org/10.1126/science.2475911

Robb, C.S., Robb, M., Nano, F.E., Boraston, A.B., 2016. The Structure of the Toxin and Type Six Secretion System Substrate Tse2 in Complex with Its Immunity Protein. Structure 24, 277–284. https://doi.org/10.1016/j.str.2015.11.012

Robledo-Avila, F.H., Ruiz-Rosado, J. de D., Brockman, K.L., Kopp, B.T., Amer, A.O., McCoy, K., Bakaletz, L.O., Partida-Sanchez, S., 2018. Dysregulated Calcium Homeostasis in Cystic Fibrosis Neutrophils Leads to Deficient Antimicrobial Responses. The Journal of Immunology 201, 2016–2027. https://doi.org/10.4049/jimmunol.1800076

Rojas Murcia, N., Lee, X., Waridel, P., Maspoli, A., Imker, H.J., Chai, T., Walsh, C.T., Reimmann, C., 2015. The Pseudomonas aeruginosa antimetabolite L -2-amino-4-methoxy-trans-3-butenoic acid (AMB) is made from glutamate and two alanine residues via a thiotemplate-linked tripeptide precursor. Frontiers in Microbiology 6. https://doi.org/10.3389/fmicb.2015.00170

Rojo-Molinero, E., Macià, M.D., Oliver, A., 2019. Social Behavior of Antibiotic Resistant Mutants Within Pseudomonas aeruginosa Biofilm Communities. Frontiers in Microbiology 10. https://doi.org/10.3389/fmicb.2019.00570

Ropy, A., Cabot, G., Sánchez-Diener, I., Aguilera, C., Moya, B., Ayala, J.A., Oliver, A., 2015. Role of Pseudomonas aeruginosa Low-Molecular-Mass Penicillin-Binding Proteins in AmpC Expression, β-Lactam Resistance, and Peptidoglycan Structure. Antimicrobial Agents and Chemotherapy 59, 3925–3934. https://doi.org/10.1128/aac.05150-14

Roy Chowdhury, P., Scott, M.J., Djordjevic, S.P., 2016. Genomic islands 1 and 2 carry multiple antibiotic resistance genes in Pseudomonas aeruginosa ST235, ST253, ST111 and ST175 and are globally dispersed. Journal of Antimicrobial Chemotherapy 72, 620–622. https://doi.org/10.1093/jac/dkw471

Roy, P.H., Tetu, S.G., Larouche, A., Elbourne, L., Tremblay, S., Ren, Q., Dodson, R., Harkins, D., Shay, R., Watkins, K., Mahamoud, Y., Paulsen, I.T., 2010. Complete Genome Sequence of the Multiresistant Taxonomic Outlier Pseudomonas aeruginosa PA7. PLoS ONE 5, e8842. https://doi.org/10.1371/journal.pone.0008842

Rubin, J.L., Thayer, S., Watkins, A., Wagener, J.S., Hodgkins, P.S., Schechter, M.S., 2017. Frequency and costs of pulmonary exacerbations in patients with cystic fibrosis in the United States. Current Medical Research and Opinion 33, 667–674. https://doi.org/10.1080/03007995.2016.1277196

Ruiz-Roldán, L., Bellés, A., Bueno, J., Azcona-Gutiérrez, J.M., Rojo-Bezares, B., Torres, C., Castillo, F.J., Sáenz, Y., Seral, C., 2018. Pseudomonas aeruginosa Isolates from Spanish Children: Occurrence in Faecal Samples, Antimicrobial Resistance, Virulence, and Molecular Typing. BioMed Research International 2018, 1–8. https://doi.org/10.1155/2018/8060178

Russo, D.M., Williams, A., Edwards, A., Posadas, D.M., Finnie, C., Dankert, M., Downie, J.A., Zorreguieta, A., 2006. Proteins Exported via the PrsD-PrsE Type I Secretion System and the Acidic Exopolysaccharide Are Involved in Biofilm Formation by Rhizobium leguminosarum. Journal of Bacteriology 188, 4474–4486. https://doi.org/10.1128/jb.00246-06

Ryall, B., Davies, J.C., Wilson, R., Shoemark, A., Williams, H.D., 2008. Pseudomonas aeruginosa, cyanide accumulation and lung function in CF and non-CF bronchiectasis patients. European Respiratory Journal 32, 740–747. https://doi.org/10.1183/09031936.00159607

Ryan, R.P., Fouhy, Y., Garcia, B.F., Watt, S.A., Niehaus, K., Yang, L., Tolker-Nielsen, T., Dow, J.M., 2008. Interspecies signalling via the Stenotrophomonas maltophilia diffusible signal factor influences biofilm formation and polymyxin tolerance in Pseudomonas aeruginosa. Molecular Microbiology 68, 75–86. https://doi.org/10.1111/j.1365-2958.2008.06132.x

Saint-Criq, V., Gray, M.A., 2016. Role of CFTR in epithelial physiology. Cellular and Molecular Life Sciences 74, 93–115. https://doi.org/10.1007/s00018-016-2391-y

Saint-Criq, V., Villeret, B., Bastaert, F., Kheir, S., Hatton, A., Cazes, A., Xing, Z., Sermet-Gaudelus, I., Garcia-Verdugo, I., Edelman, A., Sallenave, J.-M., 2017. Pseudomonas aeruginosa LasB protease impairs innate immunity in mice and humans by targeting a lung epithelial cystic fibrosis transmembrane regulator–IL-6–antimicrobial–repair pathway. Thorax 73, 49–61. https://doi.org/10.1136/thoraxjnl-2017-210298

Sala, M.A., Jain, M., 2018. Tezacaftor for the treatment of cystic fibrosis. Expert Review of Respiratory Medicine 12, 725–732. https://doi.org/10.1080/17476348.2018.1507741

Salunkhe, P., Smart, C.H.M., Morgan, J.A.W., Panagea, S., Walshaw, M.J., Hart, C.A., Geffers, R., Tummler, B., Winstanley, C., 2005. A Cystic Fibrosis Epidemic Strain of Pseudomonas aeruginosa Displays Enhanced Virulence and Antimicrobial Resistance. Journal of Bacteriology 187, 4908–4920. https://doi.org/10.1128/jb.187.14.4908-4920.2005

Sana, T.G., Berni, B., Bleves, S., 2016. The T6SSs of Pseudomonas aeruginosa Strain PAO1 and Their Effectors: Beyond Bacterial-Cell Targeting. Frontiers in Cellular and Infection Microbiology 6. https://doi.org/10.3389/fcimb.2016.00061

Sana, T.G., Soscia, C., Tonglet, C.M., Garvis, S., Bleves, S., 2013. Divergent Control of Two Type VI Secretion Systems by RpoN in Pseudomonas aeruginosa. PLoS ONE 8, e76030. https://doi.org/10.1371/journal.pone.0076030

Sauer, K., Camper, A.K., Ehrlich, G.D., Costerton, J.W., Davies, D.G., 2002. Pseudomonas aeruginosa Displays Multiple Phenotypes during Development as a Biofilm. Journal of Bacteriology 184, 1140–1154. https://doi.org/10.1128/jb.184.4.1140-1154.2002

Sawicki, G.S., Sellers, D.E., Robinson, W.M., 2009. High treatment burden in adults with cystic fibrosis: Challenges to disease self-management. Journal of Cystic Fibrosis 8, 91–96. https://doi.org/10.1016/j.jcf.2008.09.007

Scanlan, P.D., Buckling, A., 2012. Co-evolution with lytic phage selects for the mucoid phenotype of Pseudomonas fluorescens SBW25. The ISME Journal 6, 1148–1158. https://doi.org/10.1038/ismej.2011.174

Schechter, M.S., VanDevanter, D.R., Pasta, D.J., Short, S.A., Morgan, W.J., Konstan, M.W., 2018. Treatment Setting and Outcomes of Cystic Fibrosis Pulmonary Exacerbations. Annals of the American Thoracic Society 15, 225–233. https://doi.org/10.1513/annalsats.201702-111oc

Schelstraete, P., Van daele, S., De Boeck, K., Proesmans, M., Lebecque, P., Leclercq-Foucart, J., Malfroot, A., Vaneechoutte, M., De Baets, F., 2008. Pseudomonas aeruginosa in the home environment of newly infected cystic fibrosis patients. European Respiratory Journal 31, 822–829. https://doi.org/10.1183/09031936.00088907

Schick, A., Kassen, R., 2018. Rapid diversification of Pseudomonas aeruginosa in cystic fibrosis lung-like conditions. Proceedings of the National Academy of Sciences 115, 10714–10719. https://doi.org/10.1073/pnas.1721270115

Schram C. A. 2012. Atypical cystic fibrosis: identification in the primary care setting. Can Fam Physician. 58 12, 1341–e704.

Schulert, G.S., Feltman, H., Rabin, S.D.P., Martin, C.G., Battle, S.E., Rello, J., Hauser, A.R., 2003. Secretion of the Toxin ExoU Is a Marker for Highly Virulent Pseudomonas aeruginosa Isolates Obtained from Patients with Hospital‐Acquired Pneumonia. The Journal of Infectious Diseases 188, 1695–1706. https://doi.org/10.1086/379372

Secor, P.R., Michaels, L.A., Smigiel, K.S., Rohani, M.G., Jennings, L.K., Hisert, K.B., Arrigoni, A., Braun, K.R., Birkland, T.P., Lai, Y., Hallstrand, T.S., Bollyky, P.L., Singh, P.K., Parks, W.C., 2016. Filamentous Bacteriophage Produced by Pseudomonas aeruginosa Alters the Inflammatory Response and Promotes Noninvasive Infection In Vivo. Infection and Immunity 85. https://doi.org/10.1128/iai.00648-16

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Shariati, A., Azimi, T., Ardebili, A., Chirani, A.S., Bahramian, A., Pormohammad, A., Sadredinamin, M., Erfanimanesh, S., Bostanghadiri, N., Shams, S., Hashemi, A., 2018. Insertional inactivation of oprD in carbapenem-resistant Pseudomonas aeruginosa strains isolated from burn patients in Tehran, Iran. New Microbes and New Infections 21, 75–80. https://doi.org/10.1016/j.nmni.2017.10.013

Sherman, R.M., Forman, J., Antonescu, V., Puiu, D., Daya, M., Rafaels, N., Boorgula, M.P., Chavan, S., Vergara, C., Ortega, V.E., Levin, A.M., Eng, C., Yazdanbakhsh, M., Wilson, J.G., Marrugo, J., Lange, L.A., Williams, L.K., Watson, H., Ware, L.B., Olopade, C.O., Olopade, O., Oliveira, R.R., Ober, C., Nicolae, D.L., Meyers, D.A., Mayorga, A., Knight-Madden, J., Hartert, T., Hansel, N.N., Foreman, M.G., Ford, J.G., Faruque, M.U., Dunston, G.M., Caraballo, L., Burchard, E.G., Bleecker, E.R., Araujo, M.I., Herrera-Paz, E.F., Campbell, M., Foster, C., Taub, M.A., Beaty, T.H., Ruczinski, I., Mathias, R.A., Barnes, K.C., Salzberg, S.L., 2018. Assembly of a pan-genome from deep sequencing of 910 humans of African descent. Nature Genetics 51, 30–35. https://doi.org/10.1038/s41588-018-0273-y

Sherrard, L.J., Tai, A.S., Wee, B.A., Ramsay, K.A., Kidd, T.J., Ben Zakour, N.L., Whiley, D.M., Beatson, S.A., Bell, S.C., 2017. Within-host whole genome analysis of an antibiotic resistant Pseudomonas aeruginosa strain sub-type in cystic fibrosis. PLOS ONE 12, e0172179. https://doi.org/10.1371/journal.pone.0172179

Shi, H., Chen, Z., Kan, J., 2015. Development of loop-mediated isothermal amplification assays for genotyping of Type III Secretion System inPseudomonas aeruginosa. Letters in Applied Microbiology 61, 361–366. https://doi.org/10.1111/lam.12469

Shuman, J., Giles, T.X., Carroll, L., Tabata, K., Powers, A., Suh, S.-J., Silo-Suh, L., 2018. Transcriptome analysis of a Pseudomonas aeruginosa sn-glycerol-3-phosphate dehydrogenase mutant reveals a disruption in bioenergetics. Microbiology 164, 551–562. https://doi.org/10.1099/mic.0.000646

Silveira, M.C., Albano, R.M., Asensi, M.D., Carvalho-Assef, A.P.D., 2016. Description of genomic islands associated to the multidrug-resistant Pseudomonas aeruginosa clone ST277. Infection, Genetics and Evolution 42, 60–65. https://doi.org/10.1016/j.meegid.2016.04.024

Singh, M., Yau, Y.C.W., Wang, S., Waters, V., Kumar, A., 2017. MexXY efflux pump overexpression and aminoglycoside resistance in cystic fibrosis isolates of Pseudomonas aeruginosa from chronic infections. Canadian Journal of Microbiology 63, 929–938. https://doi.org/10.1139/cjm-2017-0380

Skaar, E.P., 2010. The Battle for Iron between Bacterial Pathogens and Their Vertebrate Hosts. PLoS Pathogens 6, e1000949. https://doi.org/10.1371/journal.ppat.1000949

Smith, E.E., Buckley, D.G., Wu, Z., Saenphimmachak, C., Hoffman, L.R., D'Argenio, D.A., Miller, S.I., Ramsey, B.W., Speert, D.P., Moskowitz, S.M., Burns, J.L., Kaul, R., Olson, M.V., 2006a. Genetic adaptation by Pseudomonas aeruginosa to the airways of cystic fibrosis patients. Proceedings of the National Academy of Sciences 103, 8487–8492. https://doi.org/10.1073/pnas.0602138103

Smith, L., Rose, B., Tingpej, P., Zhu, H., Conibear, T., Manos, J., Bye, P., Elkins, M., Willcox, M., Bell, S., Wainwright, C., Harbour, C., 2006b. Protease IV production in Pseudomonas aeruginosa from the lungs of adults with cystic fibrosis. Journal of Medical Microbiology 55, 1641–1644. https://doi.org/10.1099/jmm.0.46845-0

Smith, A.C., Rice, A., Sutton, B., Gabrilska, R., Wessel, A.K., Whiteley, M., Rumbaugh, K.P., 2017. Albumin Inhibits Pseudomonas aeruginosa Quorum Sensing and Alters Polymicrobial Interactions. Infection and Immunity 85. https://doi.org/10.1128/iai.00116-17

Smith, D.J., Badrick, A.C., Zakrzewski, M., Krause, L., Bell, S.C., Anderson, G.J., Reid, D.W., 2014. Pyrosequencing reveals transient cystic fibrosis lung microbiome changes with intravenous antibiotics. European Respiratory Journal 44, 922–930. https://doi.org/10.1183/09031936.00203013

Sobel, M.L., McKay, G.A., Poole, K., 2003. Contribution of the MexXY Multidrug Transporter to Aminoglycoside Resistance in Pseudomonas aeruginosa Clinical Isolates. Antimicrobial Agents and Chemotherapy 47, 3202–3207. https://doi.org/10.1128/aac.47.10.3202-3207.2003

Sobel, M.L., Neshat, S., Poole, K., 2005. Mutations in PA2491 (mexS) Promote MexT-Dependent mexEF-oprN Expression and Multidrug Resistance in a Clinical Strain of Pseudomonas aeruginosa. Journal of Bacteriology 187, 1246–1253. https://doi.org/10.1128/jb.187.4.1246-1253.2005

Song, J., Xie, G., Elf, P.K., Young, K.D., Jensen, R.A., 1998. Comparative analysis of Pseudomonas aeruginosa penicillin-binding protein 7 in the context of its membership in the family of low-molecular-mass PBPs. Microbiology 144, 975–983. https://doi.org/10.1099/00221287-144-4-975

Sotirova, A.V., Spasova, D.I., Galabova, D.N., Karpenko, E., Shulga, A., 2008. Rhamnolipid–Biosurfactant Permeabilizing Effects on Gram-Positive and Gram-Negative Bacterial Strains. Current Microbiology 56, 639–644. https://doi.org/10.1007/s00284-008-9139-3

Stannard, W., O'Callaghan, C., 2006. Ciliary Function and the Role of Cilia in Clearance. Journal of Aerosol Medicine 19, 110–115. https://doi.org/10.1089/jam.2006.19.110

Staudinger, B.J., Muller, J.F., Halldórsson, S., Boles, B., Angermeyer, A., Nguyen, D., Rosen, H., Baldursson, Ó., Gottfreðsson, M., Guðmundsson, G.H., Singh, P.K., 2014. Conditions Associated with the Cystic Fibrosis Defect Promote Chronic Pseudomonas aeruginosa Infection. American Journal of Respiratory and Critical Care Medicine 189, 812–824. https://doi.org/10.1164/rccm.201312-2142oc

Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrener, P., Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Saier, M.H., Hancock, R.E.W., Lory, S., Olson, M.V., 2000. Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. Nature 406, 959–964. https://doi.org/10.1038/35023079

Strateva, T., Yordanov, D., 2009. Pseudomonas aeruginosa – a phenomenon of bacterial resistance. Journal of Medical Microbiology 58, 1133–1148. https://doi.org/10.1099/jmm.0.009142-0

Stressmann, F.A., Rogers, G.B., Marsh, P., Lilley, A.K., Daniels, T.W.V., Carroll, M.P., Hoffman, L.R., Jones, G., Allen, C.E., Patel, N., Forbes, B., Tuck, A., Bruce, K.D., 2011. Does bacterial density in cystic fibrosis sputum increase prior to pulmonary exacerbation? Journal of Cystic Fibrosis 10, 357–365. https://doi.org/10.1016/j.jcf.2011.05.002

Suresh, M., Nithya, N., Jayasree, P.R., Vimal, K.P., Manish Kumar, P.R., 2018. Mutational analyses of regulatory genes, mexR, nalC, nalD and mexZ of mexAB-oprM and mexXY operons, in efflux pump hyperexpressing multidrug-resistant clinical isolates of Pseudomonas aeruginosa. World Journal of Microbiology and Biotechnology 34. https://doi.org/10.1007/s11274-018-2465-0

Syrmis, M.W., Kidd, T.J., Moser, R.J., Ramsay, K.A., Gibson, K.M., Anuj, S., Bell, S.C., Wainwright, C.E., Grimwood, K., Nissen, M., Sloots, T.P., Whiley, D.M., 2014. A comparison of two informative SNP-based strategies for typing Pseudomonas aeruginosa isolates from patients with cystic fibrosis. BMC Infectious Diseases 14. https://doi.org/10.1186/1471-2334-14-307

Taddei, F., Radman, M., Maynard-Smith, J., Toupance, B., Gouyon, P.H., Godelle, B., 1997. Role of mutator alleles in adaptive evolution. Nature 387, 700–702. https://doi.org/10.1038/42696

Tang, A., Marquart, M.E., Fratkin, J.D., McCormick, C.C., Caballero, A.R., Gatlin, H.P., O'Callaghan, R.J., 2009. Properties of PASP: APseudomonasProtease Capable of Mediating Corneal Erosions. Investigative Opthalmology & Visual Science 50, 3794. https://doi.org/10.1167/iovs.08-3107

Tata, M., Wolfinger, M.T., Amman, F., Roschanski, N., Dötsch, A., Sonnleitner, E., Häussler, S., Bläsi, U., 2016. RNASeq Based Transcriptional Profiling of Pseudomonas aeruginosa PA14 after Short- and

Long-Term Anoxic Cultivation in Synthetic Cystic Fibrosis Sputum Medium. PLOS ONE 11, e0147811. https://doi.org/10.1371/journal.pone.0147811

Taylor, T.B., Buckling, A., 2011. SELECTION EXPERIMENTS REVEAL TRADE-OFFS BETWEEN SWIMMING AND TWITCHING MOTILITIES IN PSEUDOMONAS AERUGINOSA. Evolution 65, 3060–3069. https://doi.org/10.1111/j.1558-5646.2011.01376.x

Telling, K., Laht, M., Brauer, A., Remm, M., Kisand, V., Maimets, M., Tenson, T., Lutsar, I., 2018. Multidrug resistant Pseudomonas aeruginosa in Estonian hospitals. BMC Infectious Diseases 18. https://doi.org/10.1186/s12879-018-3421-1

Tettmann, B., Niewerth, C., Kirschhöfer, F., Neidig, A., Dötsch, A., Brenner-Weiss, G., Fetzner, S., Overhage, J., 2016. Enzyme-Mediated Quenching of the Pseudomonas Quinolone Signal (PQS) Promotes Biofilm Formation of Pseudomonas aeruginosa by Increasing Iron Availability. Frontiers in Microbiology 7. https://doi.org/10.3389/fmicb.2016.01978

Tian, Z.-X., Yi, X.-X., Cho, A., O'Gara, F., Wang, Y.-P., 2016. CpxR Activates MexAB-OprM Efflux Pump Expression and Enhances Antibiotic Resistance in Both Laboratory and Clinical nalB-Type Isolates of Pseudomonas aeruginosa. PLOS Pathogens 12, e1005932. https://doi.org/10.1371/journal.ppat.1005932

Tonkin-Hill, G., Lees, J.A., Bentley, S.D., Frost, S.D.W., Corander, J., 2019. Fast hierarchical Bayesian analysis of population structure. Nucleic Acids Research 47, 5539–5549. https://doi.org/10.1093/nar/gkz361

Tran, V.B., Fleiszig, S.M.J., Evans, D.J., Radke, C.J., 2011. Dynamics of Flagellum- and Pilus-Mediated Association of Pseudomonas aeruginosa with Contact Lens Surfaces. Applied and Environmental Microbiology 77, 3644–3652. https://doi.org/10.1128/aem.02656-10

Treepong, P., Kos, V.N., Guyeux, C., Blanc, D.S., Bertrand, X., Valot, B., Hocquet, D., 2018. Global emergence of the widespread Pseudomonas aeruginosa ST235 clone. Clinical Microbiology and Infection 24, 258–266. https://doi.org/10.1016/j.cmi.2017.06.018

Trias, J., Nikaido, H., 1990. Outer membrane protein D2 catalyzes facilitated diffusion of carbapenems and penems through the outer membrane of Pseudomonas aeruginosa. Antimicrobial Agents and Chemotherapy 34, 52–57. https://doi.org/10.1128/aac.34.1.52

Tueffers, L., Barbosa, C., Bobis, I., Schubert, S., Höppner, M., Rühlemann, M., Franke, A., Rosenstiel, P., Friedrichs, A., Krenz-Weinreich, A., Fickenscher, H., Bewig, B., Schreiber, S., Schulenburg, H., 2019. Pseudomonas aeruginosa populations in the cystic fibrosis lung lose susceptibility to newly applied β-lactams within 3 days. Journal of Antimicrobial Chemotherapy. https://doi.org/10.1093/jac/dkz297

Tunney, M.M., Klem, E.R., Fodor, A.A., Gilpin, D.F., Moriarty, T.F., McGrath, S.J., Muhlebach, M.S., Boucher, R.C., Cardwell, C., Doering, G., Elborn, J.S., Wolfgang, M.C., 2011. Use of culture and molecular analysis to determine the effect of antibiotic treatment on microbial community diversity and

abundance during exacerbation in patients with cystic fibrosis. Thorax 66, 579–584. https://doi.org/10.1136/thx.2010.137281

Turner, K.H., Wessel, A.K., Palmer, G.C., Murray, J.L., Whiteley, M., 2015. Essential genome of Pseudomonas aeruginosa in cystic fibrosis sputum. Proceedings of the National Academy of Sciences 112, 4110–4115. https://doi.org/10.1073/pnas.1419677112

Tyrrell, J., Callaghan, M., 2016. Iron acquisition in the cystic fibrosis lung and potential for novel therapeutic strategies. Microbiology 162, 191–205. https://doi.org/10.1099/mic.0.000220

van Belkum, A., Soriaga, L.B., LaFave, M.C., Akella, S., Veyrieras, J.-B., Barbu, E.M., Shortridge, D., Blanc, B., Hannum, G., Zambardi, G., Miller, K., Enright, M.C., Mugnier, N., Brami, D., Schicklin, S., Felderman, M., Schwartz, A.S., Richardson, T.H., Peterson, T.C., Hubby, B., Cady, K.C., 2015. Phylogenetic Distribution of CRISPR-Cas Systems in Antibiotic-Resistant Pseudomonas aeruginosa. mBio 6. https://doi.org/10.1128/mbio.01796-15

van Delden, C., 2007. Pseudomonas aeruginosa bloodstream infections: how should we treat them? International Journal of Antimicrobial Agents 30, 71–75. https://doi.org/10.1016/j.ijantimicag.2007.06.015

van Hoek, A.H.A.M., Mevius, D., Guerra, B., Mullany, P., Roberts, A.P., Aarts, H.J.M., 2011. Acquired Antibiotic Resistance Genes: An Overview. Frontiers in Microbiology 2. https://doi.org/10.3389/fmicb.2011.00203

Veeraraghavan, B., Pragasam, A., Anandan, S., Narasiman, V., Sistla, S., Kapil, A., Mathur, P., Ray, P., Wattal, C., Bhattacharya, S., Deotale, V., Subramani, K., Peter, J., Hariharan, T., Ramya, I., Iniyan, S., Walia, K., Ohri, V., 2018. Dominance of international high-risk clones in carbapenemase-producing Pseudomonas aeruginosa: Multicentric molecular epidemiology report from India. Indian Journal of Medical Microbiology 36, 344. https://doi.org/10.4103/ijmm.ijmm_18_294

Veit, G., Avramescu, R.G., Chiang, A.N., Houck, S.A., Cai, Z., Peters, K.W., Hong, J.S., Pollard, H.B., Guggino, W.B., Balch, W.E., Skach, W.R., Cutting, G.R., Frizzell, R.A., Sheppard, D.N., Cyr, D.M., Sorscher, E.J., Brodsky, J.L., Lukacs, G.L., 2016. From CFTR biology toward combinatorial pharmacotherapy: expanded classification of cystic fibrosis mutations. Molecular Biology of the Cell 27, 424–433. https://doi.org/10.1091/mbc.e14-04-0935

Verchère, A., Dezi, M., Adrien, V., Broutin, I., Picard, M., 2015. In vitro transport activity of the fully assembled MexAB-OprM efflux pump from Pseudomonas aeruginosa. Nature Communications 6. https://doi.org/10.1038/ncomms7890

Verkman, A.S., Song, Y., Thiagarajah, J.R., 2003. Role of airway surface liquid and submucosal glands in cystic fibrosis lung disease. American Journal of Physiology-Cell Physiology 284, C2–C15. https://doi.org/10.1152/ajpcell.00417.2002

Vernikos, G.S., Parkhill, J., 2006. Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the Salmonella pathogenicity islands. Bioinformatics 22, 2196–2203. https://doi.org/10.1093/bioinformatics/btl369

Verstraeten, N., Braeken, K., Debkumari, B., Fauvart, M., Fransaer, J., Vermant, J., Michiels, J., 2008. Living on a surface: swarming and biofilm formation. Trends in Microbiology 16, 496–506. https://doi.org/10.1016/j.tim.2008.07.004

Vertex Pharmaceuticals, 2019. (https://investors.vrtx.com/news-releases/news-release-details/correcting-and-replacing-two-phase-3-studies-triple-combination).

Vidal, F., 1996. Epidemiology and Outcome of Pseudomonas aeruginosa Bacteremia, With Special Emphasis on the Influence of Antibiotic Treatment. Archives of Internal Medicine 156, 2121. https://doi.org/10.1001/archinte.1996.00440170139015

Visscher, P.M., Brown, M.A., McCarthy, M.I., Yang, J., 2012. Five Years of GWAS Discovery. The American Journal of Human Genetics 90, 7–24. https://doi.org/10.1016/j.ajhg.2011.11.029

Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., Yang, J., 2017. 10 Years of GWAS Discovery: Biology, Function, and Translation. The American Journal of Human Genetics 101, 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

Vongthilath, R., Richaud Thiriez, B., Dehillotte, C., Lemonnier, L., Guillien, A., Degano, B., Dalphin, M.-L., Dalphin, J.-C., Plésiat, P., 2019. Clinical and microbiological characteristics of cystic fibrosis adults never colonized by Pseudomonas aeruginosa: Analysis of the French CF registry. PLOS ONE 14, e0210201. https://doi.org/10.1371/journal.pone.0210201

VOYNOW, J., FISCHER, B., ZHENG, S., 2008. Proteases and cystic fibrosis. The International Journal of Biochemistry & Cell Biology 40, 1238–1245. https://doi.org/10.1016/j.biocel.2008.03.003

Wagner, V.E., Bushnell, D., Passador, L., Brooks, A.I., Iglewski, B.H., 2003. Microarray Analysis of Pseudomonas aeruginosa Quorum-Sensing Regulons: Effects of Growth Phase and Environment. Journal of Bacteriology 185, 2080–2095. https://doi.org/10.1128/jb.185.7.2080-2095.2003

Wall, D., Kaiser, D., 1999. Type IV pili and cell motility. Molecular Microbiology 32, 01–10. https://doi.org/10.1046/j.1365-2958.1999.01339.x

Walsh, T.R., Toleman, M.A., Poirel, L., Nordmann, P., 2005. Metallo- -Lactamases: the Quiet before the Storm? Clinical Microbiology Reviews 18, 306–325. https://doi.org/10.1128/cmr.18.2.306-325.2005

Wang, C., Liu, X., Wang, J., Zhou, J., Cui, Z., Zhang, L.-H., 2016. Design and characterization of a polyamine derivative inhibiting the expression of type III secretion system in Pseudomonas aeruginosa. Scientific Reports 6. https://doi.org/10.1038/srep30949

Wang, Z., Wu, X., Peng, J., Hu, Y., Fang, B., Huang, S., 2014. Artificially Constructed Quorum-Sensing Circuits Are Used for Subtle Control of Bacterial Population Density. PLoS ONE 9, e104578. https://doi.org/10.1371/journal.pone.0104578

Wargo, M.J., 2013. Choline Catabolism to Glycine Betaine Contributes to Pseudomonas aeruginosa Survival during Murine Lung Infection. PLoS ONE 8, e56850. https://doi.org/10.1371/journal.pone.0056850

Wargo, M.J., Hogan, D.A., 2007. Examination of Pseudomonas aeruginosa lasI regulation and 3-oxo-C12-homoserine lactone production using a heterologous Escherichia coli system. FEMS Microbiology Letters 273, 38–44. https://doi.org/10.1111/j.1574-6968.2007.00773.x

Warren, A.E., Boulianne-Larsen, C.M., Chandler, C.B., Chiotti, K., Kroll, E., Miller, S.R., Taddei, F., Sermet-Gaudelus, I., Ferroni, A., McInnerney, K., Franklin, M.J., Rosenzweig, F., 2011. Genotypic and Phenotypic Variation in Pseudomonas aeruginosa Reveals Signatures of Secondary Infection and Mutator Activity in Certain Cystic Fibrosis Patients with Chronic Lung Infections. Infection and Immunity 79, 4802–4818. https://doi.org/10.1128/iai.05282-11

Waters, C.M., Goldberg, J.B., 2019. Pseudomonas aeruginosa in cystic fibrosis: A chronic cheater. Proceedings of the National Academy of Sciences 116, 6525–6527. https://doi.org/10.1073/pnas.1902734116

Wei, Q., Le Minh, P.N., Dötsch, A., Hildebrand, F., Panmanee, W., Elfarash, A., Schulz, S., Plaisance, S., Charlier, D., Hassett, D., Häussler, S., Cornelis, P., 2012. Global regulation of gene expression by OxyR in an important human opportunistic pathogen. Nucleic Acids Research 40, 4320–4333. https://doi.org/10.1093/nar/gks017

Wei, Y., Li, Z., Chen, B., Liang, H., Duan, K., 2009. Characterization of the orf1-tolQRA operon in Pseudomonas aeruginosa. Microbiology and Immunology 53, 309–318. https://doi.org/10.1111/j.1348-0421.2009.00130.x

Wellcome Trust Case Control Consortium, 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447, 661–678. https://doi.org/10.1038/nature05911

Whelan, F. J., Surette, M. G., 2015. Clinical Insights into Pulmonary Exacerbations in Cystic Fibrosis from the Microbiome. What Are We Missing? Ann Am Thorac Soc, 12. 10.1513/AnnalsATS.201506-353AW.

Whitney, J.C., Beck, C.M., Goo, Y.A., Russell, A.B., Harding, B.N., De Leon, J.A., Cunningham, D.A., Tran, B.Q., Low, D.A., Goodlett, D.R., Hayes, C.S., Mougous, J.D., 2014. Genetically distinct pathways guide effector export through the type VI secretion system. Molecular Microbiology 92, 529–542. https://doi.org/10.1111/mmi.12571

Wi, Y.M., Choi, J.-Y., Lee, J.-Y., Kang, C.-I., Chung, D.R., Peck, K.R., Song, J.-H., Ko, K.S., 2017. Antimicrobial Effects of β-Lactams on Imipenem-Resistant Ceftazidime-Susceptible Pseudomonas

aeruginosa. Antimicrobial Agents and Chemotherapy 61. https://doi.org/10.1128/aac.00054-17

Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4

Willger, S.D., Grim, S.L., Dolben, E.L., Shipunova, A., Hampton, T.H., Morrison, H.G., Filkins, L.M., O'Toole, G.A., Moulton, L.A., Ashare, A., Sogin, M.L., Hogan, D.A., 2014. Characterization and quantification of the fungal microbiome in serial samples from individuals with cystic fibrosis. Microbiome 2, 40. https://doi.org/10.1186/2049-2618-2-40

Williams, D., Evans, B., Haldenby, S., Walshaw, M.J., Brockhurst, M.A., Winstanley, C., Paterson, S., 2015. Divergent, Coexisting Pseudomonas aeruginosa Lineages in Chronic Cystic Fibrosis Lung Infections. American Journal of Respiratory and Critical Care Medicine 191, 775–785. https://doi.org/10.1164/rccm.201409-1646oc

Williams, D., Fothergill, J.L., Evans, B., Caples, J., Haldenby, S., Walshaw, M.J., Brockhurst, M.A., Winstanley, C., Paterson, S., 2018. Transmission and lineage displacement drive rapid population genomic flux in cystic fibrosis airway infections of a Pseudomonas aeruginosa epidemic strain. Microbial Genomics 4. https://doi.org/10.1099/mgen.0.000167

Williams, D., Paterson, S., Brockhurst, M.A., Winstanley, C., 2016. Refined analyses suggest that recombination is a minor source of genomic diversity in Pseudomonas aeruginosa chronic cystic fibrosis infections. Microbial Genomics 2. https://doi.org/10.1099/mgen.0.000051

Williams, P., Cámara, M., 2009. Quorum sensing and environmental adaptation in Pseudomonas aeruginosa: a tale of regulatory networks and multifunctional signal molecules. Current Opinion in Microbiology 12, 182–191. https://doi.org/10.1016/j.mib.2009.01.005

Wilson, D.J., 2019. The harmonic mean p-value for combining dependent tests. Proceedings of the National Academy of Sciences 116, 1195–1200. https://doi.org/10.1073/pnas.1814092116

Wilson, L.M., Morrison, L., Robinson, K.A., 2019. Airway clearance techniques for cystic fibrosis: an overview of Cochrane systematic reviews. Cochrane Database of Systematic Reviews. https://doi.org/10.1002/14651858.cd011231.pub2

Winsor, G.L., Griffiths, E.J., Lo, R., Dhillon, B.K., Shay, J.A., Brinkman, F.S.L., 2016. Enhanced annotations and features for comparing thousands of Pseudomonas genomes in the Pseudomonas genome database. Nucleic Acids Research 44, D646–D653. https://doi.org/10.1093/nar/gkv1227

Winstanley, C., O'Brien, S., Brockhurst, M.A., 2016. Pseudomonas aeruginosa Evolutionary Adaptation and Diversification in Cystic Fibrosis Chronic Lung Infections. Trends in Microbiology 24, 327–337. https://doi.org/10.1016/j.tim.2016.01.008

Wisplinghoff, H., Bischoff, T., Tallent, S.M., Seifert, H., Wenzel, R.P., Edmond, M.B., 2004. Nosocomial Bloodstream Infections in US Hospitals: Analysis of 24,179 Cases from a Prospective Nationwide Surveillance Study. Clinical Infectious Diseases 39, 309–317. https://doi.org/10.1086/421946

Wittgens, A., Kovacic, F., Müller, M.M., Gerlitzki, M., Santiago-Schübel, B., Hofmann, D., Tiso, T., Blank, L.M., Henkel, M., Hausmann, R., Syldatk, C., Wilhelm, S., Rosenau, F., 2016. Novel insights into biosynthesis and uptake of rhamnolipids and their precursors. Applied Microbiology and Biotechnology 101, 2865–2878. https://doi.org/10.1007/s00253-016-8041-3

Woodford, N., Ellington, M.J., 2007. The emergence of antibiotic resistance by mutation. Clinical Microbiology and Infection 13, 5–18. https://doi.org/10.1111/j.1469-0691.2006.01492.x

Workentine, M., Surette, M.G., 2011. Complex Pseudomonas Population Structure in Cystic Fibrosis Airway Infections. American Journal of Respiratory and Critical Care Medicine 183, 1581–1583. https://doi.org/10.1164/rccm.201105-0776ed

Workentine, M.L., Sibley, C.D., Glezerson, B., Purighalla, S., Norgaard-Gron, J.C., Parkins, M.D., Rabin, H.R., Surette, M.G., 2013. Phenotypic Heterogeneity of Pseudomonas aeruginosa Populations in a Cystic Fibrosis Patient. PLoS ONE 8, e60225. https://doi.org/10.1371/journal.pone.0060225

World Health Organisation., 2017. https://www.who.int/news-room/detail/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed. Accessed August 2019.

Wozniak, D.J., Cram, D.C., Daniels, C.J., Galloway, D.R., 1987. Nucleotide sequence and characterization oftoxR: a gene involved in exotoxin A regulation inPseudomonas aeruginosa. Nucleic Acids Research 15, 2123–2135. https://doi.org/10.1093/nar/15.5.2123

Wright, E.A., Fothergill, J.L., Paterson, S., Brockhurst, M.A., Winstanley, C., 2013. Sub-inhibitory concentrations of some antibiotics can drive diversification of Pseudomonas aeruginosa populations in artificial sputum medium. BMC Microbiology 13, 170. https://doi.org/10.1186/1471-2180-13-170

Wright, L.L., Turton, J.F., Livermore, D.M., Hopkins, K.L., Woodford, N., 2014. Dominance of international 'high-risk clones' among metallo-beta-lactamase-producing Pseudomonas aeruginosa in the UK. Journal of Antimicrobial Chemotherapy 70, 103–110. https://doi.org/10.1093/jac/dku339

Wu, W., Badrane, H., Arora, S., Baker, H.V., Jin, S., 2004. MucA-Mediated Coordination of Type III Secretion and Alginate Synthesis in Pseudomonas aeruginosa. Journal of Bacteriology 186, 7575–7585. https://doi.org/10.1128/jb.186.22.7575-7585.2004

Yan, H., Wang, M., Sun, F., Dandekar, A.A., Shen, D., Li, N., 2018. A Metabolic Trade-Off Modulates Policing of Social Cheaters in Populations of Pseudomonas aeruginosa. Frontiers in Microbiology 9. https://doi.org/10.3389/fmicb.2018.00337

Yang, L., Chen, L., Shen, L., Surette, M., Duan, K., 2011a. Inactivation of MuxABC-OpmB transporter system in Pseudomonas aeruginosa leads to increased ampicillin and carbenicillin resistance and decreased virulence. The Journal of Microbiology 49, 107–114. https://doi.org/10.1007/s12275-011-0186-2

Yang, L., Jelsbak, L., Marvig, R.L., Damkiaer, S., Workman, C.T., Rau, M.H., Hansen, S.K., Folkesson, A., Johansen, H.K., Ciofu, O., Hoiby, N., Sommer, M.O.A., Molin, S., 2011b. Evolutionary dynamics of bacteria in a human host environment. Proceedings of the National Academy of Sciences 108, 7481–7486. https://doi.org/10.1073/pnas.1018249108

Yang, X., Yang, X., Xu, M., Zhou, L., Fan, Z., Jiang, T., 2015. Crystal structures of YfiR from Pseudomonas aeruginosa in two redox states. Biochemical and Biophysical Research Communications 461, 14–20. https://doi.org/10.1016/j.bbrc.2015.03.160

Yang, Z., 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution 24, 1586–1591. https://doi.org/10.1093/molbev/msm088

Yarlagadda, V., Manjunath, G.B., Sarkar, P., Akkapeddi, P., Paramanandham, K., Shome, B.R., Ravikumar, R., Haldar, J., 2015. Glycopeptide Antibiotic To Overcome the Intrinsic Resistance of Gram-Negative Bacteria. ACS Infectious Diseases 2, 132–139. https://doi.org/10.1021/acsinfecdis.5b00114

Yu, G., Smith, D.K., Zhu, H., Guan, Y., Lam, T.T.-Y., 2016. ggtree: anrpackage for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution 8, 28–36. https://doi.org/10.1111/2041-210x.12628

Zaborin, A., Gerdes, S., Holbrook, C., Liu, D.C., Zaborina, O.Y., Alverdy, J.C., 2012. Pseudomonas aeruginosa Overrides the Virulence Inducing Effect of Opioids When It Senses an Abundance of Phosphate. PLoS ONE 7, e34883. https://doi.org/10.1371/journal.pone.0034883

Zeng, L., Jin, S., 2003. aph(3')-IIb, a Gene Encoding an Aminoglycoside-Modifying Enzyme, Is under the Positive Control of Surrogate Regulator HpaA. Antimicrobial Agents and Chemotherapy 47, 3867–3876. https://doi.org/10.1128/aac.47.12.3867-3876.2003

Zerbino, D.R., Birney, E., 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Research 18, 821–829. https://doi.org/10.1101/gr.074492.107

Zhanel, G.G., Hoban, D.J., Schurek, K., Karlowsky, J.A., 2004. Role of efflux mechanisms on fluoroquinolone resistance in Streptococcus pneumoniae and Pseudomonas aeruginosa. International Journal of Antimicrobial Agents 24, 529–535. https://doi.org/10.1016/j.ijantimicag.2004.08.003

Zhang, W., Li, C., 2016a. Exploiting Quorum Sensing Interfering Strategies in Gram-Negative Bacteria for the Enhancement of Environmental Applications. Frontiers in Microbiology 6. https://doi.org/10.3389/fmicb.2015.01535

Zhang, Y., Chen, X.-L., Huang, A.-W., Liu, S.-L., Liu, W.-J., Zhang, N., Lu, X.-Z., 2016b. Mortality attributable to carbapenem-resistant Pseudomonas aeruginosa bacteremia: a meta-analysis of cohort studies. Emerging Microbes & Infections 5, 1–6. https://doi.org/10.1038/emi.2016.22

Zhao, J., Schloss, P.D., Kalikin, L.M., Carmody, L.A., Foster, B.K., Petrosino, J.F., Cavalcoli, J.D., VanDevanter, D.R., Murray, S., Li, J.Z., Young, V.B., LiPuma, J.J., 2012. Decade-long bacterial community dynamics in cystic fibrosis airways. Proceedings of the National Academy of Sciences 109, 5809–5814. https://doi.org/10.1073/pnas.1120577109

Zhao, W.-H., Hu, Z.-Q., 2010. β-Lactamases identified in clinical isolates of Pseudomonas aeruginosa. Critical Reviews in Microbiology 36, 245–258. https://doi.org/10.3109/1040841x.2010.481763

Zhou, X., Stephens, M., 2012. Genome-wide efficient mixed-model analysis for association studies. Nature Genetics 44, 821–824. https://doi.org/10.1038/ng.2310

Zulianello, L., Canard, C., Kohler, T., Caille, D., Lacroix, J.-S., Meda, P., 2006. Rhamnolipids Are Virulence Factors That Promote Early Infiltration of Primary Human Airway Epithelia by Pseudomonas aeruginosa. Infection and Immunity 74, 3134–3147. https://doi.org/10.1128/iai.01772-05

7 Appendices

7.1 Chapter 3 metadata

A list of all isolates included in both bacteraemia surveys, their accession numbers, locations, sampling dates, and MIC values, can be found at the following repository:

https://github.com/SamuelKidman/PhD_Appendix/blob/master/Chapter_3_metadata.csv

Note: The 2017/18 local isolate sequencing data will become available on 20/02/2020.

A list of all isolates included in this chapter and their assembly statistics can be found at the following repository:

https://github.com/SamuelKidman/PhD_Appendix/blob/master/Chapter_3_assembly_data.csv

7.2 Chapter 4 metadata

A list of all isolates included in this chapter and their assembly statistics can be found at the following repository:

https://github.com/SamuelKidman/PhD_Appendix/blob/master/Chapter_4_assembly_data.csv

Note: The ENA accessions for individual isolates are unavailable at time of print, but will become available under the ENA study accession number ERP022089

7.3 Chapter 5 metadata

A list of all isolates with corresponding phenotypes can be found at the following repository:

https://github.com/SamuelKidman/PhD_Appendix/blob/master/Chapter_5_phenotypes.xlsx