A comparative and phylogenetic analysis of the genome of *Pseudomonas aeruginosa*

A thesis submitted to the University of Strathclyde for the degree of Doctor of Philosophy.

By

Lewis James Stewart

Strathclyde Institute of Pharmacy and Biomedical Science

University of Strathclyde,

Glasgow, G4 0RE

July 2015

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed:

Date:

Acknowledgements

The first thanks must go to my supervisor and mentor, **Dr. Nicholas Tucker**. Thank you for your support, advice and patience over the last four years. Your guidance has been indispensable and greatly appreciated. Thank you also to my second supervisor, **Dr. Paul Hoskisson**, and to **Dr. Paul Herron** for your advice and encouragement.

A special mention must be made to the soon-to-be Dr's **Craig Shaw** and **Charles "Fredrik" Webb.** Lunch, football and shooting the old bull with you pair certainly made my studies more enjoyable. Not sure i'd be where I am without it.

Cheers to the **6th floor penthouse** members of SIPBS, both past and present, who have provided me with a source of distraction. Not so much thanks for some of the disgusting cultural delicacies that I have experienced as a result. A mention must be made for the SIPBS **football fellas** that have consistently arranged a competitive and enjoyable game of 5s of an exceptionally "high" standard in a largely pedaspherophobic environment.

My **parents**. Thank you to my Mum and Dad, **Lynn** and **Jim**. I'll forever be in debt of your support - financially, emotionally and generally keeping me alive, fed and watered. Without yous I'd be nowhere. Thank you to my Grandma, Papa, Nana and Granda, mainly for being fantastic **grandparents** and giving me an awesome family. Well done to my **Nana** on being the only member of the family that can pronounce *Pseudomonas aeruginosa*.

Lastly, I would like to thank the **University of Strathclyde** for funding my research, allowing me to realise my personal and professional goal of obtaining my doctorate.

Apologies

I'd also like to extend my heartfelt apologies to my wee brother. **Ryan**, I am eternally sorry for setting the bar so high. Cheers bro, love ya!

Abstract

Pseudomonas aeruginosa remains a constant threat both in the community and nosocomially, accounting for approximately 10% of hospital acquired infections (Davies, 2002). In particular, this free living, extremely virulent pathogen is a worry to clinicians involved in the care of patients suffering from cystic fibrosis, the genetically acquired, multi-system disease. Within the CF lung, *P. aeruginosa* is known to undergo genomic changes resulting in the progression to chronic infection and a worsening prognosis. The most notable example of this phenomenon is the switch to the mucoid phenotype resulting from frame-shift mutations in the *mucA* gene leading to overproduction of alginate.

In this work, we have identified this switch in phenotype as a marker for host adaptation by the bacterium to the CF lung environment and therefore have sequenced three mucoid strains and their non-mucoid progenitors. Using a variety of comparative genomics software, we have identified that two of these pairings are true, isogenic, host adapted counterparts. Each mucoid strain demonstrates evidence associated with such host adaption including genome shrinkage as a result of genomic island and phage loss, as well as decreased virulence and motility. In order to determine the contribution of each genomic anomaly to the change in phenotypes observed, we have utilised a variety of phenotypic analyses including a LC-MS metabolomics analysis of the CF isolates grown in artificial sputum media to mimic chronic lung infection.

Further to this, in an attempt at mapping the core genome of *P. aeruginosa* and providing the community with a greater understanding of the genomic basis of this pathogen, we have also sequenced four well known environmental isolates as well as a further clinical isolate and its rugose small colony variant. With this new data we have developed a new typing method that we believe is superior and more thorough to MLST analysis, noting that *P. aeruginosa* exists in three distinct groupings.

Publications

Stewart, Ford, Sangal, Jeukens, Boyle, Caim, Crossman, Hoskisson, Levesque and Tucker. Draft genomes of 12 host-adapted and environmental isolates of *Pseudomonas aeruginosa* and their positions in the core genome phylogeny. Pathogens and Disease (2013) vol. 71 (1) 20-5

Mackenzie, Stewart, Hoskisson and Tucker. Transcriptional analysis of *Pseudomonas aeruginosa* infected *Caenorhabditis elegans*. Methods in Molecular Biology (2014) vol. 1149 pp. 607-12

General abbreviations

^{O}C	Degree Celsius
ADP	Adenosine Diphosphate
ASM	Artificial sputum media
ATP	Adenosine Triphosphate
AI	Autoinducer
BLAST	Basic local alignment search tool
bp	Basepair
cAMP	Cyclic Adenosine Monophosphate
CCT	CGview Comparison Tool
CF	Cystic fibrosis
ddH_2O	Double distilled H_2O
DNA	Deoxyribonucleic Acid
EDTA	Ethylenediaminetetraacetic
ETA	Exotoxin A
Fe^{3+}	Ferric iron
GC	Guanine-cytosine
HCN	Hydrogen cyanide
InDel	Insertion/Deletion polymorphism
Kbp	Kilo basepair
LB	Lysogeny broth
LC-MS	Liquid Chromatography - Mass Spectrometry
LES	Liverpool Endemic Strain
M	Molar
Mbp	Million basepair
MLST	Multi Locus Sequence Typing
l	Litre
NAD	Nicotinamide Adenine Dinucleotide
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
RGP	Region of genome plasticity

RND	Resistance-Nodulation-Division
ROS	Reactive oxygen species
RNA	Ribonucleic acid
rpm	Revolutions per minute
rRNA	Ribosomal RNA
qPCR	Quantitative real time PCR
QS	Quorum sensing
scv	Small colony variant
SD	Standard deviation
SE	Standard error
SNP	Single Nucleotide Polymorphism
Sp.	Species
TLC	Thin layer chromatography
tRNA	Transcriptional RNA
w/v	Weight to volume
v/v	Volume to volume

Chapter 1 - Introduction

Chapter 2 - Materials and Methods

Chapter 3 - Population genomics of *Pseudomonas aeruginosa* lineage

Chapter 4 - Bacterial adaption as a mechanism of survival

Chapter 5 - Host adaptation between J1385 and J1532, a non-mucoid/mucoid pairing isolated from the lungs of a CF patient

Chapter 1 - Introduction

1.1 Cystic Fibrosis

Cystic fibrosis is an autosomal recessive genetic disorder that severely compromises the lives of sufferers. It is the most common lethal inherited disease in the caucasian population affecting 1 in 2500 new borns (Davis, Drumm and Konstann, 1996). This condition is a multisystem disease affecting the pancreas, liver and intestine, however, the most critically affected organs are the lungs (Davies, Alton and Bush, 2007). The term “cystic fibrosis” originated due to the scarring (fibrosis) and cyst formation within the pancreas as observed by Andersen (1938). It is now characterised by the abnormal transport of sodium and chloride ions across the lung epithelium resulting in the formation of a thick and viscous mucous with alveoli. Other symptoms include salty tasting skin, poor growth/weight gain, episodes of severe coughing and shortness of breath and chronic lung infections.

1.1.1 Cause of cystic fibrosis

CF results from the acquisition of two copies of a defective CF Transmembrane Conductance Regulator (*CFTR*) gene (Riordan *et al*, 1989). Over 1600 mutations within *CFTR* have been identified to be sufficient to cause cystic fibrosis. The most common is the deletion of a phenylalanine amino acid at position 508 (phe508del), accounting for approximately 70% of cases (Davies, Alton and Bush, 2007). The *CFTR* gene encodes a transmembrane protein that is responsible for maintaining epithelial surface liquid volume via chloride secretion and inhibition of sodium absorption. A defective protein and the accompanying reduction of this surface liquid volume results in failure of mucociliary clearance, a build up of thick mucous and an increased susceptibility to bacterial infection (Matsui *et al*, 1998).

1.1.2 Diagnosis and management of cystic fibrosis

Diagnosis of CF is primarily based on the electrolyte level of sweat owing to the fact that sufferers have raised sodium and chloride concentrations in their sweat (Rosenstein and Cutting, 1998). Cases can exist in which sufferers do not exhibit abnormal electrolyte concentrations and such persons require alternative means of diagnosis. The UK now employs the Guthrie blood spot test to screen all newborns for cystic fibrosis (Southern *et al*, 2007). Newborns are initially screened for an increased trypsinogen concentration with positive samples then tested for a variety of common *CFTR* mutations.

There is no cure for CF and sufferers experience a shortened lifespan. Management of this condition is complex and involves treating each symptom experienced individually in order to maximise the functionality of affected organs and delay their deterioration (Davies, Alton and Bush, 2007). For example, reduced nasal airflow can be overcome using steroids, osteoporosis can be prevented using vitamin D and calcium supplements and infertility can be overcome using assisted reproduction. The main aspect of managing cystic fibrosis is treating and preventing lung damage resulting from the presence of thick mucous and associated infections. Physiotherapy is used to aid in the clearance of this thick mucous allowing easier and more efficient air intake by the lung alveoli whilst improving the innate immunity of the sufferer by enabling increased mucociliary clearance. Antibiotics are heavily relied upon in the management of cystic fibrosis and are utilised in a variety of situations. They are used as a prophylactic treatment to prevent infection, to eradicated new infections and to control and limit progression of more persistent, chronic infections (Cystic Fibrosis Trust, 2009). The choice of antibiotic is based on the infecting organism.

1.1.3 Infections in cystic fibrosis

The nature of CF, including reduced innate defences and a thick mucous, results in an increased susceptibility to infection by a variety of bacteria. As mentioned, antibiotics are utilised prophylactically to prevent bacterial colonisation of the lung. However, despite this treatment, CF sufferers are often infected during infancy with microorganisms normally colonising areas of the upper respiratory tract in a commensal manner such as *Staphylococcus aureus* and *Haemophilus influenzae* (Lyczak, Cannon and Pier, 2002). Such infections are relatively innocuous and are often treated with narrow-spectrum antibiotics such as flucloxacillin. In the majority of cases, such bacteria are often outcompeted by the opportunistic pathogen *Pseudomonas aeruginosa*.

P. aeruginosa is a highly pathogenic organism and as a result, infection with this bacterium results in a 2-3 fold increase in risk of death over an eight year period (Emerson *et al*, 2002). However, in 80% of cases, with appropriate antibiotic intervention, such infections can be eradicated successfully. If eradication is not possible, *P. aeruginosa* can establish a more persistent chronic infection which is associated with a worsening of prognosis of the patient. In such circumstances, antibiotic therapy is once again utilised in order to control and prevent infection progression. Both inhaled and intravenous antibiotics are used at this stage. Using antibiotics in such a consistent manner during prophylaxis, eradication and control can lead to the emergence of multidrug-resistant isolates of these bacteria.

The microbial landscape of cystic fibrosis associated infection is related to the age of the patient (Figure 1.1). This figure, from Hauser *et al* (2011), shows that during infancy the most commonly isolated pathogen is *S. aureus* with *H. influenzae* and *P. aeruginosa* strains isolated much less commonly. As age rises, the incidence of *P. aeruginosa* increases and by the late teens, this pathogen will become the most dominant bacterium isolated from the lungs of cystic fibrosis sufferers. Previous studies have shown that 70-80% of

cystic fibrosis sufferers are infected with *P. aeruginosa* by their early teens (Lyczak, Cannon and Pier, 2002). Infections with a variety of other microorganisms are often observed in cystic fibrosis cases including the Gram negative bacteria *Stenotrophomonas maltophila* and *Burkholderia cepacia*, two genera previously classified as belonging to the genus *Pseudomonas*. The lungs of cystic fibrosis sufferers are also commonly colonised by fungi such as *Aspergillus fumigatus* (de Vrankrijker *et al*, 2011).

Due to increased pathogenicity of *P. aeruginosa* and the poor prognosis associated with such infections (Emerson *et al,* 2002), the majority of research into cystic fibrosis associated infections are focussed mainly on this pathogen. Infections with non-pseudomonal pathogens should not, however, be disregarded as unimportant and such infections remain a constant worry to clinicians.

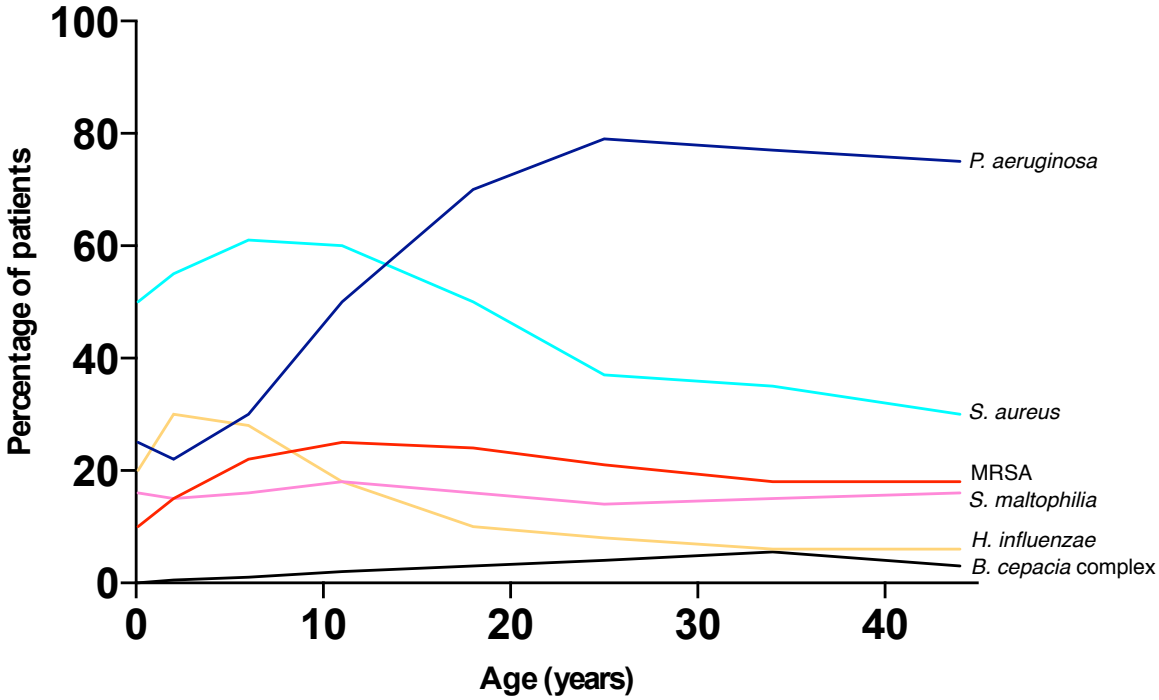


Figure 1.1 The prevalence of several common respiratory pathogens in cystic fibrosis in relation to patient age. (Adapted from Hauser *et al*, 2011)

1.2 Introduction to *Pseudomonas*

1.2.1 *Pseudomonas* Species

The genus *Pseudomonas* is a large group of Gram-negative bacteria belonging to the family Pseudomonadaceae within the gamma proteobacteria. Other members of this diverse family include the genera *Xanthomonas*, *Burkholderia* and *Ralstonia* (Anzai *et al*, 2000). The Pseudomonads are highly ubiquitous, existing throughout nature in a variety of habitats. As a result of its widespread occurrence, the genus *Pseudomonas* was first proposed early in microbiological history by Migula in 1894. As Palleroni (2010) describes, Migulas first illustration of this genus was very brief and, in hindsight, inaccurate. His description stated “Cells with polar organs of motility. Formation of spores occurs in some species, but it is rare”. Eventually, the Pseudomonads were correctly identified as being non-spore forming and the “spores” originally visualized by Migula have been postulated to be “refractile granules of reserve materials” (Palleroni, 2010).

At the most recent count, there are at least 189 species of *Pseudomonas*, as stated by Euzéby (2010) in his List of Prokaryotic names with Standing in Nomenclature (LSPN). This is a stark contrast to the 800 species associated with this genus during the middle of the twentieth century, owing to its poor description and misidentification of strains (Palleroni, 2010). Pseudomonads are rod shaped organisms with a typical diameter ranging from 0.5-2.0μm and 2.5-4.0μm in length. There are however some exceptions to this general morphology, for example, some species may have a more oval appearance as is the case in the soil bacterium *Pseudomonas putida* (Palleroni, 1984). From Migula’s original description of pseudomonads (Palleroni, 2010), one point does remain valid; *Pseudomonas* is a motile genus owing to possession of one or more polar flagella (Figure 1.2). Occasionally, some strains of several species will be isolated that appear to be non-motile however, there are very few species that permanently lack flagella and therefore motility (Palleroni, 1984). *Pseudomonas* spp. use oxygen as their

terminal electron acceptor, resulting in predominantly aerobic growth. However, certain species can alternatively use nitrate in a similar way to facilitate anaerobic growth. Other characteristic features of members of this chemoorganotrophic genus include a positive reaction to catalase test, variable reaction to oxidase test, low tolerance to acidic environments and a mol% GC content of 58-70% (Schmidt, Tummler and Romling, 1996).

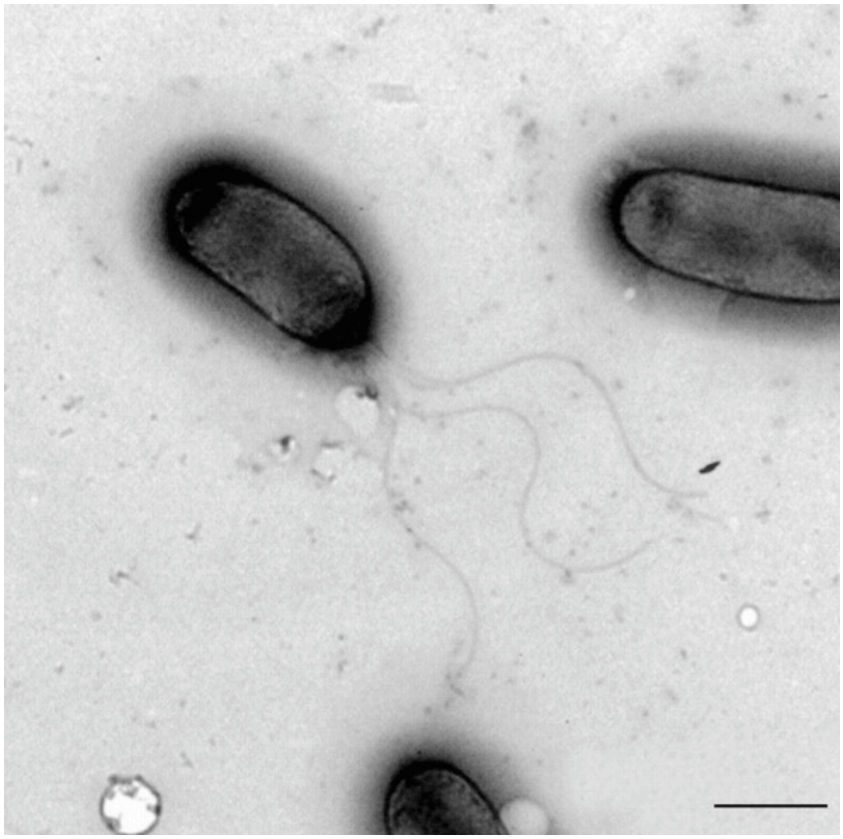


Figure 1.2 Scanning electronmicrograph demonstrating the cell morphology of a *Pseudomonas* spp. Note the presence of polar flagella. Bar represents 2µm (Kwon *et al*, 2003). Permission granted by IJSEM to republish image.

1.2.2 Taxonomy of the *Pseudomonas* genus

Since its initial discovery in 1894, the taxonomic criteria for *Pseudomonas* spp. has greatly varied, becoming substantially more comprehensive, leading to a vast reduction in genus members as previously mentioned. The present taxonomic criteria for the *Pseudomonas* genus is based upon phylogeny as well as their phenotypic characteristics as is the case for all other bacteria (Anzai *et al*, 2000). Tests can be performed to identify particular species by looking at specific characteristics, for example, cell shape, flagella type, antibiotic resistance, growth conditions and utilisation of carbon sources e.g. API testing (Madigan *et al*, 2009; Peix *et al*, 2009). Phenotypic characteristics can be used to identify more common clinical isolates but has a limited use in identification of environmental species due to the limited number of available tests (Peix *et al*, 2009). There are several other chemotaxonomic tests that are useful in discriminating between species for example, SDS-PAGE profiling, siderotyping, polar lipid profiling and biomolecule analysis using fluorescent spectroscopy fingerprinting (Peix *et al*, 2009).

Despite the wide range of tests available for typing *Pseudomonas* spp. based upon their characteristics, the most useful method for taxonomy is phylogenetic and gene sequencing studies. The key target of these studies is the genes encoding ribosomal RNAs (rRNA) that have distinct characteristics, discriminating them from all other genes. This includes their presence in all organisms, possessing the same functions required for life and have also been present since the beginning of evolution (Peix *et al*, 2009). These two characteristics allow genomic variability amongst species to be identified and allow taxonomists to accurately predict an isolates genus and species. Out of all of the bacterial ribosomal RNA, the 16s rRNA gene is most suitable and its use is now deemed to be the gold standard technique used in phylogeny. As a standard, a bacteria whose 16s rRNA sequence is not 97% identical to other sequenced microorganisms should be deemed to be a new species (Madigan *et al*, 2009). However, this method of phylogenic characterisation is limited in its use in observing intragenic relationships due

to the slow rate of evolution of the 16s rRNA (Yamamoto *et al*, 2000). In resolution to this problem, several other conserved “housekeeping” genes have been identified to be of use in typing species. Examples of these include genes encoding DNA gyrase B subunit (*gyrB*) and σ^{70} (*rpoD*) as well as several other genes including *atpD, carA* and *recA*, a gene encoding a recombinase protein (Yamamoto *et al*, 2000; Madigan *et al*, 2009; Peix *et al*, 2009). Using 16S rRNA studies, Palleroni *et al* (1973) arranged *Pseudomonas* species into five distinct RNA homology groups. Each group differs greatly from one another for example, *P. aeruginosa*, a group 1 species shares only 46% homology with the group 2 species, *P. cepacia*. This data is now largely out of date and omits several newly classified species of *Pseudomonas* however, Anzai *et al* (2000) have updated the data using the same 16S rRNA sequencing method, creating two further homology groups (Figure 1.3).

Figure 1.3 Phylogenetic tree of the true Pseudomonads based on the 16S rDNA sequence (Anzai *et al*, 2000).

1.2.3 Pathogenic pseudomonads

Due to the metabolic diversity of the *Pseudomonas* genus, they can survive in a broad range of environments including soil, fresh water and other marine environments (Michel-Briand and Baysse, 2002). They are extremely important in maintaining the ecological equilibrium due to their ability to degrade animal/plant derived soluble compounds as well as breaking down foreign compounds including pesticides. However, despite this genus' bioremedial properties, there are also several pathogenic species. Certain *Pseudomonas* species can be pathogenic to plants and animals.

Bacteria that are pathogenic to plants are referred to as phytopathogens. Such phytopathogens are rarely isolated from habitats other than the plant environment, due to their high level of adaptation to life in this niche. They can also colonise non-susceptible host plants from which they can be transported to susceptible hosts to cause infection (Madigan *et al*, 2009). A variety of factors are secreted from these pathogens in order to damage the plant including toxins, lytic enzymes and plant growth inhibitors for example syringolin A (SylA). SylA is a secreted factor from certain phytopathogenic *Pseudomonas* species that can act as a eukaryotic proteasome inhibitor, blocking proteolysis (Groll *et al*, 2008). *P. syringae*, is a model plant pathogen, not just in the genus *Pseudomonas* but throughout all plant pathogenic bacteria. Over 50 pathovars exist for this pathogen and all can infect a broad range of plant hosts. Its name originated from the tree from which it was first isolated; *Syringa vulgaris*, the lilac tree (Palleroni, 1984). *P. syringae* can be considered an opportunistic pathogen to plants as it is unable to penetrate into the leaf and instead enters via openings such as stomata or wounds (Gimenez-Ibanez and Rathjen, 2010). Once the pathogen gains entry to the plant, it remains in the intracellular spaces and multiply. Once cell concentration has reached a satisfactory level, the cells begin to secrete a variety of virulence factors via a type-III secretion system. These virulence factors will include the necrosis-inducing lipodepsipeptide phytotoxins syringomycin and syringopeptin (Hutchison and Gross, 1997),

the aforementioned plant growth inhibitor SylA (Groll et al, 2008) as well as inhibitors of the host immune system such as HopM1 (Nomura *et al*, 2006). Another important factor in *P. syringae* virulence in plants is its ability to decrease the freezing point of water to approximately -2.8^{o}C (Maki *et al*, 1974). This contributes to its leading role amongst all other organisms as the main cause of frost damage in plants. The genome of several *P. syringae* strains/pathovars have been sequenced and such data is hosted on a dedicated website – Lindeberg's *Pseudomonas*-Plant interaction (http://pseudomonas-syringae.org/). Other plant-pathogenic *Pseudomonas* include *P. tolaasii* and *P. agarici* (Palleroni, 1984).

As mentioned previously, many of the species belonging to the genus *Pseudomonas* can be pathogenic to animal hosts. These include *P. aeruginosa*, *P. oryzihabitans* and *P. plecoglossicida*. The former of these two species are able to cause infection in humans whereas the latter is a pathogen of fish able to cause haemorrhagic ascites (Nishimori *et al*, 2000). By far the most common human-infecting species of *Pseudomonas* is *P. aeruginosa*.

1.2.4 *Pseudomonas aeruginosa*

P. aeruginosa exists as a free-living, environmental organism, found in soil and water as well as in moist areas such as drains and sinks (Davies, 2002). This species could be classified amongst the plant-pathogenic species of *Pseudomonas* as it is capable of infecting the roots of several plants including Arabidopsis and sweet basil, as is described by Walker *et al* (2004). However, it is far more apt to refer to *P. aeruginosa* as a human pathogen due to its increasing prevalence as an opportunistic pathogen. In a healthy host, *P. aeruginosa* has a relatively low pathogenic effect and is rapidly removed by the innate immune system, rarely leading to infection (Davies, 2002). Therefore, being opportunistic, it requires a reduction of the host immune system however, once colonised within a such a host, it is able to infect almost all tissue types. Infections associated with this bacterium include infections of the urinary and respiratory tracts and are also very common in patients with severe burns or extensive skin injury (Madigan *et al*, 2009). Opportunistic infections associated with this pathogen now account for around 10% of nosocomial infections. *P. aeruginosa* is also the most commonly implicated pathogen in lung infection in cystic fibrosis sufferers with 80% of patients experiencing a chronic pseudomonal infection by adolescence (Davies, 2002 and Lyczak, Cannon and Pier, 2002).

1.2.5 *P. aeruginosa* colonisation of the cystic fibrosis lung

Pseudomonal infection of the lungs is the most common cause of death in sufferers of cystic fibrosis. Around 6 to 20% of cystic fibrosis sufferers asymptomatically carry *P. aeruginosa* in their enteric system without provoking a host immune response (Lyczak, Cannon and Pier, 2002). As mentioned, by adolescence *P. aeruginosa* colonises the lungs of the majority of sufferers, initialisation of which can occur at any point during the patients life. Colonisation may occur nosocomially or via social interactions.

The increased susceptibility of these patients to *P. aeruginosa* can be accounted for by the defective innate immune system associated with the characteristic failure of the lung's mucociliary clearance mechanism (Matsui *et al*, 1998). Normally, any pathogens that are inhaled and enter the respiratory tract are trapped within mucous present in lung alveoli and are expelled via mucocilary clearance. However, cystic fibrosis sufferers experience defective mucociliary clearance and are unable to effectively clear bacteria, allowing colonisation of the tract and lungs (Davies, 2002). It has also been demonstrated that cystic fibrosis will reduce the effectiveness of the host's adaptive immune system. Elevated chloride ion concentrations associated with CF results in defective phagocytotic action of neutrophils (Tager, Wu and Vermeulen, 1998). Other features associated with CF can lead to increased susceptibility to *P. aeruginosa* of the lung include increased bacterial attachment via increased expression of the epithelial cell receptor ganglioside asialo-GM_1 (Bryan *et al*, 1998) and direct binding to the defective CFTR protein (Pier, Grout and Zaidi, 1997).

1.2.6 Emergence of the mucoid phenotype

Once *P. aeruginosa* successfully establishes an infection within the cystic fibrosis lung, it often undergoes a striking change in phenotype to a mucoid appearance (Figure 1.4). This switch to mucoidy is associated with the progression from an acute to chronic infection and a significant deterioration in lung function (Pedersen *et al*, 1992). The few patients who do acquire a non-mucoid *P. aeruginosa* infection have a significantly better lung function than those patients experiencing chronic mucoid infection (Parad *et al*, 1999). Such a switch to mucoidy was first noted by Doggett *et al* (1966) who identified that this phenotype is the result of an overproduction of the polysaccharide alginate. This polysaccharide is responsible for giving *P. aeruginosa* its shiny/wet appearance when grown in the lab on solid media (Figure 1.4).

Alginate is encoded for by the alginate biosynthetic operon, beginning with *algD* (Qiu *et al*, 2007). The expression of this is directly controlled by AlgU, a stress-response sigma factor. In the absence of stress, this sigma factor is bound and inhibited by the periplasmic membrane protein MucA, an anti-sigma factor. Microarray analysis of *P. aeruginosa* strains within a mucoid phenotype has indicated that the expression of *algU* is greatly upregulated (Firoved and Deretic, 2003). As mentioned, the switch to a mucoid phenotype is initiated by changes at the genomic level. For example, mutations that mobilise AlgU from negative regulation result in the appearance of the mucoid phenotype. For example, a mutation has been identified in the anti-sigma factor *mucA*, resulting in uninhibited AlgU activity and alginate synthesis (Boucher *et al*, 1997).

P. aeruginosa makes this change in response to its host environment, a phenomenon know as host adaptation. This is an attempt by the pathogen to give itself a competitive advantage against both its host and other bacteria.

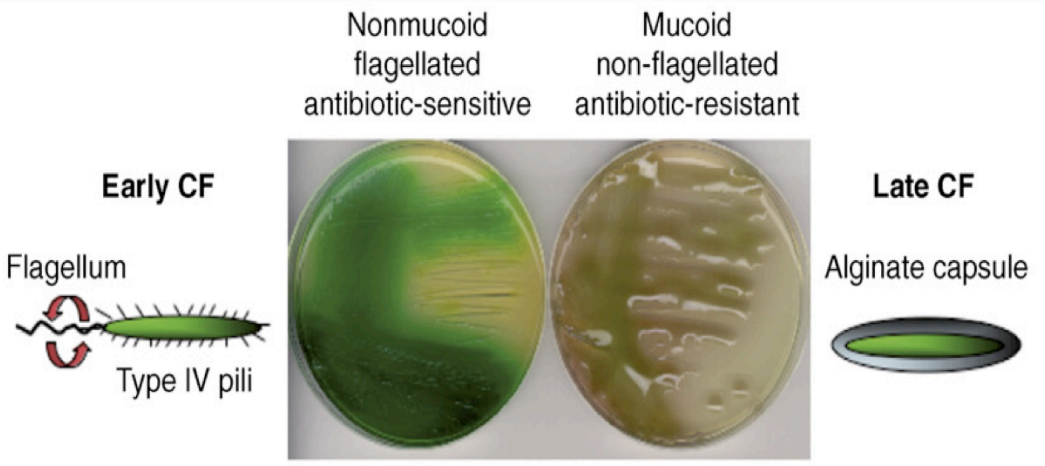


Figure 1.4 *P. aeruginosa* undergoes a variety of phenotypic changes that are associated with the switch to mucoidy. Note the shiny appearance and lack of pigmentation in the mucoid strain compared to the early non-mucoid strain. Also note the cartoon appearance of an alginate capsule around the cell replacing the flagella and pili on the non-mucoid strain (Hassett *et al*, 2009).

1.2.7 Chronic *P. aeruginosa* infection of the CF lung

The appearance of the mucoid capsule surrounding *P. aeruginosa* contributes to the establishment of chronic infection in a variety of ways including protection via acting as a barrier against phagocyctes and opsonization, a barrier against antibiotics and may also aid biofilm formation (Govan and Deretic, 1996; Hassett *et al*, 2009). Alginate is also competent at protecting the cell from the detrimental effects of oxygen radicals from activated polymorphonuclear leukocytes (Govan and Deretic, 1996). This progression to a chronic infection is also accompanied by a variety of other phenotypic changes, one example of which is increased biofilm formation

A biofilm is a grouping of surface-associated cells enclosed within an extracellular polymeric substance (EPS). Biofilms can form on both living and non-living surfaces and are used by bacterial cells to survive and grow in many locations including natural, industrial and medical environments (Hall-Stoodley et al, 2004). Survival is enabled by the inability of host immune defence mechanisms to clear the infection as a result of being ineffective against the sessile cells present within the biofilm (Costerton et al, 1999). Further survival will be conferred as a result of the increased antimicrobial resistance provided by the biofilm. The mechanism by which this resistance is provided is multifactorial. Their high cell density prevents antimicrobial agents penetrating the lower layers allowing cells to continue to survive and cause chronic infection. *P. aeruginosa* can also produced biofilm-specific periplasmic glucans that are able sequester antibiotics preventing passage through the biofilm to exert their antimicrobial effects (Mah *et al*, 2003). Metabolic heterogeneity may also contribute to the increased resistance. Due to their mechanism of action, antibiotics generally target metabolically active cells. The high cell density of the population results in nutrient and oxygen deprivation to lower-lying cells results in reduced metabolic activity and the formation of these persister cells (Drenkard, 2003).

During chronic infection, *P. aeruginosa* also reduces the expression of cell-bound adhesins such as flagella and pili as well as other extracellular pathogen associated molecular patterns (Figure 2.3) (Mahenthiralingam, Campbell and Speert, 1994 and Hassett *et al*, 2009). These changes further contribute to the establishment of a chronic infection by masking itself from host immune intervention as a result of reduced antigenic presentation. It has also been demonstrated that during chronic infection, *P. aeruginosa* reduces production of a variety of extracellular virulence factors (Kamath, Kapatral and Chakrabarty, 1998). For example, Woods *et al* (1991) showed that chronically infecting mucoid *P. aeruginosa* produces significantly reduced levels of a variety of virulence factors including exotoxin A, phospholipase C and pyochelin compared to non-mucoid controls.

Another switch in phenotype that is commonly associated with chronic CF lung infection is the appearance of small colony variants (SCV). These SCVs are selected for during chronic CF lung infection, highlighting a potential role in such an infection (Smith *et al*, 2006). They are hyperadherent, can autoaggregate and are more resistant than non-SCV counterparts (Starkey *et al*, 2009). Prolonged antibiotic exposure is associated with the appearance of SCVs (Drenkard and Ausubel, 2002). Both increased biofilm formation due to hyperadherence and the increased antibiotic resistance associated with the SCV phenotype may aid in persistence within the lung, contributing to a successful infection.

1.2.8 Virulence factors of *P. aeruginosa*

P. aeruginosa can be a highly virulent pathogen, exhibiting a vast arsenal of virulence factors. Virulence factors can be defined as any molecule expressed by a pathogen that contributes to its success in establishing an infection including initial colonisation, immunoevasion, penetrating host cells and sequestration of nutrition from its host (Casadevall and Pirofski, 2001).

The expression of all *P. aeruginosa* virulence factors, including those previously mentioned, is under the control of quorum sensing (QS). QS is the mechanism by which cells communicate with one another via production of small signalling molecules (Venturi, 2006). Cells respond to the population density of surrounding cells via monitoring the proportionate concentration of these signalling molecules, allowing them to alter expression of particular genes appropriately at the correct time. Cells signal to one another via small molecules called autoinducers (AI) (Yarwood *et al*, 2004). Normally, autoinducers are present at low concentrations, however, once cell numbers reach a sufficient density, QS is activated, increasing the autoinducer concentration and promoting QS-mediated alteration in gene expression (Venturi, 2006).

QS controls several virulence factors of *P. aeruginosa*, however, it is essential for biofilm formation with defective strains being unable to form a mature biofilm (Yarwood *et al*, 2004). Several stages of biofilm formation are influenced by this system including maturation and dispersal as a result of increased rhamnolipid production. In addition to this, QS controls *pel* transcription of polysaccharide as well as increased extracellular DNA, two significant components of a successful biofilm (Kievit, 2009).

P. aeruginosa possesses two QS systems, the Las and the Rhl systems (Ledgham *et al*, 2003). Both systems are composed of a transcriptional activator and a LuxI-type autoinducer (AI) synthase to produce a *N*-acyl homoserine lactone signal, the AI. In Las QS, the AI synthase, LasI, encoded

by the *lasI* gene, guides the synthesis of the *N*-3-oxododecanoyl homoserine lactone (PAI-1) (Pearson *et al*, 1994; Pearson *et al*, 1997; Venturi, 2006). PAI-1, when at a critical concentration, will bind to the *lasR*-encoded transcriptional activator LasR, stimulating transcriptional activation of a variety of genes encoding specific virulence factors including *lasB* (elastase), *apr* (alkaline protease) and *toxA* (Exotoxin A). LasR also has a role in positive feedback of PAI-1, stimulating transcription of *lasI* to increase AI concentrations (Pearson *et al*, 1997). The Las QS system is important in mounting a successful infection as was demonstrated by Tang *et al* (1996) in which it was shown that Δ*lasR* mutants were avirulent in pneumonia mouse models.

The second QS system present in *P. aeruginosa*, the Rhl system is directed by activation of the AI synthase-encoding *rhlI* to produce two AI molecules, predominantly *N*-(butanoyl)-L-homoserine lactone but also N-butyryl homoserine lactone (PAI-2) (Pearson *et al*, 1997; Ledgham *et al*, 2003; Venturi, 2006). These AI will bind to the RhlR transcriptional activator, encoded by *rhlR*, and will induce transcription of its target genes, primarily the *rhlAB* genes that encode for production of the biosurfactant rhamnolipid, as previously described. It has also been shown that there is an intimate relationship existing between both systems, connected via a regulatory cascade (Ledgham *et al*, 2003). Pearson *et al* (1997) demonstrated that the Las system can activate both *lasB* and *rhlA*. However, despite this association between QS systems of *P. aeruginosa*, Las preferentially activates its own natural target of *lasB* over *rhlA*.

1.2.8.1 Adhesins

A variety of factors – adhesins - are involved in the initial adherence of the bacterium during colonisation. The majority of *P. aeruginosa* strains will express flagella, extracellular appendages responsible for providing the cell with motility allowing migration to areas suitable for infection. These polar flagella are composed of many subunits encoded by the *fliC* gene.

Transcription of this gene and resulting production of the flagellar subunits is under control of σ^{28}, the key σ-factors involved in controlling motility in most bacteria (Farinha *et al*, 1993). Another adhesin possessed by *P. aeruginosa* is its unipolar pili (Davies, 2002). *P. aeruginosa* possesses type 4 (*N*-methyl-Phe) pili, composed of *pilA*-encoded monomeric protein subunits (Farinha *et al*, 1993). Both of these adhesins will attach to disaccharides (GalNAcb1-4Gal) present on host cells. For example, in cystic fibrosis sufferers, the specific disaccharide is asialoGM1 which is present in large numbers on the lung epithelial cells, particularly in cystic fibrosis sufferers, providing an alternative explanation for their increased susceptibility (Tang *et al*, 1995; Davies, 2002). Animal models have been used to demonstrate that these adhesins are required for *P. aeruginosa* infection. It has been noted that in strains that are defective for pili, a less virulent infection is observed (Tang *et al*, 1995). As previously mentioned, mucoid *P. aeruginosa* cells rarely express such cell-bound adhesins (Hassett *et al*, 2009). Reduced expression of these antigenic adhesins will help mask the cells from host immune cells, contributing to the establishment of a chronic infection.

1.2.8.2 Metalloproteases

P. aeruginosa secretes a variety of different metalloproteases including elastase (pseudolysin) and alkaline protease (aeruginolysin). Elastase is a 39.5kDa protein toxin encoded by the *lasB* gene. It will degrade the elastin present in the human lung and can also degrade other matrix proteins including laminin and collagen (types III and IV). Degradation of these proteins will damage the lung tissue, resulting in changes in the lung structure (Yanagihara *et al*, 2003). Elastase can also worsen infection by blocking opsonization via cleavage of the complement C3b receptor (Tosi *et al*, 1990) and can also degrade immunoglobulins A and G, coagulation factors, cytokines (interferon gamma and tumour necrosis factor) and the alpha proteinase inhibitor (Chmiel and Davis, 2003). Elastase significantly increases the quantity of lymphocytes in an infected lung as demonstrated in murine models, indicating an important role in establishing a chronic infection

(Yanagihara *et al*, 2003). The exact role of the second metalloprotease, alkaline protease, is unclear however, it also cleaves the same targets mentioned for elastase. This 50kDa protein is encoded by the *apr* gene its secretion is essential for mediating corneal infection (Guzzo *et al*, 1990). As previously mentioned, chronically infecting mucoid *P. aeruginosa* cells will reduce production of the metalloprotease elastase (Woods *et al*, 1991 and Kamath, Kapatral and Chakrabarty, 1998). It is expected that reduced elastase and therefore reduced lung tissue damage will contribute to the establishment of a chronic infection and prolonged infection of the lung via reducing the antagonistic effects of *P. aeruginosa* and thereby its "visibility" to the host.

1.2.8.3 Haemolysins

P. aeruginosa can produce heat-resistant rhamnolipid biosurfactants, amphiphilic molecules that have several roles in bacterial virulence. As "biosurfactant" suggests, they can solubilse and promote uptake of hydrophobic substrates e.g. *n*-alkanes via reducing surface tension and emulsification (Soberón-Chávez *et al*, 2005). Rhamnolipids can also give the organism a significant competitive advantage over other bacteria via inducing toxicity. They have antibacterial activity against many other organisms but in particular, Gram positive pathogens (Soberón-Chávez *et al*, 2005). Aside from its first putative role as a haemolysin, it was also suggested that it can solubilse lung surfactant, exposing the lungs to cleavage by another haemolysin – phospholipase C (Kurioka and Liu, 1967). Other roles of rhamnolipids in mediating virulence include disruption of the polymorphonuclear chemotactic response, release of cytokines, hinder MCC and blocking of correct macrophage functioning (Soberón-Chávez *et al*, 2005). Production of rhamnolipids is a complex process involving a series of sequential glycosyl transferase reactions mediated by rhamnosytransferases with the whole process encoded by the *rhlABRI* gene cluster (Ochsner *et al*, 1995).

Another haemolysin of *P. aeruginosa* is phospholipase C (PLC), a heat-labile toxin. All virulent strains are capable of secreting this haemolytic toxin although those isolated from the lungs and urinary tract are present at higher concentrations (Berka and Vasil, 1982). Two separate homologues of this enzyme exist, the haemolytic PLC-H and the non-haemolytic PLC-N, encoded by the *plcSR* operon and *plcN* gene, respectively (Ostroff *et al*, 1995). The exact role of these toxins is not fully understood. The mature 72.8kDa PLC-H is capable of cleaving a variety of phospholipids possessing an ammonium group including phosphatidycholine, sphingomyelin and lysophosphatidylcholine (Berka and Vasil, 1982).This accounts for PLC-H's haemolytic properties as cleavage of these phospholipids will damage the host tissue cell walls. PLC-N also cleaves these phospholipids but at a much lower efficiency (30-50%). This will damage the host cell membrane but is not sufficient to cause haemolysis and is therefore non-haemolytic (Ostroff *et al*, 1995).

During acute infection, the antimicrobial effect of such haemolysins (Soberón-Chávez *et al*, 2005) will aid in outcompeting co-colonising pathogens, especially the predominant Gram positive bacterium *S. aureus* (Lyczak, Cannon and Pier, 2002). During later cystic fibrosis associated infections, haemolysins are not essential for a successful infection due to the reduced presence of competing pathogens (Hauser *et al*, 2011). Instead, these enzymes will mainly be utilised to cause host cell damage. Therefore, like metalloproteases, it may be in the interests of *P. aeruginosa* to reduce production of these enzymes as is the case regarding phospholipase C (Woods *et a*, 1991).

1.2.8.4 Exotoxin A

Another exoproduct produced and secreted by *P. aeruginosa* is the toxin exotoxin A (ETA). ETA is encoded by the *toxA* gene and again, its exact role is unclear (Gallant *et al*, 2000). It is, however, one of the most potent cytotoxins secreted by *P. aeruginosa* and almost all clinically significant strains are capable of secreting it. This 613 amino acid toxin is able to kill cells via interfering with protein synthesis (El Hage *et al*, 2010). It is able to

carry out this task due to possession of ADP-ribosyltransferase (ADPRT) activity (Gallant *et al*, 2000). ETA will catalyse the transfer of the ADP-ribose from NAD^+ to elongation factor 2 (EF-2), a key protein involved in translational elongation. As a result of ADP-ribosylation, EF-2 will cease to function, inhibiting protein synthesis and killing the cells since no essential polypeptide chains are being elongated. El Hage *et al* (2010) also indicated that ETA may have a role in inducing programmed cell death. It was shown that treatment of cells with ETA will cause mitochondrial release of cytochrome c, caspase-9 and caspase-3 activation. These factors are key intermediates in intracellular signaling intended to cause apoptosis. Therefore, ETA exerts its cytotoxic effects and cause cell death via two methods – inhibiting protein synthesis and promoting programmed cell death.

P. aeruginosa also possesses several type-III cytotoxins (ExoS, ExoT, ExoU and ExoY) which are directly delivered from the bacterial cell to the host cell via a type-III secretion system. These toxins will act antagonistically to the host via inhibition of their innate immune system (Barbieri and Sun, 2004). ExoU has phosphlipase activity and can promote septic shock as well as having a major role in accelerated lung injury (Sato *et al*, 2003). ExoY, an adenylate cyclase, causes a rise in intracellular cyclic adenosine monophosphate (cAMP), leading to reorganisation of the host cytoskelton leading to instabiity and cell death. It may also have a role in protecting *P. aeruginosa* from phagocytic removal (Yahr *et al*, 1998). Like ETA, ExoS provides ADPRT activity whilst contributing to the virulence of *P. aeruginosa* (Iglewski *et al*, 1978). However, unlike ETA, ExoS does not modify EF-2 but instead ADP-ribosylates a variety of proteins involved in the Ras signalling pathway. This will interfere with intracellular signalling exerting its anti-apoptotic and anti-phagocytic functions (Barbieri and Sun, 2004). ExoT also has ADPRT activity, indicating that ExoS and ExoT originate from a common ancestor, with gene duplication the likely cause of differentiation. ExoT will also exert its anti-phagocytic funtcions on the host cell via blocking

intracellular signalling but instead, it targets Crk and its associated proteins (Barbieri and Sun, 2004).

Woods *et al* (1991) demonstrated that mucoid strains produce significantly reduced levels of exotoxin A. As mentioned, ETA is one of the most potent toxins produced by *P. aeruginosa* and is highly effective at damaging host cells. Therefore, in regards to the attempts of *P. aeruginosa* to establish a chronic infection within the lung, reducing production of this toxin, and similar exotoxins, will contribute to its evasion of the host immune system via masking itself and preventing provocation.

1.2.8.5 Siderophores

Siderophores are iron chelators that are responsible for sequestration of iron from the bacterial environment. It is in direct competition with the human iron-binding protein transferrin in order to provide the bacterium with sufficient iron. *P. aeruginosa* will secrete two major siderophores, pyoverdine and pyochelin.

Pyoverdine, sometimes referred to as pseudobactins, is a yellow-green, fluorescent pigment. *P. aeruginosa* can produce and secrete three different pyoverdines with each having different peptide chains and associated chromophores (Cornelis, 2010). There have been over 30 genes identified throughout the *Pseudomonas* genus that have the task in directing the synthesis of the pyoverdines (Meyer, 2000). Genes implicated in its synthesis include the *pvcABCD* operon and *pvdA* and *pvdD*. These genes are not part of the core genome of fluorescent strains of *Pseudomonas*, suggesting that they have been acquired later, most probably via horizontal gene transfer (Cornelis, 2010). They are composed of a fluorescent chromophore (dihydroxyquinoline) attached to a peptide, specific for each strain. Initially, they are synthesised as immature precursor peptides, called ferribactin that can then be adapted to give rise to the mature, functional pyoverdine molecule (Cornelis, 2010). These mature pyoverdines are potent

siderophores. Once secreted from the cell, these molecules will bind to ferric (Fe^{3+}) iron to form a ferripyoverdine complex that then binds to the FpvA receptor located on the bacterial cell outer membrane (Imperi *et al*, 2009). This protein will facilitate the uptake of ferric iron into the cell where it is converted to its ferrous state and can be utlilised or stored for later use (Cornelis, 2010). It also initiates a signalling cascade involving a variety of proteins including FpvA and FpvR culminating in the activation of PvdS that is then free to interact with RNA polymerase to initiate transcription of a variety of target genes, namely pyoverdine synthesis genes (Figure 1.5) (Lamont *et al*, 2002).

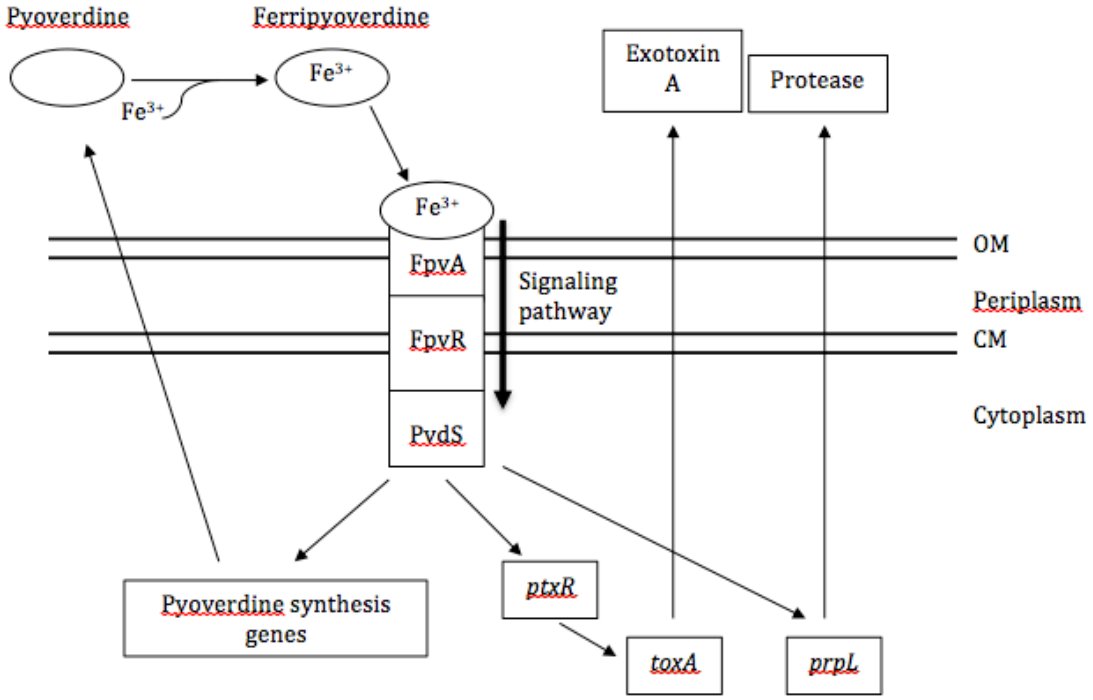


Figure 1.5 Diagram of the ferripyoverdine activated FpvA/FpvS/PvdS signaling pathway. Ferripyoverdine will bind to its cell surface receptor protein FpvA, initiating a signal cascade. Activated FpvR will release PvdS allowing its activation and subsequent transcription of the genes required for pyoverdine synthesis, and exotoxin A and protease synthesis (adapted from Lamont *et al*, 2002).

In addition to its role in iron uptake, pyoverdine also has several roles in mediating infection. Its synthesis is essential for infection and pyoverdine-defective strains have attenuated virulence in several animal models (Visca *et al*, 2007; Cornelis, 2010). Pyoverdine controls a signal transduction pathway that, when complexed with ferric iron, stimulates production of a variety of virulence factors including exotoxin A, protease and further pyoverdine (Figure 2.4)(Lamont *et al*, 2002 and Visca *et al*, 2007). Within *P. syringae*, pyoverdine is essential for production of the quorum sensing signalling molecules *N*-acyl-homoserine lactones (Cornelis, 2010). Another important role of pyoverdine mediated iron-chelation is formation of a mature biofilm. Banin *et al* (2005) showed that mutant *P. aeruginosa* strains defective for pyoverdine were unable to form the mushroom-shaped structure characteristic of biofilms, even in high iron concentrations. Therefore, pyoverdine and its role in gathering iron for the organism is essential in biofilm formation and maturation as well as for several other mechanisms of virulence in *P. aeruginosa*.

Production of pyoverdine is significantly increased in mucoid *P. aeruginosa* strains (Woods *et al*, 1991). This differs from the general trend in which mucoid cells have been associated with reduced virulence factor production, possibly to aid in concealment from the host and evasion of host defenses. Perhaps, in the case of pyoverdine, the pros, namely increased biofilm formation and iron acquisition, may outweigh the cons and provide the bacteria with a more significant advantage in establishing a chronic infection.

P. aeruginosa secretes a second, less abundant siderophore, pyochelin. Pyochelin is similar in its actions to pyoverdine but utilises the outer membrane transporter FptA instead of the pyoverdine receptor FpvA. Like pyoverdine, pyochelin is required for successful infection by *P. aeruginosa* (Brandel *et al*, 2012). Compared to pyoverdine, pyochelin is produced by clinical strains only sparingly. In contrast to pyoverdine, pyochelin production

is significantly reduced by mucoid strains falling in line with the trend that virulence factors of *P. aeruginosa* are reduced by mucoid strains.

1.2.8.6 Pyocyanin

P. aeruginosa strains will characteristically secrete a variety of pigments including the blue-green pyocyanin, the red-brown pyorubin and the previously mentioned siderophore pyoverdine, a fluorescent/yellow-green pigment. In fact, pyocyanin is so characteristic of *P. aeruginosa* that this organisms name originates from it. The name *aeruginosa* is derived from the Latin term for the blue-green colour of copper rust (Liu and Nizet, 2009). These pigments will provide nomenclature and can be used for diagnostic purposes but most importantly, they contribute to the pathogenesis of this organism. Pyocyanin (PCN) is a zwitterion, capable of penetrating biological membranes and is readily identifiable in the sputum of infected CF sufferers, supporting its role as a virulence factor in infection (Lau *et al*, 2004).

The synthesis of PCN is extremely complex and is directed by two copies of the *phzABCDEFG* operon, as well as the genes *phzH*, *phzM* and *phzS*. Its production is influenced by the *las* and *rhl* QS systems. PCN can act as a potent producer of reactive oxygen species (ROS) due to its ability to accept electrons from NADPH, glutathione and other reducing reagents. It will then transfer them to oxygen to generate these ROS including hydrogen peroxide (H_2O_2) (Liu and Nizet, 2009). This may be the key factor of PCN in mediating host damage. For example, oxidative stress will increase intracellular levels of inositol triphosphate leading to disruption in Ca^{2+} homeostasis (Lau *et al*, 2004). Ca^{2+} is important for several processes including ion transport, ciliary beating and mucous secretion with alterations in these providing an explanation for the defective MCC observed in *P. aeruginosa* lung infections (Liu and Nizet, 2009). ROS also induce apoptosis in neutrophils. Usher *et al* (2002) showed that PCN can induce ROS formation within neutrophils, producing sufficient quantities for inducing apoptosis. It was additionally shown that PCN, via reducing intracellular cAMP levels, can lead to a time

and concentration dependant induction of apoptosis (Usher *et al*, 2002; Liu and Nizet, 2009). ROS can also boost cytokine levels via alteration of Nuclear Factor κβ (NF-κβ) and other signalling pathways. This will lead to an increased innate immune response, exacerbating the infection and clinical symptoms (Liu and Nizet, 2009). There are several other suggested mechanisms of virulence of PCN and it has been shown that production of this pigment is essential for successful lung infection in mice, highlighting its importance in infection (Lau *et al*, 2004).

1.2.9 Infection Models of *P. aeruginosa*

Infection models can be used to determine the virulence of a particular species of *Pseudomonas,* a particular strain of *P. aeruginosa*, or a particular gene product produced by an individual strain. Animal infection models are used with the intention of simulating the course of infection of a particular pathogen that would normally be observed in a human host. Infection models have been used for many years for this purpose for a variety of bacteria, viruses, parasites and fungi. Infection models are also commonly utilised in plant hosts, for example the leaves of lettuce and Arabidopsis have been used to observe pathogenesis of *P. aeruginosa* strains PA01 and PA14 (Walker *et al*, 2004). Models can also be used to determine the efficacy of a particular antimicrobial drug and have been used for this purpose for many years. For example, in 1940, Chain *et al*, used mice, rat and cat models infected with a variety of bacteria, including *Streptococcus pyogenes* and *Staphylococcus aureus*, to test the antimicrobial activity of the newly identified penicillin.

Regarding *P. aeruginosa*, several infection models have been used over the years. Ideally, mammalian hosts would be used for infection models due to their similarities in physiology and clinical response to infection. However, their use is limited due to ethical, practical and economical limitations (Alibaud *et al*, 2008). As a result of these restrictions, most research regarding *P. aeruginosa* virulence is performed using non-mammalian hosts. In 2001, D'Argenio *et al* used an insect model to identify mutants of *P. aeruginosa* that possessed reduced virulence. In this study, the fruit fly, *Drosophila melanogaster* was used as a host to identify these strains with reduced virulence compared to the relatively virulent PA02. Other studies have used amoeba as an infection model for *P. aeruginosa*. For example, *Dictyostelium discoideum* can be used as a model host for *P. aeruginosa*, being susceptible to lysis by PA01 infection (Cosson *et al*, 2002). A variety of plants can be used as an alternative to mammalian infection models including lettuce, *Arabidopsis* and sweet basil leaves (Walker *et al*, 2004).

Another emerging non-mammalian infection host that is becoming more commonly used is the nematode *Caenorhabditis elegans*. This nematode has not just been used for mapping the virulence of *P. aeruginosa* but can also be used for infection models of a variety of bacteria including *S. aureus*, *S. pyogenes*, *Enterococcus faecalis*, *Burkholderia sp*. and *Serratia marcescens* to name but a few (Sifri *et al*, 2005). More recently, researchers have began to use the greater wax moth larvae *Galleria mellonella* as an effective infection model.

1.2.9.1 *Galleria mellonella*

The caterpillar larvae of the greater wax moth larvae, *Galleria mellonella*, is found feeding within bee nests feeding on the honey present there. They are also often used as food for terrarium pets such as spiders and a variety of reptiles. In research, they are often used as models for investigating pathogenicity and toxicity. Compared to other non-mammalian infection models, such as the nematode *C. elegans*, infection models of *G. mellonella* are far more suited due to their far more complex antimicrobial defense systems and as a result will produce results far more relevant to the infection process of mammalian organisms (Ramarao, Nielsen-Leroux and Lereclus, 2012). Their advanced innate immune system includes similar defenses to that of mammalian hosts including the ability of cells to phagocytose infecting bacteria and also produce antibacterial products such as lysozyme. Also, epithelial cells present within both *G. mellonella* and mammalian hosts show a degree of homology allowing bacterial cells to initiate colonisation via epithelial cell surface adhesion. Also, larvae can be easily infected via injecting a fixed number of bacterial cells and killing is easily identified via black pigmentation of the invertebrate. Therefore, *G. mellonella* are suitable model organisms for studying the pathogenicity of bacterial organisms.

Jander, Rahme and Ausubel (2000) showed that *G. mellonella* is a good model organism for studying the mammalian virulence factors of *P. aeruginosa*. They showed that there was a direct correlation between the

virulence observed in mammalian cells and *G. mellonella*. In contrast, similar studies focussing on plant and nematode infection models did not show such a correlation. Subsequent studies using this infection model have shown that greater wax moth larvae can be used to study the role of type lll secretion system effectors in infection (Miyata *et al*, 2003), to investigate the role of elastase B in stimulating the host humoral response (Andrejko and Mizerska-Dudka, 2011) and has been used to show that *P. aeruginosa* is capable of killing the invertebrate phagocyte cells, haemocytes via inducing apoptosis and autophagal cell death (Mizerska-Dudka and Andrejko, 2013).

1.3 *P. aeruginosa* genomics

1.3.1 Introduction to genomic analysis

The discipline of genomics is the study of an organisms entire genome via DNA sequencing and subsequent bioinformatic analysis. There are currently four major manufacturers within the “next generation" DNA sequencer market; Roche, Illumina, Life Technologies, and Pacific Biosciences.

The first commercially successful “next generation" DNA sequencer was the 454 by 454 Life Sciences, a company later purchased by Roche. This technology uses a method of sequencing called pyrosequencing. Pyrosequencing follows a “sequencing by synthesis” method that is able to sequence a single strand of DNA. The complementary strand is synthesised and chemiluminescence is utilised to detect which nucleotide is added to the strand one base pair at a time. Incorporation of a nucleotide will release pyrophosphate that will, based on the nucleotide incorporated, quantitatively convert PPi to ATP. Converted ATP will power the conversion of luciferin to the chemiluminescent oxyluciferin via luciferase (Ronaghi, Uhlen and Nyren, 1998). This technology is packaged in a variety of machines including both bench-top and stand alone systems. 454 sequencers are regarded as quick analysers and automation of library preparation allows for reduced manpower.

Like Roche, Illumina is responsible for bringing a variety of bench-top and stand alone sequencers to the market all of which utilise a “sequence by synthesis” method. Rather than relying on the release of detectable light via enzyme action, Solexa sequencing by Illumina detects incorporating bases via fluorescent labelling of each nucleotide. Based on the fluorescent light detected by the internal camera, the machine can determine the individual nucleotide added one base at a time (Bentley *et al*, 2008). This technology is capable of sequencing up to 1Mbp per second (Mardis, 2008).

A third “next generation” sequencer that is based on a “sequence by synthesis” method is the Ion Torrent system manufactured by Life Technologies. In contrast to the optical detection systems of 454 and Solexa sequencers, the Ion Torrent utilises a semiconducter based system (Rusk, 2011). When a complimentary nucleotide is incorporated onto the single stranded target DNA molecule, a hydrogen ion is released. This hydrogen ion, and subsequent increase in pH, is detected by the machine allowing the sequence to be determined. Available only in a bench-top format, the Ion Torrent offers researchers a low cost and fast option when choosing an appropriate DNA sequencer. Life Technologies is also responsible for a second sequencing platform; SOLiD (Sequencing by Oligonucleotide Ligation and Detection). SOLiD sequencing utilises a sequence by ligation method wherein oligonucleotides of a fixed length are labelled then annealed and ligated to the target DNA sequence. Oligonucleotides ligated will release a signal that allows determination of the DNA sequence of the target.

The fourth company that contributes significantly to the “next generation” sequencer market is Pacific Biosciences. They have developed a method of sequencing termed Single Molecule Real Time sequencing (SMRT). SMRT utilises a technique known as zero-mode waveguide (ZMW) (Levene *et al*, 2003). A ZMW is a structure within the platform that forms an illuminated observation volume that allows observation of only one single nucleotide of DNA being incorporated onto the DNA template via DNA polymerase. Each of the four individual nucleotides are differentially tagged with fluorescent dyes. Upon incorporation, the machine will detect which base has been added based upon the fluorescence detected (Levene *et al*, 2003).

The variety of platforms available and competition between each has lead to the improvement in quality and reduction in cost of DNA sequencing. This has resulted in the greater use of DNA sequencing in scientific research and as a consequence, an increased availability of DNA sequence information to the general research population.

1.3.2 The *P. aeruginosa* genome

The single chromosome of *P. aeruginosa* is relatively large, ranging from 5.2 to 7 million base pairs (Mbp) depending upon the strain (Wiehlmann *et al*, 2007). The *P. aeruginosa* genome has a relatively high GC content of around 66%. The first strain to have its genome fully sequenced was PA01 in 2000 by Stover *et al*. It is 6.3Mbp in length, which at the time, was significantly larger than most of the other fully sequenced bacterial genomes including that of the previously largest sequenced genome *E. coli* (4.6Mbp) and *B. subtilis* (4.2Mbp) (Stover *et al*, 2000). Its sequence is highly complex and with approximately 5570 open reading frames (ORF) whose number nears that of simple eukaryotes, for example the yeast *Saccharomyces cerevisiae* whose genome contains around 6200 ORF. As previously mentioned, *P. aeruginosa* is ubiquitous in the environment and is found in a wide range of niches. This can be attributed to the fact that, based on the sequence of PA01, its significantly larger genome is more complex and sophisticated than most other bacteria. It possesses a large number of genes that are involved in regulation, catabolism, transport/efflux of organic compounds and also a variety of chemotaxis systems (Stover *et al*, 2000; Lee *et al*, 2006). In fact, it has been predicted that almost one in ten genes possessed by PA01 is involved in regulatory function, which is one of the highest proportions observed amongst bacterial genomes (Stover *et al*, 2000; Battle *et al*, 2009).

Wolfgang and colleagues (2003) have previously demonstrated, using a whole genome microarray of 18 clinically and environmentally important strains, that a high percentage (~90%) of the *P. aeruginosa* genome is conserved amongst strains. The remainder of these genes, approximately 10%, are not conserved amongst strains and are expected to be present due to acquisition of a variety of genomic islands (Battle *et al*, 2009). These segments of DNA are acquired via horizontal gene transfer and will contain a GC content distinct to the core genome (Stover *et al*, 2000; Battle *et al*, 2009). These islands of “foreign” DNA will confer particular properties to the host strain, for example, genes encoding flagellar biosynthesis are acquired

this way (Lee *et al*, 2006). The gene encoding the ExoU cytotoxin, *exoU*, was acquired within a 80kb genomic island (Sato *et al*, 2003; Wolfgang *et al*, 2003). Although certain genomic islands may be specific to one or a few strains, others have been observed throughout many strains, for example, a 45kb island called *P. aeruginosa* genomic island-1 (PAGI-1) has been observed in 85% of clinically tested strains (Liang *et al*, 2001).

1.3.3 Current genomic understanding of the *P. aeruginosa* lineage

The current genomic understanding of *P. aeruginosa* is relatively limited. All fully sequenced and published strains of this bacterium are hosted on a dedicated online database; *Pseudomonas* genome database (Windsor *et al*, 2011). This database was first established to only host genomic data from *P. aeruginosa* strains but since 2005, it also provides data on other *Pseudomonas* species genomes.

Since the genome of PA01 was first published in 2000, a further 12 strains have had their genomes sequenced, annotated and uploaded to the *Pseudomonas* genome database (Windsor *et al*, 2011). The first of these was the genome of the highly virulent strain PA14 (UCBPP-PA14) (Lee *et al*, 2006). PA14 is a clinically significant, mulithost pathogen that demonstrates a far greater level of pathogenicity in a variety of organisms including humans, the nematode *C. elegans* and the plant *A. thaliana* (Lee *et al*, 2006; Liberati *et al*, 2006). The genome of PA14 is slightly larger than that of PA01 at 6.5Mbp containing 5973 ORFs, over 300 more than PA01. It was demonstrated that PA14 contained 95.8% of the PA01 genome and PA01 contained 92.6% of the PA14 genome. Therefore the enhanced virulence of PA14 compared to PA01, as well as other strains, has been attributed to the acquisition of a variety of horizontally acquired regions of genomic information, accounting for the larger genome and extra ORFs as observed by Lee *et al* (2006). Two such regions acquired by PA14 are the pathogenicity islands PAPI-1 and PAPI-2 (He *et al*, 2004). These 108kb and 11kb islands, respectively, are absent in PA01 and carry a variety of plant and animal virulence genes, 19 of which are required for full plant or animal virulence. Interestingly, many cystic fibrosis isolates carry one or both of these pathogenicity islands (Het *et al*, 2004).

Another of these fully sequenced strains, and the third sequenced *P. aeruginosa* strain, is the clinically significant LESB58 (Winstanley *et al,* 2009). This strain was the first archived isolate from the Liverpool Epidemic

Strain (LES). LES isolates were involved in the spread of a beta-lactam resistant *P. aeruginosa* amongst patients within a cystic fibrosis unit in Liverpool (Cheng *et al*, 1996). These LES isolates are the most frequently isolated clonal-type of *P. aeruginosa* isolated from patients in England and Wales and cases have also been observed in Scotland (Scott and Pitt, 2004, Edenborough *et al*, 2004). LES strains are extremely virulent and are associated with increased patient morbidity compared to other *P. aeruginosa* isolates (Al Aloul *et al*, 2004). Cases have existed where a LES strain was transmitted from a cystic fibrosis sufferer to non-sufferers and even a family cat causing significant morbidity (McCallum *et al*, 2002 and Mohan *et al*, 2008). Sequencing of the LESB58 isolate revealed that the increased persistence and transmissibility is due to the presence of several large genomic islands, including five prophage clusters, five non-phage clusters and a single defective (pyocin) prophage cluster (Figure 1.6). It was shown that these clusters conferred a competitive advantage to LES isolated in a chronic infection model (Winstanley *et al*, 2009).

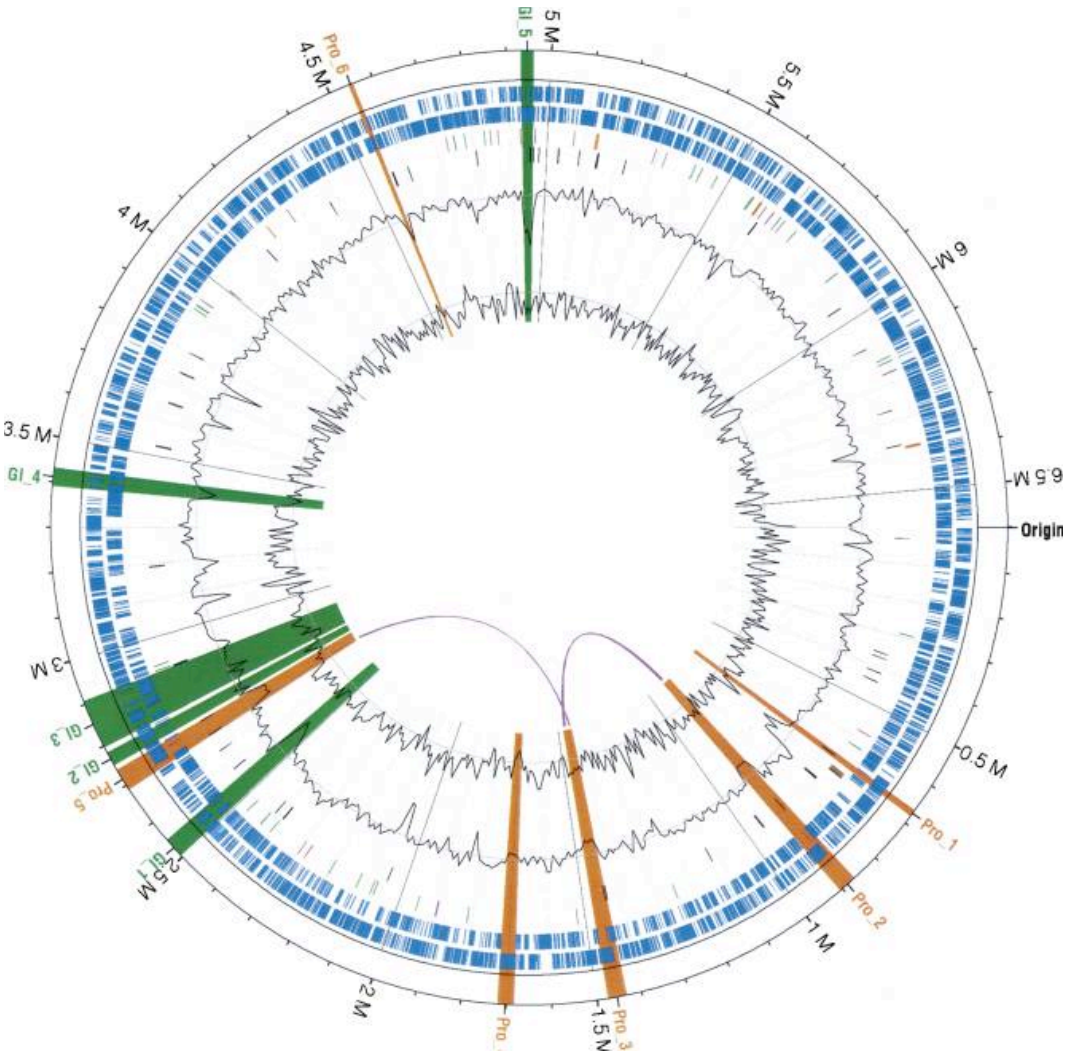
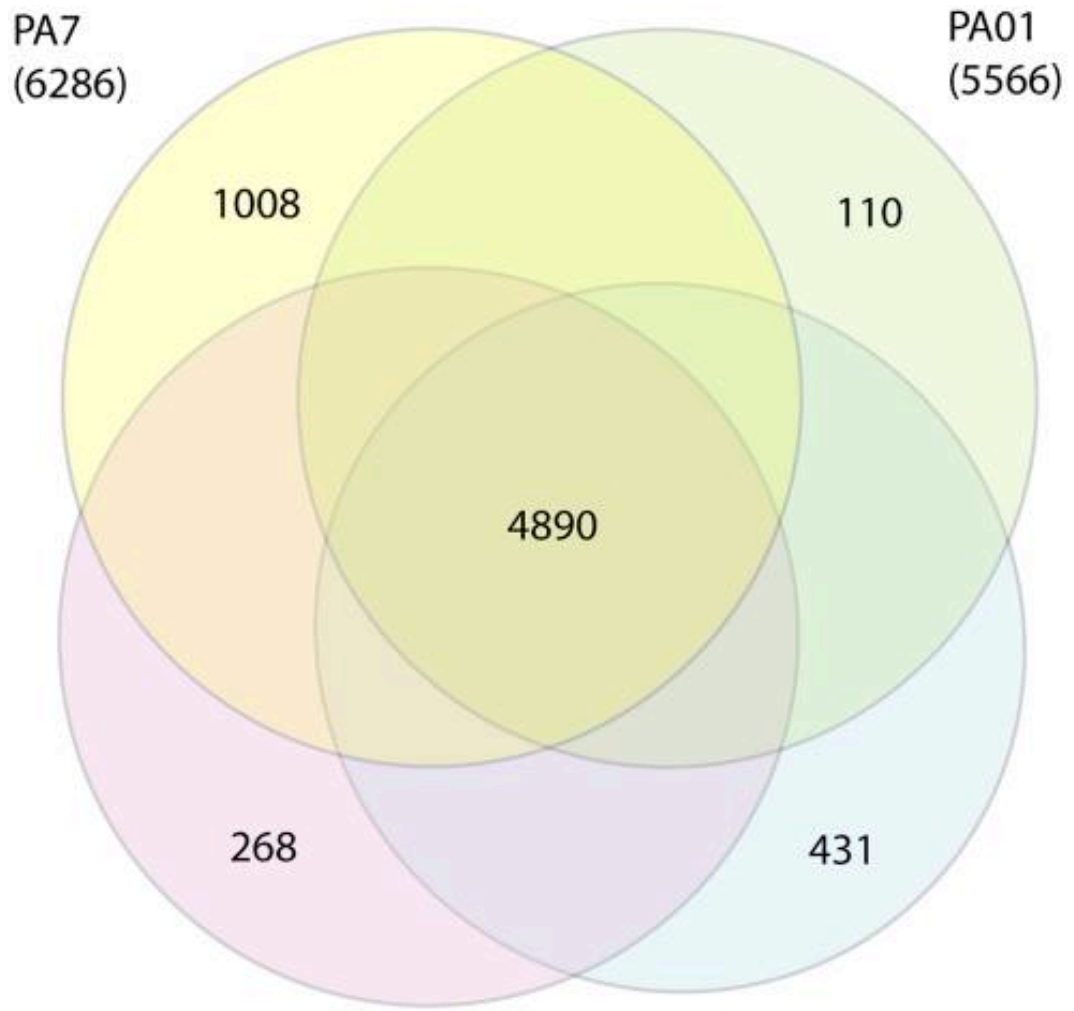


Figure 1.6 Circular map of the LESB58 genome. Prophage (orange) and genomic islands (green) that contribute to the competitiveness of LESB58 are identified. Also shown are protein coding genes (blue) on outer and inner strands, tRNAs (green), rRNAs (orange) and other non-coding RNA genes (purple) (Winstanley *et al*, 2009)

The fourth *P. aeruginosa* strain to have its genomic sequence determined was the taxonomic outlier PA7 (Roy *et al*, 2010). Despite being significantly distinct from other *P. aeruginosa* strains in evolutionary terms, the genome of PA7, like the three previously sequenced genomes of *P. aeruginosa*, has a GC content of around 66%. It possesses around 5-10% more ORFs (6286) compared to these strains (Figure 1.7) and has acquired 18 novel genomic islands, several of which are involved in antibiotic resistance. As a result of the acquisition of these novel genomic islands, PA7 possesses over 1000 ORFs that do not have significant homologues in either PA01, PA14 or LESB58 which have only 110, 268 and 431 unique ORFs, respectively (Figure 1.7). In fact, approximately only 80-90% of the ORFs identified in each of these strains are present in all genomes. PA7 is not as virulent as other strains due to the lack of a variety of potent virulence factor genes including *exoS, exoU, exoT* and *exoY* as well as *toxA* (Roy *et al*, 2010). The genomes of these strains do however share some similarities in that their percentage coding is consistent at around 89% and possess a similar amount of structural RNAs.

The sequence of PA7, bringing the total number of sequenced *P. aeruginosa* genomes to four, was published in 2010, the year this project began. Since then a further nine strains have been sequenced; B136-33, DK2, M18, NCGM2.S1, RP73, 2192, 39016, C3719 and PACS2. Although thirteen is significantly more than the original four, there is still a need to push for many more strains to be sequenced in order to achieve a fuller understanding of the *P. aeruginosa* genome.



	PA01	PA14	LESB58	PA7
Genome Size (bp)	6,264,404	6,537,648	6,601,757	6,588,339
GC content (%)	66.6	66.3	66.3	66.5
ORFs	5566	5892	5925	6286
% coding	89	89	88	89
Structural RNAs	77	72	81	75
Pseudogenes	5	None	34	8

Figure 1.7 **A** Venn diagram showing the number of common and unique ORFs between the four original sequenced *P. aeruginosa* genomes. The total number of ORFs in each strain are in parenthesis (Roy *et al*, 2010). **B** Table showing general genome features of each of the four original sequenced *P. aeruginosa* genomes including genome size, GC content, number of ORFs and % coding. Adapted from Roy *et al* (2010).

1.3.4 Benefits of unravelling the *P. aeruginosa* genome

Bacterial genome sequencing is becoming extremely popular amongst researchers with submissions to GenBank exponentially growing year on year. This increase is justified due to their use in clinical microbiology. Initially, genome sequencing was used to map bacterial evolution, a study that is still useful in *P. aeruginosa* research. Genomics could be utilised to identify evolutionary changes that cells may undergo during host adaptation. For example, the switch to mucoidy associated with chronic *P. aeruginosa* cystic fibrosis lung infection are attributed to mutations in the *mucA* anti-sigma factor gene (Boucher *et al*, 1997). Normal functioning copies of the protein product of this gene are responsible for repression of alginate production via inhibition of the AlgU sigma factor. Gene studies have shown that over 80% of mucoid strains isolated from the lungs of cystic fibrosis patients possess a mutation in the *mucA* gene (Pulcrano *et al*, 2012).

Genomic sequencing of bacterial genomes can also be used in molecular detection/genotyping to aid in molecular epidemiology, helping clinical researchers trace the origin of a particular outbreak - information that is central to the epidemiological surveillance performed by public health organisiations (Fournier, Drancourt and Raoult, 2007). This has also aided the development of molecular diagnostic tools, allowing researchers to identify a variety of specific DNA signatures that can be used as targets for PCR detection of pathogens from clinical specimens. One of the most widely used molecular tests in routine clinical microbiology diagnostic labs is a Real Time PCR system that allows for rapid detection of both *Chlamydia trachomatis* and *Neisseria gonorrhoea*, reducing the time of diagnosis from up to several days to only a few hours (Cheng, Qian and Kirby, 2011).

Genotype determination via genomic sequencing can be used to extrapolate the phenotypic profiles of bacteria. Researchers can use this data to identify and unravel metabolic pathways and determine essential nutrients required by the bacterium to survive. Data like this may be used to create recipes for

media providing essential nutrients for the cultivation of fastidious organisms. For example, culture medias have been created for optimal growth of *Haemophilus influenzae* (Schilling *et al*, 2000) and *Helicobacter pylori* (Schilling *et al*, 2002) using metabolic pathway data modelled using genomic sequences. Genomic sequences can be utilised to overcome the increasing emergence of antimicrobial-resistant pathogens via identification of resistance mechanisms. Identification of such mechanisms can be used by clinicians to identify resistant patterns in clinical isolates allowing for more rapid treatment. Phenotypic elucidation via genomic sequencing will also contribute to the understanding of the repertoire of virulence factors of a specific bacterium and its interactions with a host further contributing to the design of future antimicrobial agents as well as providing a greater understanding to the role of bacteria in infection and their clinical impact.

Aside from the prediction of phenotypic profiles of bacteria, genomic sequencing can be utilised for proteomic analysis allowing identification of all protein-encoding genes within a particular bacteria. Proteomic analysis can be utilised in a variety of ways, contributing significantly to the impact of care to a patient within a clinical setting. Serological tests can be developed based on the antigenic profile and subsequent immune response allowing rapid identification of infection via specific immunoglobulin capture from the serum of patients. Further, antigenic profiling via proteomics may be used in the design of vaccines.

At present, the use of genomic sequencing in clinical laboratories is limited however, with the rapidly declining overheads associated with this technology, sequencing and molecular diagnostics will become established as routine tests in infection management.

1.3.5 The “omics” revolution

In an attempt to utilise genomic data to enhance clinical understanding and control of infection, genomic sequencing has been combined with a variety of other “omic” studies. The “omic” suffix can be defined as “all constituents considered collectively”. In other words, genomics can be described as the collective study of all genes within a particular organism. Researchers now apply “omic” studies to other aspects of an organisms biology including their metabolic, protein and lipid profiles, termed metabolomics, proteomics and lipidomics, respectively. Collectively, these “omic” studies will provide researchers with a very thorough understanding of a particular organisms biology, potentially allowing the development of novel diagnostic, prevention and treatment techniques.

1.3.5.1 Proteomics

Proteomics is defined as the large-scale study of proteins, focussing mainly on their functions, structures and interactions with each other (Anderson and Anderson, 1998). Proteomic studies can be applied in a variety of situations including the identification of biomarkers, proteins which are characteristic of a particular biological response or pathogenic process. Biomarkers are often used in diagnostic laboratories to detect bacterial antigens that act as targets for the human immune response (Ceciliani *et al*, 2014). It is predicted that in the near future, diagnostic microbiology laboratories will heavily rely upon proteomics as an alternative to biochemical analysis in the identification of bacterial isolates from patient samples. Technologies are emerging that combine matrix-assisted laser desorption/ionisation (MALDI) with time-of-flight mass spectrometry (TOF) to provide a rapid and relatively cheap identification of bacterial isolates (Seng *et al*, 2009).

1.3.5.2 Metabolomics

Metabolomics is the study of the metabolome of a particular organism that consists of all intermediate and end products of all cellular processes within the cell. The metabolome is a highly complex network wherein an end product of one process may feed directly into another meaning the task of accurately mapping such data is an extremely complex approach. As such, metabolomic studies tend to be approached in a more targeted manner, looking for specific metabolites and their changes in response to a variety of stimuli. For example, metabolomics can be utilised in a variety of fields including evaluation of the condition of human health and the toxic effect of drugs during preclinical research (Ramirez *et al*, 2013).

Metabolomic analysis is also now being used alongside genomic sequencing in a field termed “functional genomics”. As the name suggests, functional genomics focusses on the functions of the DNA products including gene transcription and translation as well as protein-protein interactions as opposed to focussing on the DNA sequence and its structures. In these studies, metabolomic analysis can be utilised to identify any phenotypic changes that may arise from changes at the genomic level e.g. gene deletion/insertion (Bryant, Faruqi and Pinney, 2013). In this way, metabolomic analysis can be utilised to map the effect of genomic evolution in response to an environment or host on the metabolic processes of a specific organism. For example, metabolomic analysis may be useful during the evaluation of the effects of host adaptation of *P. aeruginosa* observed during cystic fibrosis lung infection and its effect on patient prognosis.

1.4 The present work

P. aeruginosa remains a constant threat both in the community and nosocomially, accounting for approximately 10% of hospital acquired infections (Davies, 2002). In particular, this free living, extremely virulent pathogen is a worry to clinicians involved in the care of patients suffering from cystic fibrosis, the genetically acquired, multi-system disease. The primary cause of death in such patients is respiratory failure resulting from chronic pulmonary infection and in over 70% of cases, this can can be attributed to *P. aeruginosa* (Lyczak, Cannon and Pier, 2002). As a result, the management of cystic fibrosis is largely dependent upon prevention and control of *P. aeruginosa* colonisation via antibiotic intervention. Due to the highly resistant and evolving nature of *P. aeruginosa*, such a heavy reliance on antibiotics for the management of cystic fibrosis can no longer be sustained. There is a need to develop new, novel treatments, a task only accomplishable by first expanding the understanding of this pathogen.

Currently, the understanding of *P. aeruginosa* and its role in infection is mainly limited to phenotypic aspects of the bacterium. For example, it is known to produce and secrete a vast array potent of virulence factors including a variety of adhesins, toxins, enzymes and molecules tasked with sequestration of a variety of nutrients. It is also known to be extremely competent at biofilm formation, a state of growth that makes *P. aeruginosa* significantly more resistant to antibiotic and host immune intervention (Costerton et al, 1999). Researchers have identified a striking phenotypic alteration that *P. aeruginosa* can undertake once a successful infection has been established; the switch to mucoidy associated with overproduction of the polysaccharide alginate (Doggett *et al*, 1999). This switch is most commonly observed in isolates from cystic fibrosis lung infections and has been shown to be associated with the progression from an acute to chronic infection and a significant deterioration in lung function (Pedersen *et al*, 1992).

During progression to chronic infection, the switch to a mucoid phenotype is accompanied by a variety of further phenotypic alterations including loss of antigenic appendages such as adhesins and pili, reduced expression of potent virulence factors and increased biofilm production (Hassett *et al*, 2009). Researchers have suggested that, due to the nature of these changes - beneficial to the bacterium in host immune and antibiotic evasion - the phenotypic alterations observed during the progression to chronicity occur in response to the host in a phenomenon knows as “host adaptation”. Such changes have been shown to be accounted for at the genomic level. For example, a mutation has been identified in the operon responsible for the production of alginate biosynthesis. Mutations have been identified within the anti-sigma factor *mucA*, resulting in uninhibited AlgU activity, alginate synthesis and the appearance of the mucoid phenotype (Boucher *et al*, 1997). It is expected that the bacterium will alter aspects of its genome in order to give itself a competitive advantage in its attempts at surviving within the cystic fibrosis lung. This, coupled with the fact that the progression to chronicity, the stage during which these changes occur, is associated with significant deterioration of lung functioning and the subsequent worsening of prognosis, it is important to fully elucidate the genomic changes that are occurring at this point in order to improve management of cystic fibrosis-associated infection.

When this project was visualised, only four *P. aeruginosa* strains had their genome fully sequenced. The first to be completed was PA01 in 2000 by Stover *et al*, followed by PA14 (Lee *et al*, 2006), LESB58 (Winstanley *et al*, 2009) and most recently, PA7 (Roy *et al*, 2010). Each of these strains are clinically significant pathogens and therefore, despite extensive analysis on these four genomes, the data obtained may not actually be representative of the “average” *P. aeruginosa* strain genome. This lack of data resulted in a poor understanding of the genome of this organism and a significant need for many more strains to be sequenced. Since then, a further nine *P. aeruginosa* genomes have been sequenced. Whilst this is a significant step towards

developing our understanding of the genome of this organism, the majority of the strains sequenced were isolated from clinical samples. As a result, the diversity of *P. aeruginosa* is still not truly understood.

1.4.1 Aims of the work

The ultimate aim of developing a more thorough understanding of the *P. aeruginosa* genome is to contribute to the discovery of new, novel treatments to be used in the case of clinical infections, most notably those affecting the lungs of cystic fibrosis patients. To achieve this, we aim to provide the *P. aeruginosa* and cystic fibrosis community a more thorough understanding of the genome of this pathogen and more specifically, the role of this pathogen in infections of the cystic fibrosis lung. With regards to this aim, the project revolves around the sequencing and analysis of a further twelve strains. Each strain, the details of which can be found in Table 3.3, was chosen based upon their significance in nature and their relevance in the literature. Four clinical pairs have been chosen, three non-mucoid early CF lung isolates and their subsequently isolated mucoid pairs as well as PA17 and its lab evolved small colony variant. Four environmental strains have also been sequenced to provide the project with a significant aspect of strain diversity.

We will look at the placing of these twelve genomes in the overall population structure of *P. aeruginosa*. To do this our aim is to perform a comprehensive analysis of the core genome including all currently available *P. aeruginosa* genomes on GenBank. The core genome contains the genes essential for the basic functioning of the bacterium as well as its major phenotypic characteristics. This eliminates nonessential genes and those gained/lost via horizontal gene transfer and therefore focuses on the most conserved portions of the genome (Medini *et al*, 2005). As a result, our study will hopefully provide a more insightful understating of the diversity of the core genome of *P. aeruginosa*.

We will also use our new genomic data to evaluate the role of host adaptation during infections of the cystic fibrosis lung. Bacteria often utilise this phenomenon as a mechanism of survival and growth in hostile conditions. *P. aeruginosa* is particularly efficient at this and can hypermutate in response to internal and external challenges such as host immunity and antibiotic intervention (Folkesson *et al*, 2012). This process is essential to the success of the bacterium within its environment, facilitating the switch from an acute infection to a more persistent chronic infection. A variety of changes in the bacterium can be observed phenotypically during this time including deflagination, increased antibiotic sensitivity and perhaps most commonly and significantly, a switch from a non-mucoid to a mucoid phenotype as previously detailed (Hassett *et al*, 2009). Strains appear to be highly selective, readily disposing of any factors that are deemed surplus to requirements. This contributes to the high level diversity observed within the species *P. aeruginosa*. We hypothesise that such phenotypic changes will be attributable to changes at the genomic level. As mentioned, we have sequenced three non-mucoid early CF lung isolates and their subsequently isolated mucoid pairs. We will attempt to identify if these mucoid pairings have arisen as a result of host adaptation or strain succession. Following on from this, we will use any true host adapted pairs to thoroughly assess the effect of host adaptation on the phenotypic and genotypic diversity of *P. aeruginosa*. We will attempt to map some of the genomic changes that arise as a result of this process to any changes observed in the phenotypic profile. In this regard, a variety of phenotypic analyses will be utilised including virulence models, biofilm growth assays, antibiotic resistance assays and large scale metabolomic analysis.

We hypothesise that during infection of a cystic fibrosis lung, *P. aeruginosa* undergo a variety of phenotypic changes as a result of host adaption that will be accounted for at the genomic level. In addition to this, sequencing of a batch of diverse strains it is expected that we will develop a more thorough understanding of the genome of *P. aeruginosa*.

Chapter 2 - Materials and Methods

2.1 Strains and plasmids

All strains used for this study are listed in tables 2.1 and 2.2. When required, strains were streaked out onto an appropriate agar, containing relevant antibiotics where necessary (Section 2.4.2). Long-term storage of each strain was attained using MicroBank vials (VWR Scientific), prepared following manufacturers instruction and stored indefinitely at -20^{O}C. Plasmids used in this work are listed in Table 5.3.

Strain Identifier	Comments/Origin	Reference
DH5α	*sipE44* ∆(*lacU169* (Φ80d*lacZ*∆M15) *hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	Hanahan, 1983

Strain Identifier	Accession number	Isolation Scenario	Phylogenetic Group	Comments/Origin	Reference
J1385	ASRB00000000	CF	2	CF, Progenitor of J1532	This Work, Stewart *et al*, 2013
J1532	ASQZ00000000	CF	2	CF, Mucoid	This Work, Stewart *et al*, 2013
C1426	ASRD00000000	CF	1	CF, Progenitor of C1433	This Work, Stewart *et al*, 2013
C1433	ASRC00000000	CF	1	CF, Mucoid	This Work, Stewart *et al*, 2013
PA17	ASQY00000000	CF	1	CF, Progenitor of PA17SCV	This Work, Stewart *et al*, 2013
PA17SCV	ASRA00000000	CF	1	Lab evolved from CF Isolate	This Work, Stewart *et al*, 2013
C763	ASQS00000000	CF	1	CF, from same patient as C1334	This Work, Stewart *et al*, 2013
C1334	ASQT00000000	CF	1	CF, Mucoid, from same patient as C763	This Work, Stewart *et al*, 2013
E2UOS	ASQV00000000	Envir	2	Soil	This Work, Stewart *et al*, 2013
PA62	ASQW00000000	Envir	1	Soil	This Work, Stewart *et al*, 2013
MSH3	ASQU00000000	Envir	1	Lake, Mount St. Helens	This Work, Stewart *et al*, 2013
MSH10UoS	ASQX00000000	Envir	1	Lake, Mount St. Helens	This Work, Stewart *et al*, 2013
PA14	NC_008463.1	Envir	2	Plant, USA	Lee *et al*, 2006
PA01	NC_002516.2	Tissue	1	Burn	Stover *et al*, 2000

Plasmid	Comments/origin	Reference
pMiniCTX	Integration proficient plasmid for cloning into *P. aeruginosa*	Hoang *et al*, 2000
pGEM-T Easy	Blunt end cloning vector	Promega
pRK2013	Helper plasmid for triparental mating	Deretic *et al*, 1986

2.2 Suppliers

All chemicals and reagents, unless otherwise stated, were purchased from one of the following suppliers: Sigma-Aldrich, Fisher Scientific or BDH Chemicals.

2.3 Enzyme suppliers

All Restriction Enzymes, supplied with corresponding buffer and Bovine Serum Albumin, were purchased from either Promega or New England Biolabs. DNA polymerases such as Accuzyme^{Tm}, MyTaq^{Tm} SensiFAST^{Tm} SYBR Hi-ROX kit, were obtained from Bioline. A subsequent DNA polymerase, T_4 DNA polymerase, used for its 3' exonuclease activity, was supplied by Promega, as was T_4 DNA Ligase.

2.4 Buffers and solutions

2.4.1 Media

All media were prepared by dissolving the appropriate quantity of reagents, as listed below, in distilled water. These blends were then autoclaved for 15 minutes at 121^{O}C and 15 psi. If solid media were required, 2% (w/v) of laboratory grade agar (Melford) was added to mixture prior to autoclaving. Agar plates were prepared by pouring molten agar, cooled to 50^{O}C, into sterile Petri dishes in aliquots of around 20ml each. Plates were then left in a laminar flow hood until they had set and excess liquid had evaporated.

Lysogeny Broth (pH 7.0)	1% (w/v) Tryptone 0.5% (w/v) Yeast extract 0.5% (w/v) NaCl
Nutrient Yeast Medium (pH7.0)	1.3% (w/v) Nutrient broth base 0.5% (w/v) Yeast extract

King's A Medium (pH7.0)	2% (w/v) Enzymatic digest of gelatin 0.14% (w/v) $MgCl_2$ 1% (w/v) K_2SO_4 1% (v/v) Glycerol
King's B Medium (pH7.0)	2% (w/v) Protease peptone 0.15% (w/v) K_2HPO_4 0.15% (w/v) $MgSO_4 \bullet 7H_2O$ 1% (v/v) Glycerol
Mueller-Hinton Medium (pH7)	30% (w/v) Beef extract 1.75% (w/v) Casein hydrosylate 0.15% (w/v) Starch
M9 Medium	50 mM Na_2HPO_4-$7H_2O$ 20 mM KH_2PO_4 10 mM NaCl 20 mM NH_4Cl 0.2% (v/v) Glycerol 2 mM $MgSO_4$ 0.1 mM $CaCl_2$
M63 Medium (pH 7.0)	15 mM $(NH_4)2SO_4$ 100 mM KH_2PO_4 9 µM $FeSO_4 \cdot 7H_2O$ 1 mM $MgSO_4$ 0.2% (w/v) Glucose 0.1% (w/v) Casamino acids

MacConkey Medium	1.7% (w/v) Enzymatic digest of gelatin 0.15% (w/v) Enzymatic digest of casein 0.15% (w/v) Enzymatic digest of animal tissue 0.15% (w/v) Bile salts 0.5% (w/v) NaCl 0.003% (w/v) Neutral red 0.0001% (w/v) Crystal violet 1% (w/v) Maltose

2.4.1.1 Artificial Sputum Media

Artificial sputum media (ASM) is designed to replicate growth conditions experienced by bacterial cells within the cystic fibrosis lung (Kirchner *et al*, 2012). This media contains components of CF sputum including amino acids, free DNA and mucin. ASM was prepared according to Kirchner *et al* (2012) with the following modification; Salmon sperm DNA and mucin were sterilised via ethanol soaking and were added to the media post-filtering.

2.4.2 Antibiotics

Antibiotics were added to the media as and when required in the following final concentrations:

Rifampicin	100 µg/ml
Tetracycline	12.5 µg/ml (*E. coli*) 50 µg/ml (*P. aeruginosa*) 125 µg/ml (*P. aeruginosa* + pMiniCTX)
Carbenicillin	100 µg/ml
Kanamycin	50 µg/ml

Gentamicin	15 µg/ml
	30 µg/ml

2.4.3 DNA and RNA buffers

These buffers were used for the manipulation of either DNA or RNA:

TE Buffer	10 mM Tris-HCl (pH8.0)
	1 mM EDTA (pH8.0)
50X TAE Buffer	24.2% (w/v) Tris-base
	5.71% (v/v) Glacial Acetic Acid
	50 mM EDTA (pH8.0)
5X Loading Dye	0.25% (w/v) Bromophenol Blue
	0.25% (w/v) Xylene Cyanol FF
	50% (v/v) Glycerol

2.4.4 Protein buffers

These buffers were used for the manipulation of protein

2X Laemmli Buffer	60 mM Tris-HCL (pH6.8)
	2% (w/v) SDS
	10% (v/v) Glycerol
	5% (v/v) β-mercaptoethanol
	0.01% (w/v) Bromophenol Blue
1X Running Buffer	25 mM Tris
	192 mM Glycine
	0.1% (w/v) SDS
1X Transfer Buffer	192 mM Glycine
	25 mM Tris
	10% (v/v) Methanol

1X TBST	100 mM Tris-HCL (pH7.8) 150 mM NaCl 0.1% (v/v) Tween 20

2.4.5 Assay buffers

Sörensen buffer (pH 8.0)	59.5 mM KH_2PO_4 277.8 µM Na_2HPO_4

2.5 Microbiological methods

2.5.1 Preparation of competent *E. coli* cells

E. coli cells were grown up overnight at 37°C in a 5 ml starter culture in LB. This overnight culture was used to inoculate a 250 ml conical flask containing 100 ml of LB. The flask was returned to the 37°C incubator and cells were allowed to grow until an OD_{600} of 0.3-0.4 was reached. Cells were immediately chilled down on ice to inhibit growth. Cells were harvested by centrifugation at 4000 rpm for 5-10 minutes at 4°C in 50 ml Falcon tubes. After carefully discarding the resulting supernatant, cells were slowly resuspended in 12.5 ml of chilled 100 mM $MgCl_2$. Resuspended cells were then subjected to further centrifugation under the conditions previously described. The supernatant was discarded and cells were resuspended in 25 ml of chilled 100 Mm $CaCl_2$. Yet again, the suspended cells were harvested via centrifugation at 4000 rpm for 10 minutes at 4°C. Cells were finally resuspended in 3 ml of 100 mM $CaCl_2$ in 20% (w/v) glycerol and then stored in aliquots of 100 µl at -80°C.

2.5.2 Transformation of competent *E. coli* cells

Transformation reactions were carried out in 1.5 ml centrifuge tubes containing 50 µl of competent cells mixed with approximately 0.5 µg of plasmid DNA. The reaction was initially incubated on ice for 30 minutes, followed by a heat-shock treatment at 42°C for 90 seconds before being incubated on ice for a further 2 minutes. To the reaction, 450 µl of LB broth was added and mixed prior to incubation at 37°C for 1 hour to allow recovery of the cells. Following incubation, 200 µl of the transformed cells were plated out onto LB agar containing the appropriate antibiotic. After an overnight incubation at 37°C, plates were check for cell growth.

2.5.3 Triparental mating of *P. aeruginosa*

For conjugation of *P. aeruginosa*, triparental mating using pRK2013 (Deretic *et al*, 1986) was used. A single colony of the donor strain (DH5α containing recombinant pMiniCTX), helper (DH5α containing pRK2013) and recipient (*P.*

aeruginosa strain) were mixed on a LB agar plate with a sterile inoculating loop and incubated overnight at 37°C. The resultant growth was then transferred to an LB agar plate containing 150 µg/ml tetracycline and 100 µg/ml carbenicillin, added to kill DH5α containing recombinant pMiniCTX and DH5α containing pRK2013, respectively. This plate was then incubated for a further 24-48 hours at 37°C and growth was checked to confirm transformation.

2.5.4 Biofilm assay

Fresh overnight cultures were diluted 1:100 in fresh M63 media. 100 µl of this dilution was pipetted into the wells of a 96-well microtitre plate. Plates were incubated at 37°C for 20 hours in a static incubator. Post-incubation, cells were discarded and the plate was gently submerged in ddH_2O to remove any non-attached cells. The ddH_2O was removed via blotting on tissue paper and 150 µl 0.3% (w/v) crystal violet was added to each well and left at room temperature for 10 minutes to allow adhered cells to take up the stain. Crystal violet was removed and the plate was washed several times as described previously. Excess water was removed via rigorous blotting on tissue paper and wells were allowed to air dry for one hour. To each well, 150 µl 100% ethanol was added and allowed to incubate at room temperature for 10 minutes to solubilise the crystal violet retained by the cells. Biofilm growth was determined via reading the absorbance of the solubilised crystal violet in ethanol at 582 nm.

2.5.5 Protease assay

Brain heart infusion agar was prepared as described by the manufacturer (Oxoid, UK) and was supplemented post-autoclaving with sterile 1% (w/v) skimmed milk. Carbenicillin was added to a final concentration of 500 µg/ml to prevent growth of contaminating bacteria. Plates were poured and holes with a 5 mM diameter were drilled in the media. Bacterial cultures were grown in LB for 48 hours at 37°C with shaking at 250 rpm. Cultures were centrifuged at 4000 rpm for 10 minutes to pellet cells. Cell supernatants were removed and kept. Plates were retrieved and 50µl culture supernatant was

added to the holes previously drilled. Inoculated plates were incubated at 37^{o}C until a zone of lysis appeared. Protease activity was determined based upon the diameter of the zone of lysis.

2.5.6 Lipase Assay

The lipase assay was adapted from Winkler and Stuckmann (1979). This assay utilises *p*-nitrophenylpalmitate (*p*NPP) as a substrate. 30 mg of *p*NPP was dissolved in 10ml isopropanol and was then mixed with 90 ml of 0.05 M Sörensen buffer containing 207 mg of sodium deoxycholate and 100 mg of gum arabic. Bacterial cultures were grown in LB for 48 hours at 37^{o}C with shaking at 250 rpm. Cultures were centrifuged at 4000 rpm for 10 minutes to pellet cells. Cell supernatants were removed and kept. 80 µl culture supernatant was mixed with 920 µl prewarmed (37^{o}C for 5 minutes) substrate solution. After a 30 minute incubation at 37^{o}C, the OD_{410} was recorded. One enzyme unit was defined as 1 nmol of *p*NP/ml/minute released from the substrate. Under these conditions, the extinction coefficient of *p*NP is 15000 cm^2/mg.

2.5.7 Esterase Assay

The esterase assay was adapted from Wilhelm, Tommassen and Jaeger (1999). This assay utilises *p*-nitrophenyl caproate (*p*NPC) as a substrate. 23.7 mg of *p*NPC was dissolved in 5 ml 99% (v/v) ethanol and was then mixed with 95 ml of 100 Mm potassium phosphate buffer containing 10 mM $MgSO_4$. Bacterial cultures were grown in LB for 48 hours at 37^{o}C with shaking at 250 rpm. Cultures were centrifuged at 4000 rpm for 10 minutes to pellet cells. Cell supernatants were removed and kept. 80 µl culture supernatant was mixed with 920 µl substrate solution and the OD_{410} was recorded at time points 5-10 minutes One enzyme unit is defined as the amount of enzyme which forms 1 µM of *p*NP/ml/minute released from the substrate. Under these conditions, the extinction coefficient of *p*NP is 10400 M^{-1}.

2.5.8 Pyocyanin assay

Cultures were grown in liquid King's A medium. After a 26 hour incubation at 37^{o}C with shaking (250 rpm), the OD_{600} was recorded. Cultures were then centrifuged for 10 minutes at 4000 rpm and the supernatant was collected. The OD_{695} of the supernatant was recorded and pyocyanin production levels were determined by normalising to the previously recorded OD_{410} values.

2.5.9 Siderophore secretion assay

The siderophore secretion assay was taken from Schwyn and Neilands (1987) utilising chrome azurol S (CAS) and hexadecyltrimethylammonium bromide (HDTMA) as indicators. A CAS stock solution was prepared by mixing 9 ml of 2 mM CAS in ddH_2O with 1ml of 1.7 mM $FeCl_3$ in 10 mM HCl. This mix was then added to 40 ml of 5 mM HDTMA in ddH_2O prior to sterilisation by autoclaving. PIPES agar was prepared by mixing 870 ml M9 minimal media with 30 ml of 10X LB and 100 ml of the previously described CAS stock solution and was subsequently poured into individual petri-dishes. Fresh overnight cultures were prepared in LB and incubated at 37^{o}C with shaking (250 rpm) and 2 µl aliquots were spotted on onto the CAS/PIPES media. The formation of an orange halo indicated siderophore production.

2.5.10 Nitrocefin Assay

Fresh overnight cultures in LB were diluted 1:20 with prewarmed, cation-adjusted Mueller-Hinton broth. Cultures were incubated for 90 minutes at 37^{o}C with shaking (250 rpm). β-lactamase production was induced with the addition of clavulanic acid to a final concentration of 50 µg/ml and cells were incubated under the same conditions for a further 2.5 hours. Cells were harvested via centrifugation at 4000 rpm for 10 minutes and cells were washed once with 100 Mm phosphate buffer (pH7.0) and finally resuspended in 1ml phosphate buffer. Using dry-ice, cells were lysed via 3 cycles of freeze/thawing. Residual cells were removed via centrifugation at 20000 ×g (4^{o}C) for 20 minutes. The resultant extract was diluted 1:5 in phosphate buffer and nitrocefin (Calbiochem) was added to a final concentration of 100 µM in 100 mM phosphate buffer (pH 7.0) and 1 mM EDTA (pH 7.0). The

assay was left to incubate at 30^{o}C for 30 minutes prior to recording the absorbance at OD_{486}. The protein concentration of the crude extracts were determined using a standard colorimetric bradford assay (Bio-Rad, UK).

2.6 Metabolomic analysis

2.6.1 Metabolomic extractions

Overnight cultures were grown in ASM at 37^{o}C with shaking at 250 rpm. Post incubation, cultures were diluted to an OD_{600} of 0.05 (±0.01) and were used to inoculate Artificial Sputum Media (ASM) 1/100. Inoculated cultures were incubated at 37^{o}C for 72 hours with shaking at 75 rpm. Immediately after this period of incubation, cells were quenched by the addition of 1/2 volume of cold methanol (-20^{o}C). Cells were transferred to suitable polypropylene tubes and a further 1/2 volume of cold ethanol was added. Cells were snap-frozen using liquid nitrogen and were then placed on ice for one hour. Cells were then centrifuged at 4000 rpm for ten minutes and the supernatant was recovered. The cells were then re-extracted and centrifuged under the same conditions and the supernatants were pooled. Extractions were concentrated using a vacuum centrifuge until <1 ml remained. Extractions were then corrected to exactly 1 ml with the addition of 100% methanol to normalise metabolite concentrations.

2.6.2 Thin Layer Chromatography

To confirm that metabolite extractions were a success, TLC was used to visualise the amino acid profile of each extraction. Standard 20x20 cm Whatman TLC plates were purchased from Fisher Scientific. The mobile phase used was prepared as follows. n-Butanol, acetic acid (purity 98-100%) and ddH_2O were mixed in a 5:1:5 ratio. The mixture was left to separate and the upper layer was recovered and used as the eluent. 5 µl of each extraction was spotted onto the plate at equal distances and left to dry. The plate was introduced into the eluent in an appropriate chamber. Elution was stopped once the solvent front had travelled up the plate to within 7-10 mM of the end of the plate. The plate was removed from the chamber and left to dry. Visualisation was achieved by finely spraying the plate with 0.3% (w/v)

ninhydrin in n-butanol supplemented with 3% (v/v) acetic acid. The plate was incubated at 60^{o}C until staining occurred.

2.6.3 Liquid Chromatography - Mass Spectrometry

LCMS on all samples was performed using a Dionex Ultimate 3000 (Thermo Fisher Scientific) system consisting of a pump, autosampler and column compartment. Samples were run on a SeQuant ZIC®-pHILIC (Merck) column. The column was connected to a Thermo Scientific Exactive Mass Spectrometer (Thermo Fisher Scientific) running in positive/negative scanning mode using the parameter settings as detailed in Tables 2.4 and 2.5. Acetonitrile and Ammonium carbonate (20 mM, pH 9.2) in water were used as solvents.

Scan Parameter	Setting
Scan range (m/z)	100 – 1200
Resolution	50,000
Microscans	3
Lock masses (positive/negative)	195.0877 / 265.1479
ACG target	1,000,000
Maximum inject time	250 ms

Table 2.4 Scan parameter setting used for LC-MS analysis

ESI Source Parameter	Setting
Sheath gas flow rate (ml/min)	50
Aux gas flow rate (ml/min)	17
Sweep gas flow rate (ml/min)	0
Spray voltage (positive/negative) (|kV|)	4500 / 4000
Capillary temperature (ºC)	275
Capillary voltage (positive/negative) (V)	28 / -30
Tube lens voltage (positive/negative) (V)	95 / -110
Skimmer voltage (positive/negative) (V)	21 / -15

Table 2.5 Electrospray Ionisation source parameters used for LC-MS

2.6.4 Data analysis of LC-MS output

Processing of raw LC-MS outputs and subsequent putative metabolite identification and analysis were all performed using the graphical user interface of IDEOM (Creek *et al*, 2012)

2.7 Nucleic acid purification

2.7.1 Genomic DNA preparation

Genomic DNA from *P. aeruginosa* strain PA14 was used as a template for PCR reactions. These preparations were carried out from 1 ml overnight cultures using the Genomic DNA and total RNA Purification kit (Macherey-Nagal), as directed by the manufacturer.

2.7.2 Plasmid DNA purification

Small volume plasmid isolations were performed from 5 ml overnight cultures using a NucleoSpin Plasmid kit (Macherey-Nagal), following manufacturers instructions. Larger plasmid purifications were performed on 50 ml overnight cultures using a QIAfilter Plasmid Midi kit (Qiagen) as per the manufacturers instructions.

2.7.3 RNA extraction

Bacterial cultures are added to 1 X volume of RNAprotect (Qiagen), mixed vigorously and left to incubate at ambient temperature for five minutes. The mix is centrifuged at 13000 rpm for 10 minutes and the supernatant was discarded. The cell pellet was resuspended in 100 µl phosphate buffered saline (PBS) (Sigma). Lysozyme (Sigma) was added to a final concentration of 1 mg/ml, mixed and left to incubate at ambient temperature for two minutes. RNA was extracted and isolated using TRI reagent (Sigma) as per manufacturers guidelines. The resulting RNA pellet was resuspended in 50 µl RNase-free water prior to DNase I (Qiagen) treatment. RNA extractions were cleaned using an ISOLATE II RNA Mini Kit (Bioline). Eluted RNA samples were stored at -80^{o}C until required for downstream applications.

2.7.4 Nucleic acid quantification

Nucleic acids purified from methods described in sections 5.7.1 to 5.7.3 were quantified using either a NanoDrop 2000 (Thermo Scientific) or Qubit (Life Technologies) fluorometric quantitation.

2.8 Manipulation of DNA

2.8.1 Polymerase Chain Reaction

Primers for all PCR experiments were designed so that all reactions could be performed using the following conditions: One cycle initial denaturation at 95 ^{o}C for 5 minutes. Thirty-five cycles of denaturation (95 ^{o}C for 30 seconds), annealing (55 ^{o}C for 30 seconds) and extension (72 ^{o}C for 15 seconds/250 bp). A final one cycle extension at 72 ^{o}C for 5 minutes. All PCR experiments were performed using MyTaq (Bioline) and were set up according to manufacturers instructions. Primers used in PCR are listed in Table 2.6.

2.8.2 Restriction endonuclease digestions

DNA was digested at 37^{o}C for 1.5 hours using an appropriate concentration of enzyme (1 U/µg substrate DNA) in the correct buffer, as is indicated by the manufacturer.

2.8.3 Agarose Gel Electrophoresis

DNA was visualised using agarose gel electrophoresis in order to determine its size and also to purify specific fragments. The gel was prepared by melting 1% agarose in TAE buffer in a microwave. To aid in DNA visualisation, once the melted agarose had cooled sufficiently, UV-sensitive ethidium bromide was added to a final concentration of 1 µg/ml and the gel was then poured into a mould and allowed to set. DNA was mixed with 5X loading dye (5.4.4) and loaded to the gel, alongside a suitable molecular weight marker. Gels were typically run at 80 V for between 30 minutes and 1.5 hours, depending on the size of the DNA fragments.

2.8.4 DNA ligations

T4 DNA ligase was utilised in ligation reactions to catalyse the formation of the phosphodiester bonds between the exposed 5' and 3' DNA ends. Typically, such reactions were performed via incubation at 16^{o}C overnight in the buffer supplied by the manufacturers. Occasionally, Quick-Stick Ligase (Bioline) was used, significantly reducing reaction time, only requiring a five minute incubation at room temperature.

2.8.5 MiniCTX promoter fusions

MiniCTX, is an integration-proficient plasmid that was used to clone genes that could be incorporated into the *P. aeruginosa* genome via conjugal transfer.

2.8.6 Phage transduction

High titre lysates were prepared by serial dilutions of φPA3 (Monson *et al*, 2011) on the donor strain using 4 ml of 0.35% LB top agar. Cultures were incubated overnight at 37^{o}C. Top agar from plates displaying confluent lysis was scraped and the plate was washed using 1 ml LB before pooling both.

500 µl chloroform was added to the recovered top agar and the mixture was vortex vigorously until confluent. The mix was centrifuged at 4000 rpm for 20 minutes at 4^{o}C. The supernatant was recovered and centrifuged for a further five minutes at 14000 rpm. The supernatant containing the donor phage was filtered and then stored at 4^{o}C with additional chloroform until required. Overnight cultures of an appropriate recipient strain were grown at 37^{o}C. Estimating a MOI of 0.1, cultures were infected with the previously prepared high titre lysates and were mixed thoroughly prior to a 30 minute incubation at ambient temperature without shaking followed by a 20 minute incubation at 37^{o}C with shaking (250 rpm). Post incubation, cells were pelleted by centrifugation at 4000 rpm for ten minutes, washed in LB and centrifuged a second time under the same conditions. The supernatant was discarded and the cells were resuspended in 300 µl. Two equal aliquots (150 µl) were spread onto LB agar plates containing an appropriate concentration of antibiotic based upon the resistance pattern of the gene being transduced. Plates were incubated at 37^{o}C until colonies appeared.

2.9 Real-time Polymerase Chain Reaction methods

2.9.1 Reverse transcription

Reverse transcription was performed using M-MLV Reverse Transcriptase (Promega) and random primers (Promega). 10 Mm of each dNTP was also added.

2.9.2 qPCR

qPCR was performed using a StepOnePlus (Life Technologies) thermocycler using SYBR Green I as a nucleic acid stain. For ease of use and reproducibility, a SensiFAST SYBR Hi-ROX (Bioline) mastermix was used and reactions were set up following manufacturers instructions. Thermocycling conditions were as follows; 95^{o}C for two minutes followed by 40 cycles of denaturation at 95^{o}C for 5 seconds and combined annealing/ extension at 62^{o}C for 20 seconds. Primers used for this assay are detailed in Table 2.7.

Primer	**Sequence**
16S forward	5'-cggyccagactcctacgg-3'
16S reverse	5'-ttaccgcggctgctggca-3'
ampR forward	5'-tgctgtgtgactccttcgac-3'
ampR reverse	5'-agatcgatgaagggatggc-3'
ampC forward	5'-ctcgacctcgcgacctatac-3'
ampC reverse	5'-gatgctcgggttggaataga-3'

2.10 Whole genome sequencing methods

2.10.1 Genomic DNA preparation

Genomic DNA was prepared from each strain for the purpose of genome sequencing using the same method previously described in section 5.7.1

2.10.2 Shearing of genomic DNA

Prior to entering the sequencing reaction, genomic DNA it was necessary to shear the DNA into fragments of approximately 200 bp in length. This was performed using two separate machines, each was used for the preparation of DNA from six organisms; the Covaris S220 and Bioruptor Plus.

2.10.3 Preparation of 50bp paired end fragments

Sequencing was performed using the Illumina Genome Analyser II that utilises paired end sequencing with read pairs having a length of 50 bases. The sequencing library, including barcoding of genomes, was prepared following instructions provided with the machine and all twelve genomes were sequenced in one flow cell lane. Individual, unique barcoding of each genome facilitated separation of each strain sequence.

2.11 Bioinformatic work flows

2.11.1 Mapping to reference genome

Each 50bp paired-end read, contained within .FASTQ files, was aligned against the genome sequence of PA14 using Mapping and Assembly with Qualities (MAQ) software. The output file from this analysis (.map) was converted into a Sequence Alignment/Map (SAM) format using SAMtools (Li *et al*, 2009). Using the same software, this SAM file was then converted to a Binary Alignment/Map (BAM). The Sanger Institute software Artemis: Genome Browser and Annotation Tool (Rutherford *et al*, 2000) was used to view the BAM file read against the annotated PA14 sequence.

2.11.2 *De novo* assemblies

De novo assembly was performed using the CLC Genomics Workbench software (CLCbio, Aarhus, Denmark). The files produced by this software were in the .fasta format. The assembled contigs contained within these files were reordered using Mauve (Darling *et al*, 2004), producing a .fas.fa file comprising fewer and larger contigs in a more ordered manner. These files were then annotated using the xBase bacterial genome annotation service (Altschul *et al*, 1997; Chaudhuri *et al*, 2007; Delcher *et al*, 2007; Lagesen *et al*, 1997; Kurtz *et al*, 2004; Lowe and Eddy, 1997). Annotated files produced by xBase were in the GeneBank (.gbk) format and could be used for genome browsing on the Artemis: Genome Browser and Annotation Tool (Rutherford *et al*, 2000) as well as further comparative genomics such as SNP and InDel detection on the CLC Genomics Workbench software (CLCbio, Aarhus, Denmark), identification of genomic islands using IslandViewer (Langille and Brinkman, 2009) and secondary metabolite prediction using antiSMASH (Medema *et al*, 2011). BLAST comparison was performed and visualised using the CGview tool (Stothard and Wishart, 2005).

2.11.3 Multilocus Sequence Typing

MLST was performed following instructions from the *P. aeruginosa* MLST scheme (Jolley and Maiden, 2010); this allowed the comparison of the

sequences of seven house-keeping genes (*acsA*, *aroE*, *guaA*, *mutL*, *nuoD*, *ppsA* and *trpE*). The internal fragments of these genes were extracted from the *de novo* assemblies of each of the sequenced strains. One exception was the internal fragment for acsA which was not possible from several strains. Therefore, PCR was utilised using the corresponding genes listed in table 5.4. Fragment sequences were obtained from these PCR products by external sequencing at Eurofins MWG Operon (Eurofins Scientific Group, Germany) using the following sequencing primers present in Table 5.4.

Tree drawing based on this scheme was performed using the Phylodendron tool within the PHYLIP package (Felsenstein, 1989). An eBURST figure was produced using the BURST algorithm (Feil *et al*, 2004).

2.12 Statistical analysis of results

Statistical analysis was performed using GraphPad Prism 6 (GraphPad Software) and Microsoft Excel (Microsoft Corporation).

Chapter 3 - Population genomics of the *Pseudomonas aeruginosa* lineage

3.1 Introduction

When this project was first envisaged, there was a limited understanding of the genomic structure of the *P. aeruginosa* species, with only four fully sequenced strains; PA01, PA14 LESB58 and PA7. Our aim was to contribute significantly to the limited understanding of the genomic structure of the *P. aeruginosa* lineage. Working towards this aim, we fully sequenced the genomes of a further twelve *P. aeruginosa* strains (Table 3.1). As previously mentioned, this includes four host-adapted mutants and their revertant pairings, a further clinical strain and its lab evolved small colony variant as well as four environmental strains. Sequencing was performed using Illumina technology, utilising a 50 bp paired end read method. Upon retrieving the raw sequence data produced using this sequencing technology, the first challenge to overcome was assembling the reads into a contiguous sequence.

Sequence assembly was performed via two methods. The first of these involved mapping the reads to a closely related strain whose sequence has already been published. The *P. aeruginosa* strain PA14 was chosen for this (Figure 3.1). This method has been utilised for further analysis on a very limited basis. The second, and perhaps most useful, assembly method was *de novo* assembly using the CLC genomics workbench. *De novo* assembly matches paired end reads to form a large number of contigs of varying size. These contigs must be reordered and grouped to produce larger contigs in an order more representative of the true sequence. This task was performed using the Mauve Aligner (Rissman *et al*, 2009). This software reorders and groups the contig files for each strain several times, creating larger contigs whilst reducing the total number of contigs within each file significantly. Working draft genomes were compiled by creating annotated versions of

these files using the freely available automated annotation service provided by xBase (Chaudhuri *et al*, 2008). Annotations of our genomes was performed using the previously annotated *P. aeruginosa* strain PA14 as a reference genome (Lee *et al*, 2006).

The resulting assembled and annotated draft GenBank files were of sufficient quality to be used for genomic comparison between the *P. aeruginosa* strains sequenced. An example of one of these .gbk files can be seen in Figure 6.1. Where necessary, contig boundaries were bridged using PCR and subsequent Sanger sequencing. Table 3.1 demonstrates the efficiency of our sequencing and assembly methods, listing the number of contigs and the n50 statistic for each of our draft genomes. Also described within this table is the genomic length of each of our sequenced strains. As can be seen the majority of our strains fall within the previously described *P. aeruginosa* genome size of between 5.2-5.7 million base pairs (mbp) (Mathee *et al*, 2008). However, one of our strains, C763 is significantly larger than this expected range with a genome size of over 7.3 mbp.

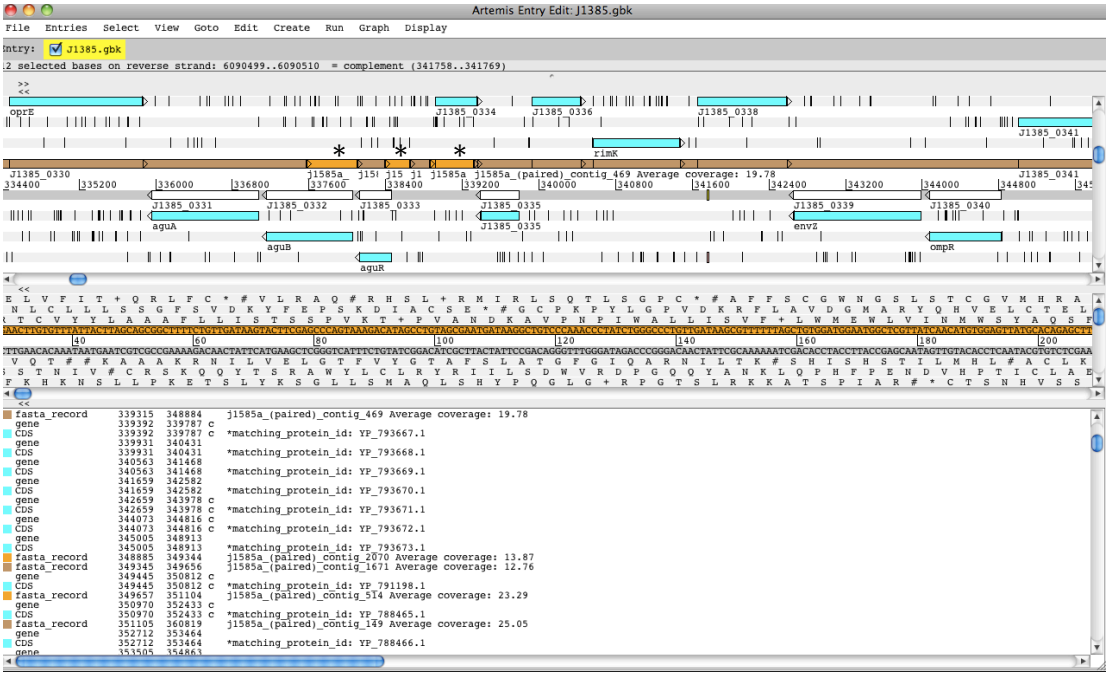


Figure 3.1 Screenshot showing an example of .gbk output from xBase. Several genes were annotated based on the published annotations of *P. aeruginosa* strain PA14. Contig boundaries are visualised within this screenshot as alternating brown and orange regions of genome, indicated by *.

Strain	Number of contigs	n50	Total Bases
J1385	1212	9672	6775891
J1532	1177	9726	6619350
C1426	186	72819	6663710
C1433	1016	10919	6440071
C763	307	41571	7340720
C1334	254	44300	6515869
MSH3	392	32458	6693052
MSH10	280	45133	6765229
E2	189	72170	6753193
PA62	227	54963	6968208
PA17	353	35716	6595428
PA17scv	695	16962	6439398

Table 3.1 Assembly efficiencies of each of our sequenced strains, showing the length, total number of contigs and the n50 statistic for each genome sequence.

3.2 Multi Locus Sequence Typing (MLST)

The first comparative genomics analysis undertaken was MLST analysis (Jolley *et al*, 2004). This is a global scheme that is a “nucleotide sequence based approach for the unambiguous characterisation of isolates of bacteria and other organisms”. There are around 40 individual bacterial schemes, each focusing on one particular bacterium. The *P. aeruginosa* scheme is based on the sequences of 7 housekeeping genes (*acsA, aroE, guaA, mutL, nuoD, ppsA, trpE*). Researchers amplify and sequence each of these genes for their strain of interest and then submit the data to the scheme database. So far the profiles of over 2000 isolates have been submitted. Using the previously mentioned .gbk files and the MLST server at the Centre for Biological Sequence Analysis (CBS) in Denmark (Larsen *et al*, 2012), we extracted the sequences of each of these genes, eliminating the need for PCR amplification and sequencing. We determined the allelic profile of each gene and finally, each strains sequence type (Table 3.2) using the software made available on the MLST website (Jolley *et al*, 2004).

Strain	*acsA*	*aroE*	*guaA*	*mutL*	*nuoD*	*ppsA*	*trpE*	Sequence Type
J1385	4	4	16	12	1	6	3	253
J1532	4	4	16	12	1	6	3	253
C1426	6	20	1	3	4	4	2	132
C1433	6	20	1	3	4	4	2	132
C763	36	27	28	3	4	13	7	179
C1334	6	68	20	11	4	4	7	439
PA17	105	5	30	3	3	10	14	810
PA17scv	105	5	30	3	3	10	14	810
PA62	11	5	1	4	4	6	17	97
E2	38	11	3	13	1	2	4	235
MSH3	39	5	11	30	4	15	94	689
MSH10	39	5	11	30	4	15	94	689

Table 3.2 Table showing the allelic profile of each gene from each strain and the sequence type as determined by the MLST database (Jolley *et al*, 2004)

Such MLST data can be utilised in several ways, including phylogenetic analysis. We utilised the phylogenetic tree drawing tool Phylodendron (Gilbert, 1999) to help identify any similarities that exists between our strains. This tree, viewable in Figure 3.2, shows the relationship of our 12 sequenced strains to each other and also to several previously sequenced reference strains including PA14, PA01, LESB58 and the significantly distinct, genomic outlier PA7. From this figure it can be seen that amongst our strains we have four highly similar pairings; MSH3 & MSH10, PA17 & PA17scv, C1426 & C1433 and J1385 & J1532. This does not come as a surprise as each strain was isolated from the same location as its partner and in the case of PA17scv, we know that this was evolved from PA17 in laboratory conditions. Perhaps the most interesting and surprising outcome from this analysis is the fact that the non-mucoid/mucoid pairing of C763 and C1334 was found to be non-isogenic. We know that these strains were isolated from the lungs of the same CF sufferer and that each strain was the dominant strain at the time of isolation. As a result of this, we expected and C763 had undergone genotypic and phenotypic change, “mutating” into the mucoid C1334. However, our MLST data and subsequent phylogenetic analysis suggests that rather than the occurrence of a host adaptation event as predicted, the switch in strain dominance was most likely due to C763 being succeeded in the lung by C1334. One possible explanation for such a switch is the fact that mucoid strains of *P. aeruginosa* are more persistent in the CF lung than non-mucoid variants (Hewitt *et al*, 2005). It is possible that the extensive antibiotic intervention that is associated with CF was sufficient only at attenuating the growth of C763 allowing the more persistent mucoid C1334 to flourish and establish a chronic infection.

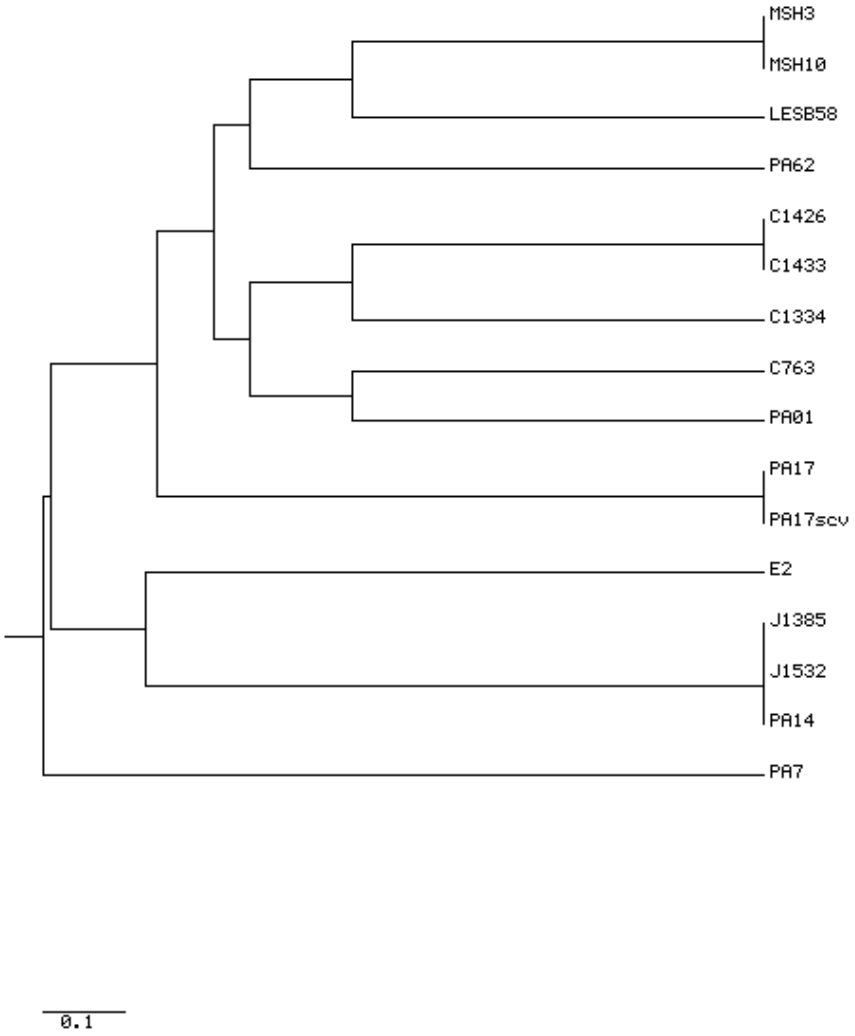


Figure 3.2 A phylogentic tree drawn using MLST data reveals that our collection includes four highly similar pairings; MSH3 & MSH10, PA17 & PA17scv, C1426 & C1433 and J1385 & J1532.

The MLST data shown in Table 3.2 was also used to visualise the positioning of our strains within the lineage of the *P. aeruginosa* strains already submitted to the MLST database (Jolley & Maiden, 2010). This was performed using the online eBUSRT algorithm (Fell *et al*, 2004). The BURST algorithm identifies and matches the genotypes of related strains and then predicts the founding genotype of the profiles included in the analysis. Based on this predicted founding genotype, eBURST then produces a radial diagram with the genotypes of each profile placed in relation to the founding genotype. As can be seen in the diagram produced by eBURST (Figure 3.3), our strains are highly diverse and are scattered throughout the predicted current population structure. The majority of our submitted profiles already exist within the dataset and their allelic profiles match existing profiles exactly. However, rather interestingly, two of our sequence types 810 (PA17 & PA17scv) and 689 (MSH3 & MSH10) are unique to our dataset and have never previously been submitted to the existing MLST scheme.

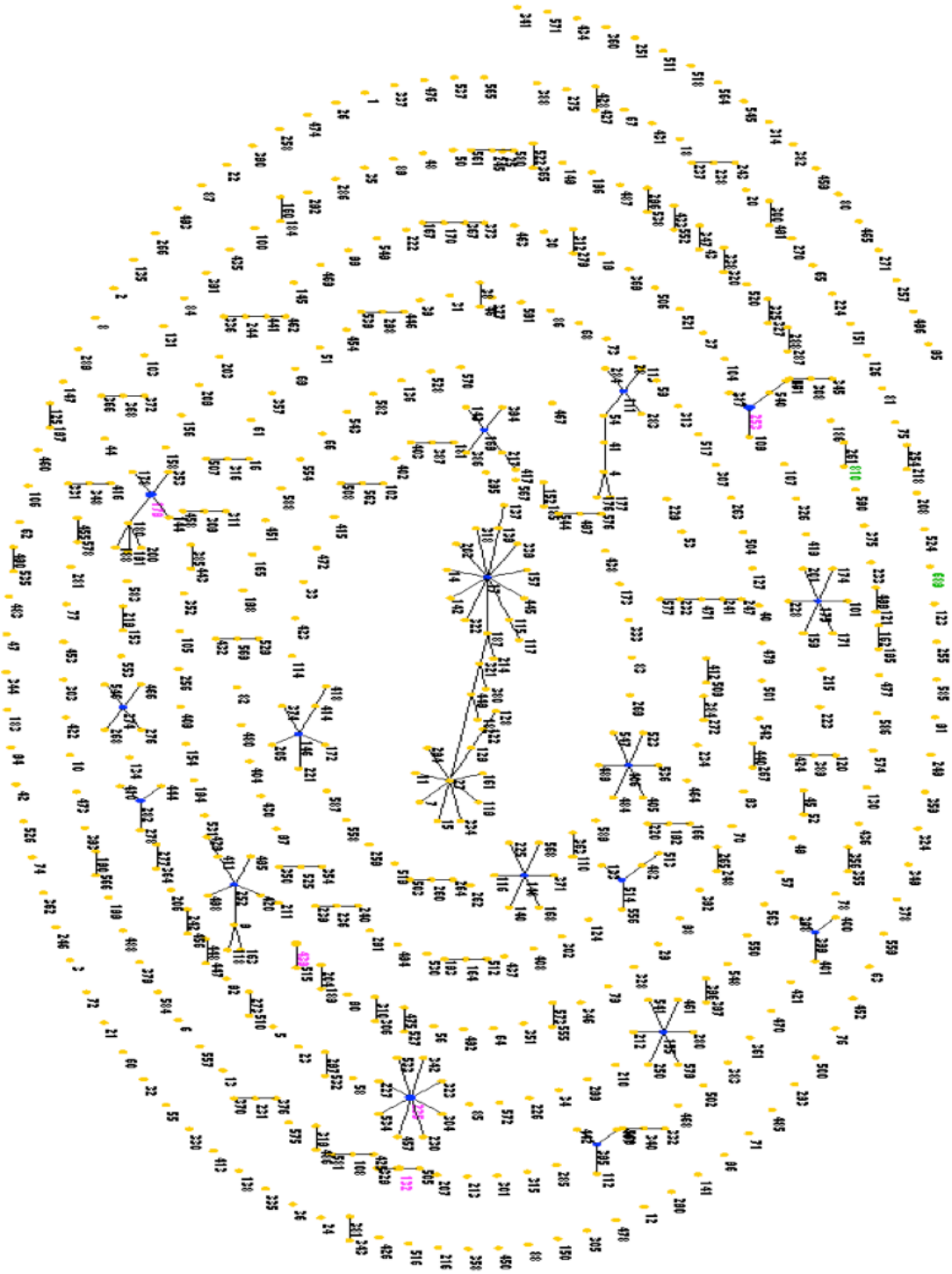


Figure 3.3 Sequenced strains from this study are scattered throughout the currently predicted species population as demonstrated by eBURST diagram analysis. Profiles present in both datasets are in PINK whereas those unique to our dataset are in GREEN.

3.3 Diversity of the core genome

Although MLST data can be manipulated in a variety of different ways as has been demonstrated here, the observations generated using this data are of limited use due to the highly interpretative and predicted nature of subsequent analyses. To obtain a better understanding of the position of our strains in the population of *P. aeruginosa*, we carried out a more extensive investigation into the diversity of their core genomes. The core genome contains the genes essential for the basic functioning of the bacterium as well as its major phenotypic characteristics. This eliminates nonessential genes and those gained/lost via horizontal gene transfer and therefore focuses on the most conserved portions of the genome (Medini *et al*, 2005). As a result, this evaluation and population studies of the core genome provides a more useful understanding of the diversity of the core genome of *P. aeruginosa*.

We utilised Panseq (Laing *et al*, 2010) to analyse the core genome of 55 *P. aeruginosa* strains including the genome sequences of the 12 strains used throughout this work (Table 3.3). To ensure that only essential genes and therefore “true” core genes were included in the analysis, the genome threshold was set at 55 i.e. all genomes must contain the gene in question for it to be considered a true member of the core genome and be used in the analysis. Sequence identity was set at 90%. The SNPs in the core genomes identified by Panseq (v2.0) were then used for phylogenetic analysis using SplitsTree (v4.13.1) that can be viewed in Figure 3.4.

Strain Identifier	Comments/Origin	Phylogenetic Group	Accession	Reference
J1385	CF, Progenitor of J1532	2	ASRB00000000	This Study
J1532	CF, Mucoid	2	ASQZ00000000	This Study
C1426	CF, Progenitor of C1433	1	ASRD00000000	This Study
C1433	CF, Mucoid	1	ASRC00000000	This Study
PA17	CF, Progenitor of PA17SCV	1	ASQY00000000	This Study
PA17SCV	Lab evolved from CF Isolate	1	ASRA00000000	This Study
C763	CF, from same patient as C1334	1	ASQS00000000	This Study
C1334	CF, Mucoid, from same patient as C763	1	ASQT00000000	This Study
E2UOS	Soil	2	ASQV00000000	This Study
PA62	Soil	1	ASQW00000000	This Study
MSH3	Environmental, Mount St. Helens	1	ASQU00000000	This Study
MSH10 UoS	Environmental, Mount St. Helens	1	ASQX00000000	This Study
2192	CF	1	AAKW00000000	(Mathee et al. 2008)
39016	Keratitis	2	AEEX01000000	(Stewart et al. 2011)
138244	Sputum, non-CF	1	AEVV00000000	(Soares-Castro et al. 2011)
152504	Sputum, non-CF	2	AEVW00000000	(Soares-Castro et al. 2011)
18A	CF	1	CAQZ00000000	Unpublished
AES-1R	Australia, CF	1	AFNF00000000	(Naughton et al. 2011)
AH16	China, Sputum, non-CF	1	ALJH00000000	(Wu et al. 2012)
ATCC 14886	Soil	1	AKZD00000000	(Chugani et al. 2012)
ATCC 25324	Glass crusher air plate	1	AKZE00000000	Unpublished
ATCC 700888	Industrial water system	2	AKZF00000000	(Chugani et al. 2012)
B136.33	Diarrheoa	2	CP004061	Unpublished
C3719	Manchester Epidemic Strain	1	AAKV00000000	(Mathee et al. 2008)
CI27	Sputum, CF	2	AKZG00000000	(Chugani et al. 2012)
CIG1	Sputum, CF	1	AKBD00000000	(Chugani et al. 2012)
DK2	CF, Denmark	1	NC_018080.1	(Rau et al. 2012)
DK2_CF510	CF, Denmark	1	AJHI00000000	(Rau et al. 2012)
DQ8	Crude oil, China	1	ALIO00000000	(Gai et al. 2012)
E2UoW	Plant, USA	1	AKZH00000000	(Chugani et al. 2012)
LCT-PA102	Blood	1	AJKG00000000	(Fang et al. 2012)
LESB58	Liverpool Epidemic Strain	1	NC_011770.1	(Winstanley et al. 2009)
M18	Plant isolate, China	1	CP002496	(Wu et al. 2011)
MRW44.1	Mutator lineage 44	1	ALBW00000000	(Weigand & Sundin 2012)
MSH-10	Environmental	1	ASWW00000000	Unpublished, Broad Institute
N002	Crude oil, India	1	ALBV00000000	(Abhjit Sarma Roy et al. 2013)
NCGM2.S1	Urinary tract infection	2	AP012280	(Miyoshi-Akiyama et al. 2011)
NCMG1179	Respiratory, Japan	1	BADP00000000	(Tada et al. 2011)
PA01	Burn	1	NC_002516.2	(Stover et al. 2000)
PA14	Plant, USA	2	NC_008463.1	(Lee et al. 2006)
PA19BR	Polymixin resistant, Brazil	1	AFXJ00000000	(Boyle et al. 2012)
PA21_ST175	Blood, MDR	1	AOIH00000000	(Viedma et al. 2013)
PA213BR	Polymixin resistant, Brazil	1	AFXK00000000	(Boyle et al. 2012)
PA45	Blood, Italy	1	APMD00000000	(Segata et al. 2013)
PA7	Non-respiratory clinical isolate	7	NC_009656.1	(Paul H Roy et al. 2010)
PA9BR	Brazil, non-CF	1	AFXI00000000	(Boyle et al. 2012)
PAb1	Frostbite	2	ABKZ00000000	(Salzberg et al. 2008)
PABL056	Blood, USA	2	ALPS00000000	(Ozer et al. 2012)
PACS2	CF	1	NZ_AAQW00000000	(Mathee et al. 2008)
PAK	PAK	1	ASWU00000000	Unpublished, Broad Institute
PAO579	Mucoid	1	ALOF00000000	(Withers et al. 2012)
SJTD-1	Crude oil, China	1	AKCM00000000	(Liu et al. 2012)
VRFPA01	Blood, India	7	AOBK00000000	Malathi & Madhavan, unpublished
VRFPA02	Eye, India	1	AQHM00000000	Malathi & Madhavan, unpublished
XMG	Soil, China	1	AJXX00000000	(Gao et al. 2012)

Table 3.3 *P. aeruginosa* genomes and corresponding accession numbers used in this study

3.3.1 Three phylogenetic groups exist in the *P. aeruginosa* species lineage

To the best of our knowledge, such a comprehensive study into the diversity of the core genome of *P. aeruginosa* has never been performed. Previous studies determined the species diversity by either focussing on specific genes, a technique similar to that utilised by MLST (Pirnay *et al*, 2009 and Kiewitz & Tummler, 2000) or including analysis of markers acquired via horizontal transfer events (Wiehlmann et al, 2007). The advantage of our method is that it separates only the most conserved portions of the genome such that the resulting phylogenies are not influenced by horizontal gene transfer events in the accessory genome.

We demonstrated that, based on the strains included in the analysis, strains of *P. aeruginosa* fall into one of three clearly evident groups (Figure 3.4). This unrooted UPGMA tree shows that this species possesses two major groups with a much smaller and extremely distinct third group containing the taxonomic outliers PA7 and VRFPA01. Focussing on our twelve sequenced strains the majority exist within Group one, whilst only three strains lie within Group two. It also further confirms that the previously identified pairings of J1385 & J1532 and C1426 & C1433 are indeed isogenic pairs. Figure 3.4 also shows that the environmental strains MSH3 and MSH10, both isolated from the same location on Mount St. Helens, are near identical strains with only a few genomic difference, corroborating the findings of our MLST analysis.

As mentioned, and as can be seen in Figure 3.4b, strains PA7 and VRFPA01 are taxonomic outliers and lie at a striking distance from the remainder of the *P. aeruginosa* strains in this study. Our analysis using Panseq identified SNPs in 118934 positions of the core genome of this species. However, running the same analysis, under the same conditions but excluding these outliers, this number fell to 73206. This demonstrates just how distant these

isolates are from the general *P. aeruginosa* population with almost 40% of the total SNPs accounted for by only 4% of the strains included. Based on this, we have shown that *P. aeruginosa* strains can be broadly divided into two major subgroups. For simplicity, we have named these groups Group 1 and Group 2 with the third outlying group named Group 3. Of the two main groups, Group 1 is clearly the larger of the two, containing approximately 75% of strains analysed. This group contains many notable strains such as DK2, PA01, LESB58 and PAK. Group 2 is significantly less well populated containing only twelve strains including the highly virulent PA14. Our MLST analysis, as well as whole genome BLAST analysis, identified that our non-mucoid/mucoid pairing of J1385/J1532 shows a high degree of homology to PA14, a finding that is clearly supported by the the core genome phylogeny present in Figure 3.4.

No clear pattern has emerged in regards to the grouping of strains in relation to their source of isolations. For example, strains isolated from CF sputum are well represented in both Groups 1 and 2. The only exception is that all hydrocarbon remediation strains are located in Group 1, however this cannot be assumed to be a true reflection of such strains due to their limited inclusion. We also must remember that only 55 strains have been included within this study. Therefore, we cannot assume that Group 1 strains are definitely more abundant than Group 2 and may have to consider that such a pattern has emerged due to the selection of strains included in our analysis. Further analysis using a larger and broader dataset is necessary, however the limited number of strains of *P. aeruginosa* sequenced applies some limitations to our conclusions.

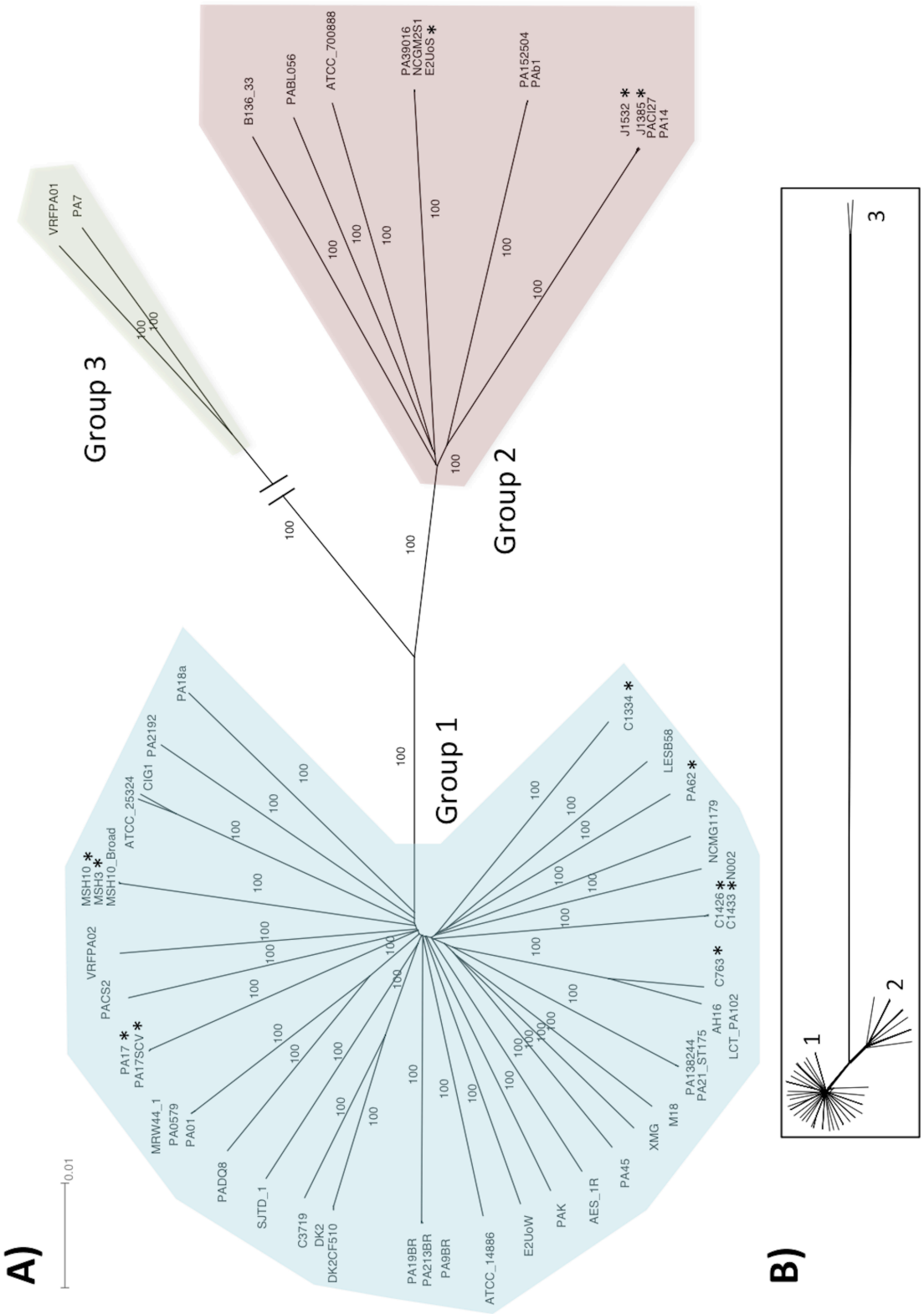


Figure 3.4 **A** Unrooted UPGMA tree of 55 *P. aeruginosa* genomes based on SNPs within the core genome as defined by panseq with a core genome threshold of 55 and required sequence identity of 90%. The three groups are indicated in blue (group 1), pink (group 2) and green (group 3). The tree was prepared in SplitsTree 4 and was bootstrapped with 1000 replicates. Because group three is so distant from groups one and two, it has been truncated for clarity. Strains sequenced in our study have been identified by ✱. **B** Unrooted UPGMA tree, produced under same conditions as **A** but showing the true distance of group three in relation to groups one and two.

3.4 Chapter 3 discussion

Initially, at the beginning of this project, the understanding of the genome of *P. aeruginosa* was extremely limited with only four sequenced strains published. Since then, the *Pseudomonas* Genome Database (Windsor *et al*, 2011) now has been updated to include a further nine published and fully sequenced genomes taking the total to thirteen with many more partially sequenced strains including the twelve presented in this study. Although still not sufficient, we now have a significant dataset to begin to predict the population structure of the *P. aeruginosa* lineage based on the currently sequenced strains.

The majority of analyses trying to elucidate a greater understanding of the genome of this species currently focus on the widely used Multilocus Sequence Typing technique. Using MLST analysis, we provisionally identified several isogenic pairings in our twelve strain collection including the non-mucoid/mucoid pairings of J1385/J1532 and C1426/C1433. Based on this observation we predict that these pairings arose as a result of host adaptation and the resulting genomic changes. Interestingly, our third non-mucoid/mucoid pairing of C763/C1334 was shown to be non-isogenic. This finding suggests that rather than host adaptation, the switch in strain dominance within this CF sufferers lung was as a result of strain succession, a possible explanation for which is the increased persistence observed in mucoid strains when challenged with antibiotic intervention (Hewitt *et al*, 2005). Other isogenic pairs identified in our collection include the Mount Saint Helens isolates MSH3 and MSH10 as well as PA17 and its lab evolved small colony variant strain PA17scv. Upon comparison to the complete *P. aeruginosa* MLST database, it was identified that both of these pairings have never been previously submitted and can be considered new isolates.

As mentioned, MLST analysis is extremely limited in its use in predicting the population structure of *P. aeruginosa*. This technique predicts strain sequence similarities based on the DNA sequence of only seven genes. In

other words, MLST analysis predicts genome similarities based on the sequence of only ~0.12% of the total *P. aeruginosa* genes, a phenomenally low percentage of genome meaning that many significant genomic anomalies may be overlooked. Therefore, there is a need to develop a more thorough method to predict strain sequence similarities and overall population structure. This project aimed to and succeeded in developing such a method. Our method, utilising Panseq (Laing *et al*, 2010) considers only genes considered to be in the core genome of *P. aeruginosa* excluding any genes that may have been horizontally acquired. The core genome was defined as the collection of all genes that exist in all 55 strains analysed with no exceptions and no less than 90% homogeneity across these strains.

This method has revealed that strains of the *P. aeruginosa* species fall into two major groups as well as a very distant third group containing two taxonomic outliers (Figure 3.4). This figure shows that the largest of these groups is group one, containing 75% of the strains analysed whereas group 2 is significantly smaller containing only twelve strains. Unfortunately, due to the small study population of only 55 strains, it cannot be assumed that these groupings and their proportions are a true reflection of the *P. aeruginosa* population. However, we believe that this analysis has contributed significantly to the limited understanding of the species structure described in previous studies (Pirnay *et al*, 2009, Kiewitz & Tummler, 2000 and Wiehlmann et al, 2007).

This analysis of the core genome has also proved invaluable in our analysis of our twelve sequenced strains. It demonstrates that the non-mucoid/mucoid pairing of J1385 and J1532 is indeed a near isogenic pair further confirming the observations made using MLST data. This is also the case of the second non-mucoid/mucoid pair C1426 and C1433. This supports our predictions that the switch in strain dominance is due to host adaptation, in other words, J1385 and C1426 have “mutated” or “evolved” into J1532 and C1443, respectively. It is important to not rule out a second possibility that within the

CF lung both early and late strains existed concurrently but were not isolated at both time points. Our third pairing of this nature, C763 and C1334 have again been confirmed to be distinct from one another and this further supports our belief that the switch is due to strain succession rather than host adaptation.

Our method to map and visualise the population structure of *P. aeruginosa* is deficient in one are in that its analysis has included relatively few *P. aeruginosa* strains. However, as can be seen in Table 3.3 our strains represent a broad range of host niches including an array of medically and environmentally isolated strains. We believe that upon the inclusion of a significant number of additional strains, our method of investigating the population structure of *P. aeruginosa* will yield comparable findings and it is our aim to prove so in the future. Firstly we need the *P. aeruginosa* genomics community to significantly increase the amount of data currently available.

Chapter 4 - Bacterial adaptation as a mechanism of survival

4.1 Introduction

A large part of this thesis focuses on the phenomenon commonly observed in infections of the CF lung called host adaptation. Under hostile growth conditions, bacteria undergo a variety of changes as part of host adaptation as a mechanism of survival and growth. *P. aeruginosa* is particularly adept at this, hypermutating in response to internal and external challenges such as host immunity and antibiotic intervention (Folkesson *et al*, 2012). Host adaptation is an essential process that must be undertaken prior to the switch from an acute infection to a more persistent chronic infection. A variety of changes occur in the bacterium during this time including deflagination, increased antibiotic sensitivity and perhaps most commonly, a switch from a non-mucoid to a mucoid phenotype as previously detailed (Hassett *et al*, 2009). Strains may display a degree of selectivity, potentially disposing of any factors that are considered surplus to requirements by the cell. This contributes to the high level diversity observed within the species *P. aeruginosa*. We have acquired two such pairings (C1426 & C1433 and J1385 & J1532) that have potentially undergone host adaptation whilst infecting the lungs of cystic fibrosis patients, resulting in a switch to the mucoid phenotype. The following two chapters will delve further into the changes occurring during these periods and will attempt to link some of the phenotypic changes observed to changes at the genomic level.

4.2 Phenotypic analyses of C1426 and C1433

The first of our two pairings is the non-mucoid C1426 and its isogenic mucoid strain C1433. Both of these strains were isolated from the sputum aspirated from the lungs of the same CF patient one month apart. During this time, the dominant strain switched from C1426 to C1433 (Professor John Govan, personal correspondence). As part of the work carried out in Chapter 3, the genomic sequences of these strains have shown that the switch between these strains is most likely a result of host adaptation, and the resulting genomic changes, rather than co-infection or strain succession.

A variety of phenotypic differences have been observed between these strains of which, the most striking difference is the appearance of a mucoid phenotype in the late strain C1433. However, this is not the only difference observed between the pairs. C1433 is also proficient at forming biofilms (Figure 4.1) and producing pyocyanin, esterase and lipase (Table 4.1). On the other hand, this strain produces lower level of protease and siderophore (Table 4.1). Previous work within our group has shown that C1433 is more resistant to antibacterial intervention (Rita Lagido, unpublished results), has reduced virulence and has reduced motility (Amy Ford, unpublished results). Over the course of this chapter, the genomic basis for any phenotypic differences existing between these strains will be explored and discussed.

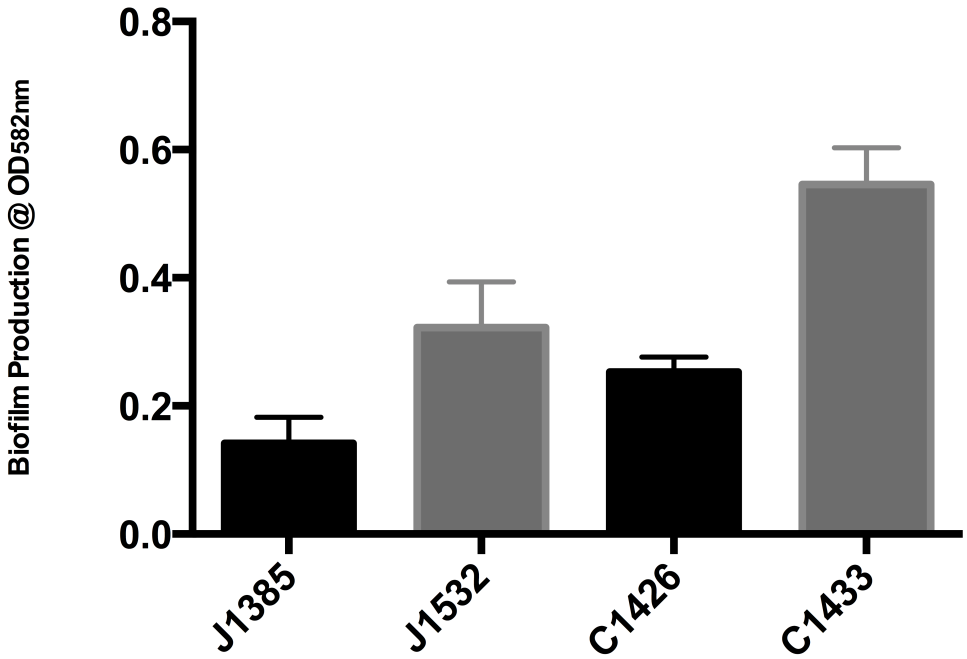


Figure 4.1 Conversion to mucoidy as a result of host adaptation will confer increased proficiency at biofilm formation. Mucoid strains from both pairs produce around twice as much biofilm as measured using a 96-well assay.

Strain	Mucoid/Non-mucoid	Production of:				
		Esterase (nM/ml/min)	Lipase (nM/ml/min)	Pyocyanin	Protease	Siderophore
C1426	Non-mucoid	378 (±36)	250 (±12)	2.338 (±0.010)	+++	+++
C1433	Mucoid	622(±57)	374 (±23)	2.901 (±0.030	++	++
PA01	Non-mucoid	55 (±25)	477 (±1)	1.000 (±0.009)	++	++

Table 4.1 Comparison of production of a variety of enzymes/secondary metabolites between C1426 and its near isogenic partner C1433. Lipase and esterase enzyme production is expressed as nM/ml/min (±SE of mean). Pyocyanin is expressed in arbitrary units normalised to the amount produced by PA01. The numbers in parenthesis are the standard errors of the mean. Protease production was determined by measuring the zones of clearance on an appropriate agar (+++ - >18mm, ++ - 12-18mm & + - <12mm). Siderophore production was determined measuring the zones of colour change (blue to orange) on a CAS containing agar (+++ - >15mm, ++ - 10-15mm & + - <10mm).

4.3 C1426 and C1433 contain a number of genomic regions not conserved throughout the species

Whilst this investigation focuses on the vertical transfer of genetic information between C1426 and its mutant C1433 it is important to examine these strains for any evidence of horizontal gene transfer that may contribute to their ability to successfully infect the CF lungs. Horizontal gene transfer is a common occurrence across all bacteria and genes acquired in this way will normally confer a selective advantage to the recipient strain under certain conditions (Koonin, Makarova and Aravind, 2001). However, it is important to note that the presence of potentially foreign DNA within a genome does not necessarily mean that it confers any advantage or is involved in cellular process in any way. Bacteria will acquire some “selfish” coding regions that are there purely for their own benefit such as bacteriophage and transposons (Lawrence, 1999). As a result of DNA acquisition, one would expect that bacterial genomes are steadily growing in size. However, this is not the case and DNA acquisition is often counterbalanced by the loss of genomic information. The DNA lost in these circumstances normally encode for functions that provide less of a selective advantage (Lawrence, 1999). These “loss-of-function” mutations will in fact confer a selective advantage to the bacterium by diverting resources to more important functions.

We have attempted to identify both acquisition and loss of genomic information in C1426 and C1433. We have utilised CGView Comparison Tool (CCT) (Grant, Arantes and Stothard, 2012) to visualise both strains in relation to the 53 other fully sequenced *P. aeruginosa* strains detailed in the previous chapter. This CCT map has been used to identify any regions unique to C1426/C1433 or those regions not conserved throughout the species (Figure 4.2). Large areas of horizontal gene transfer and areas of hyper-variability can easily be visualised in this image as areas with no or low homology, respectively. Based on our subsequent analysis, we have elected to name those regions unique to either C1426 or C1433.

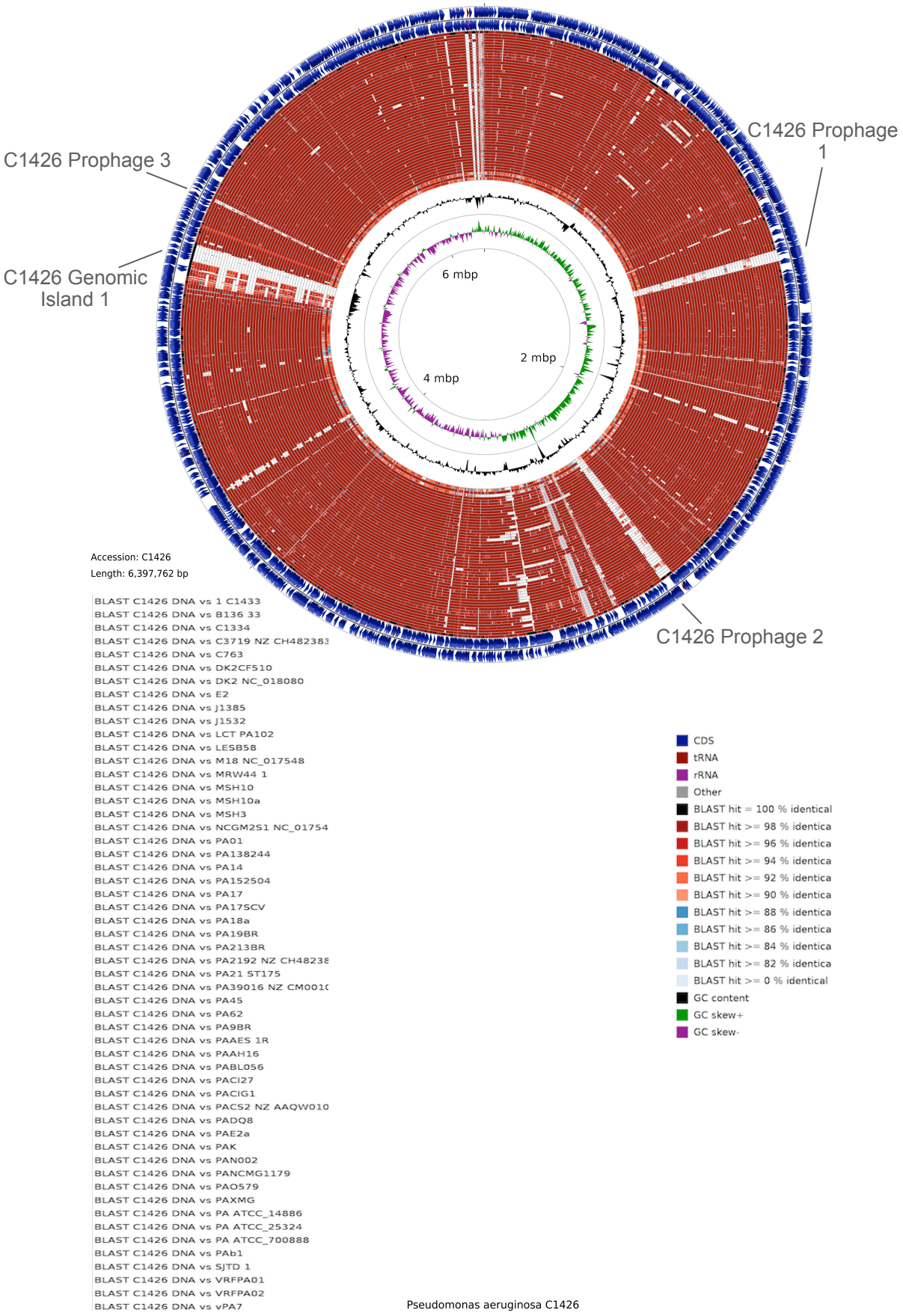


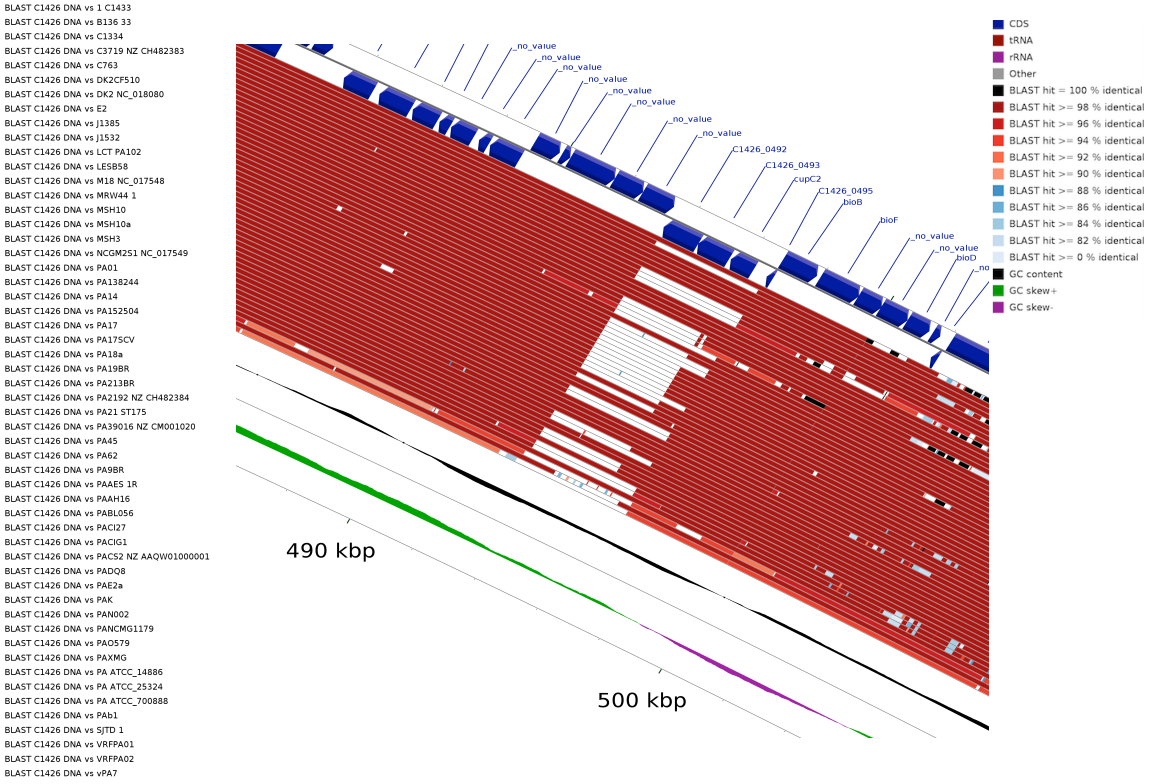
Figure 4.2 CCT map showing C1426 versus C1433 and 53 other fully sequenced or draft *P. aeruginosa* genomes as listed in the key below. BLAST hits in comparison to C1426 are colour coded by percentage match as detailed. Large areas not present in other genomes bar the reference C1426 can be seen as white columns. Accession numbers for strains included in this figure can be found in Table 6.3.

4.4 C1426 and C1433 contain an extra set of fimbrial genes

The first large genomic difference observed between C1426/C1433 and the 53 other *P. aeruginosa* strains included in the analysis is a 3 kbp region located at 492 kbp (Figure 4.3). The orientation of the genes within this region compared to surrounding genes is indicative that these genes have been acquired via horizontal gene transfer (Koonin, Makarova and Aravind, 2001). As can be seen by the BLAST results in this figure, the genomic information contained within this region is only present in around 40% of the analysed strains of which, over 50% of those strains have been isolated from the lungs of CF patients and overall, ~75% are clinical isolates. This may be entirely coincidental however, it may also suggest that the genes acquired from this gene transfer event have been selected to give the bacterium a selective advantage in their niche. Interestingly, from our phylogenetic analysis in Figure 3.4, each of the strains possessing this region fall into *P. aeruginosa* Group 1. The xBase annotation of this genome reveals that this region contains four genes (Figure 4.3). Performing a BLAST analysis on the protein sequences of these genes reveals that three (C1426_0492, C1426_0493 and *cupC2*) are involved in fimbriae formation. More specifically, they are putative *cup* family proteins involved in the synthesis of fimbriae tasked with bacterial attachment during biofilm formation (Vallet *et al*, 2004). Orthologs of these genes have previously been identified in PA01 by Vallet *et al* (2004), sharing over 99% homology with the genes present in C1426 and C1433. In order to establish a chronic infection within the CF lung, *P. aeruginosa* must be competent at biofilm formation (Hoiby *et al,* 2001). This extra set of fimbrial genes acquired by both C1426 and C1433 may aid in biofilm formation within the CF lung. This is supported by Figure 4.1 that demonstrates that both C1426 and C1433 are more proficient at biofilm formation than a second set of non-mucoid/mucoid strains, J1385 and J1532. These strains do not posses this extra set of fimbrial genes and therefore their efficiency at biofilm formation may be compromised in relation to C1426 and C1433.

The fourth gene in this region, C1425_0495 is extremely short and due to the fact that it has no described orthologs, the most likely explanation is that xBase has incorrectly called this segment of DNA as a gene instead of an intergenic region.

A



B

Gene id	Putative function	Homology (Identity %)	Comments
C1426_0492	Fimbral family protein	*P. aeruginosa* c7447m (99%), *P. aeruginosa* PA01 (100%), *P. aeruginosa* LESB58 (100%)	Possible *cup* gene
C1426_0493	Fimbral family protein	*P. aeruginosa* c7447m (100%), *P. aeruginosa* LESB58 (100%), *P. aeruginosa* E2 (100%)	Possible *cup* gene
cupC2	Pili formation chaperone	*P. aeruginosa* LESB58 (100%), *P. aeruginosa* E2 (100%), *P. aeruginosa* M18 (100%)	
C1426_0495	Hypothetical protein	*P. aeruginosa* 18A (98%), *P. aeruginosa* 39016 (88%), *P. aeruginosa* NCGM2.S1 (88%)	May be a miscalled gene.

Figure 4.3 **A** C1426, C1433 and 40% of the other *P. aeruginosa* strains included in this analysis possess an extra set of fimbrial genes as predicted by xBase annotations. **B** Table showing the gene id of these genes, their putative functions and their homology to other bacterial genomes.

4.5 C1426 and C1433 contain a R-type pyocin related to Phage P2

A second large genomic difference in C1426 and C1433 in relation to the 53 *P. aeruginosa* currently available genomes is a 13 kbp region containing 15 predicted genes (Figure 4.4). BLAST analysis reveals that these genes have a strong homology to bacteriophage P2 (Table 4.2). However, compared to the previously described genetic map of P2 (Christie *et al*, 2002) it can be seen that the genes present in C1426 and C1433 lie in a very different order with no signs of gene synteny. Therefore, it is unlikely that this set of genes will result in the formation of a productive temperate phage.

Instead, it has previously been shown that Phage P2 shares a common ancestral origin with the *P. aeruginosa* R-type pyocin of PA01 (Nakayama *et al*, 2000). These pyocins are rod-like particles that resemble a bacteriophage protein. They posses bactericidal activity and can kill other *P. aeruginosa* strains as well as a variety of other bacterial species (Michel-Briand and Baysse, 2002). The genetic map of the cluster of genes in Figure 4.4 show a high degree of synteny with those described for *P. aeruginosa* strain PA01 by Nakayama *et al* (2000) and it is therefore expected that these genes encode a R-type pyocin (R2). There are several features present in both genetic maps that corroborate this hypothesis: 1) Both syntenic regions are 13 kbp in length, 2) Genes are located between the amino acid transport and metabolism genes *trpE* and *trpGCD,* and 3) Both clusters are preceded by a transcriptional regulator of pyocins *prtN* and its regulator *prtR*. This cluster has also been described as one of the six prophage clusters (LES-prophage 1) present in LESB58 that are responsible for its *in vivo* competitiveness (Winstanley *et al*, 2009). This gene cluster in LESB58, like that in C1426, was also predicted to be a defective prophage encoding the pyocin R2. Klockgether *et al* (2011) have identified that the region that this defective prophage has occupied is a known region of genome plasticity. In PA01, this location plays host to two individual gene clusters encoding pyocins R2 and

F2. It has been noted that C1426, like LESB58, has not acquired the second of these pyocins, F2.

In total, 73% of the strains included in this study posses a copy of these genes. There is no pattern as to the niche from which these strains were isolated so it is impossible to theorise where these genes were acquired by the strains. However, due to the bactericidal effects of pyocins, it is expected that acquisition of these genes will have given each strain a selective advantage, providing an increased ability to establish strain dominance in their respective niches.

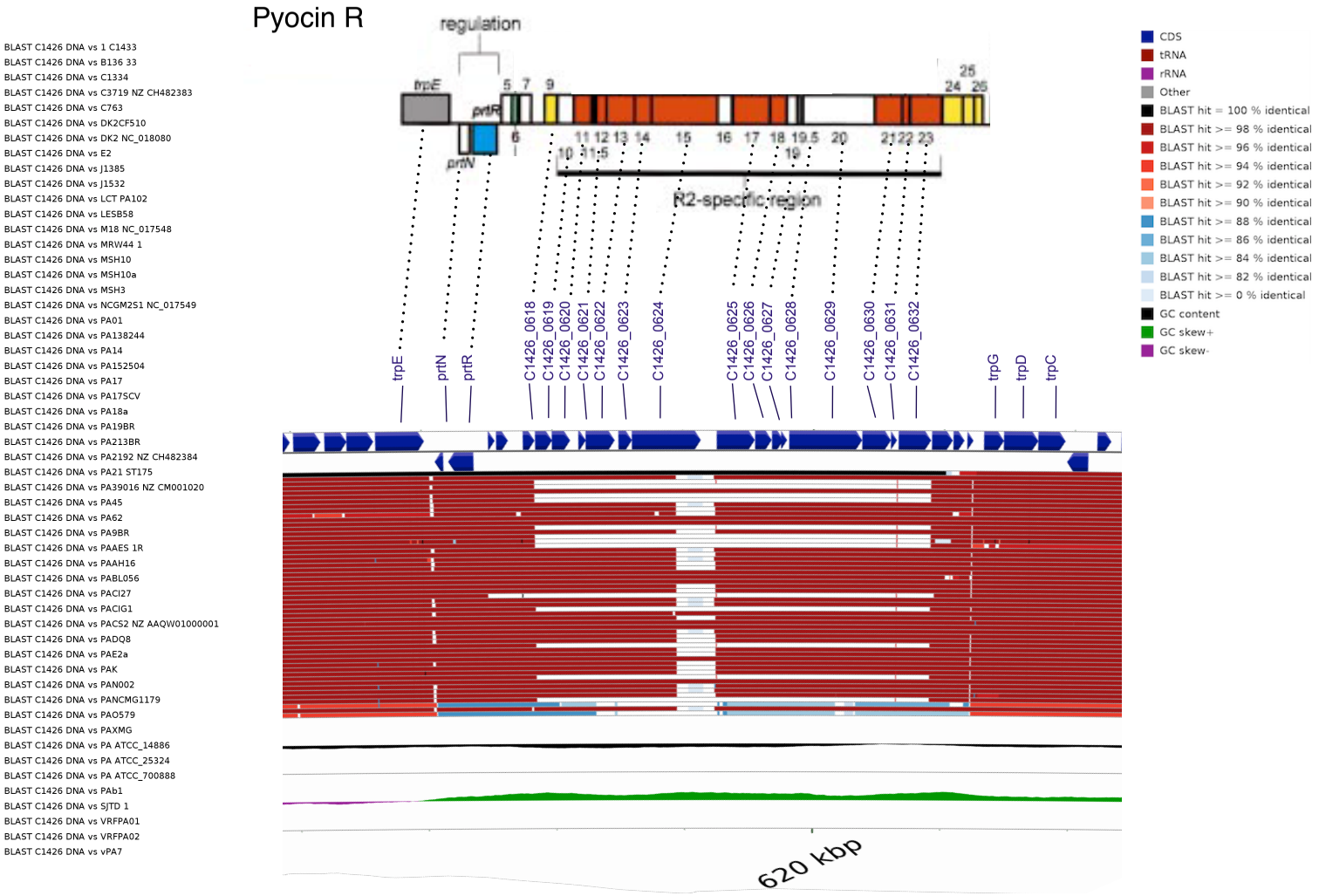


Figure 4.4 Zoomed CCT map showing the presence of a 13 kbp regions containing R-type pyocin genes present in C1426, C1433 and 73% of included strains. Genes present in this region are compared to the gene in the R-type pyocin cluster of PA01 (adapted from Nakayama *et al*, 2000).

Gene ID	Putative Function	Homology (Identity %)	Comments
C1426_0618	Holin	*P. aeruginosa* NCGM1179 (100%), *P. aeruginosa* 18A (100%), *P. aeruginosa* PAK (100%)	R-type pyocin protein 9 - lytic protein
C1426_0619	Hypothetical protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* 18A (100%), *P. aeruginosa* PAK (100%)	R-type pyocin protein 10
C1426_0620	Phage baseplate assembly protein V	*P. aeruginosa* LESB58 (100%), *P. aeruginosa* PAK (100%), *P. aeruginosa* 18A(100%)	R-type pyocin protein 11
C1426_0621	Bacteriophage protein W	*P. aeruginosa* LESB58 (100%), *P. aeruginosa* PA14 (100%), *P. aeruginosa* NCGM2.S1 (100%)	R-type pyocin protein 12
C1426_0622	Phage baseplate assembly protein J	*P. aeruginosa* LESB58 (99%), *P. aeruginosa* 18A (99%), *P. aeruginosa* c7447m (99%)	R-type pyocin protein 13
C1426_0623	Phage tail protein I	*P. aeruginosa* LESB58 (100%), *P. aeruginosa* 18A (100%), *P. aeruginosa* c7447m (100%)	R-type pyocin protein 14
C1426_0624	Phage tail fibre protein H	*P. aeruginosa* PAK (100%), *P. aeruginosa* 18A (100%), *P. aeruginosa* (99%)	R-type pyocin protein 15
C1426_0625	Phage tail sheath protein FI	*P. aeruginosa* NCGM2.S1 (100%), *P. aeruginosa* 39016 (100%), *P. aeruginosa* LESB58 (99%)	R-type pyocin protein 17
C1426_0626	Phage tail tube protein FII	*P. aeruginosa* PA14 (100%), *P. aeruginosa* LESB58 (100%), *P. aeruginosa* NCGM2.S1 (100%)	R-type pyocin protein 18
C1426_0627	Hypothetical protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* LESB58 (100%), *P. aeruginosa* NCGM2.S1 (100%)	R-type pyocin protein 19
C1426_0628	Hypothetical protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* VRFPA02 (100%), *P. aeruginosa* VRFPA03 (100%)	R-type pyocin protein 19.5
C1426_0629	Phage tail length tape measure protein	*P. aeruginosa* 138244 (99%), *P. aeruginosa* PA21_ST175 (99%), *P. aeruginosa* NCMG1179 (99%)	R-type pyocin protein 20
C1426_0630	Phage tail formation protein U	*P. aeruginosa* LESB58 (100%), *P. aeruginosa* PAK (100%), *P. aeruginosa* NCGM2.S1 (100%)	R-type pyocin protein 21
C1426_0631	Phage tail protein X	*P. aeruginosa* LESB58 (100%), *P. aeruginosa* PA14 (100%), *P. aeruginosa* PA01 (100%)	R-type pyocin protein 22
C1426_0632	Phage late control protein D	*P. aeruginosa* LESB58 (99%) *P. aeruginosa* c7447m (99%) *P. aeruginosa* PA01 (99%)	R-type pyocin protein 23

Table 4.2 Table showing the gene id, putative function and percentage homology as predicted by BLAST analysis of genes present within the 13 kbp region possessed by C1426, C1433 and 73% of other *P. aeruginosa* strains included in this study.

4.6 C1426 and C1433 possess a key two-component regulatory system involved in biofilm maturation

Both C1426 and C1433 contain a ~4.5 kbp segment of genome that is present in only 65% of *P. aeruginosa* genomes included in this analysis (Figure 4.4). The majority of these genes encode hypothetical proteins with no known functions (Table 4.3). However, two genes stand out, C1426_0839 and the putatively annotated *gltR*. It is important to note that the gene labelled *gltR* by xBase has been incorrectly annotated. Both strains do possess a true *gltR* later in their genomes that match the *gltR* genes present in both PA01 and LESB58. This highlights the need to check annotations by this automated service and demonstrates that annotations are not always correct.

Instead, BLAST analysis has predicted *gltR* and C1426_0839 to be a two-component system annotated as *bfmSR*, respectively (PA4101 and PA4102 in PA01). The proteins encoded by these genes encode a signal transduction histidine kinase BfmS and its regulator BfmR. No other copies of this two-component system have been identified in either strain. These proteins are essential in the maturation stage of biofilm formation hence their nomenclature, **b**io**f**ilm **m**aturation **R**egulator (Petrova and Sauer, 2009). In fact, when *bfmR* is knocked out, biofilm dispersal readily occurs.

This may be yet another example of positive selection by C1426 and C1433 as possession of this two-component regulatory system may give these strains a competitive advantage in establishing dominance and chronic infection within the CF lung via increased biofilm formation (Hoiby *et* al*,* 2001). This is supported by Figure 4.1 that in comparison to the second pair of non-mucoid/mucoid strains J1385 and J1532 which do not possess this regulatory system, C1426 and C1433 are clearly more proficient at biofilm formation.

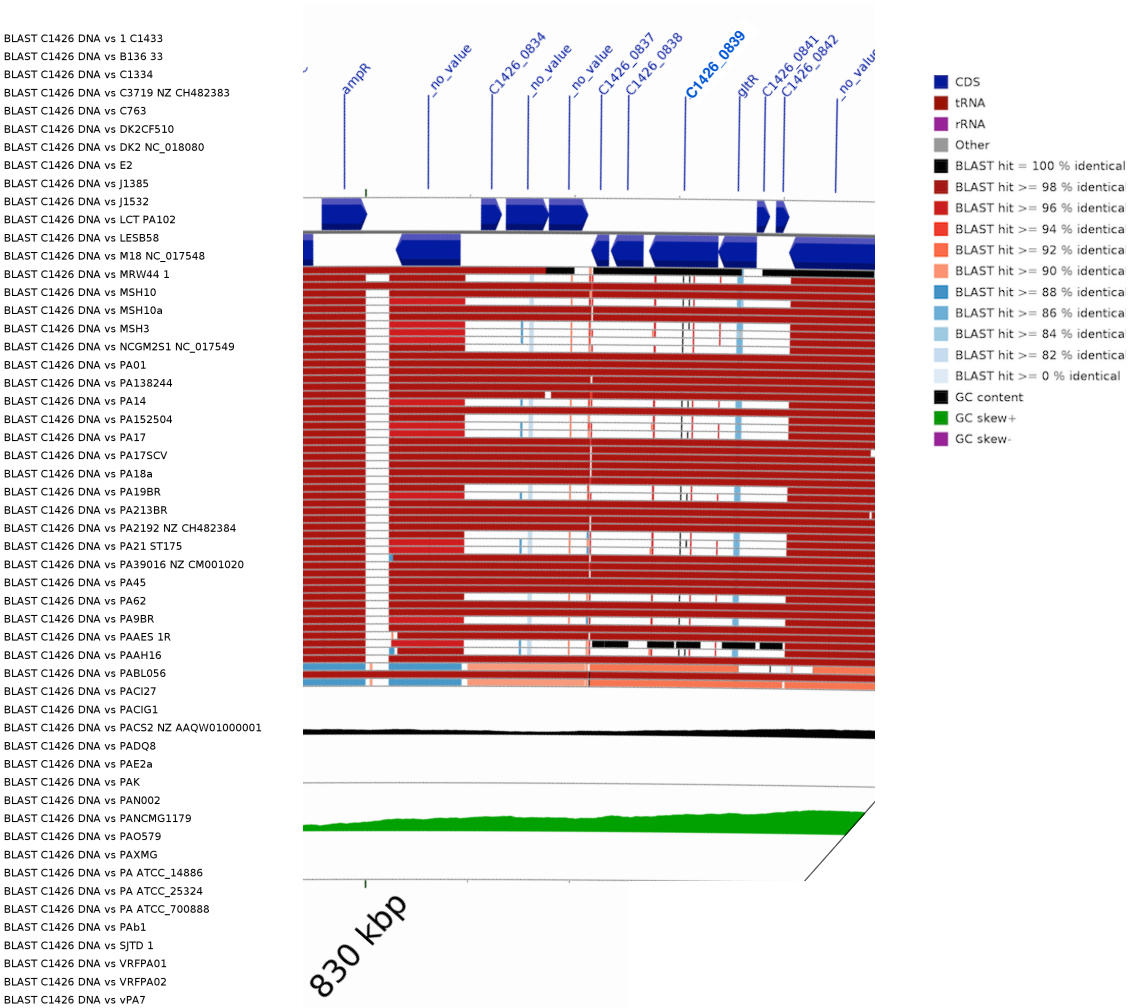


Figure 4.4 Both C1426 and C1433 possess the genes *gltR* and C1426_0839 that BLAST analysis reveals are homologous to the two component regulatory system *bfmSR* that is essential for successful biofilm formation (Petrova and Sauer, 2009).

Gene ID	Putative Function	Homology (Identity %)	Comments
C1426_0834	Hypothetical protein	*P. aeruginosa* LESB58 (100%), *P. aeruginosa* M18 (100%), *P. aeruginosa* NCM1179 (100%)	
C1426_0835	Hypothetical protein	*P. aeruginosa* DK2 (100%), *P. aeruginosa* M18 (100%), *P. aeruginosa* PAK (100%)	
C1426_0836	Hypothetical protein	*P. aeruginosa* LESB58 (100%), *P. aeruginosa* M18 (100%), *P. aeruginosa* NCM1179 (100%)	
C1426_0837	Hypothetical protein	*P. aeruginosa* LESB58 (98%), *P. aeruginosa* M18 (98%), *P. aeruginosa* PA01 (98%)	
C1426_0838	Membrane protein	*P. aeruginosa* LESB58 (100%), *P. aeruginosa* M18 (100%), *P. aeruginosa* PA01 (100%)	
C1426_0839	Two-component sensor	*P. aeruginosa* 18A (99%), *P. aeruginosa* M18 (99%), *P. aeruginosa* PA01 (99%)	Protein BfmS. Signal transduction histidine kinase
gltR	Two-component response regulator	*P. aeruginosa* PA01 (100%), *P. aeruginosa* M18 (100%), *P. aeruginosa* NCM1179 (100%)	Protein BfmR. Signal transduction regulator
C1426_0841	Hypothetical protein	*P. aeruginosa* M18 (99%)	
C1426_0842	Hypothetical protein	*P. aeruginosa* LESB58 (100%), *P. aeruginosa* M18 (100%), *P. aeruginosa* NCM1179 (100%)	

Table 4.3 Table showing the gene id, putative function and percentage homology as predicted by BLAST analysis of genes present within the ~4.5 kbp region possessed by C1426, C1433 and 60% of other *P. aeruginosa* strains included in this study.

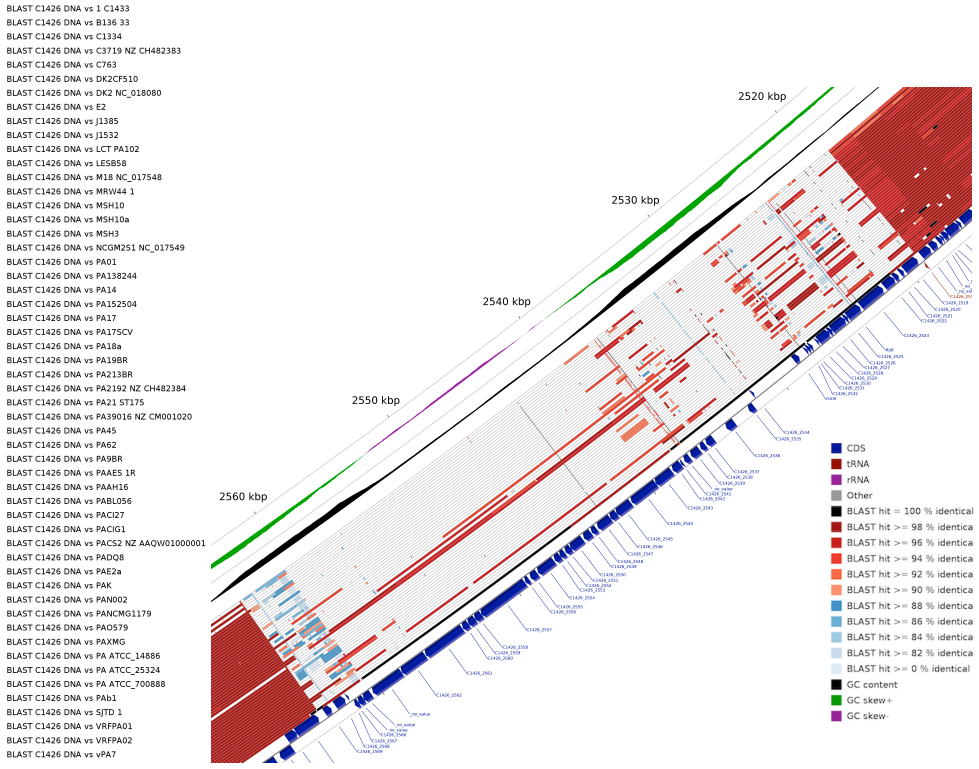
4.7 C1426 and C1433 possess a unique LESB58 Prophage 5-like bacteriophage

A large 44 kbp prophage is present in C1426 and its mucoid partner C1433 but is not present in any other strain included in this analysis in its entirety (Figure 4.5). Table 4.4 demonstrates that the genes present in this region share similar homology with a variety of *Pseudomonas* phages including H66, YMC/01/01P52_PAE_BP, D3 and φ297 leading to the hypothesis that this region of the C1426 genome was acquired via phage integration prior to adapting into C1433. This hypothesis is supported by the fact that genes within this region lie in the opposing direction to flanking genes.

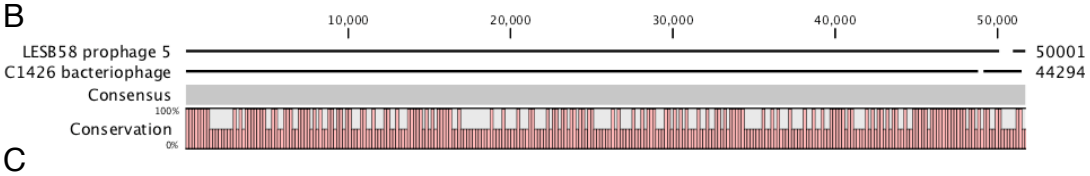
Initial observations suggests that this phage was in fact the bacteriophage LES-prophage 5 from LESB58 (Winstanley *et al*, 2009). Both phage share a 99% identical site specific integrase (C1426_2519 and PALES_25021) (Table 4.4). However, this is where the similarities end. Prophage 5 is 6 kbp longer and as can be seen in both the zoomed CTT and the nucleotide alignment in Figure 7.5, the homology between these phages is extremely low. A separate BLASTn analysis also shows that this phage has only 15% query coverage compared to LESB58 showing that the synteny is not conserved. Therefore, the phage observed in C1426 and C1433 cannot be LES-prophage 5, despite sharing some similar properties. We propose that this phage identified in both C1426 and C1433 is a unique to these strains and has not been observed in the literature to date. Consequently, this phage has been named C1426 Prophage 2.

Annotation of this new phage has failed to assign functions to the genes present and the majority have a putative function of “hypothetical protein”. There is no evidence to suggest that the acquisition of this phage will confer any selective advantage to C1426 and C1433.

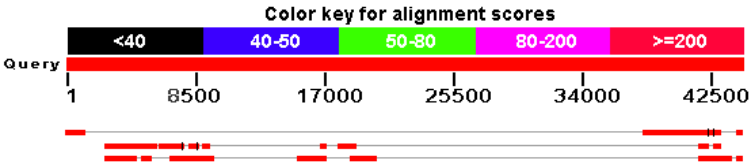
A



B



C



Description	Max score	Total score	Query cover	E value	Ident	Accession
Pseudomonas aeruginosa LESB58 complete genome sequence	6207	10753	15%	0.0	93%	FM209186.1
Pseudomonas phage YMC/01/01/P52_PAE_BP, complete genome	5576	13903	21%	0.0	96%	JX403939.1
Pseudomonas aeruginosa PA7, complete genome	4050	15218	27%	0.0	92%	CP000744.1

Figure 4.5 **A** The 44 Kbp C1426 Prophage 2 is present in C1426 and C1433 but absent from the genomes of all other *P. aeruginosa* strains included in this analysis. **B** Nucleotide alignment of the bacteriophage in C1426 versus prophage 5 from LESB58 demonstrating that they do not share a high degree of homology **C** BLAST analysis of the bacteriophage in C1426 demonstrates that only 15% of the C1426 bacteriophage matches prophage 5.

Gene ID	Putative Function	Homology (Identity %)	Comments
C1426_2519	Site specific integrase	*P. aeruginosa* PA45 (99%), *P. aeruginosa* DK2 (99%), *P. aeruginosa* LESB58 (99%)	Phage integrase family protein
C1426_2520	Hypothetical protein	*P. aeruginosa* LESB58 (100%), *P. aeruginosa* DK2 (100%), *P. aeruginosa* PA45 (100%)	
C1426_2521	NA	No significant matches	Miscalled gene?
C1426_2522	NA	No significant matches	Miscalled gene?
C1426_2523	Hypothetical protein	*P. aeruginosa* ATCC 700888 (95%), *P. aeruginosa* MSH-10 (96%), *Pseudomonas* phage YMC/01/01/P52_PAE_BP (95%)	Phage related
ftsK	DNA segregation ATPase	*Pseudomonas* phage YMC/01/01/P52_PAE_BP (100%)	Cell division protein
C1426_2525	Hypothetical protein	*Pseudomonas* phage YMC/01/01/P52_PAE_BP (100%), *P. aeruginosa* PA21_ST175 (98%), *P. aeruginosa* 138244 (98%)	
C1426_2526	Hypothetical protein	*P. aeruginosa* 2192 (99%), *P. aeruginosa* MSH-10 (99%)	
C1426_2527	Hypothetical protein	*P. aeruginosa* 2192 (100%), *P. aeruginosa* MSH-10 (100%), *Pseudomonas* phage YMC/01/01/P52_PAE_BP (100%)	
C1426_2528	Hypothetical protein	*P. aeruginosa* MSH-10 (99%), *P. aeruginosa* PACL (98%), *P. aeruginosa* DK2 (97%)	
C1426_2529	Hypothetical protein	*P. aeruginosa* PA7 (100%), *P. aeruginosa* MSH-10 (100%), *Pseudomonas* phage YMC/01/01/P52_PAE_BP (100%)	
C1426_2530	Hypothetical protein	*P. aeruginosa* M18 (100%), *Pseudomonas* phage D3 (100%), *P. aeruginosa* VRFPA03 (98%),	
C1426_2531	Hypothetical protein	*P. aeruginosa* MSH-10 (100%), *P. aeruginosa* RP73 (100%), *Pseudomonas* sp. P179 (100%)	
C1426_2532	Hypothetical protein	*Pseudomonas* sp. P179 (88%), *P. aeruginosa* 2192 (95%), *Pseudomonas* phage H66 (73%)	
rsmA	Carbon storage regulator	*P. aeruginosa* PA21_ST175 (99%), *P. aeruginosa* 138244 (99%), *Pseudomonas* phage YMC/01/01/P52_PAE_BP (99%)	Incorrect annotation
C1426_2534	Hypothetical protein	*P. aeruginosa* DK2 (100%)	
C1426_2535	NA	No significant matches	
C1426_2536	Hypothetical protein	*Pseudomonas* sp. 2_1_26 (100%), *P. aeruginosa* 2192 (99%), *P. aeruginosa* DK2 (99%)	
C1426_2537	Hypothetical protein	*P. aeruginosa* DK2 (98%), *P. aeruginosa* 2192 (98%), *Pseudomonas* sp. 2_1_26 (98%)	
C1426_2538	Endodeoxyribonuclease rusA	*P. aeruginosa* PA21_ST175 (100%), *P. aeruginosa* 138244 (96%), *Pseudomonas* phage YMC/01/01/P52_PAE_BP (100%)	Holliday junction resolvase
C1426_2539	Hypothetical protein	*P. aeruginosa* ATCC 700888 (100%), *P. aeruginosa* DK2 (100%), *Pseudomonas* phage phi297 (100%)	
C1426_2540	Hypothetical protein	*P. aeruginosa* 2192 (100%), *P. aeruginosa* VRFPA03 (100%), *Pseudomonas* phage phi297 (100%)	
C1426_2541	Hypothetical protein	*P. aeruginosa* DK2 (96%), *Pseudomonas* phage YMC/01/01/P52_PAE_BP (96%), *Pseudomonas* phage phi297 (96%)	
C1426_2542	Terminase	*P. aeruginosa* VRFPA03 (99%), *P. aeruginosa* VRFPA04 (99%), *Pseudomonas* sp. 2_1_26 (99%)	
C1426_2543	Terminase	*P. aeruginosa* DK2 (98%), *P. aeruginosa* 2192 (98%), *Pseudomonas* sp. 2_1_26 (97%)	
C1426_2544	Phage related protein	*P. aeruginosa* DK2 (97%), *P. aeruginosa* 2192 (96%), *Pseudomonas* sp. 2_1_26 (96%)	HI1049 family
C1426_2545	Hypothetical protein	*P. aeruginosa* 2192 (99%), *Pseudomonas* sp. 2_1_26 (99%)	
C1426_2546	Hypothetical protein	*P. aeruginosa* DK2 (99%), *Pseudomonas* sp. 2_1_26 (96%)	

C1426_2547	Hypothetical protein	*P. aeruginosa* DK2 (99%), *P. aeruginosa* 2192 (99%), *Pseudomonas* sp. 2_1_26 (99%)	
C1426_2548	Hypothetical protein	*P. aeruginosa* 2192 (91%), *Pseudomonas* sp. 2_1_26 (92%), *P. aeruginosa* DK2 (90%)	
C1426_2549	Hypothetical protein	*P. aeruginosa* DK2 (95%), *P. aeruginosa* 2192 (95%), *Pseudomonas* sp. 2_1_26 (94%)	
C1426_2550	Hypothetical protein	*P. aeruginosa* DK2 (97%), *P. aeruginosa* 2192 (97%), *Pseudomonas* sp. 2_1_26 (97%)	
C1426_2551	Hypothetical protein	*P. aeruginosa* 2192 (100%), *P. aeruginosa* DK2 (99%), *Pseudomonas* sp. 2_1_26 (98%)	
C1426_2552	Hypothetical protein	*Pseudomonas* sp. 2_1_26 (96%), *P. aeruginosa* 2192 (87%), *P. aeruginosa* DK2 (87%)	
C1426_2553	Hypothetical protein	*Pseudomonas* sp. 2_1_26 (99%), *P. aeruginosa* DK2 (99%), *P. aeruginosa* 2192 (98%)	
C1426_2554	Hypothetical protein	*Pseudomonas* sp. 2_1_26 (99%), *P. aeruginosa* 2192 (99%), *P. putida* F1 (76%)	
C1426_2555	Hypothetical protein	*Pseudomonas* sp. 2_1_26 (99%), *P. aeruginosa* DK2 (99%), *P. aeruginosa* 2192 (98%)	
C1426_2556	Hypothetical protein	*P. aeruginosa* DK2 (95%), *P. aeruginosa* 2192 (95%)	
C1426_2557	Putative tail length tape measure protein	*P. aeruginosa* DK2 (93%), *Pseudomonas* sp. 2_1_26 (90%), *P. aeruginosa* 2192 (90%)	Lambda phage minor tail
C1426_2558	Putative alkanesulfonate ABC transporter membrane protein	*P. aeruginosa* DK2 (99%)	
C1426_2559	Hypothetical protein	*P. aeruginosa* DK2 (100%)	
C1426_2560	Hypothetical protein	*P. aeruginosa* DK2 (99%), *Pseudomonas* phage PAJU2 (61%)	
C1426_2561	Hypothetical protein	*P. aeruginosa* DK2 (98%)	
C1426_2562	Hypothetical protein	*P. aeruginosa* LESB58 (87%)	
C1426_2563	Hypothetical protein	*P. aeruginosa* LESB58 (98%), *P. aeruginosa* M18 (66%), *P. aeruginosa* PA45 (61%)	
C1426_2564	Lytic enzyme	*P. aeruginosa* NCGM2.S1 (93%), *P. aeruginosa* PA21_ST175 (97%), *P. aeruginosa* 138244 (97%)	Predicted chitinase class 1
C1426_2565	Hypothetical protein	*P. aeruginosa* MSH-10 (89%), *P. aeruginosa* NCMG1179 (88%), *P. aeruginosa* 39016 (84%)	
C1426_2566	Hypothetical protein	*Pseudomonas* phage PAJU2 (92%, *P. aeruginosa* 138244 (89%), *P. aeruginosa* VRFPA02 (87%)	
C1426_2567	Hypothetical protein	*P. aeruginosa* LESB58 (78%)	
C1426_2568	Putative phage protein	*P. aeruginosa* DK2 (99%), *Pseudomonas* phage F116 (92%), *P. aeruginosa* PA7 (90%)	
C1426_2569	Hypothetical protein	*P. aeruginosa* PA45 (95%), *P. aeruginosa* DK2 (94%), *P. aeruginosa* NCGM2.S1 (91%)	

Table 4.4 Table showing the gene id, putative function and percentage homology as predicted by BLAST analysis of genes present within the 44 kbp C1426 Prophage 2 possessed by only C1426 and C1433

4.8 C1426 and C1433 possess a second unique bacteriophage that resembles *Pseudomonas* phage Pf1

An 11.5 kbp segment of DNA has been visualised in C1426 and C1433 but, is not found in any other *P. aeruginosa* strains included in this study in its entirety. The zoomed CCT map in Figure 4.6 shows that flanking genes run in opposing directions to those within this region. This, along with a reduced GC content, suggests that this 11.5 kbp region is yet another phage acquired by C1426. Performing a BLAST analysis of the genes present in this putative phage show that the majority of genes are phage related (Table 4.4). This includes a phage integrase (C1426_5167), two coat proteins (C1426_5172 and *coaB*) and a ssDNA binding protein (C1426_5176), possibly responsible for host helix-destabilising, enabling phage DNA integration. This ssDNA binding protein, if correctly annotated, may be responsible for acquisition of this phage-like region.

Many of these genes show a high degree of homology to *Pseudomonas* filamentous phage Pf1 (Hill *et al*, 1991) genes suggesting that this new phage is in fact Pf1 (Table 4.4). However, a second BLASTn analysis shows that this is not the case and instead, only 47% of the nucleotide sequence of the C1426 phage showed any degree of homology to the complete nucleotide sequence of Pf1 (Figure 4.6). Pf1 is also ~800 bp smaller. Therefore, this new phage, unique to C1426 and C1433, is not the filamentous *Pesudomonas* phage Pf1 but is instead a Pf1-like phage and is hereby referred to a C1426 Prophage 3. This region has previously been identified as a region of genome plasticity (RGP10) by Klockgether *et al* (2011) in PA14, PA7 and LESB58 (LES-prophage 6).

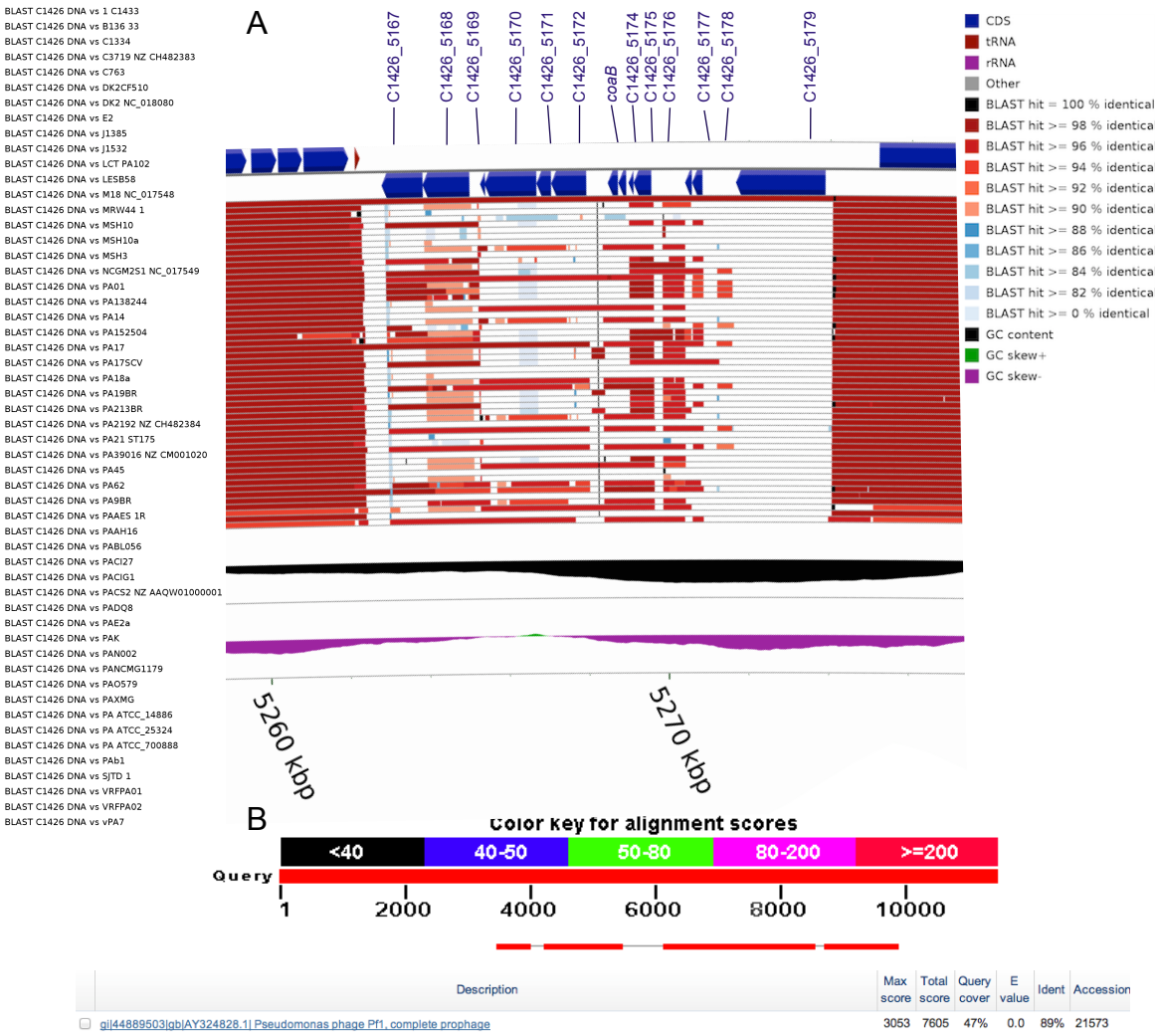


Figure 4.6 **A** The 11.5 Kbp C1426 Prophage 3 is unique to C1426 and C1433. **B** Only 47% of the C1426 Prophage 3 shows any degree of homology to the complete nucleotide sequence of the filamentous *Pseudomonas* phage Pf1 as demonstrated by BLAST analysis

Gene ID	Putative Function	Homology (Identity %)	Comments
C1426_5167	Phage integrase	*P. aeruginosa* 18A (99%), *P. aeruginosa* ATCC 700888 (99%), *P. aeruginosa* PAK (97%)	
C1426_5168	Hypothetical protein	*P. aeruginosa* M18 (100%), *P. aeruginosa* 2192 (100%), *P. aeruginosa* 18A (99%)	From bacteriophage Pf1
C1426_5169	Hypothetical protein	*P. aeruginosa* VRFPA01 (100%), *P. aeruginosa* PA45 (100%), *Pseudomonas* sp. P179 (100%)	
C1426_5170	Hypothetical protein	*P. aeruginosa* PA01 (100%), *P. aeruginosa* PAK (98%), *Psuedomonas* phage Pf1 (98%)	From bacteriophage Pf1 - Zonula occludens toxin
C1426_5171	Hypothetical protein	*P. aeruginosa* PA01 (100%), *P. aeruginosa* PA14 (100%), *Pseudomonas* sp. 2_1_26 (100%)	From bacteriophage Pf1 - head virion protein
C1426_5172	Bacteriophage coat protein	*P. aeruginosa* PA14 (64%), *P. aeruginosa* CIG1 (64%)	Pf1 bacteriophage coat protein A
coaB	Bacteriophage coat protein	*P. aeruginosa* PA01 (98%), *P. aeruginosa* PAK (98%), *Psuedomonas* phage Pf1 (98%)	Pf1 bacteriophage coat protein B
C1426_5174	Hypothetical protein	*P. aeruginosa* PA01 (100%), *P. aeruginosa* PAK (100%), *Psuedomonas* phage Pf1 (100%)	From bacteriophage Pf1
C1426_5175	Hypothetical protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* PA01 (100%), *P. aeruginosa* LESB58 (100%)	From bacteriophage Pf1
C1426_5176	ssDNA binding protein	*P. aeruginosa* PA14 (99%), *P. aeruginosa* LESB58 (99%), *Psuedomonas* phage Pf1 (99%)	Helix destabilising protein of bacteriophage Pf1
C1426_5177	Hypothetical protein	*P. aeruginosa* VRFPA04 (98%)	
C1426_5178	Hypothetical protein	*P. aeruginosa* MSH-10 (100%), *P. aeruginosa* 2192 (100%), *P. aeruginosa* PA7 (99%)	
C1426_5179	Hypothetical protein	*Bacteroides* sp. 3_1_33FAA (45%), *P. stutzer*i CCUG 29243 (52%), *Capnocytophaga cynodegmi* (43%)	

Table 4.4 Table showing the gene id, putative function and percentage homology as predicted by BLAST analysis of genes present within the 11.5 kbp C1426 Prophage 3 possessed by only C1426 and C1433.

4.9 A large, unique Genomic Island has been acquired by C1426 and C1433

We have inspected C1426 for the presence of the key virulence-determining pathogenicity island PAPI-1 identified in PA14, the insertion location of which has previously been identified as a region of genome plasticity (RGP41) (Klockgether *et al*, 2011). PAPI-1 encodes for a variety of proteins that contribute substantially to the virulence of PA14 (He *et al*, 2004 & Carter, Chen and Lory, 2010). In PA01, this island integrates into the tRNA($_{lys}$) genes PA4541.1 and PA4541.2 (Qiu, Gurkar and Lory, 2006). In PA14 this results in the 108 kbp island inserting between genes PA14_58900 and *clpB* (PA14_60190). In this location in C1426, between genes C1426_4878 and *clpB*, there is a large insertion, however, it is much larger at 139 kbp compared to the 108 kbp of PAPI-1 (Figure 4.7). A separate BLASTn analysis of both islands shows that only 51% of the 139 kbp insertion is homlogous with the nucleotide sequence of PAPI-1 (Figure 4.7). Based on this, we have named this large island C1426 Genomic Island 1.

As can be seen in Figure 4.7, there is a region of ~56 kbp that is unique to C1426 and C1433. This region contains an additional 49 genes, of which 29 have putative functions assigned due to homology with other bacterial strains/phage (Table 4.5). Several of these genes may contribute to the phenotype observed in these strains and their overall success as pathogens in the CF lung. For example, the first gene within this region has been annotated by xBase as *fpvA*, encoding for a ferripyoverdin receptor. The protein sequence of this gene has some degree of homology to the FpvA proteins in *Pseudomonas pychrophila* and *Pseudomonas putida*. Ferriopyoverdin receptors such as FpvA, are responsible for iron sequestration by the cell from the surrounding environment (Meyer, Stintzi and Poole, 1999). A second ferripyoverdin receptor may be encoded by gene C1426_4965 (Figure 4.7). Possessing an additional two ferripyoverdin receptors may confer a competitive advantage to C1426 and C1433, allowing an increase in iron uptake from their environments.

Based on xBase annotations and sequence homology to other bacterial species, this region also contains two operons that encode for two distinct, tripartite multidrug efflux pumps of the RND type (Table 4.5). The first of these, encoded for by genes C1426_4979 to C1426_4981, shows a high degree of homology to proteins observed in *Sphingomonas* sp. ERG5. The second multidrug efflux pump, encoded for by genes C1426_4992 to C1426_4994, shows homology to drug-efflux proteins observed in *Xanthomonas* sp.. Previous work by Poole *et al* (1993) has shown that RND mutants were more susceptible to a variety of antibiotics such as tetracycline, chloramphenicol and ciprofloxacin. These types of efflux pumps export a variety of other substances such as acriflavine, ethidium bromide and SDS, as well as other organic solvents (Piddock, 2006). These pumps may confer an increased antibiotic resistance to a variety of drugs and as a result, C1426 and C1433 may have a competitive advantage and persistence during infection of the CF lung.

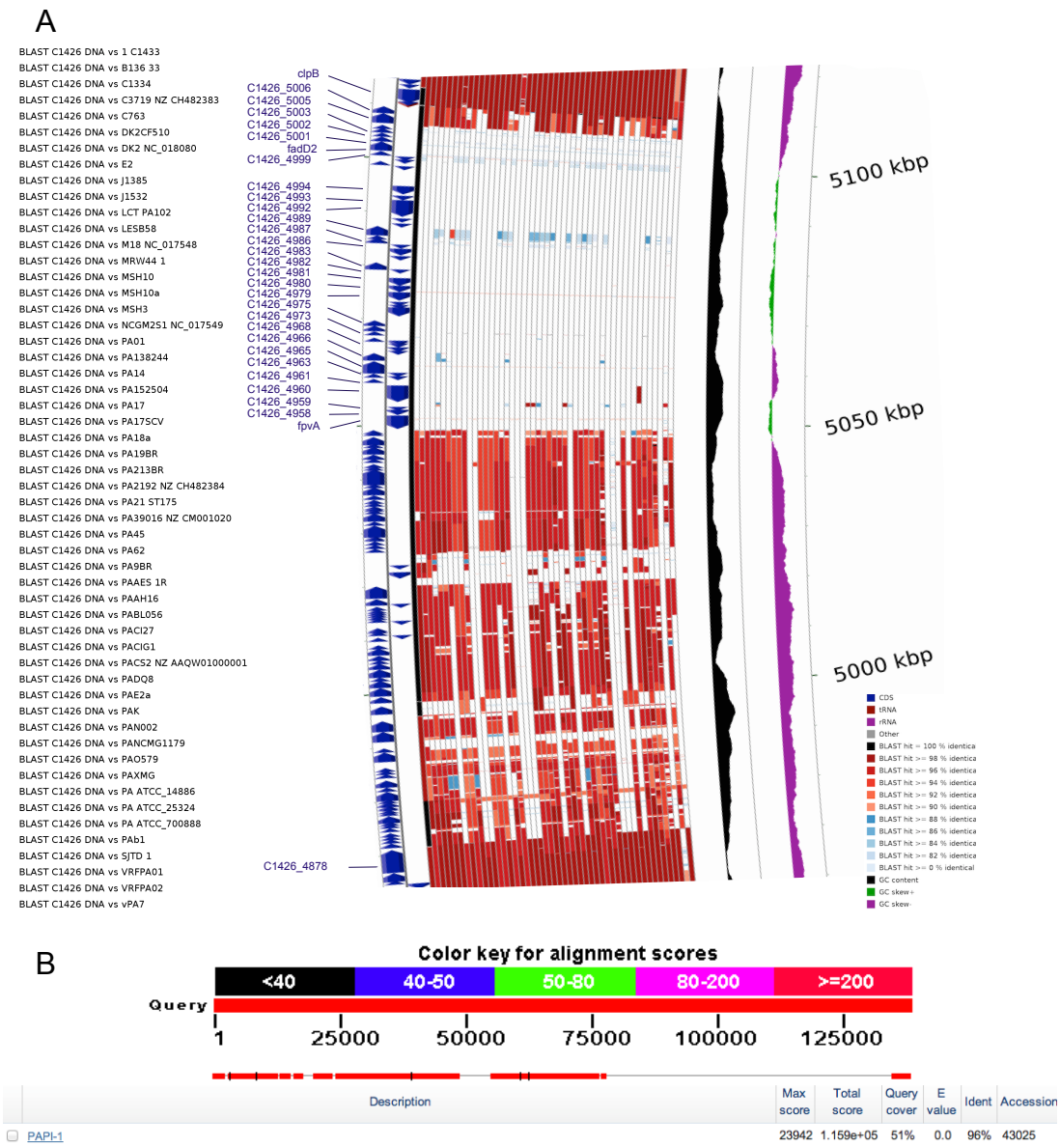


Figure 4.7 **A** C1426 and C1433 have both acquired the large, 139kbp C1426 Genomic Island 1 in place of the 108 Kbp PAPI-1 island present in PA14. This island has a 56 Kbp region unique to these strains. **B** Only 51% of the C1426 island shows any degree of homology to PAPI-1.

Gene ID	Putative function	Homology (Identity %)	Comments
fpvA	ferripyoverdin receptor	*P. psychrophila* (71%), *P. putida* (69%)	
C1426_4958	Membrane receptor (ABC system)	*P. nitroreducens* (89%), *Pseudomonas* sp. M1 (76%)	
C1426_4959	ECF Sigma factor	*Pseudomonas* sp. M1 (100%), *P. nitroreducens* (87%)	Sigma 70
C1426_4960	Type III restriction enzyme	*P. aeruginosa* ATCC 14886 (99 %)	Subunit R
C1426_4961	RNA polymerase sigma factor	*P. putida* (92%), *P. fluorescens* (70%)	
C1426_4963	Membrane protein, Anti Sigma?	*P. putida* (77%)	
C1426_4965	Hydroxamate-type ferrisiderophore receptor	*P. putida* (85%), *P. stutzeri* (84%)	
C1426_4966	Putative transposase	*P. aeruginosa* DK2 (96%)	
C1426_4968	AraC family regulator	*P. fluorescens* (78%)	
C1426_4973	Chloride channel core protein	*Pseudomonas* sp. M1 (82%), *P. nitroreducens* (82%), *P. fuscovaginae* (79%)	
C1426_4975	Short chain dehydrogenase	*Xanthomonas* (98%), *Cronobacter* (96%)	
C1426_4979	RND drug efflux transporter	*Sphingomonas* sp. ERG5 (81%)	Inner membrane protein
C1426_4980	Multidrug resistance protein	*Sphingomonas* sp. ERG5 (78%)	Membrane fusion component
C1426_4981	Multidrug resistance protein	*Sphingobium japonicum* (71%), *Sphingomonas* sp. KA1 (70%)	Outer membrane protein
C1426_4982	Alcohol dehydrogense	*Cronobacter* (93%), *Xanthomonas* (94%)	
C1426_4983	Oxidoreductase	*Cronobacter* (100%), *Xanthomonas* (97%)	
C1426_4986	Transcriptional regulator	*Xanthomonas* (98%)	*tetR* family
C1426_4987	Transcriptional regulator	*Cronobacter* (99%), *Xanthomonas* (97%)	
C1426_4989	TraG, conjugal transfer coupling protein	*P. putida (*99%), *Xanthomonas* (98%), *P. aeruginosa* BWHPSA026 (97%)	Allow DNA transfer
C1426_4992	RND Transporter	*Xanthomonas* (81%)	Acriflavin resistance type
C1426_4993	Secretion protein	*Xanthomonas* (71%)	HylD, type one secretion
C1426_4994	Efflux transporter OprM	*Xanthomonas* (53%)	
C1426_4999	Enoyl-CoA-hydratase	*Pseudomonas* sp. CF150 (92%)	
fadD2	Ligase	*P. stutzeri* (78%), *Pseudomonas* sp. CF150 (73%)	Long chain fatty acid-CoA ligase
C1426_5001	Hydrolase	*P. stutzeri* (81%), *Pseudomonas* sp. CF150 (78%)	
C1426_5002	Nitroreductase	*P. stutzeri* (79%), *Pseudomonas* sp. CF150 (72%)	
C1426_5003	Enoyl-CoA-hydratase	*P. stutzeri* (92%), *Pseudomonas* sp. CF150 (88%)	
C1426_5005	Relaxase	*P. aeruginosa* PA45 (99%), *P. aeruginosa* BL04 (99%), *P. aeruginosa* BL09 (99%)	Integrating conjugative element
C1426_5006	Phage Integrase	*P. aeruginosa* 152504 (99%), *Pseudomonas* sp. 2_1_26 BL04 (99%), *P. aeruginosa* PA14 (98%)	

Table 4.5 Table showing the gene id, putative function and percentage homology as predicted by BLAST analysis of genes present within the C1426/C1433 specific 56 kbp region within the 139 kbp island possessed by only C1426 and C1433.

4.10 A 37 kbp bacteriophage unique to C1426 has been lost from C1433 during host adaptation

All previously identified large genomic differences described in this chapter are present in both C1426 and its mutant C1433. In this next example, we have identified a further bacteriophage of around 37 kbp that is present in C1426 only (Figure 4.8). It is hypothesised that this bacteriophage, C1426 Prophage 1, was lost by C1426 during host adaptation to its partner strain C1433. Therefore, it is expected that the loss of this bacteriophage will contribute to the altered phenotype exhibited by the late strain C1433. As can be seen in Table 4.6, the majority of proteins encoded by genes present in this bacteriophage are of unknown function. However, several significant proteins have been lost along with this bacteriophage (Table 4.6).

The first of these proteins is predicted to be a chitinase (C1426_1336). Chitinases are responsible for degrading chitin, a polysaccharide ubiquitous in nature (Thompson *et al*, 2001). It is the second most abundant polysaccharide and is a key component of the carbohydrate cell wall of fungi. Chintinases produced by bacterial strains will therefore aid in the killing of fungal species and degradation of fungal material in the rhizosphere. In the context of this discussion, more importantly, chitinases have been shown to be antigenic. That is, a substance produced by C1426 that will result in the production of an antibody by a host (Thompson *et al*, 2001). Mucoid strains such as C1433 lack antigenic molecules such as pili/flagella (Hassett *et al*, 2009). Therefore, it is expected that C1433 has lost the gene encoding this chitinase protein in an attempt to evade its host by being less antigenic, potentially masking its presence within the CF lung.

The second key protein lost along with this C1426 unique bacteriophage is predicted to be a histone H1 protein (C1426_1341). Histone H1 is a key component of the eukaryotic chromatin (Perez-Montero *et al*, 2013). Therefore, it is highly unlikely that this gene in *P. aeruginosa* will encode for

histone H1. However, it has previously been shown that *P. aeruginosa* possess a protein, AlgR3, that resembles eukaryotic histone H1 (Kato, Misra and Chakrabarty, 1990). AlgR3 has been shown to regulate alginate biosynthesis in a positive manner. As C1433 has a mucoid phenotype due to constitutive production of alginate, there is no need for for activation, proving a reasonable explanation for the loss of this protein.

It seems likely that C1426 Prophage 1 entered the genome between genes corresponding to PA14_16970 and PA14_16980 in PA14 and PA3663 and PA3664 in PA01. This region has been identified as a region of genome plasticity (RGP82) (Klockgether *et al*, 2011) that, to date, has only been found in the LES-like genomes. In LESB58, this region plays host to the 42.8 kbp LES-prophage 3. Compared to the 53 genes present within LES-prophage 3 (Winstanley *et al*, 2009), C1426 Prophage 1 is compromised of only 50 genes based on our xBase annotations. This may be accounted for by the smaller size of our identified bacteriophage. A separate BLASTn analysis comparing these two bacteriophages shows that there is only a 39% homology between the pair (Figure 4.8) showing that C1426 Prophage 1, truly is unique to C1426 when compared to others in this study.

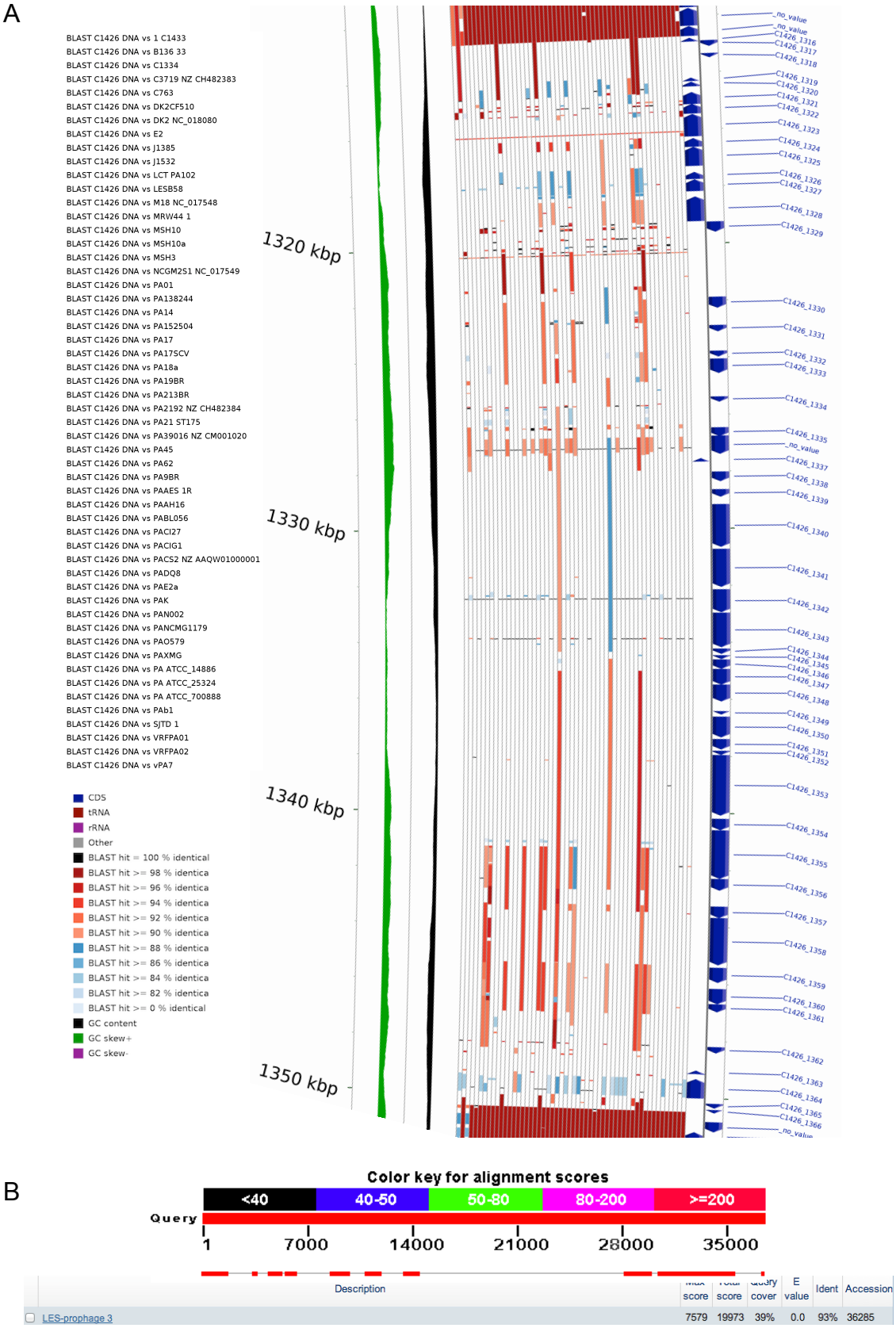


Figure 4.8 **A** The 37 Kbp C1426 Prophage 1 is present in C1426 only and occupies a region in LESB58 that is reserved for LES-prophage 3. **B** C1426 Prophage 1 and LES-prophage 3 share only a small degree of homology to each other.

Gene id	Putative function	Homology (Identify %)	Comments
C1426_1316	Phage integrase	*P. aeruginosa* VRFPAO2 (98%) ,*P. aeruginosa* NCMG1179 (98%), *P. aeruginosa* WC55 (98%)	
C1426_1317	Hypothetical protein	*Alicyclobacillus acidoterrestris* ATCC 49025 (53%)	
C1426_1319	Hypothetical protein	*Pseudomonas* sp. P179 (97%), *P. aeruginosa* PACL (97%), *P. aeruginosa* NCMG1179 (94%)	
C1426_1320	Hypothetical protein	*Pseudomonas* phage vB_PaeS_PMG1 (94%), *P. aeruginosa* NCMG1179 (91%)	
C1426_1321	Hypothetical protein	*P. aeruginosa* M18 (76%), *P. aeruginosa* PABE173 (86%), *P. putida* KT2440 (68%)	
C1426_1322	Hypothetical protein	*P. aeruginosa* PA7 (95%), *P. aeruginosa* Stone 130 (99%), *P. aeruginosa* PAB173 (78%)	
C1426_1323	Methyltransferase	*Rhodocyclaceae bacterium* RZ94 (61%), *Delftia* sp. Cs1-4 (58%)	
C1426_1324	Hypothetical protein	*P. aeruginosa* LESB58 (99%), *P. aeruginosa* 2182 (98%), *P. aeruginosa* PACL (98%)	
C1426_1325	Hypothetical protein	*P. aeruginosa* 138244 (96%), *P. aeruginosa* PACL (96%), *Psuedomonas* sp. M1 (37%)	
C1426_1326	Hypothetical protein	*P. aeruginosa* WC55 (98%), *P. aeruginosa* 138244 (87%), *P. aeruginosa* LESB58 (84%)	
C1426_1327	Transcriptional regulator	*P. aeruginosa* WC55 (98%), *P. aeruginosa* LESB58 (98%), *P. aeruginosa* 138244 (96%)	LuxR family
C1426_1328	Hypothetical protein	*P. aeruginosa* WC55 (98%), *P. aeruginosa* PACL (91%), *P. aeruginosa* M18 (81%)	
C1426_1330	Hypothetical protein	*P. aeruginosa* PACL (99%), *P. aeruginosa* LESB58 (95%), *P. aeruginosa* 138244 (95%)	
C1426_1331	Hypothetical protein	*P. aeruginosa* WC55 (96%), *P. aeruginosa* NCMG1179 (93%), *P. aeruginosa* (95%)	
C1426_1332	NinG protein	*P. aeruginosa* WC55 (100%), *P. aeruginosa* NCMG1179 (100%)	Derived from Pseudomonas phage D3
C1426_1333	NinG protein	*P. aeruginosa* WC55 (99%), Bacteriophage lambda *ninG* (99%), *P. aeruginosa* LESB58 (96%)	Phage Derived
C1426_1334	Hypothetical protein	*P. aeruginosa* PACL (100%), *P. aeruginosa* WC55 (100%), *P. aeruginosa* 138244 (98%)	
C1426_1335	Holin	*P. aeruginosa* PACL (99%), *P. aeruginosa* WC55 (99%), *P. aeruginosa* VRFPAO2 (99%)	Phage derived - involved in cell lysis.
C1426_1336	Chitinase	*P. aeruginosa* 138244 (93%), *P. aeruginosa* M18 (93%), *P. aeruginosa* LESB58 (91%)	Break down chitin in the cell wall of fungi and eukaryotes.
C1426_1338	Hypothetical protein	*P. aeruginosa* 138244 (98%), *P. aeruginosa* WC55 (98%), *P. aeruginosa* M18 (54%)	
C1426_1339	Hypothetical protein	*P. aeruginosa* 138244 (100%), *P. aeruginosa* WC55 (100%), *Ralstonia* sp. PBA (74%)	
C1426_1340	Terminase	*P. aeruginosa* WC55 (99%), *P. aeruginosa* 138244 (96%), *Bordetella bronchiseptica* RB50 (73%)	Phage derived. Involved in DNA packaging and lambda phage assembly.
C1426_1341	Histone H1	*P. aeruginosa* WC55 (100%), *P. aeruginosa* 138244 (95%), *Pseudomonas* sp. TJI-51 (84%)	Eukaryotic chromatin protein
C1426_1342	Peptidase	*P. aeruginosa* WC55 (99%), *P. aeruginosa* 138244 (95%), *Pseudomonas* sp. TJI-51 (74%)	Clp protease - degrades incorrectly formed proteins
C1426_1343	Capsid protein	*P. aeruginosa* WC55 (99%), *P. aeruginosa* 138244 (95%), *P. psychrotolerans* L19 (88%)	Bacteriophage HK97 related

C1426_1344	Hypothetical protein	*P. aeruginosa* WC55 (100%), *P. aeruginosa* 138244 (73%), *Chronobacter turicensis* z3032 (42%)	
C1426_1345	Hypothetical protein	*P. aeruginosa* WC55 (100%), *P. aeruginosa* 138244 (71%), *Polaribacter* sp. MED152 (35%)	
C1426_1346	DNA packaging protein	*P. aeruginosa* WC55 (98%), *P. aeruginosa* 138244 (81%), *P. stutzeri* CCUG 29243 (64%)	Bacteriophage HK97 derived
C1426_1347	Hypothetical protein	*P. aeruginosa* WC55 (99%), *P. aeruginosa* M18 (87%), *P. aeruginosa* 138244 (88%)	
C1426_1348	Hypothetical protein	*P. aeruginosa* WC55 (98%), *P. aeruginosa* 138244 (97%), *P. mendocina* ymp (58%)	
C1426_1349	Hypothetical protein	*P. aeruginosa* 138244 (100%), *P. aeruginosa* WC55 (97%), *P. stutzeri* CCUG 29243 (50%)	
C1426_1350	Hypothetical protein	*P. aeruginosa* WC55 (96%), *P. aeruginosa* M18 (96%), *P. aeruginosa* 138244 (96%)	
C1426_1351	Hypothetical protein	*P. aeruginosa* M18 (98%), *P. aeruginosa* WC55 (98%), *P. aeruginosa* 138244 (97%)	
C1426_1352	Hypothetical protein	*P. aeruginosa* 138244 (96%), *P. aeruginosa* WC55 (94%), *P. stutzeri* (38%)	
C1426_1353	Phage tail protein	*P. aeruginosa* 138244 (99%), *P. aeruginosa* M18 (99%), *P. stutzeri* CCUG 29243 (45%)	Phage related minor tail protein
C1426_1354	Hypothetical protein	*P. aeruginosa* M18 (100%), *P. aeruginosa* WC55 (100%), *P. aeruginosa* 138244 (100%)	
C1426_1355	Hypothetical protein	*P. aeruginosa* M18 (98%), *P. aeruginosa* WC55 (97%), *P. aeruginosa* 138244 (96%)	
C1426_1356	Hypothetical protein	*P. aeruginosa* WC55 (97%), *P. aeruginosa* M18 (96%), *P. aeruginosa* PACL (96%)	
C1426_1357	Hypothetical protein	*P. aeruginosa* PACL (99%), *P. aeruginosa* WC55 (99%), *Pseudomonas* Phage F10 (99%)	Phage derived
C1426_1358	Hypothetical protein	*P. aeruginosa* (98%), *Pseudomonas* Phage F10 (97%), *P. aeruginosa* PACL (96%)	Phage derived
C1426_1359	Hypothetical protein	*P. aeruginosa* WC55 (98%), *P. aeruginosa* 138244 (96%), *P. aeruginosa* NCMG1179 (96%)	
C1425_1360	Hypothetical protein	*P. aeruginosa* LESB58 (99%), *Pseudomonas* Phage F10 (99%), *P. aeruginosa* (99%)	Phage derived
C1426_1361	Hypothetical protein	*P. aeruginosa* NCMG1179 (99%), *P. aeruginosa* (99%), *Pseudomonas* Phage F10 (99%)	Phage derived
C1426_1362	Hypothetical protein	*P. aeruginosa* M18 (97%), *P. aeruginosa* LESB58 (97%), *Pseudomonas* Phage DMS3 (51%)	Phage derived
C1426_1363	Hypothetical protein	*Cryptobacterium curtum* DSM 15641 (58%), *Acidaminococcus intestini* CAG:325 (48%)	
C1426_1364	Phage protein	*P. aeruginosa* (not identified) (95%), *P. aeruginosa* E2 (93%), *P. aeruginosa* MSH10 (93%)	
C1426_1365	Bacteriophage protein	*P. aeruginosa* WC55 (100%), *P. aeruginosa* (not identified) (100%), *P. aeruginosa* 138244 (98%)	

Table 4.6 Table showing the gene identity, putative function and percentage homology as predicted by BLAST analysis of genes present within the 37 kbp C1426 specific bacteriophage, C1426 Prophage 1. The majority of proteins are of hypothetical function, many are phage derived.

4.11 Limitations of using xBase to annotate C1426

C1426 and C1433 both contain a large segment of DNA that is only present in 30% of the genomes included in this analysis. One of these genomes is that of LESB58. In this 13 kbp portion of DNA between genes *ihfB* and *wbpM*, only two genes of significant size have been annotated in C1426 (Figure 4.9). In contrast, in the same region of DNA in LESB58, four individual genes have been annotated (Figure 4.9). An alignment of the nucleotide sequences of the DNA between *ihfB* and *wbpM* for both C1426 and LESB58 has shown that there is a significant level of homology between the two, albeit C1426 contains an extra 1.8 kbp of genomic information at the 5' end. We therefore predict that the C1426 annotation predicted by xBase for this region is inaccurate and that the true annotation is extremely similar to that described for LESB58 being comprised of four significantly sized genes (Figure 4.9).

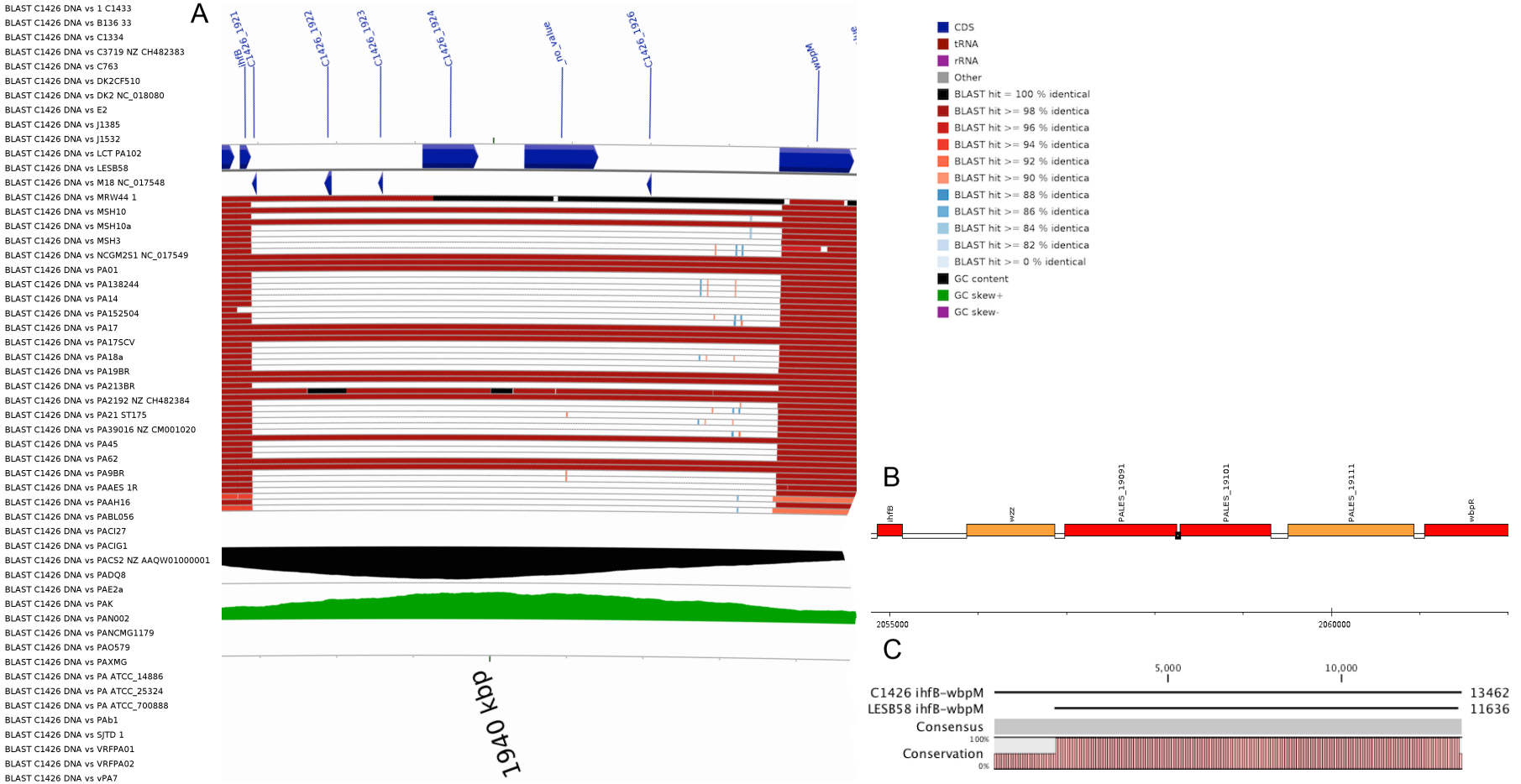


Figure 4.9 **A** Annotations by xBase has predicted that this 13 kbp segment contains only two significant genes. **B** The same segment of DNA in LESB58 shows that, in fact, this segment should be annotated with four genes of significant size. **C** Alignment demonstrating that these segments of DNA have a significant level of homology with each other.

4.12 Genomic differences between C1426 and C1433 at the single nucleotide level

Previously, the major genomic macrodifferences between C1426 and C1433 were investigated during this chapter. However, it is essential to take a closer look at their genomes and investigate any minor genomic changes occurring at the single nucleotide level e.g. single nucleotide polymorphisms (SNP) and insertion/deletion polymorphisms (InDel). It is possible that such a small polymorphism might result in a dramatic change in phenotype. For example, the switch of *P. aeruginosa* strains from non-mucoid to mucoid occurs through loss-of-function mutations in the *mucABCD* cluster that regulates *algU* expression and subsequent alginate biosynthesis, particularly within the anti-sigma factor MucA (Schurr *et al*, 1996, Boucher *et al*, 1997 & Jones *et al*, 2010). Within this mucoid pairing, we hope to identify any such loss-of-function mutations that may contribute to the altered phenotype of C1433. Over the course of one month during host adaptation, C1433 has acquired 27 non-synonymous SNPs and 10 non-synonymous InDels.

4.12.1 C1433 has acquired 27 non-synonymous SNPs relative to C1426 during host adaptation

Over the course of one month within the CF lung, C1426 underwent a variety of mutations including acquiring 27 non-synonymous SNPs during its switch to the mucoid C1433. These non-synonymous SNPs are listed in Table 4.7 and many of these have occurred in genes with no predicted function (hypothetical protein). However, many SNPs do occur in genes whose function has been predicted by xBase or based on the functions of identical proteins in other *P. aeruginosa* strains.

A SNP has been identified in *wspF* (Table 4.7), the regulator of *wspA*, a gene encoding a chemotaxis protein involved in cell motility. Mutants of *wspF* are deficient in the twitching, swarming and swimming motility commonly observed in *P. aeruginosa* strains (D'Argenio *et al*, 2002). Therefore, potential inactivation of *wspF* by the introduction of a non-synonymous SNP may contribute to the reduced motility observed in C1433 (Amy Ford - unpublished results). Colvin *et al* (2013) previously showed that introduction of a mutation into *wspF* also results in increased biofilm formation. It is therefore hypothesised that the SNP present in the *wspF* gene of C1433 contributes to the increased biofilm formation of this strain as demonstrated in Figure 4.1.

Non-synonymous SNPs are also found in three genes involved in pyoverdin synthesis; *pvdL*, C1426_2790 and C1426_2791 (Table 4.7). Pyoverdin is one of two siderophores produced by *P. aeruginosa* that act as powerful iron chelators, sequestering iron from the environment surrounding the bacterium (Meyer *et al*, 1996). Work within our lab has shown that compared to C1426, C1433 has a reduced level of siderophore production (Table 4.1).

Reference Position	Consensus Position	Reference	Allele Variations	Gene ID	Amino Acid Change	Putative function
799608	799421	A	T	C1426_0807	p.Leu200His	Cytochrome C oxidase subunit I (partial)
829640	829453	T	G	*ampR*	p.Leu161Arg	Regulator of β-lactamase
908732	908500	T	G	C1426_0920	p.*40Glu	Two component response regulator
1185412	1185098	T	A	C1426_1198	p.Leu417His	Hypothetical protein
1271191	1270877	G	A	*wspF*	p.Arg229His	Chemotaxis response regulator
1315342	1312420	A	T	C1426_1322	p.Cys34Ser	Hypothetical protein
1315377	1312455	G	A	C1426_1322	p.Ala22Val	
1315395	1312473	G	T	C1426_1322	p.Ala16Asp	
1377779	1339976	C	T	*dnaE*	p.Ala751Val	DNA polymerase III, α-subunit
1379716	1341913	C	G	*accA*	p.Pro161Arg	Acetyl-CoA carboxylase, α-subunit
1911006	1873101	C	G	C1426_1900	p.Gly30Ala	Hypothetical protein
1999439	1961461	G	A	C1426_1976	p.Pro185Leu	Hypothetical protein
2282589	2244555	G	A	C1426_2257	p.Ala118Val	Secretion protein
2795880	2757759	C	T	*pvdL*	p.Pro4233Leu	Pyoverdin synthesis
2823213	2785092	ACC	ATG	C1426_2790	p.Thr979Met	pvdI (III)
2832518	2794397	C	A	C1426_2791	p.Leu1398Ile	Peptide synthase (pvd (III))
3489829	3451482	A	T	C1426_3411	p.Gln137Leu	Two component sensor - parS
4183763	4145237	G	A	C1426_4098	p.Gly175Asp	Alcohol dehydrogenase
4337422	4298894	G	T	*dapA*	p.Asn75Lys	Dihydrodipicolinate synthase
4673978	4635416	C	T	C1426_4592	p.Gly18Ser	vreA
4921662	4883059	A	G	C1426_4829	p.Glu102Gly	Hypothetical protein
5210014	5171350	G	T	C1426_5113	p.Thr42Asn	Hypothetical protein
5341464	5302790	A	C	*acsB*	p.Met371Leu	Acyl-CoA synthetase
5594432	5555746	G	T	C1426_5486	p.Pro85Thr	Hypothetical protein
5864683	5825958	G	C	C1426_5731	p.Ala313Gly	Acyl-CoA dehydrogenase
6237581	6198773	A	G	C1426_6088	p.Tyr46His	Hypothetical protein
6238791	6199983	G	C	C1426_6090	p.Thr199Ser	Hypothetical protein

Table 4.7 Table showing the non-synonymous SNPs present in C1433 and the putative function of the genes in which they have occurred. 27 non-synonymous SNPs have occurred in C1433 over the space of one month in the CF lung.

4.12.2 A SNP in *ampR* may result in a loss-of-function of the global regulator AmpR

A non-synonymous SNP was identified in *ampR* of C1433 resulting in an arginine replacing a leucine at amino acid 161 (Table 4.7). AmpR is a global transcriptional regulator in *P. aeruginosa* that regulates expression of a variety of targets. This includes positive regulation of *ampC* and quorum sensing genes *lasB* and *rhlR* and negative regulation of *poxB* and quorum sensing genes *lasA*, *lasI* and *lasR* (Kong *et al*, 2005) Overall, AmpR influences the expression of over 500 genes (Balasubramanian *et al*, 2012). In order to investigate what, if any, effects this non-synonymous SNP has on the phenotype of C1433 a quantitative Real-time PCR analysis was performed on the expression levels of *ampR* and *ampC* between strains C1426 and C1433 in relation to a housekeeping gene. Gene expression was quantified using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). (Figure 4.10).

Real-time PCR analysis revealed that compared to C1426, the expression of *ampR* and *ampC* in C1433 are down regulated 2.45-fold and 5.23-fold, respectively. It is therefore hypothesised that the *ampR* SNP results in a loss-of-function mutation. Several phenotypic changes exhibited by C1433 support this suggestion. For example, AmpR has been shown to negatively regulate alginate biosynthesis and biofilm formation (Balasubramanian *et al*, 2012). The mucoid C1433 displays increased alginate production and is more proficient at biofilm formation compared to C1426 (Figure 4.1), phenotypes possibly contributed to by decreased AmpR activity. Kong *et al* (2005) have also shown that Δ*ampR* mutants have increased pyocyanin production, a phenotype consistent in C1433 (Table 4.1)

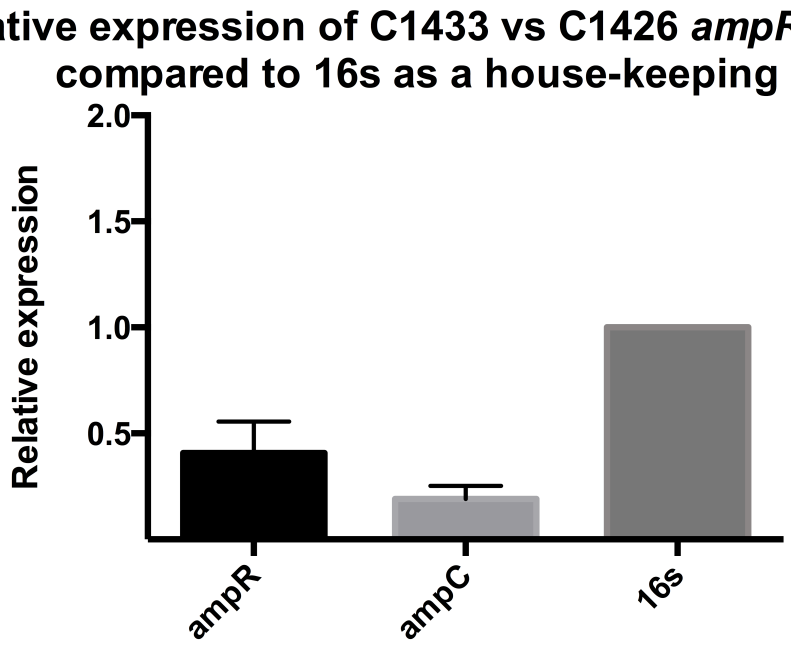


Figure 4.10 Graph showing the relative expression of *ampR* and *ampC* in C1433 versus C1426 in comparison to a 16S house-keeping gene. A relative expression of <1 is indicative of reduced expression in C1433 compared to C1426.

AmpR regulates β-lactam resistance via upregulation of expression of the β-lactamase, AmpC (Balasubramanian *et al*, 2012). A loss-of-function mutation in C1433, as expected, may result in increased sensitivity to β-lactam antibiotics. This hypothesis was tested using a standard nitrocefin assay. Nitrocefin is a chromogenic β-lactam antibiotic that changes from red to yellow when cleaved by a β-lactamase (O'Callaghan *et al*, 1972). Loss of a functioning AmpR protein is expected to result in decreased production of β-lactamase. As can be seen in Figure 4.11, this is not the case. Both C1426 and C1433 are similarly proficient at production of this enzyme.

Pseudomonads characteristically produce inducible β-lactamases when their production is challenged by an inhibitors such as clavulanic acid (Livermore *et al*, 1989). We therefore explored the ability of C1426 and C1433 to increase their production of β-lactamase in response to challenge by clavulanic acid. As can be seen in Figure 4.11, C1433 has lost the ability to induce β-lactamase production in response to clavulanic acid-mediated inhibition. In contrast, C1426 has significantly (p-value <0.001) increased production as a result of induction as demonstrated by paired T-test analysis. Therefore, we hypothesise that the loss-of-function mutation in *ampR* of C1433 may result in increased sensitivity to β-lactam antibiotics due to a loss of inducible production of β-lactamase enzymes.

Although the loss-of-function mutation in *ampR* may contribute to increased sensitivity to β-lactam antibiotics, this mutation may also contribute to the increased resistance to antibiotic intervention associated with the mucoid phenotype (Hassett *et al*, 2009). AmpR has been shown to have a negative regulatory effect on MexT, a key protein in a RND drug-efflux system involved in resistance to fluroquinalones, chloramphenicol and trimethoprim (Balasubramanian *et al*, 2012). This study demonstrated that genes involved in these efflux pumps, *mexEF-oprN*, were upregulated significantly in Δ*ampR* mutants. Therefore, it is suggested that via a loss-of-function mutation in

C1433 *ampR* will contribute to increased antibiotic resistance via overexpression of these RND efflux pumps.

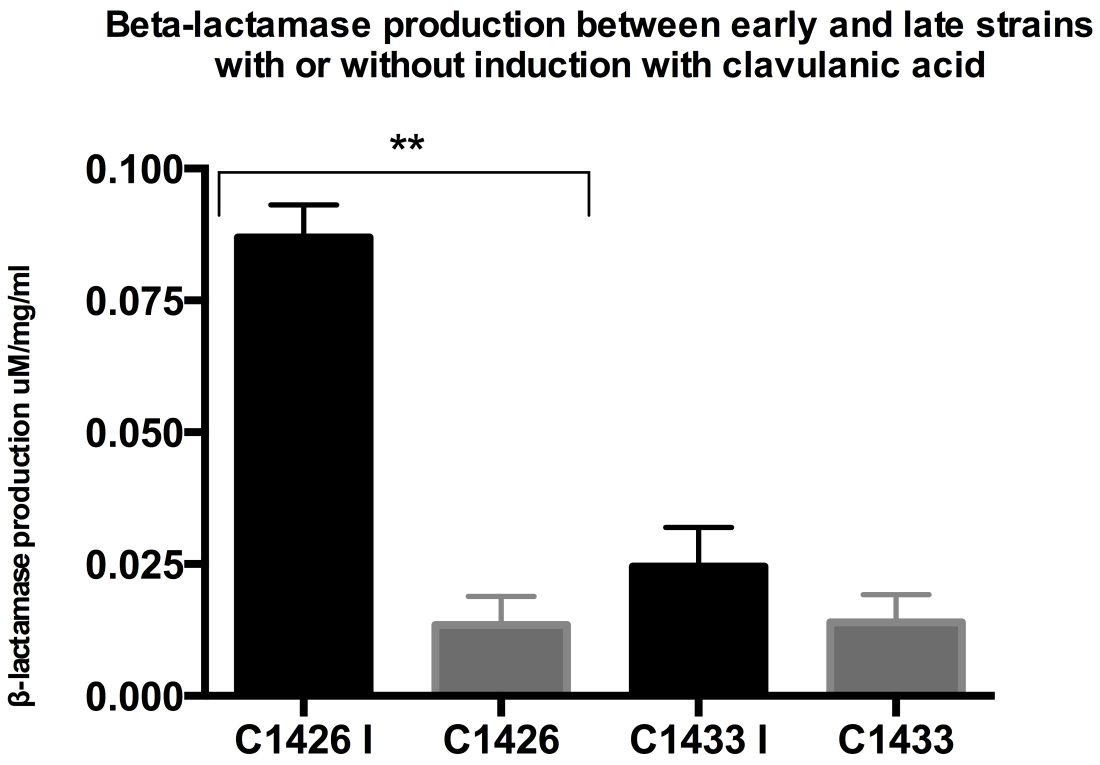


Figure 4.11 Graph showing the results of the nitrocefin assay performed on C1426 and C1433 with and without induction with clavulanic acid. β-lactamase production for each strain and condition is standardised based on the protein concentrations of the crude extracts. (** p-value <0.001)

4.12.3 C1433 has undergone 10 non-synonymous InDel polymorphisms

During host adaptation in the CF lung, compared to its non-mucpid counterpart C1426, C1433 has acquired 10 InDel mutations (Table 4.8). 60% of these are deletion polymorphisms with the remainder being insertion polymorphisms. Around half of the genes affected by InDels code for enzymes involved in amino acid, lipid or carbohydrate modifications. This may have implications on the metabolic profile of C1433, a hypothesis that will be considered during a significant metabolomic analysis that will be discussed in Chapter 6.

Reference Position	Consensus Position	Reference	Allele Variations	Gene ID	Amino Acid Change	Putative function
506921	506806	AA	--	C1426_0505	p.Lys258fs	Acyl-CoA dehydrogenase
1185411	1185098	C	-	C1426_1198	p.Leu417fs	Hypothetical Protein
2515274	2477193	C	C/-	C1426_2510	p.Gly154fs	Glutathione s-transferase
3285528	3247215	-	A	*exaA*	p.Pro304fs	Glucose dehydrogenase
3290528	3252219	-	G	C1426_3213	p.Ala64fs	Two-component system sensor
4047854	4009360	A	-	C1426_3974	p.Val153fs	Proline Racemase
4047859	4009364	-	G	C1426_3974	p.Arg155fs	
4588301	4549745	C	-	*mucA*	p.Ala144fs	Anti-sigma factor
4589177	4550620	A	A/-	*algU*	p.Ile56fs	Sigma factor
5210185	5171519	-	-/T	C1426_5114	p.Ala1025fs	Hypothetical Protein

Table 4.8 Table showing the non-synonymous InDels present in C1433 and the putative function of the genes in which they have occurred. 10 non-synonymous InDels have occurred in C1433 over the space of one month in the CF lung.

4.12.4 Two InDels present in C1433 may contribute to the mucoid phenotype

Two InDels particularly stand out from Table 4.8 in context with the mucoid phenotype exhibited by strain C1433. The mucoid phenotype results from an overproduction of the exopolysaccharide alginate, the production of which is encoded by around 30 genes (Ramsey and Wozniak, 2005). Overall control of alginate biosynthesis is controlled by the sigma factor AlgU (AlgT) and its antagonistic anti-sigma factor MucA (Mathee, McPherson and Ohman, 1997). Interestingly, InDels have been identified in both AlgU and MucA (Table 4.8).

Inactivation of the *mucA* gene results in the overproduction of alginate as a result of uncontrolled AlgU activity (Govan and Deretic, 1996). Subsequently, it has been identified that 80% of mucoid *P. aeruginosa* strains isolated from the lungs of CF patients harbour a variety of mutations in *mucA* (Boucher *et al*, 1997). This suggests that *mucA* is a “hot-spot” for mutations resulting in the appearance of a mucoid phenotype. We therefore hypothesised that the InDel identified in *mucA* of C1433 results in alginate overproduction and the subsequent appearance of the mucoid phenotype observed in this strain. This InDel is the result of the deletion of a single cytosine nucleotide and an amino acid change of Ala144Arg (GCC→G_C) (Figure 4.12). This frameshift mutation introduces a premature stop codon at amino acid 148, truncating the protein from its full size of 194 amino acids (Figure 4.12). This significant reduction in protein size will most likely result in a truncated, defective MucA protein and a loss of function. Consequently, inactivation of this anti-sigma factor might result in uncontrolled expression of the AlgU sigma factor, overproduction of alginate and the appearance of the mucoid phenotype.

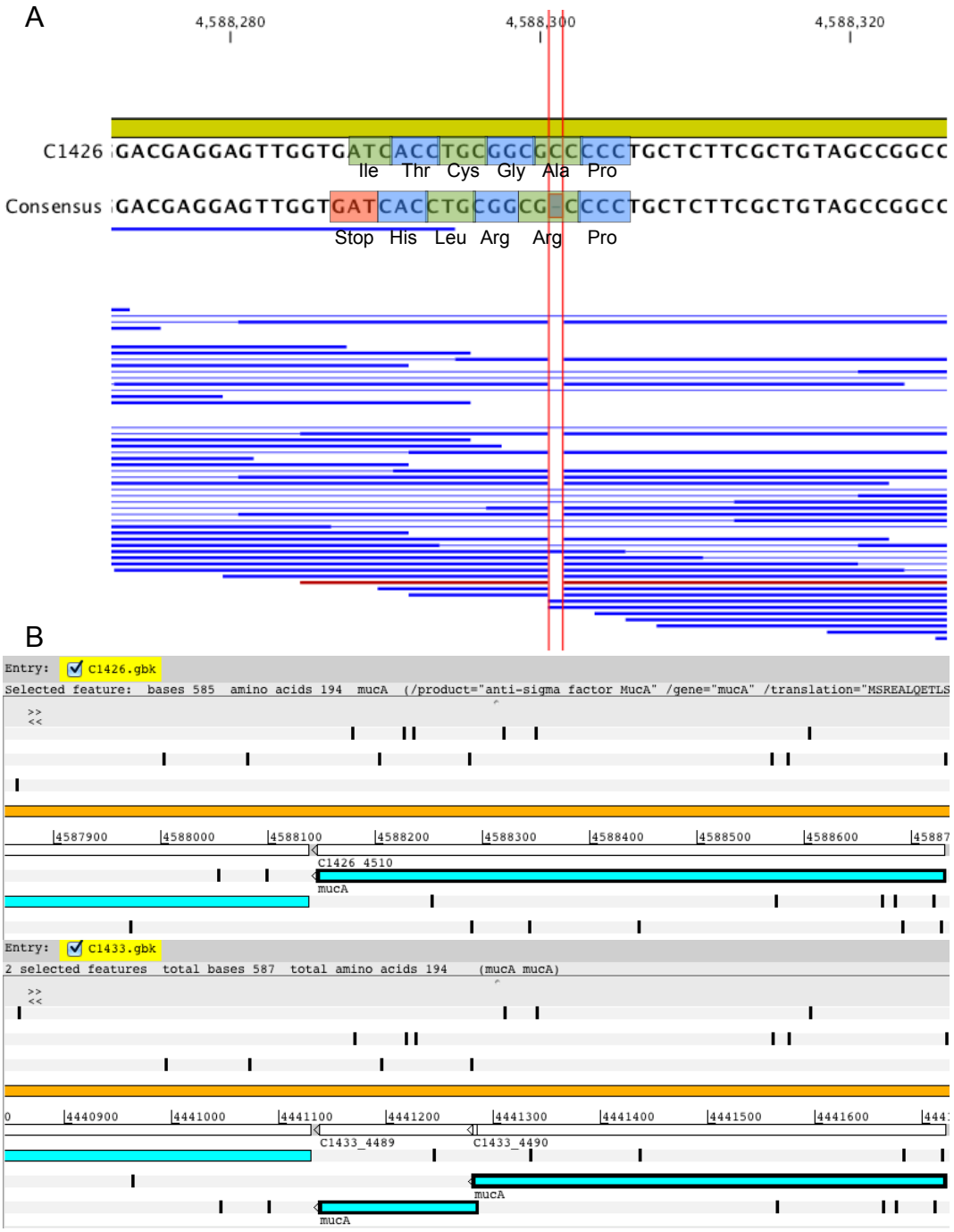


Figure 4.12 **A** CLC Genomics Viewer read map analysis showing a deletion of a cytosine within the *mucA* gene in C1433 compared to its progenitor C1426. This results in a frameshift mutation in C1433 and insertion of a premature stop codon (red). **B** Artemis screen grab showing the truncated MucA protein in C1433 compared to C1426 as a result of the single nucleotide deletion in the *mucA* gene.

As mentioned, an InDel has been observed in the sigma factor *algU*. To the best of our knowledge, no previous polymorphisms causing a switch to mucoidy have been associated with this sigma factor. Any mutation within this protein would have to cause constitutive activation, without which excessive alginate biosynthesis would not occur. The InDel identified in C1433 *algU* results in a frameshift mutation and a truncated protein (Figure 4.13). If this was the case, C1433 could not be a mucoid strain as alginate biosynthesis could not occur. Further investigation into this InDel demonstrates that this polymorphism is only present in seven out of twenty-two reads and therefore, this should not be considered a true InDel (Figure 4.13). Instead, it can be assumed that annotations are incorrect and C1433 possesses a fully functioning AlgU protein.

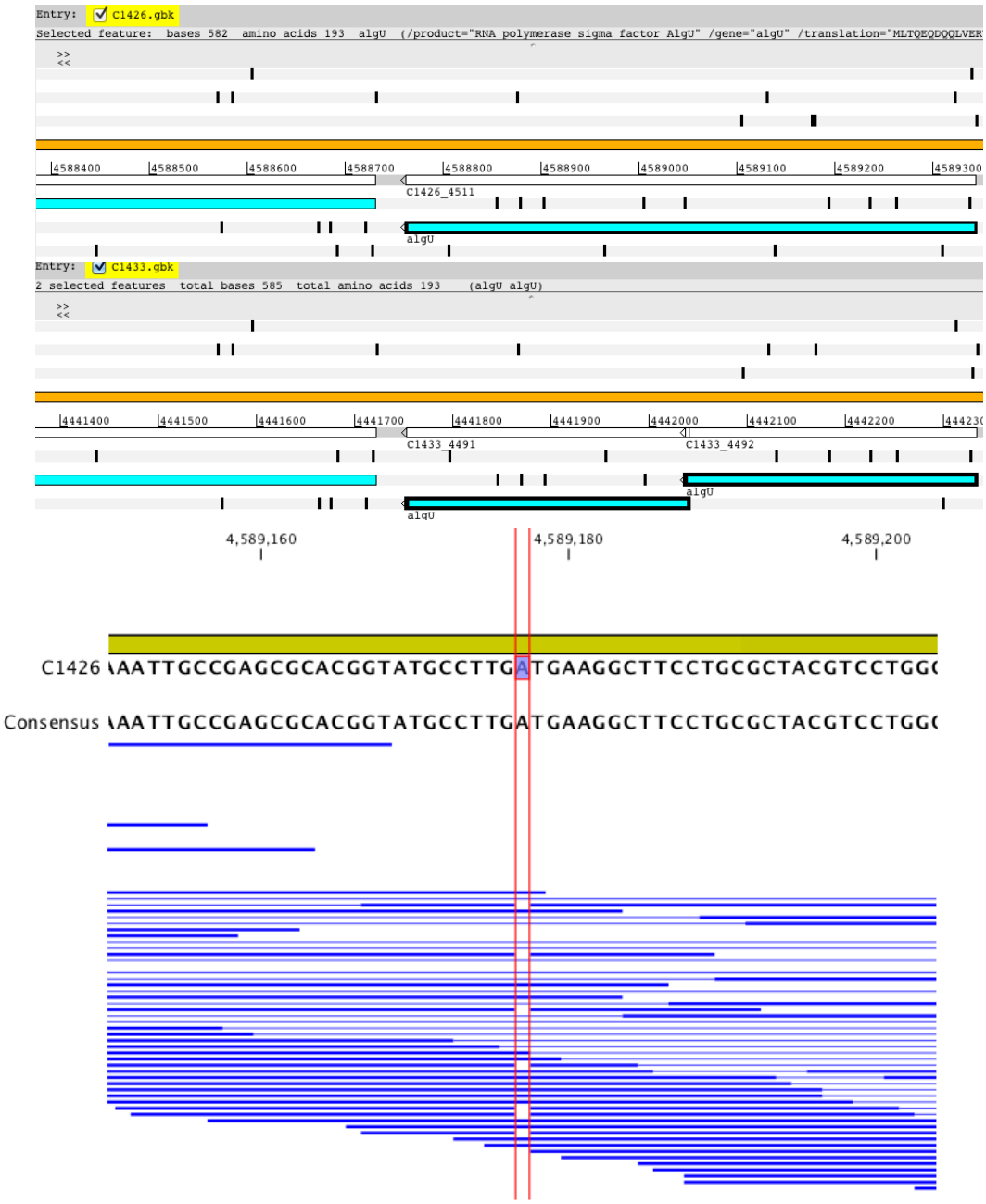


Figure 4.13 **A** C1433 has been annotated to show that a SNP within the *algu* gene results in a truncated AlgU protein compared to C1426. **B** Closer analysis shows that *algU* in C1433 does not contain a deletion polymorphism and the annotation of AlgU in C1433 is incorrect and actually possesses a fully functioning AlgU protein.

4.13 Chapter 4 discussion

Throughout this chapter, we have discussed both macro and micro differences observed between the non-mucoid C1426 and its mucoid mutator strain C1433. We have also compared both strains to 53 other fully sequenced *P. aeruginosa* strains from a broad range of niches. It appears that both C1426 and C1433 contain several large genomic regions acquired via horizontal gene transfer present in either none or very few *P. aeruginosa* strains included in this study. It has been shown that C1426 and C1433 have acquired a variety of bacteriophages including a LESB58 Prophage 5-like bacteriophage C1426 Prophage 2 (Winstanley *et al*, 2009) and a Pf1-like bacteriophage C1426 Prophage 3 (Hill *et al*, 1991). There is no evidence to suggest that these phages will confer any selective advantage to either C1426 or C1433.

In contrast, three further horizontally acquired regions present in C1426 and C1433 do appear to confer an advantage to these strains in their host environment of the CF lung. The first of these, a small 3 kbp region, contains several *cup* genes involved in the synthesis of fimbriae, appendages on the surface of cells that are involved in biofilm formation. The second horizontally acquired region is a 4.5 kbp region that contains nine genes of which, two in particular stand out; a two-component regulatory system. This system, consists of a signal transduction histidine kinase BfmS and its regulator BfmR, two proteins are essential in the maturation stage of biofilm formation (Petrova and Sauer, 2009). Both acquisition of *cup* genes and the genes encoding the BfmSR two-component system by C1426 and C1433 may result in the increased biofilm formation observed in these strains compared to the second non-mucoid/mucoid pairing of J1385 and J1532 (Figure 4.1). This may also confer a selective advantage to C1426/C1433 in establishing a successful infection within the CF lung. The third of these regions resembles bacteriophage P2. However, the genomic map of this bacteriophage does not resemble that of P2 (Christie *et al*, 2002). Instead, the genes within the bacteriophage show significant homology to an R-type pyocin, a descendant

of phage P2 (Nakayama *et al*, 2000). Pyocins resemble phage tail proteins are proficient at killing other bacterial species as well as *P. aeruginosa* (Michel-Briand and Baysee, 2002). Therefore, acquisition of this pyocin may confer a selective advantage by allowing C1426/C1433 to establish dominance in the CF lung, killing other infecting *P. aeruginosa*.

A final, horizontally acquired region in C1426 and C1433 is located in the genomic location normally occupied by the PA14 pathogenicity island PAPI-1 (Qui, Gurkar and Lory, 2006). This genomic island, C1426 Genomic Island 1, shows some degree of homology to PAPI-1 (Figure 4.7) however, it is significantly larger at 139 kbp and contains a 56 kbp region that is specific to C1426 and C1433. This additional region encodes for a further 49 genes, some of which may confer an advantage to these strains. This includes two putative ferripyoverdine receptors allowing increased iron sequestration from their environment, giving them a competitive edge against other co-colonising bacterial strains. This region also contains two operons that encode for two, distinct tripartite RND multidrug efflux pumps. Possessing these pumps may contribute to the increased antibiotic profiles of these strains (Rita Lagido, unpublished results) and may also confer an increased chance of successfully infecting the CF lung where they will no doubt experience increased antibiotic interventions.

Each of these previously described macro differences in both C1426 and C1433 are present in either very few of or are entirely absent from the other *P. aeruginosa* strains included in the study. However, investigating the role of host adaptation within the CF lung, we must look specifically at macro and single nucleotide differences existing between C1426 and its mucoid descendant C1433. It has been identified that a bacteriophage, C1426 Prophage 1 has been excised from the chromosome by C1433 during the one month period of host adaptation within the lung. It is expected that C1433 has selected against this bacteriophage in an attempt to gain a selective advantage for establishing a successful chronic infection. In C1426,

this bacteriophage contains several significant proteins including a putative chitinase. Chitinases are responsible for breaking down chitin in the cell wall of fungi (Thompson *et al*, 2001) and interestingly, have also been shown to be antigenic. This leads to the suggestion that C1433 will have selected against the gene encoding this enzyme in an attempt to ravoid recognition by the innate immunity intervention and allowing continued survival of the bacterium. A second key gene lost along with the bacteriophage present in C1426 encodes for a protein that has a putative function of a key eukaryotic chromatin protein, histone H1. Kato, Misra and Chakrabarty (1990) have demonstrated that histone h1 resembles a *P. aeruginosa* protein involved in alginate synthesis AlgR3 and therefore, it is expected that rather than encoding for histone H1, this gene actually encodes for AlgR3. AlgR3 regulates alginate biosynthesis in a positive manner but as C1433 is already mucoid due to alginate overproduction, there is no need for further promotion via AlgR3, providing a reasonable explanation for its loss.

We have also identified a variety of differences between C1426 and C1433 at the single nucleotide level including 27 nonsynonymous SNPs and 10 InDels. This includes a non-synonymous SNP within the global regulator *ampR*. This regulator is involved in controlling antibiotic resistance to β-lactams via promoting expression of the β-lactamase AmpC. qPCR analysis of *ampR* has demonstrated that the SNP has resulted in a loss-of-function in this protein. A β-lactamase assay revealed that C1433 has lost its ability to induce β-lactamase activity in response to an inhibitor. The loss-of-function observation is supported by several phenotypic analyses including increased alginate biosynthesis and biofilm formation and increased pyocyanin production, two processes normally regulated by functioning AmpR proteins (Kong *et al*, 2005 & Balasubramanian *et al*, 2012). The mutant AmpR may alternatively contribute to increased antibiotic resistance via an entirely different mechanism involving its inability to negatively regulate the RND drug-efflux protein MexT (Balasubramanian *et al*, 2012).

SNPs were also identified in genes encoding proteins involved in motility (Chemotaxis protein WspF) and pyoverdin production (PvdL, C1426_2790 and C1426_2791). The SNP in *wspF* may contribute to the deficient twitching, swarming and swimming forms of motility observed in C1433 (Amy Ford, unpublished results). SNPs present in the proteins involved in pyoverdin synthesis may be solely responsible for the reduction observed in siderophore production in C1433 compared to C1426 (Table 4.1).

An InDel was identified that may contribute to the mucoid phenotype of C1433. This strain possesses a significantly truncated MucA protein as a result of a single deletion polymorphism. MucA, in its normal role, is an anti-sigma factor tasked with negatively regulating alginate biosynthesis via antagonistically binding to the sigma-factor AlgU. Inactivation of this anti-sigma factor will result in uncontrolled expression of AlgR that in turn, will cause overproduction in alginate biosynthesis and the appearance of the mucoid phenotype observed in strain C1433. An InDel was also identified in the sigma factor *algU* resulting in a truncated protein. However, closer inspection revealed that this was an incorrect annotation.

Throughout this chapter we have discussed a variety of macro and single nucleotide differences that may contribute to the altered phenotype observed in C1433 compared to its progenitor strain C1426. However, our investigations have only initiated this process and further analysis is required to fully map these differences. As can be seen in each table in this the chapter, the vast majority of genes/proteins identified are of unknown functions. It is almost certain that the mutations within or the entire loss of genes contribute to the overall phenotype of C1433. To fully understand the effect of host adaptation between C1426 and C1433, the function of these proteins must be investigated further.

Chapter 5 - Host adaptation between J1385 and J1532, a non-mucoid/mucoid pairing isolated from the lungs of a CF patient

5.1 Introduction

The second of our two pairings is the non-mucoid J1385 and its counterpart mucoid strain J1532. Both of these strains were isolated from the sputum aspirated from the lungs of the same CF sufferer, approximately three months apart. Over the course of these months, it is predicted that the most abundant strain infecting the CF lung switched from the non-mucoid J1385 to the mucoid J1532 (Professor John Govan, personal correspondence). Previous results (Figure 6.2) based on the genomic sequences of these strains showed that the switch between these strains is a result of host adaptation, and the resulting genomic changes, rather than strain succession or co-infection.

5.2 Phenotypic analyses of J1385 and J1532

These strains exhibit a high degree of phenotypic variability with each other. The most striking phenotypic difference is the appearance of the mucoid phenotype in the late J1532. However, there are several other striking differences between these strains. J1532 is more proficient at biofilm formation (Figure 5.1), pyocyanin production and esterase production (Table 5.1). In contrast, J1532 is less proficient at producing siderophores, lipase and protease (Table 5.1). Previous work within our group has demonstrated that the mucoid strain J1532 has reduced virulence and reduced twitching motility compared to the non-mucoid J1385 (Amy Ford, unpublished results). Over the course of this chapter, the genomic basis for any phenotypic differences existing between these strains will be explored and discussed.

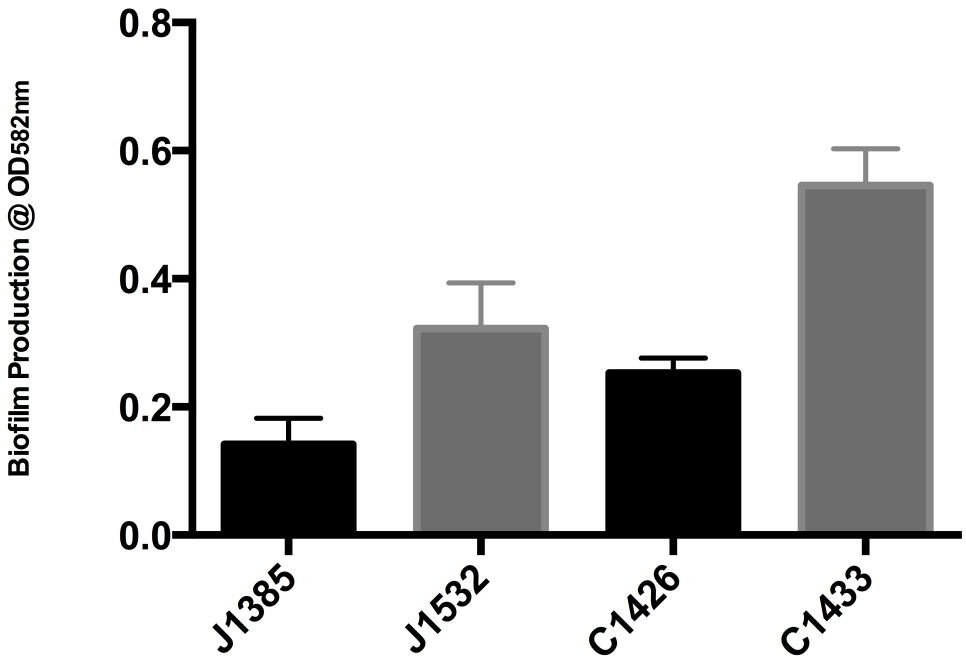


Figure 5.1 Conversion to mucoidy as a result of host adaptation will confer increased proficiency at biofilm formation. Mucoid strains from both pairs produce around twice as much biofilm as measured using a 96-well assay.

Strain	Mucoid/ Non-mucoid	Production of:				
		Esterase (nm/ml/min)	Lipase (nM/ml/min)	Pyocyanin	Protease	Siderophore
J1385	Non-mucoid	744 (±28)	285 (±7)	0.425 (±0.007)	+++	+++
J1532	Mucoid	814 (±36)	267 (±10)	0.812 (±0.006)	++	++
PA01	Non-mucoid	55 (±25)	477 (±1)	1.000 (±0.009)	++	++

Table 5.1 Table showing the non-mucoid/mucoid pairing of J1385 and J1532 and their respective efficiencies at producing a variety of enzymes/secondary metabolites. Lipase and esterase production is expressed as nm/ml/min (±SE of mean). Pyocyanin is expressed in arbitrary units normalised to the amount produced by PA01. The numbers in parenthesis are the standard errors of the mean. Protease production was determined by measuring the zones of clearance on an appropriate agar (+++ - >18mm, ++ - 12-18mm and + - <12mm). Siderophore production was determined measuring the zones of colour change (blue to orange) on a CAS containing agar (+++ - >15mm, ++ - 10-15mm and + - <10mm).

5.3 J1385 and J1532 contain a number of genomic regions not conserved throughout all strains of *P. aeruginosa*

This chapter will focus on the second non-mucoid/mucoid pairing of J1385 and J1532. Here, using a CCT (Grant, Arantes and Stothard, 2012) map showing J1385/J1532 versus 53 other *P. aeruginosa* strains, we aim to identify any regions present within the genomic maps of these strains that act as evidence of horizontal or vertical gene transfer events e.g. bacteriophage acquisition/loss. We will also attempt to identify any macro and micro genomic discrepancies that contribute to the phenotypic differences observed between J1385 and J1532 only and also between 53 other *P. aeruginosa* strains. The CCT map has highlighted potential horizontally-acquired genomic regions that are either unique to J1385/J1532 or not conserved throughout the species (Figure 5.2). Based on the subsequent analysis described in this chapter, we have named a variety of these acquired gene clusters.

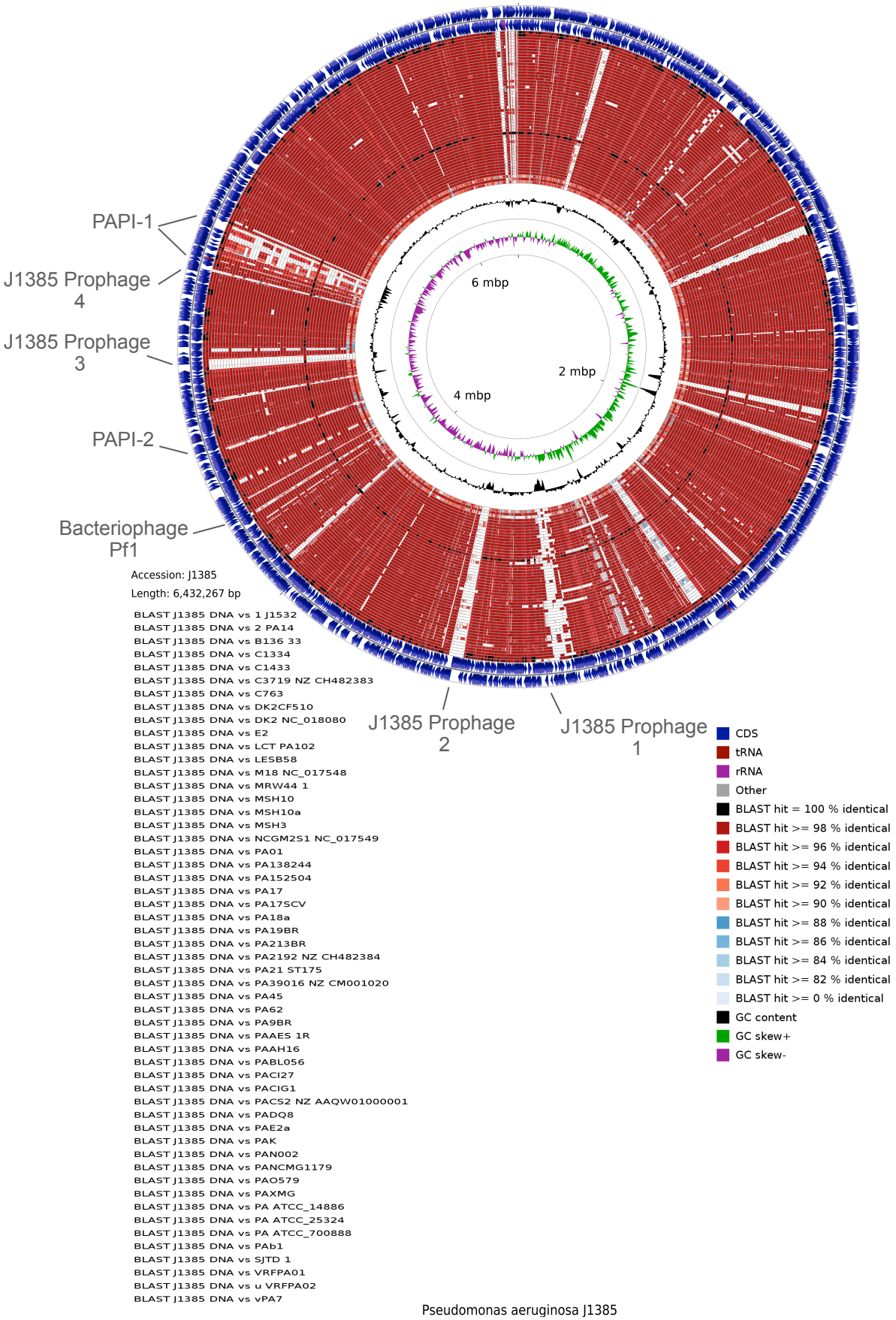


Figure 5.2 CCT map showing J1385 versus J1532 and 53 other fully sequenced *P. aeruginosa* genomes as listed in the key. BLAST hits in comparison to J1385 are colour coded by percentage match as detailed. Large areas not present in other genomes bar the reference J1385 can be seen as white columns. We have elected to name all currently identified acquired regions as based on our analysis.

5.4 Genomic regions in J1385 present evidence of previous environmental colonisation

The non-mucoid, CF isolate J1385 previously acquired a variety of genomic regions that contain genes that confer an advantage to this strain in an environmental setting such as genes encoding proteins involved in heavy metal resistance. The first of these genomic differences is a large genomic region of around 33 kbp that is, in its entirety, present in only J1385, its highly similar partner J1532 and the highly similar PA14 (Figure 5.3). This region is fully conserved between J1385 and J1532 and differs from PA14 by less than 1%. Within this region is a transposon that shows a high degree of similarity to the transposon Tn*21* (Liebert, Hall and Summers, 1999). Tn*21*, and the Tn*21* family of transposons, play an important role in the dissemination of antibiotic resistance in Gram-negative bacteria. These transposons also contain a mercury resistance (*mer*) operon including the regulatory genes *merR* and *merD*, the structural genes *merT*, *merP*, *merC* and *merA* and a gene of unknown function *merE*. The annotations of the genes present within this region in J1385 shows that this strain also posses six out of seven of these genes with only *merC* absent (Table 5.2). This does not hinder the mercury resistance conferred by this operon as this gene is nonessential for this process (Hamlett *et al*, 1992).

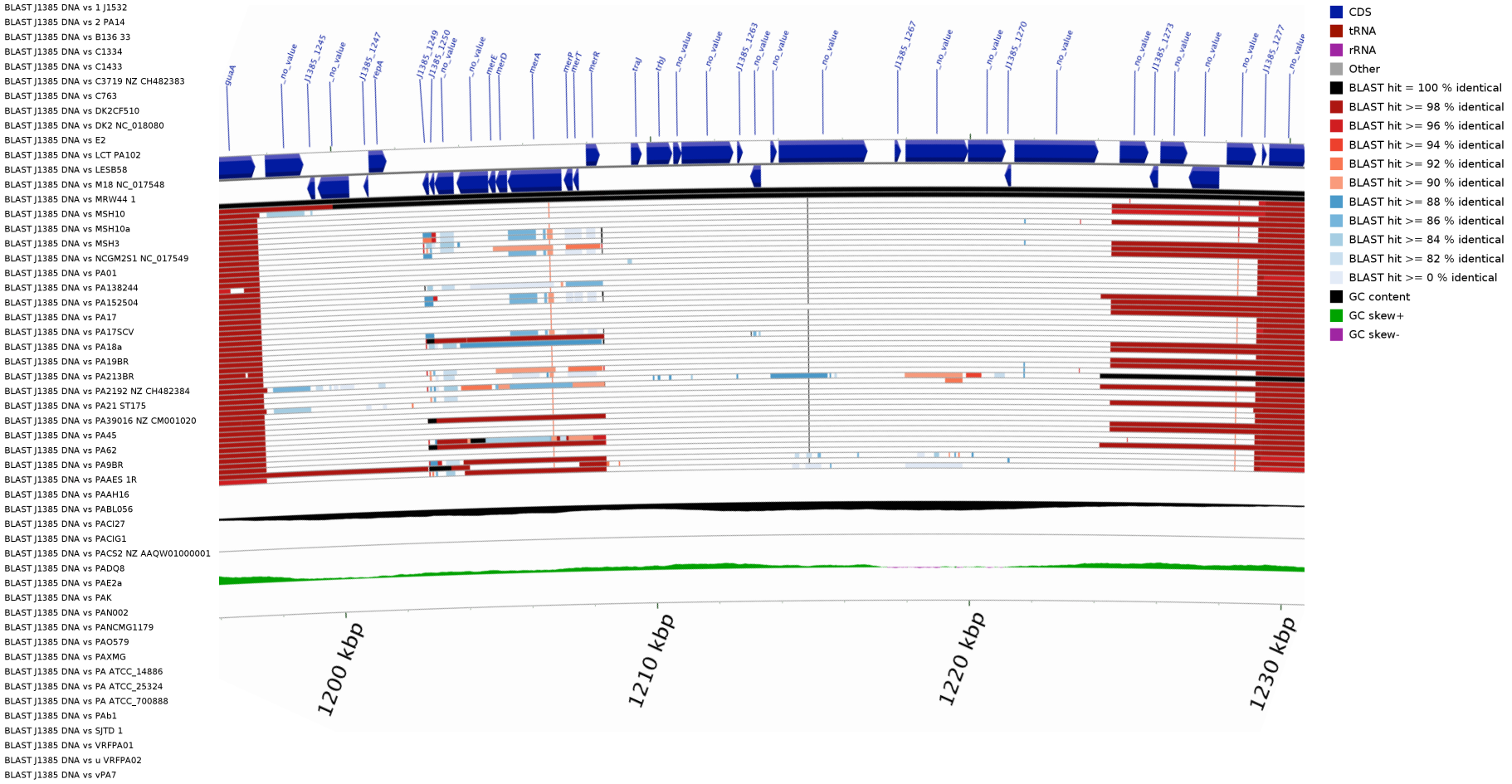


Figure 5.3 A large ~33 kbp containing a Tn21-like transposon is present in only J1385, J1532 and PA14. Based on gene annotations, this transposon may confer mercury resistance to these strains, providing a selective advantage in niches experiencing high mercury/heavy metal concentrations.

Gene ID	Putative Function	Homology (%)	Comments
J1385_1244	Integrase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	
repA	Replicative helicase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	
J1385_1249	Transposase	*P. aeruginosa* VRFPA03 (100%), *P. aeruginosa* VRFPA05 (100%), *P. aeruginosa* VRFPA01 (100%)	
J1385_1250	Transposase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	For transposon Tn21, TnpA
J1385_1251	Resolvase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	For transposon Tn21
merE	Mercury resistance protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	Mercury resistance, Tn21
merD	Transcriptional regulator	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	Mercury resistance, Tn21
merA	Mercuric reductase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	Mercury resistance, Tn21
merP	Periplasmic mercuric ion binding protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	Mercury resistance, Tn21
merT	Mercuric transport protein	*P. putida* S16 (100%), *P. aeruginosa* NCMG1179 (100%), *P. aeruginosa* PA14 (100%)	Mercury resistance, Tn21
merR	Transcriptional regulator	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	Mercury resistance, Tn21
traJ	Conjugal transfer protein	*P. aeruginosa* PA14 (98%), *P. aeruginosa* BWHPS027 (98%), *P. aeruginosa* BL16 (98%)	
trbJ	Conjugal transfer protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	
J1385_1261	Entry/exculsion protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	*trbK*
J1385_1262	Mating pair formation protein	*P. aeruginosa PA14 (100%), P. aeruginosa BWHPS027 (100%), P. aeruginosa BL16 (100%)*	*trbL*, P-type
J1385_1265	Phage transcriptional regulator	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	
J1385_1266	Type II restiction enzyme	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	Methylase subunit
J1385_1272	Oxidoreductase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	
J1385_1274	MFS transporter	*P. aeruginosa* PA14 (100%), *P. aeruginosa* CI27 (100%), *P. aeruginosa* (100%)	
J1385_1276	Oxidoreductase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* CI27 (100%), *P. aeruginosa* (100%)	

Table 5.2 Gene id, putative function and percentage homology as predicted by BLAST analysis of genes present within the ~33 kbp region possessed by J1385, J1532 and PA14. The highlighted cells show those exhibiting a high degree of similarity to the Tn*21* transposon (Liebert, Hall and Summers, 1999). All other genes within this region but not included here are of hypothetical function and show a high degree of homology to several *P. aeruginosa* strains.

A second large genomic difference within J1385 that is suggestive of previous environmental colonisation is the presence of a large ~42 kbp plasmid at 2570 kbp. As indicated by BLAST analysis (Table 5.3), this region contains many genes required for conjugal transfer. However, it does not appear to contain the entire complement outlined by Haase *et al* (1995). This region consists primarily of genes encoding proteins of hypothetical function. However, the functions of the proteins encoded by several genes (J1385_2651, J1385_2652 and J1385_2653) warrant further discussion. These genes encode proteins that function as cation efflux proteins, in other words, integral membrane proteins involved in providing a resistance to divalent metal ions, or perhaps antibiotics, by pumping such compounds out of the cell (Xiong and Jayaswal, 1998). BLAST analysis indicates that these three proteins have a high degree of homology to the cation-proton antiporter CzcABC that confers cobalt, zinc and cadmium resistance to *Alcaligenes eutrophus* and *E. coli* (Nies *et al*, 1995).

These previous two examples of large genomic regions that confer resistance to a variety of metals such as mercury, zinc, cobalt and cadmium is supplemented with a third genomic region that confers resistance to copper as will be shown later in this chapter (section 5.12). Resistance to metals such as these will provide no benefit to J1385 within the CF lung, a niche lacking toxic concentrations of such metals. Possession of these resistance mechanisms act as evidence to support the hypothesis that J1385 previously colonised an environmental niche where they may encounter high and potentially toxic concentrations of such metals (Cánovas, Cases and de Lorenzo, 2003). It is important to note that a strong correlation between heavy metal resistance and antibiotic resistance exists. Antibiotic resistance is increased in bacteria experiencing high environmental heavy metal concentrations (Knapp *et al*, 2011). Possession of these heavy metal resistance genes may therefore contribute to antibiotic resistance in these strains.

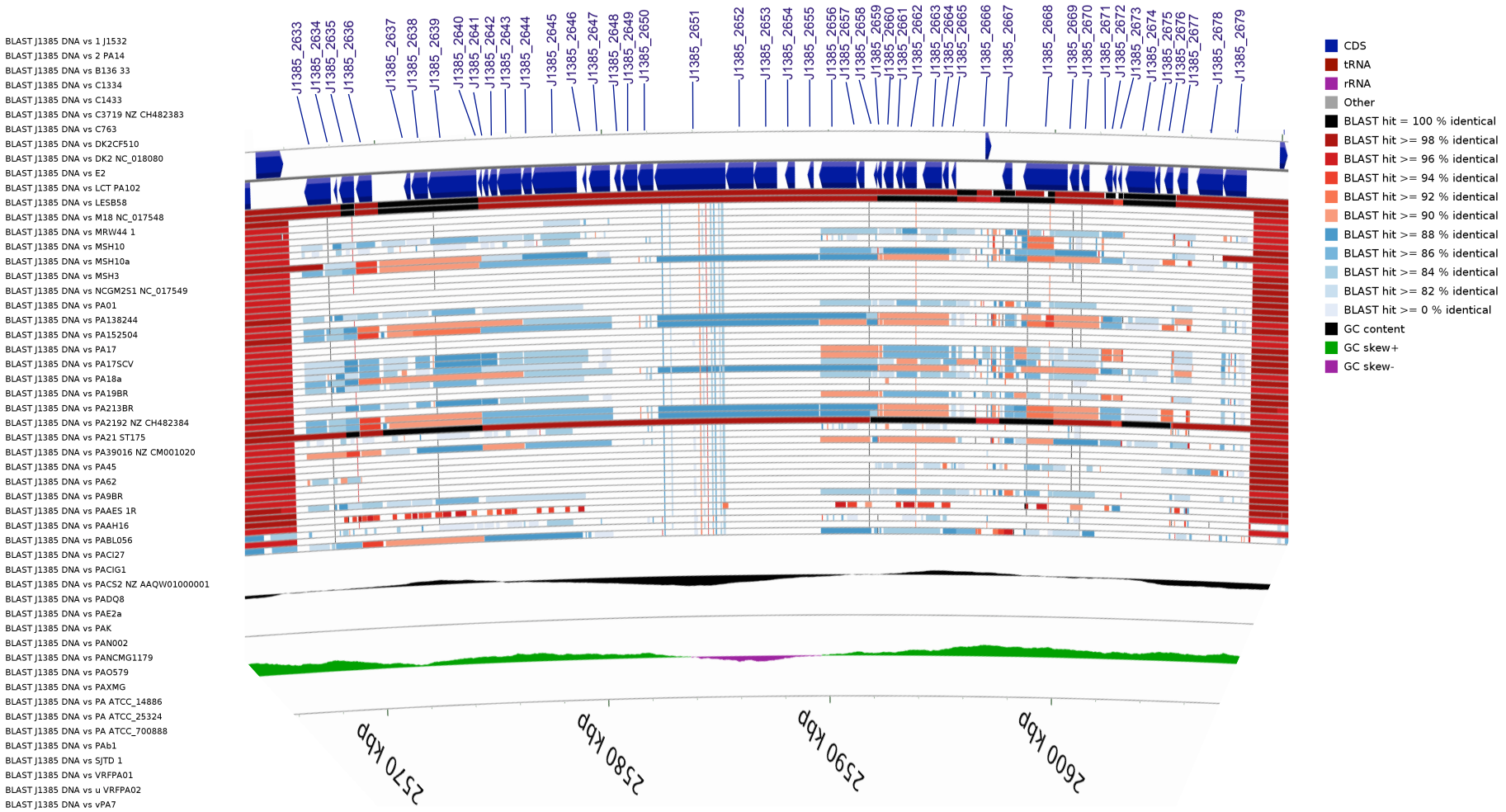


Figure 5.4 A large ~42 kbp plasmid containing a *CzcABC*-type operon putatively encoding a cation-proton antiporter potentially confers resistance to cobalt, zinc and cadmium in J1385, J1532 and PA14

Gene ID	Putative Function	Homology (%)	Comments
J1385_2633	Conjugative transfer protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	TrbL
J1385_2634	Conjugative transfer protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	TrbG, P-type
J1385_2635			
J1385_2636	Conjugative transfer protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	TrbF
J1385_2638	Conjugative transfer protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	TrbJ-like, P-type
J1385_2639	Conjugal transfer ATPase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	TrbE
J1385_2640			
J1385_2641	Conjugal transfer protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	TrbD
J1385_2642	Conjugative transfer protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	TrbC
J1385_2643	Conjugative transfer protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	TrbB, P-type
J1385_2644	Transcriptional regulator	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	CopG family
J1385_2645	Conjugal transfer protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	TraG
J1385_2651	Cation efflux protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	CzcA
J1385_2652	Cation efflux protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	CzcB
J1385_2653	Cation efflux protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	CzcB
J1385_2658	Conjugal transfer signal peptidase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	TraF
J1385_2659			
J1385_2662	Plasmid partitioning protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	ParA
J1385_2663	Replication initiator and transcriptional repressor protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	RepA
J1385_2668	Plasmid stabilisation protein	*P. aeruginosa* PA14 (95%), *P. aeruginosa* BWHPS027 (95%), *P. aeruginosa* BL16 (95%)	
J1385_2674	MFS transporter	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	NrbE-like
J1385_2677	DNA repair protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	RadC-like
J1385_2679	Integrase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	

Table 5.3 Table showing the gene id, putative function and percentage homology as predicted by BLAST analysis of genes present within the ~42 kbp region possessed by J1385, J1532, PA14 and one other *P. aeruginosa* strain included in this study. All other genes within this region but not included here are of hypothetical function and show a high degree of homology to several *P. aeruginosa* strains.

5.5 The hypervariable pyoverdin locus of J1385 contains two rare regions present in very few *P. aeruginosa* genomes

The pyoverdin locus is the “most divergent alignable” locus present in the genome of *P. aeruginosa* strains (Spencer *et al*, 2003). This locus spans over 50 kbp of genome and the codon directionality and altered GC-content suggests a history of horizontal transfer events (Smith *et al*, 2005). This is supported by a ~10 kbp region present within the pyoverdin locus of J1385 but only 14 other *P. aeruginosa* strains included in this analysis (Figure 5.5). This region consists of seven genes (J1385_2865 to J1385_2871) all of which, according to BLAST analysis, encode for CRISPR-associated (cas) proteins/enzymes (Table 5.4). These *cas* genes are involved in the mediation and implementation of the antiviral response to counteract infection by phages and conjugation of plasmids (Brouns *et al*, 2008). Possession of these genes suggest that previous phage infection or plasmid integration has occurred within this region and supports the hypothesis by Smith *et al* (2005).

This pyoverdin locus within J1385 also possesses a second, larger region that is present in 21 of the *P. aeruginosa* genomes in the present analysis (Figure 5.5). This ~21 kbp region contains several genes whose protein products are involved in pyoverdin biosynthesis including *pvdJ*, *pvdD* and *pvdE* (Table 5.4). Also within this region is the ferripyoverdine receptor *fpvA*. It is these genes involved in peptide synthesis and transport, along with *pvdI*, that are the most divergent amongst the pyoverdin genes (Smith *et al*, 2005). The presence of these genes presumably results in enhanced iron uptake due to both increased siderophore biosynthesis and increased siderophore uptake by the cognate receptor. This might provide a selective advantage to J1385 within the CF lung, allowing it to compete against any co-colonising bacterium in iron sequestration. However, as Dingemans *et al* (2014) have shown, adaption to the CF lung by *P. aeruginosa* is also associated with deletion of similar TonB-dependent receptor genes. It is therefore not possible to attribute the changes observed here to increased siderophore uptake without further work.

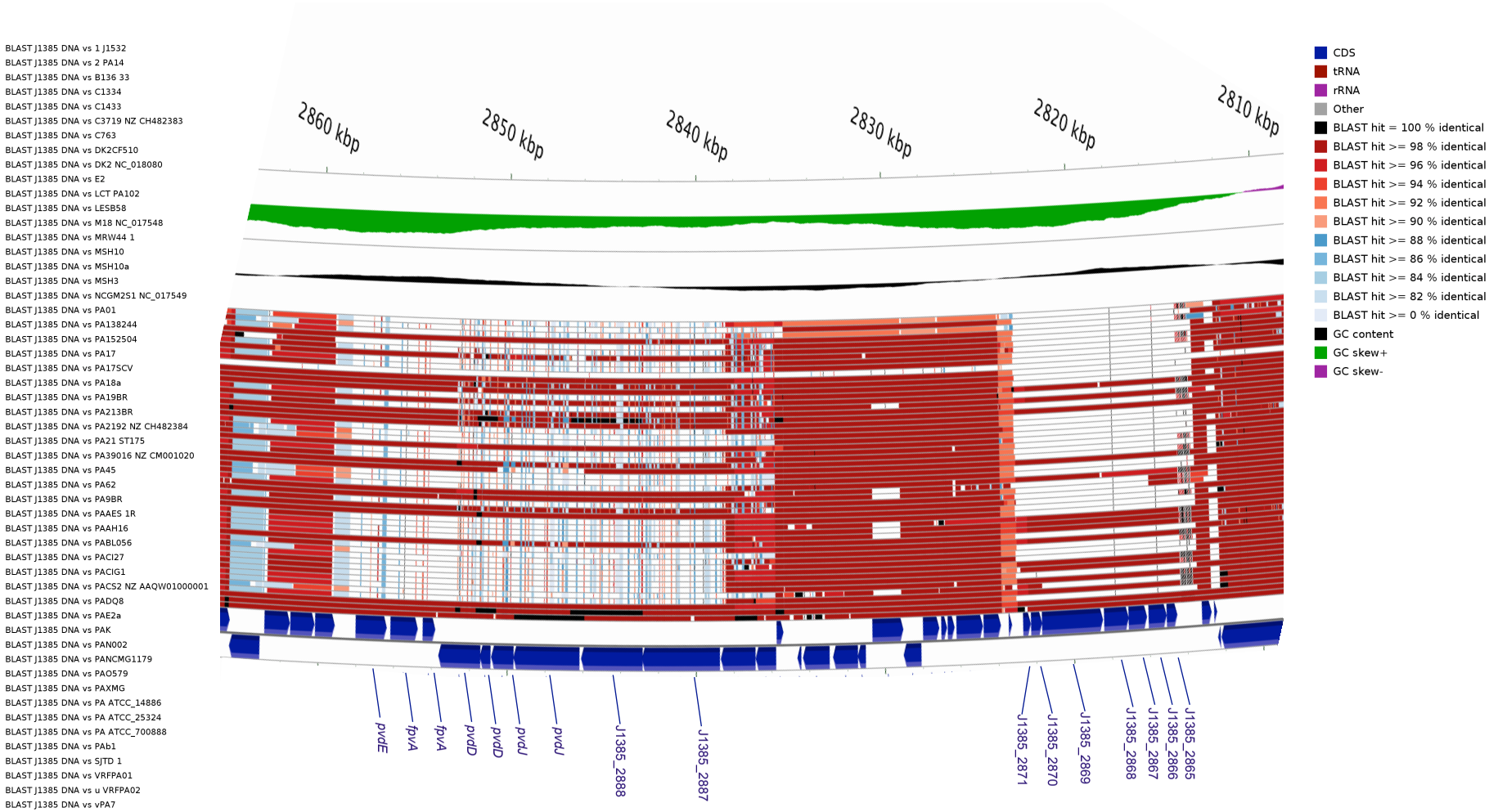


Figure 5.5 The pyoverdin cluster in J1385 possess two genomic regions not present in many other *P. aeruginosa* genomes. The first of these contains several CRISPR-associated (*cas*) genes involved in conferring bacterial resistance to plasmid/phage infection. The second regions contains several genes involved in pyoverdin synthesis and transport previously shown to be the most divergent genes in the *P. aeruginosa* genome.

Gene ID	Putative Function	Homology (%)	Comments
J1385_2865	CRISPR-associated protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	
J1385_2866	CRISPR-associated protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	
J1385_2867	CRISPR-associated protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	
J1385_2868	CRISPR-associated protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	
J1385_2869	CRISPR-associated helicase	*P. aeruginosa* PA14 (99%), *P. aeruginosa* BWHPS027 (99%), *P. aeruginosa* BL16 (99%)	
J1385_2870	CRISPR-associated endonuclease	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	
J1385_2871	CRISPR-associated endonuclease	*P. aeruginosa* PA14 (99%), *P. aeruginosa* BWHPS027 (99%), *P. aeruginosa* BL16 (99%)	
J1385_2887	Peptide synthase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	Pyoverdin biosynthesis
J1385_2888	Hypothetical protein	*P. aeruginosa* PA14 (99%), *P. aeruginosa* BWHPS027 (99%), *P. aeruginosa* BL16 (99%)	
pvdJ	Partial protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	Pyoverdin biosynthesis
pvdJ	Partial protein		
pvdD	Partial protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	Pyoverdin biosynthesis
pvdD	Partial protein		
fpvA	Ferripyoverdine receptor	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	Partial
fpvA			Partial
pvdE	Cyclic peptide transporter	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	Pyoverdin biosynthesis

Table 5.4 Table showing the gene id, putative function and percentage homology as predicted by BLAST analysis of genes present within two genomic regions present within the pyoverdin locus of J1385 but not all other *P, aeruginosa* genomes included in this study.

5.6 J1385 possess another extra set of siderophore biosynthesis and transport genes

J1385 possesses a large segment of DNA spanning ~34 kbp of genome and compromising of 22 coding sequences that is present in only 7.5% of strains included in this analysis (Figure 5.6). It is predicted that this region is in fact a bacteriophage, it exhibits characteristics suggestive of a bacteriohage including a reduced GC content, change in direction of flanking genes and the presence of a phage integrase at the 5'-end of the sequence (Table 5.5). We have subsequently named this bacteriophage as J1385 Prophage 3. BLAST analysis reveals that over half of the genes within this bacteriophage are involved with either siderophore biosynthesis or transport. Increased siderophore production will allow more solubilisation of the much sought after ferric ion. Once solubilised the increased expression of siderophore receptors and transporters are likely to allow s greater uptake of iron by the bacterium. This confers J1385 with a competitive advantage amongst any co-colonising bacterium, increasing its own uptake of iron and limiting the source of iron for other competing bacterial strains/species.

Another gene present within this bacteriophage is J1385_4794 that encodes for a thioesterase (Table 5.5). Thioesterases can optimise the non-ribosomal synthesis of the siderophore pyochelin (Reimmann *et al*, 2004). Therefore, the presence of this gene will further increase iron uptake by J1385, enhancing the selective advantage of increased iron acquisition conferred by this bacteriophage. Aside from this, as the name suggests, thioesterase enzymes exhibit esterase activity. This may contribute to the 14-15 fold increase in esterase activity previously observed in J1385 and J1532 compared to the reference strain PA01 (Table 5.1).

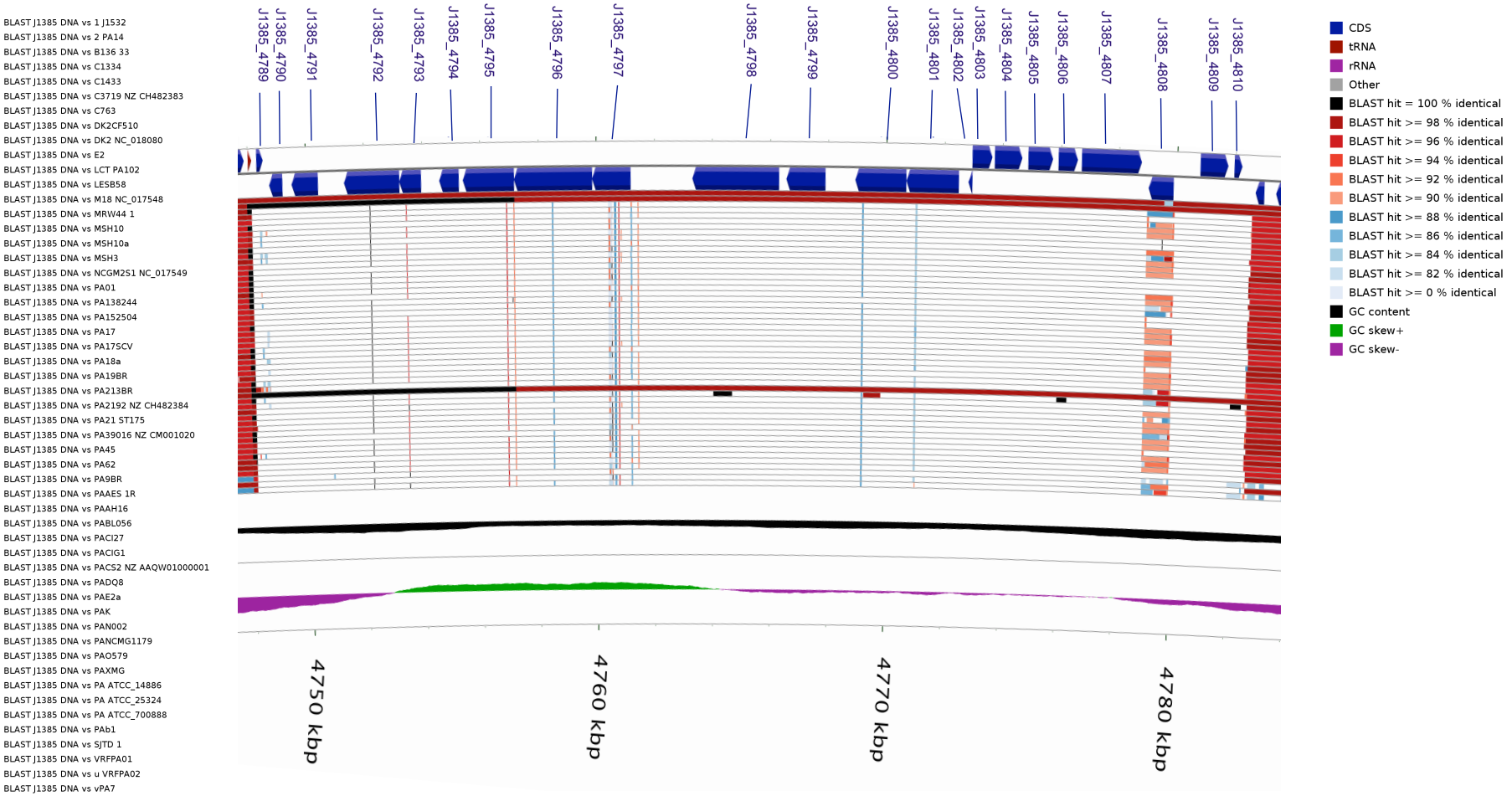


Figure 5.6 The large, ~34 kbp J1385 Prophage 3 possessed by J1385 and only three other *P. aeruginosa* strains included in this study contains 22 genes, over half of which encode proteins involved in siderophore production or transport (Table 5.5).

Gene ID	Putative Function	Homology (%)	Comments
J1385_4789	Integrase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	
J1385_4790	Hypothetical protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	
J1385_4791	Transcriptional regulator	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	AraC family
J1385_4792	Saccharopine dehydrogenase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	
J1385_4793	Saccharopine dehydrogenase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	
J1385_4794	Thioesterase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	
J1385_4795	Peptide synthetase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	Siderophore biosynthesis
J1385_4796	Peptide synthetase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	Siderophore biosynthesis
J1385_4797	Peptide synthetase	*P. aeruginosa* PA14 (99%), *P. aeruginosa* BWHPS027 (99%), *P. aeruginosa* BL16 (99%)	Siderophore biosynthesis
J1385_4798	Adenylate synthase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	Siderophore biosynthesis
J1385_4799	TonB-dependent siderophore receptor	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	
J1385_4800	ABC transporter	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	Fe3+ siderophore transport
J1385_4801	ABC transporter	P. *aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	Fe3+ siderophore transport
J1385_4803	ABC transporter periplasmic component	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	Fe3+ siderophore transport
J1385_4804	ABC transporter permease	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	Fe3+ siderophore transport
J1385_4805	ABC transporter permease	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	Fe3+ siderophore transport
J1385_4806	Ferric enterobactin transporter	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	Fe3+ siderophore transport
J1385_4807	TonB-dependent siderophore receptor	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	
J1385_4808	Transposase	*P. aeruginosa* PA14 (99%), *P. aeruginosa* BWHPS027 (99%), *P. aeruginosa* BL16 (99%)	
J1385_4809	Hypothetical protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BWHPS027 (100%), *P. aeruginosa* BL16 (100%)	
J1385_4810	Hypothetical protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* CI27 (100%), *P. aeruginosa* (100%)	

Table 5.5 Table showing the gene id, putative function and percentage homology as predicted by BLAST analysis of genes present within the ~34 kbp J1385 Prophage 3 acquired by J1385 and only three other *P. aeruginosa* strains included in this study. BLAST analysis reveals that over half of the genes within this bacteriophage are involved in either siderophore biosynthesis or transport.

5.7 J1385 possess both PAPI-1 and PAPI-2 pathogenicity islands

PA14 is a highly virulent strain of *P. aeruginosa* that can infect many hosts infecting invertebrate and vertebrate animals as well as plant species (Kukavica-Ibrulj *et al*, 2008). The increased virulence of this strain is attributed to two pathogenicity islands, PAPI-1 and PAPI-2 (He *et al*, 2004). PAPI-1 spans 108 kbp and carries 115 open reading frames (ORFs) and PAPI-2 spans 11 kbp and carries 14 ORFs. More recently, it has been shown that these islands work both individually and synergistically in mediating the virulence of PA14 (Harrison *et al*, 2010). Both of these islands were acquired, in their entirety, by J1385 and subsequently retained by J1532 (Figure 5.7).

As mentioned, the pathogenicity island PAPI-1 possesses 115 coding regions including both animal and plant virulence genes (He *et al*, 2004). Over 80% of the DNA sequence is, based on BLASTn analysis, unique to this island and the large majority of the genes present code for proteins of unknown function. However, it codes for a variety of virulence factors including a full complement of type IVb pili genes (Figure 5.7). It is likely that possession of an extra set of type IVb pili may result in increased motility of J1385 conferring a selective advantage in its attempt at colonising the CF lung. It is important to note however, that hyperpilliated mutants have been shown to be defective in motility as a result of their increased adherence (Deziel *et al*, 2001). Therefore, possession of this extra set of pili genes may actually result in reduced motility. Similarly, the increased adherence attributed to type IVb pili may contribute to biofilm formation at the attachment stage, improving the infectivity of the bacterium (Barken *et al*, 2008). A further selective advantage may be conferred due to the possession of the *cupD1* to *cupD5* fimbrial biosynthesis gene cluster and *pvrR*, a two-component response regulator-encoding gene, both of which are involved in biofilm formation. The latter of these acquired virulence factors is also involved in antibiotic resistance (Harrison *et al*, 2010). Both increased biofilm formation and antibiotic resistance most likely aid J1385 in establishing a successful

chronic infection within the lung via protection from host and drug interventions.

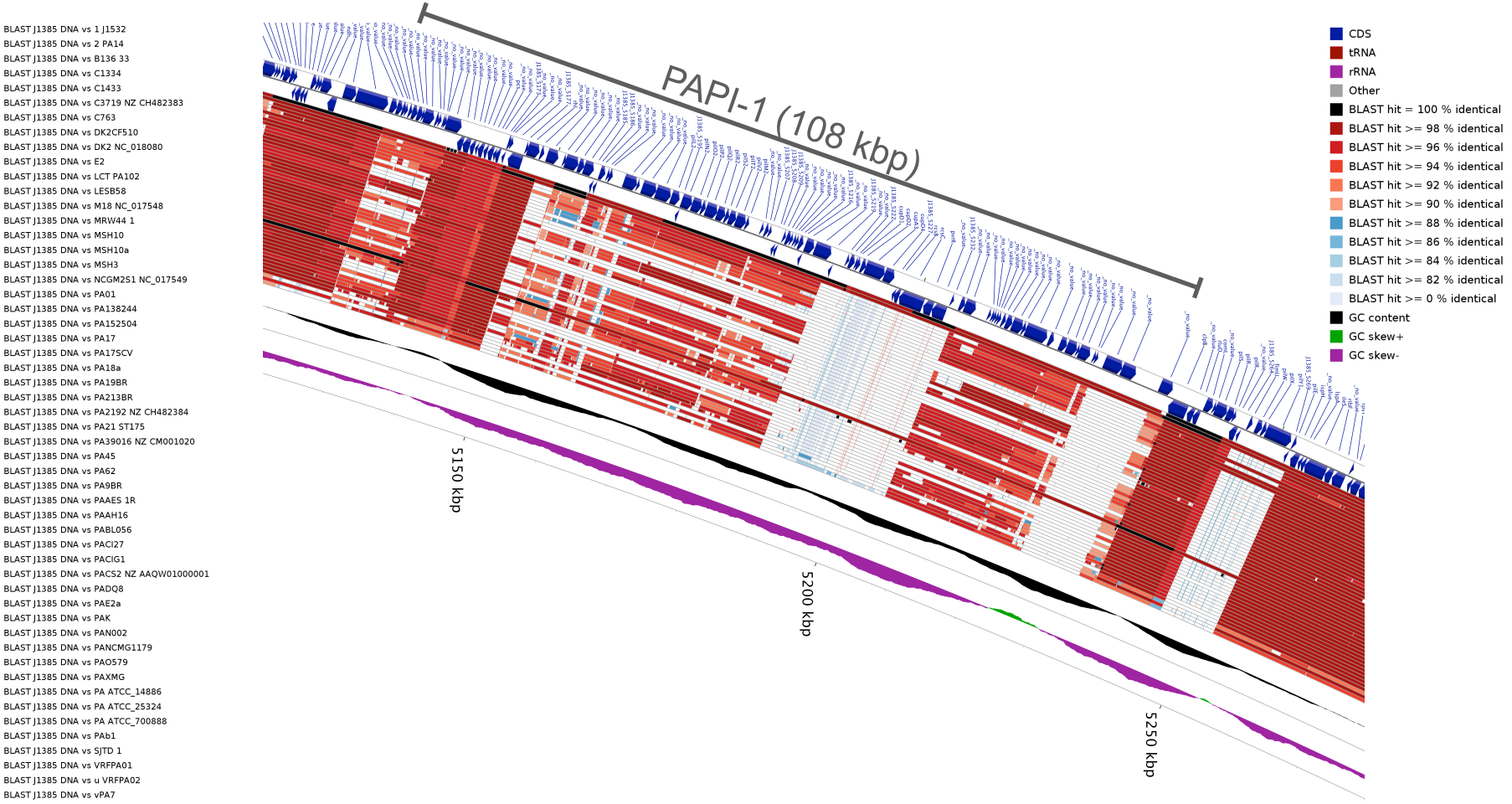


Figure 5.7 J1385 has acquired the pathogenicity island PAPI-1 which includes a variety of virulence genes including a type IVb pili locus, *cupD1-cupD5* genes and the *pvrR* two component regulator gene (Harrison *et al*, 2010)

J1385 has also acquired the second pathogenicity island PAPI-2 (Figure 5.8). The literature indicates that this island codes for 14 proteins spanning approximately 11 kbp (He *et al*, 2004). However, as can be seen in Figure 5.8, PAPI-2 in J1385 only possess four genes compared to the 14 visualised for PA14. This suggests that our automated annotation of this island in J1385 is entirely and xBASE has underestimated the number of genes present.

That aside, encoded within this island is the potent cytotoxin ExoU and its chaperone protein SpcU. ExoU is one of the many effector proteins that is part of the type III secretion systems of *P. aeruginosa* that delivers such toxins into eukaryotic cells (Kurahashi *et al*, 1999). Within eukaryotic cells, ExoU causes rapid cell lysis as an effect of its potent phospholipase A_2 activity (Philips *et al*, 2003). ExoU is capable of limiting the host immune response via infiltration and killing of host immune factors such as phagocytes (Diaz and Hauser, 2010). Therefore, possession of this ExoU cytoxin may confer a selective advantage to such strains, assisting with colonisation with minimal intervention from the host immune system. However, the selective advantage conferred by ExoU does not end here. This cytotoxin is also toxic to eukaryotic cells other than mammalian cells including yeasts (Rabin and Hauser, 2003). This will allow ExoU-expressing *P. aeruginosa* isolates to overcome the co-colonisation events commonly observed in lung infections.

Therefore, it seems likely that J1385 benefits from its ability to express the potent cyotoxin ExoU within the CF lung. ExoU may cause reduces host immune intervention via inducing cell lysis of immune effectors such as phagocytes and potentially overcome and kill co-colonising eukaryotic infectors such as *Candida albicans* freeing up essential carbon sources. Both of these functions of ExoU may allow successful colonisation and infection by J1385.

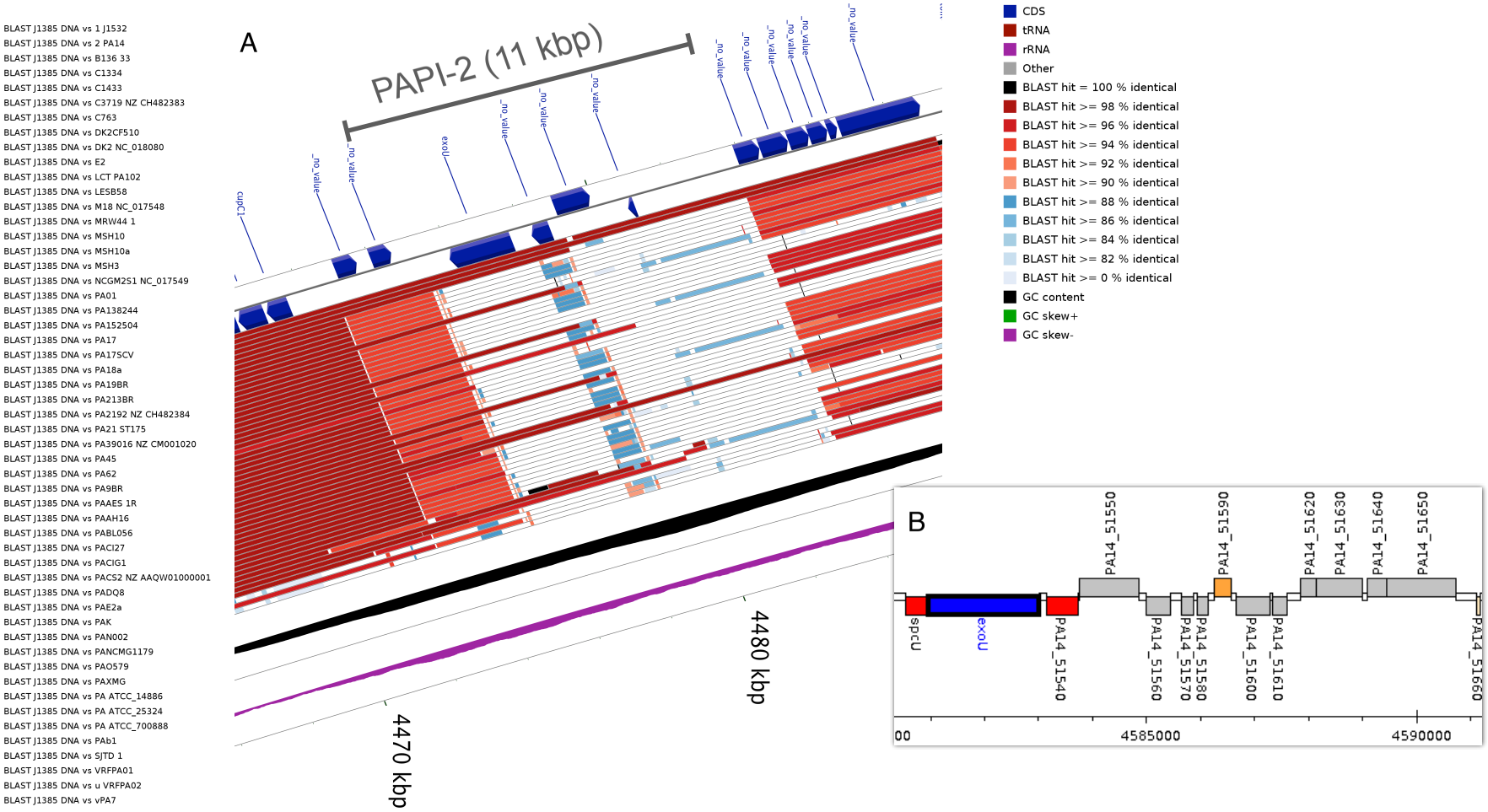


Figure 5.8 **A** J1385 has acquired the pathogenicity island PAPI-2 which encodes the potent cytotoxin ExoU. xBASE has incorrectly predicted that this island contains only four genes. **B** The same island in PA14 shows that PAPI-2 actually encodes for fourteen genes which is supported by He *et al* (2004).

5.8 J1532 has lost four large regions of genomic DNA during its host adaptation from J1385

Throughout this chapter so far, we have focused on large scale genomic differences existing between J1385 and the remainder of the strains included in the BLAST analysis. However, as each of these regions are also present in its mucoid descendant J1532 and therefore such analyses have investigated the general diversity of both strains amongst the *P. aeruginosa* species. This following section will instead focus on the phenomenon that is host adaptation within the CF lung, investigating the genomic basis for some of the phenotypic anomalies arising during the switch between the non-mucoid J1385 and the mucoid J1532.

Looking closely at the CCT map of J1385 versus 54 other *P. aeruginosa* genomes, there are four genomic regions present within J1385 that are not present in J1532. These regions are found at the following locations - 3080 kbp (Figure 5.12), 3395 kbp (Figure 5.11), 4245 kbp (Figure 5.9) and 5070 kbp (Figure 5.10). The later two are not well characterised and no clear role in altering phenotype can be detected. Both show high degree of homology to bacteriophages (Tables 5.6 and 5.7, respectively) and contain no genes that encode any proteins of characterised function. Phage can be linked to DNA inversions and phenotypic variation (Mooij *et al*, 2007) and as a result neither of these regions can be dismissed as innocuous in the change in phenotype exhibited between J1385 and J1532. The first of these, at 4245 kbp, spans approximately 10.8 kbp of genome (Figure 5.9) and shows a high degree of homology to the filamentous phage Pf1 (Hill *et al*, 1991). Additional investigation into the degree of homology between the sequence of this bacteriophage removed by J1532 and Pf1 was performed via a BLASTn. This analysis demonstrated that over 99% of the DNA sequences of each was identical (Figure 5.9). Phage Pf1 has not been shown to carry any genes that encode proteins of known functions. However, genes are amongst the most highly upregulated genes during biofilm development and as a result

may play a role during such an event (Webb *et al*, 2003 and Webb, Lau and Kjelleberg, 2004).

The second of these bacteriophages missing from J1532, spanning ~8 kbp of genome, contains seven genes (Figure 5.10), six of which encode proteins of unknown function (Table 5.6). This bacteriophage, J1385 Prophage 4, shows some degree of homology to a variety of sequenced bacteriophages, but it is unclear which of these this is derived from. Based on the putative functions of the genes within both of these bacteriophage, it is predicted that their loss would have no bearing on the overall change in phenotype exhibited by J1532. However, without further investigation this hypothesis cannot be definitively proven.

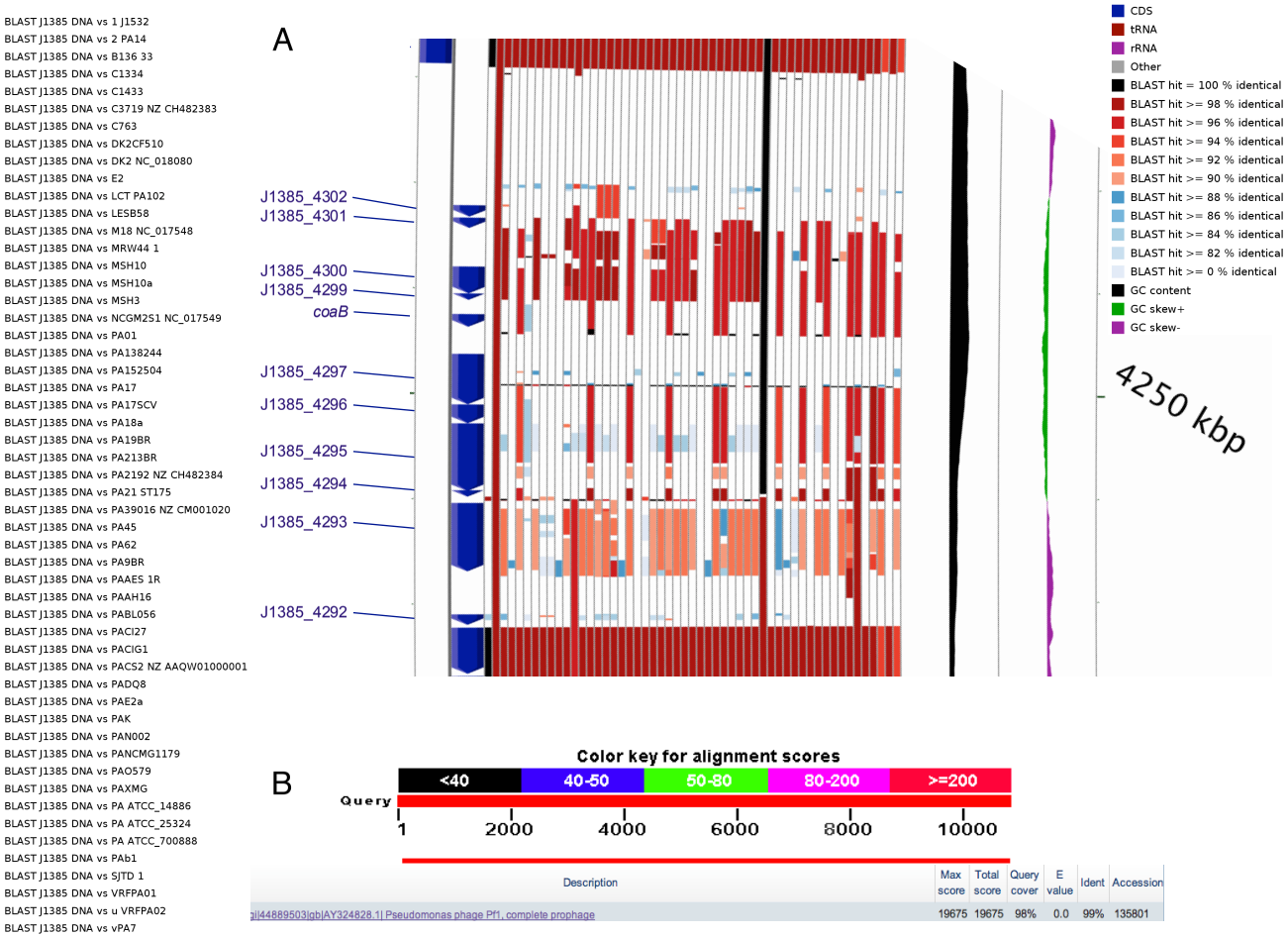


Figure 5.9 **A** J1532 has lost a bacteriophage spanning 10.8 kbp that exhibits a high degree of homology to the filamentous phage Pf1. **B** This is supported using a BLASTn analysis that demonstrates that the sequence of the bacteriophage lost by J1532 is over 99% identical to the DNA sequence of phage Pf1.

Gene ID	Putative Function	Homology (%)	Comments
J1385_4292	Integrase	*P. aeruginosa* PA14 (100%), *Pseudomonas* phage Pf1 (100%)	Phage Pf1
J1385_4293	Hypothetical protein	*P. aeruginosa* PA14 (100%), *Pseudomonas* phage Pf1 (100%)	Phage Pf1
J1385_4294	Hypothetical protein	*P. aeruginosa* PA14 (100%), *Pseudomonas* phage Pf1 (100%)	Phage Pf1
J1385_4295	Hypothetical protein	*P. aeruginosa* PA14 (100%), *Pseudomonas* phage Pf1 (100%)	Phage Pf1
J1385_4296	Head virion protein	*P. aeruginosa* PA14 (100%), *Pseudomonas* phage Pf1 (100%)	Phage Pf1
J1385_4297	Coat protein A	*P. aeruginosa* PA14 (100%), *Pseudomonas* phage Pf1 (100%)	Phage Pf1
coaB	Coat protein B	*P. aeruginosa* PA14 (100%), *Pseudomonas* phage Pf1 (100%)	Phage Pf1
J1385_4299	Hypothetical protein	*P. aeruginosa* PA14 (100%), *Pseudomonas* phage Pf1 (100%)	Phage Pf1
J1385_4300	ssDNA binding protein	*P. aeruginosa* PA14 (100%), *Pseudomonas* phage Pf1 (100%)	Phage Pf1
J1385_4301	Hypothetical protein	*P. aeruginosa* PA14 (100%), *Pseudomonas* phage Pf1 (100%)	Phage Pf1
J1385_4302	Hypothetical protein	*P. aeruginosa* PA14 (100%), *Pseudomonas* phage Pf1 (100%)	Phage Pf1

Table 5.5 Table showing the gene id, putative function and percentage homology as predicted by BLAST analysis of genes present within a 10.8 kbp bacteriophage lost by J1532. Genes present within this phage show a high degree of homology to filamentous phage Pf1 genes.

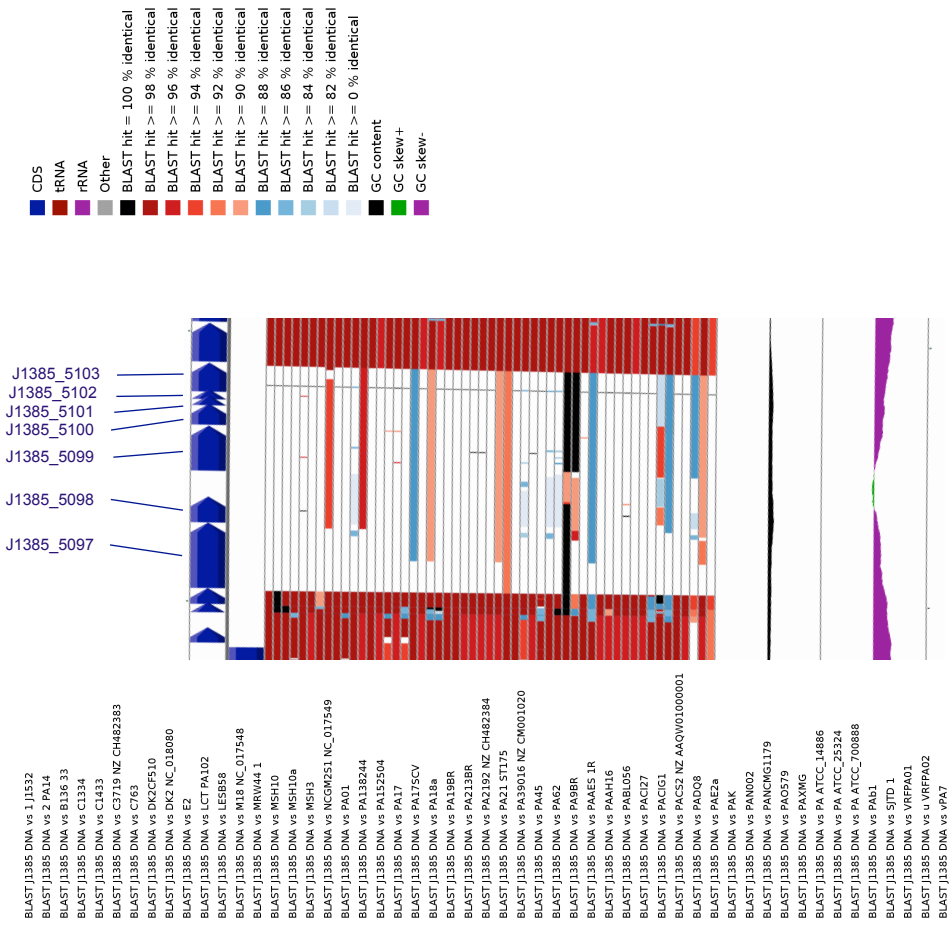


Figure 5.10 J1532 has lost a bacteriophage spanning 8 kbp, named J1385 Prophage 4.

Figure 5.10 CCT map showing J1385 versus J1532 and 53 other fully sequenced *P. aeruginosa* genomes as listed in the key in Figure 5.2. J1532 has lost a bacteriophage spanning 8 kbp, named J1385 Prophage 4.

Gene ID	Putative Function	Homology (%)	Comments
J1385_5097	Phage tail component	*Pseudomonas* phage MP42 (99%), *Pseudomonas* phage JBD30 (100%)	
J1385_5098	Hypothetical protein	*Pseudomonas* phage PA1/KOR/2010 (99%), *Pseudomonas* phage MP29 (99%)	
J1385_5099	Hypothetical protein	*Pseudomonas* phage JBD5 (100%), *P. aeruginosa* PADK2_CF510 (100%), *P. aeruginosa* LESB58 (100%)	
J1385_5100	Hypothetical protein	*P. aeruginosa* CI27 (100%), *P. aeruginosa* CIG1 (100%), *P. aeruginosa* 19660 (99%)	
J1385_5101	Hypothetical protein	*Pseudomonas* phage MP29 (100%), *Pseudomonas* phage JBD5 (100%), *Pseudomonas* phage JBD24 (100%)	
J1385_5102	Hypothetical protein	*Pseudomonas* phage LPB1 (100%), *P. aeruginosa* BL12 (100%), *P. aeruginosa* VRFPA04 (98%)	
J1385_5103	Hypothetical protein	*P. aeruginosa* CI27 (98%), *P. aeruginosa* CIG1 (98%), *Pseudomonas* phage D3112 (97%)	

Table 5.6 Table showing the gene id, putative function and percentage homology as predicted by BLAST analysis of genes present within the 8 kbp J1385 Prophage 4 lost by J1532. Genes present within this phage show a high degree of homology to a variety of *Pseudomonas* phage.

5.8.1 The loss of a gene encoding a putative homologue of the transcriptional regulator AlgR within a large bacteriophage may contribute to the increased alginate biosynthesis observed in J1532

J1532 no longer possesses a third bacteriophage at position 3395 kbp (Figure 5.11). This bacteriophage, named J1385 Prophage 2, spans 48 kbp and contains 57 ORFs, with the majority of products of which are of hypothetical function (Table 5.7) and several involved in phage formation including tail proteins and lytic and replicative enzymes. Perhaps the only gene to stand out from the 57 genes within this bacteriophage is a homologue of *algR*. This gene encodes the transcriptional regulator AlgR that is involved in the regulation of a variety of different systems including alginate biosynthesis, type IV pili functioning and other virulence genes (Lizewski *et al*, 2004). AlgR induces activation of the *fimU-pilVWXY1Y2* operon, the operon that controls twitching motility in *P. aeruginosa*. It is suggested that the loss of this putative second copy of *algR* from J1532 within this bacteriophage contributes to the reduced twitching observed in J1532 compared to J1385 (Amy Ford, unpublished results) as a result of a lack of activation of the aforementioned operon (Lizewski *et al*, 2004). AlgR is also essential for the virulence of *P. aeruginosa* using a murine model of septicaemia (Lizewski *et al*, 2003). The switch to mucoidy within the CF lung is associated with a reduction in virulence and this is supported by results using a *G. mellonella* infection model wherein J1385 exhibited increased virulence compared to J1532 (Figure 5.15 and Amy Ford, unpublished results). It is hypothesised, therefore, that since AlgR is essential for virulence in *P. aeruginosa*, the loss of this regulator may contribute to the reduced virulence observed in J1532.

AlgR is involved in alginate biosynthesis; it is required for alginate production via positive regulation of the *algD* promoter. As has been described, J1532 has increased alginate production that is responsible for the appearance of the mucoid phenotype. The loss of *algR* within this bacteriophage is contradictory to this fact. Therefore, it is expected that there are further

genomic differences present between the two strains that result in the increased production of alginate.

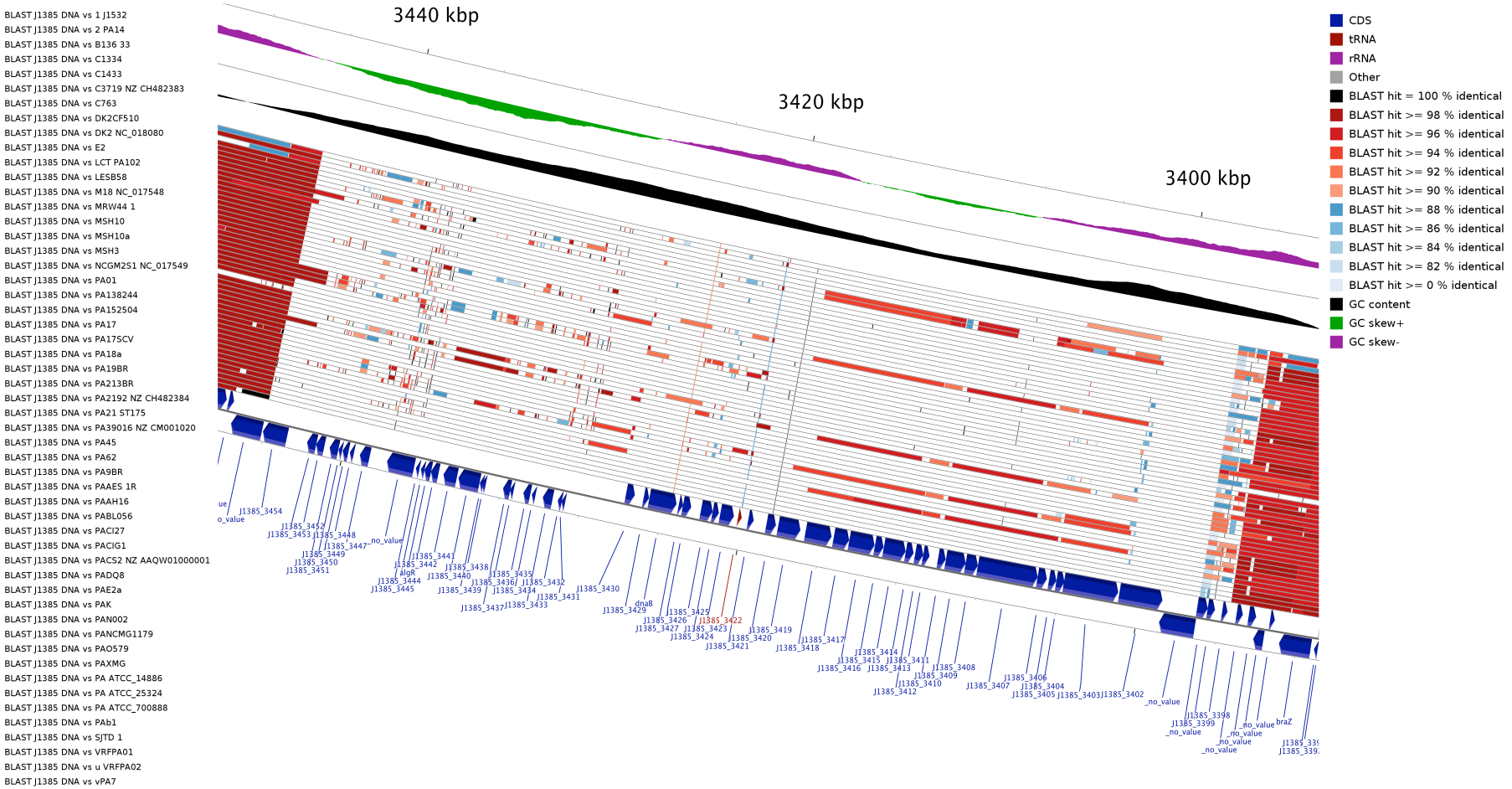


Figure 5.11 J1532 has lost the J1385 Prophage 2 spanning 48 kbp and encoding 57 proteins. The majority of these are uncharacterised function but one gene in particular stands out, *algR*. AlgR is a transcriptional regulator involved in positive and negative regulation of a variety of virulence systems within *P. aeruginosa*.

Gene ID	Putative Function	Homology (%)	Comments
J1385_3400	Lytic enzyme	*P. aeruginosa* (98%), *P. aeruginosa* M9A.1 (97%), *P. aeruginosa* (95%)	
J1385_3403	Phage tail protein	*P. aeruginosa* BL16 (99%), *Pseudomonas* phage phi297 (99%). *Pseudomonas* phage vB_PaeS_PMG1 (99%)	
J1385_3407	Glycoprotein	*P. aeruginosa* NCGM2.S1 (97%), *Pseudomonas* phage phi297 (98%). *P. aeruginosa* (98%)	
J1385_3408	Glycoprotein	*P. aeruginosa* BWHPSA027 (100%), *P. aeruginosa* BL12 (100%), *P. aeruginosa* BWHPSA022 (100%)	
J1385_3419	Terminase	*P. aeruginosa* BWHPSA027 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BL04 (100%)	Large subunit
J1385_3425	NinG protein	*P. aeruginosa* BL04 (92%), *P. aeruginosa* DHS01 (93%), *P. aeruginosa* PA21_ST175 (92%)	
J1385_3426	NinB protein	*P. aeruginosa* NCGM2.S1 (99%), *Pseudomonas* phage PAJU (99%). *P. aeruginosa* (99%)	
dnaB	Helicase	*P. aeruginosa* BWHPSA022 (99%), *P. aeruginosa* CI27 (99%), *P. aeruginosa* BL12 (99%)	
J1385_3441	ssDNA binding protein	*P. aeruginosa* HB13 (100%), *Pseudomonas* phage D3 (99%). *P. aeruginosa* NCGM2.S1 (100%)	ERF family
J1385_3442	Exonuclease	*P. aeruginosa* BWHPSA027 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* HB13 (99%)	
algR	Transcriptional regulator	*P. aeruginosa* BWHPSA027 (100%), *P. aeruginosa* BL16 (100%), *Pseudomonas* phage vB_PaeS_PMG1 (98%)	Putative homologue
J1385_3454	Integrase	*P. aeruginosa* BL16 (100%), *P. aeruginosa* 2192 (99%), *P. aeruginosa* PA21_ST175 (99%)	

Table 5.7 Table showing the gene id, putative function and percentage homology as predicted by BLAST analysis of genes present within the 48 kbp J1385 Prophage 2 lost by J1532. The majority of genes present within this phage encode proteins of uncharacterised function (not shown). The most significant gene lost along with this bacteriophage was a putative homologue of *algR* which encode a transcriptional regulator involved in the regulation of a variety of cellular processes.

5.10.2 The loss of a large bacteriophage by J1532 may contribute to the reduced virulence observed in this strain

We identified a large bacteriophage of around 79 kbp present in J1385 but not in J1532, that has been confirmed using PCR (Figure 5.12). This region is characteristic of bacteriophage acquisition with flanking genes transcribed in opposing directions, a drop in GC content and the presence of phage integrase TnpS and TnpT at the prophage origin (Figure 5.12). This large region, hereby referred to as J1385 Prophage 1, contains 73 genes (Table 5.7), the putative functions of which suggest that their loss may contribute to the altered phenotype of J1532 and act as evidence for host adaptation. For example, loss of J1385 Prohage 1 has resulted in the loss of a variety of genes involved in heavy metal resistance including a two-component system (J1385_3099 and J1385_3100), a copper oxidase (*pcoA*) and a copper chaperone *copZ* (J1385_3108). As can be seen, these genes share significant homology with environmental *Pseudomonas* species such as *P. putida*, *P. mendocina* and *P. thermotolerans*. This was to be expected as environmental isolates encounter high and potentially toxic concentrations of heavy metals in the environments in which they colonise e.g. soil/plants (Cánovas, Cases and de Lorenzo, 2003). It is not known where J1385 acquired such genes. However, within the CF lung J1385 is unlikely to encounter heavy metals at toxic levels and therefore has no reason to retain these genes. This potentially explains why J1532 has lost this bacteriophage containing heavy metal resistance genes, providing evidence of host adaptation and negative selection.

The loss of this large bacteriophage by J1532 has also resulted in the loss of a variety of genes that might provide this strain with a selective advantage in establishing a chronic infection within the CF lung. For example, removal of J1385 Prophage 1 has resulted in the loss of a gene (J1385_3133) which encodes the outer membrane protein OpdG (Table 5.7). This porin is involved in transporting amino acids into the cell. However, it is also required for import of the structurally related antibiotics imipenem and meropenem

(Hancock and Brinkman, 2002). Both of these antibiotics are used in treating *P. aeruginosa* infections of the CF lung (Banerjee and Stableforth, 2000) and as a result, J1385 will most likely have encountered at least one of these antibiotics whilst colonising the CF lung. The loss of OpdG by J1532 will result in imipenem and meropenem being unable to enter the bacterium and therefore exerting no antimicrobial effects. It is expected therefore, that loss of this porin will confer a selective advantage to J1532, increasing its resistance to commonly used antibiotics.

Removal of this bacteriophage by J1532 has also resulted in a loss of a variety of genes coding for virulence factors associated with *P. aeruginosa* infection. For example, J1385_3148 encodes a putative protease. Proteases are key virulence determinants in *P. aeruginosa*, possessing the ability to cleave a number of host proteins (Engel *et al*, 1998). The loss of this protease-encoding gene may contribute to the reduced protease production previously observed in J1532 compared to J1385 (Table 5.1), a change in phenotype commonly associated with conversion to mucoidy (Kamath, Kapatral and Chakrabarty, 1998). Reduced protease production will aid J1532 in establishing a chronic infection within the CF lung due to reduced antagonistic effects on its host and the resulting increased host immune intervention.

A second virulence factor from this phage is the potent toxin ExoY (Table 5.7), an extracellular adenylate cyclase (Yahr *et al*, 1998). It is not entirely clear how this secreted enzyme contributes to *P. aeruginosa* virulence, however it has been shown to cause host epithelial retraction and rounding (Prasain *et al*, 2009 and Vallis *et al*, 1999) and conferred increased virulence *in vivo* in a mouse model (Hritonenko *et al*, 2011). As with the protease previously described, loss of ExoY may confer a selective advantage to J1532 by reducing virulence that will, in turn, aid host immune system evasion.

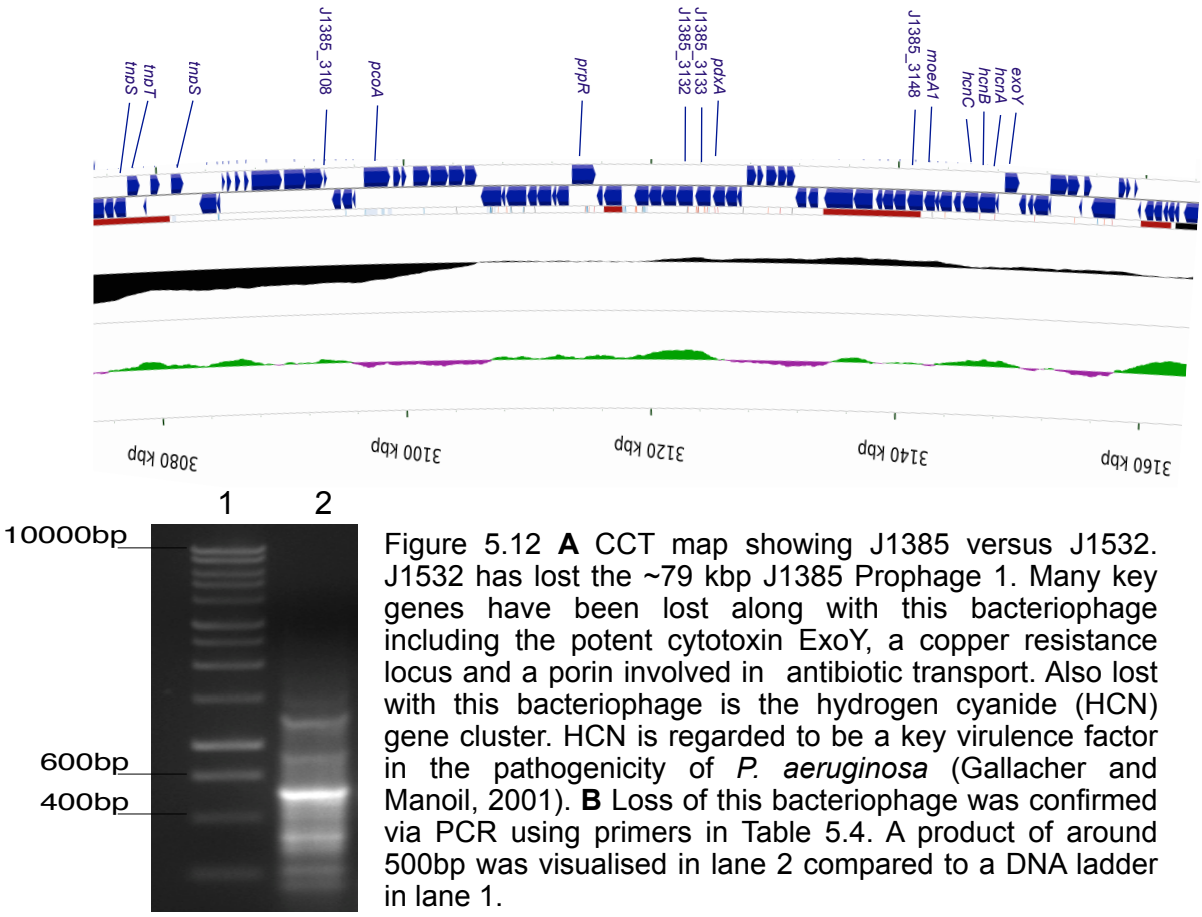


Figure 5.12 **A** CCT map showing J1385 versus J1532. J1532 has lost the ~79 kbp J1385 Prophage 1. Many key genes have been lost along with this bacteriophage including the potent cytotoxin ExoY, a copper resistance locus and a porin involved in antibiotic transport. Also lost with this bacteriophage is the hydrogen cyanide (HCN) gene cluster. HCN is regarded to be a key virulence factor in the pathogenicity of *P. aeruginosa* (Gallacher and Manoil, 2001). **B** Loss of this bacteriophage was confirmed via PCR using primers in Table 5.4. A product of around 500bp was visualised in lane 2 compared to a DNA ladder in lane 1.

Gene ID	Putative Function	Homology (%)	Comments
tnpS	Cointegrate resolution protein S	*P. aeruginosa* PA14 (99%), *P. aeruginosa* DK2 (99%), *P. aeruginosa* PA7 (99%)	
tnpT	Cointegrate resolution protein T	*P. aeruginosa* CI27 (100%), *P. aeruginosa* ATCC 25324 (100%), *P. aeruginosa* PA14 (75%)	
J1385_3097	Integrase catalytic subunit	*P. aeruginosa* C52 (100%), *P. aeruginosa* HB15 (100%), *P. aeruginosa* BL21 (100%)	
tnpS	Phage integrase	*P. aeruginosa* DK2 (100%), *P. aeruginosa* CI27 (100%), *P. aeruginosa* BL20 (100%)	
J1385_3099	ATPase	*P. aeruginosa* BL16 (100%), *P. aeruginosa* CI27 (100%), *P. putida* (99%)	Heavy metal sensor signal transduction histidine kinase
J1385_3100	Two component regulator	*P. mendocina* NK-01 (100%), *P. putida* H8234 (100%), *P. alcaligenes* OT 69 (100%)	Heavy metal response transcriptional regulator
J1385_3104	Plastocyanin/Azurin	*P. aeruginosa* CI27 (100%), *P. mendocina* ymp (100%), *P. putida* (100%)	Copper domain-containing protein
J1385_3105	ATPase	*P. aeruginosa* BL16 (100%), *P. aeruginosa* BL09 (100%), *P. mendocina* ymp (96%)	Copper-translocating P-type
J1385_3106	Chromosome segregation protein SMC	*P. aeruginosa* BL16 (100%), *P. aeruginosa* BL09 (100%), *P. aeruginosa* BL20 (100%)	
J1385_3108	Copper chaperone copZ	*P. aeruginosa* CI27 (100%), *P. aeruginosa* BL16 (100%), *P. putida* (100%)	
pcoA	Copper oxidase	*P. putida* H8234 (97%), *P. putida* CSV86 (97%), *P. putida* TR01 (99%)	copA family copper resistance protein
J1385_3115	Amino acid permease	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	
J1385_3116	Gamma-aminobutyraldehyde dehydrogenase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	
J1385_3117	Diaminobutyrate-2-oxoglutarate aminotransferase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* WC55 (100%)	
J1385_3118	Dehydrogenase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	
J1385_3119	Acetate permease	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	
J1385_3121	Acyl-CoA synthetase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	
J1385_3122	Dehydrogenase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	
J1385_3123	Acyl-CoA dehydrogenase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	
J1385_3124	Acyl-CoA dehydrogenase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	
J1385_3125	Aminoglycoside phosphotransferase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	*fadE36*
prpR	Propionate catabolism operon regulatory protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	
J1385_3128	Paraquat inducible protein B	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	Mamalian cell entry protein
J1385_3129	Aldehyde dehydrogenase	*P. aeruginosa* PA01 (82%), *P. aeruginosa* MPAO1/P1 (82%), *P. aeruginosa* MPA01/P2 (82%)	
J1385_3132	MFS transporter, YfaV	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* PA45 (100%)	

J1385_3133	Porin, OpdG	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	
pdxA	4-hydroxythreonine-4-phosphate dehydrogenase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* LESB58 (100%), *P. aeruginosa* RP73 (100%)	
J1385_3136	MFS transporter	P. *aeruginosa* PA14 (100%), *P. aeruginosa* MSH3 (100%), *P. aeruginosa* DK2 (100%)	
J1385_3137	LsyR trancriptional regulator	*P. aeruginosa* PA14 (100%), *P. aeruginosa* DK2 (100%), *P. aeruginosa* PA01 (100%)	
J1385_3139	ABC-transporter substrate binding protein	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	
J1385_3140	Amino acid permease	*P. aeruginosa* PA14 (100%), *P. aeruginosa* DK2 (100%), *P. aeruginosa* PA01 (100%)	
J1385_3141	Amino acid ABC transporter permease	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	
J1385_3143	Signal transduction histidine kinase	P. *aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	
J1385_3144	Extracellular nuclease	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	
J1385_3145	Alkaline phosphotase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* DK2 (100%), *P. aeruginosa* PA01 (100%)	Extracellular DNA degradation protein, EddA
J1385_3148	Protease	*P. aeruginosa* PA14 (100%), *P. aeruginosa* DK2 (100%), *P. aeruginosa* PA01 (100%)	
moeA1	molybdenum cofactor biosynthetic protein A1	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	
J1385_3150	Dehydrogenase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	
J1385_3151	Antibiotic biosynthesis monooxygenase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	
J1385_3153	Transcriptional regulator	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	TetR family
hcnC	Hydrogen cyanide synthase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), P. aeruginosa BWHPSA027 (100%)	
hcnB	Hydrogen cyanide synthase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), P. aeruginosa BWHPSA027 (100%)	
hcnA	Hydrogen cyanide synthase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	
exoY	Adenylate cyclase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	Secreted toxin
J1385_3158	Clp protease	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	
J1385_3160	Carboxylate-amine ligase	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	
J1385_3161	Ligase	*P. aeruginosa* 39016 (96%), *P. aeruginosa* NCGM2.S1 (96%), *P. aeruginosa* NCMG1179 (96%)	
J1385_3166	Two component sensor/response regulator hybrid	*P. aeruginosa* PA14 (100%), *P. aeruginosa* BL16 (100%), *P. aeruginosa* BWHPSA027 (100%)	PAS sensor protein

Table 5.7 Table showing the gene id, putative function and percentage homology as predicted by BLAST analysis of genes present within the ~79 kbp J1385 Prophage 1 lost by J1532. Many of the genes present within this phage encode proteins of uncharacterised function (not shown). Also lost within this bacteriophage are a variety of virulence factor genes e.g *hcnABC* and *exoY*

5.10.3 J1532 lacks a functioning hydrogen cyanide biosynthesis cluster

Figure 5.12 and Table 5.7, demonstrate that the loss of the ~79 kbp bacteriophage also results in the loss of the *hcnABC* gene cluster. This gene cluster is solely responsible for the production of potent virulence factor hydrogen cyanide (HCN). HCN is produced by *P. aeruginosa*, particularly during periods of low oxygen and high cell densities (Pessi and Haas, 2000) and is known to be significantly reduced in mucoid phenotypes (Gilchrist *et al*, 2011). Previously, using a *C. elegans* infection model, Gallagher and Manoil (2001) showed that cyanide poisoning is a key virulence factor in killing by *P. aeruginosa*. We hypothesised that a loss of this gene cluster may result in the reduced virulence associated with host adapted strains and observed in J1532 compared to J1385 (Figure 5.15 and Amy Ford, unpublished results). To test this theory we replicated the loss of HCN production observed in J1532 in the early J1385. Using the PA14 non-redundant mutant library (Liberati *et al*, 2006) we inserted a transposon into J1385 *hcnA* with the intention of blocking HCN production. Once confirmed, the new mutant J1385::*hcnA*$^{-}$ was used in a *G. mellonella* infection assay to identify any changes in virulence of the strain.

Transduction of the transposon insertion mutation was performed using the transducing phage ΦPA3 to transfer MAR2xT7 from PA14::*hcnA*$^{-}$ to J1385. This was confirmed by PCR (Figure 5.13). We designed two sets of primers which together could definitively prove that a transposon had been inserted in the correct location (Table 2.6). The first primer pair (universal) amplified an internal segment of the transposon of 296 bp . The second pair (specific) consisted of an internal transposon primer and a primer specific to the *hcnA* gene. Therefore, amplification of the 389 bp product will only occur if the transposon was inserted in the correct location within the gene. As can be seen in Figure 5.13, both our new mutant (lanes 2 and 3) and the original PA14 mutant (lanes 6 and 7) produce the required bands of 296 bp and 389 bp, respectively. Although the negative control of WT J1385 (lanes 4 and 5) produce PCR products they are not both at the correct size and instead can

be deemed to be non-specific products and have no bearing upon the credibility of the transduction event.

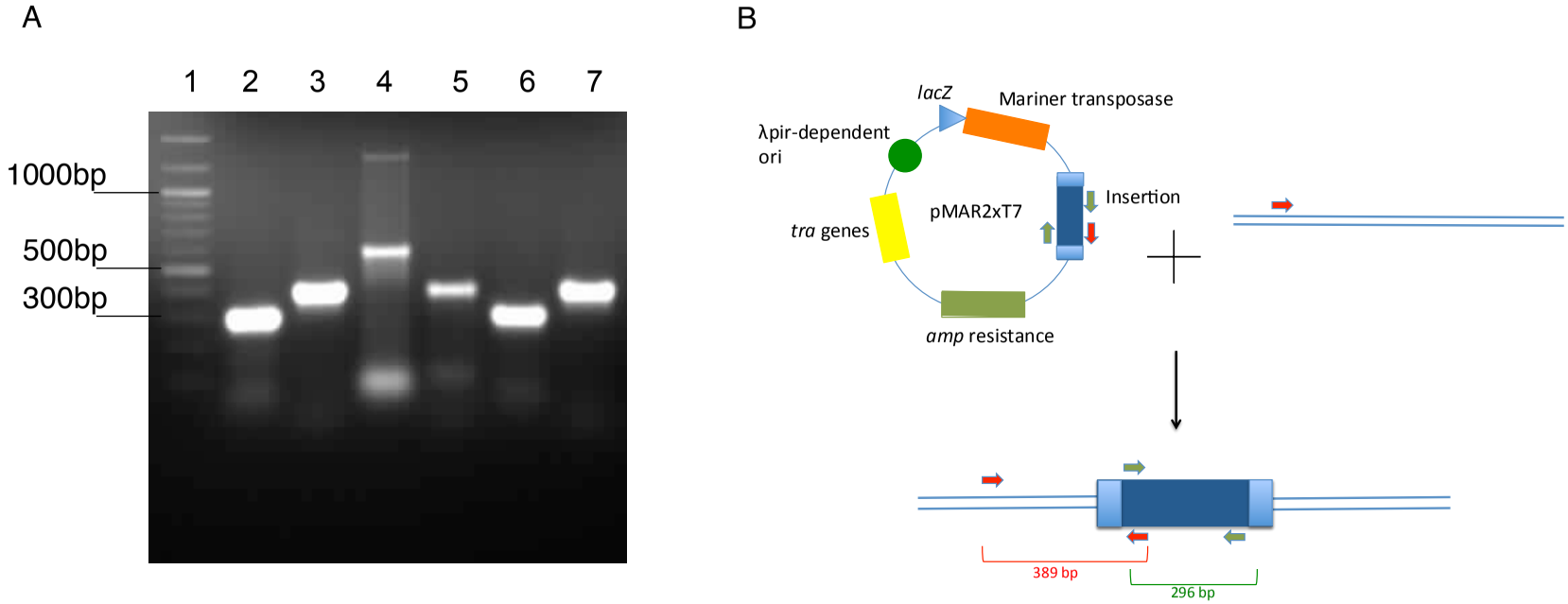


Figure 5.13 **A** PCR agarose gel showing confirmation of *hcnA* knockout via transposon insertion. A universal primer pair designed to amplify internal transposon sequence can be seen to amplify products of the desired size of 296 bp in both J1385::*hcnA*$^{-}$ (lane 2) and PA14::*hcnA*$^{-}$ (lane 6) but not in the WT J1385 (lane 4). A specific primer pair designed to amplify a product only when the transposon is inserted in the correct location of *hcnA* can be seen to yield a product of the desired size of 389 bp in both J1385::*hcnA*$^{-}$ (lane 3) and PA14::*hcnA*$^{-}$ (lane 7) and a much weaker band in WT J1385 (lane 5). Therefore, as bands of the correct size are present in both lanes for the mutant samples but not the WT, transduction is deemed to be a success. **B** Primer map showing location of primers used to detect a successful transposon event

Following confirmation that the correct transposon insertion mutation had been transduced into J1385 *hcnA*, it was necessary to confirm that HCN production was significantly reduced. A standard HCN assay (Figure 5.14) identified that HCN production of the mutant had been reduced to levels similar to those of the WT late strain J1532.

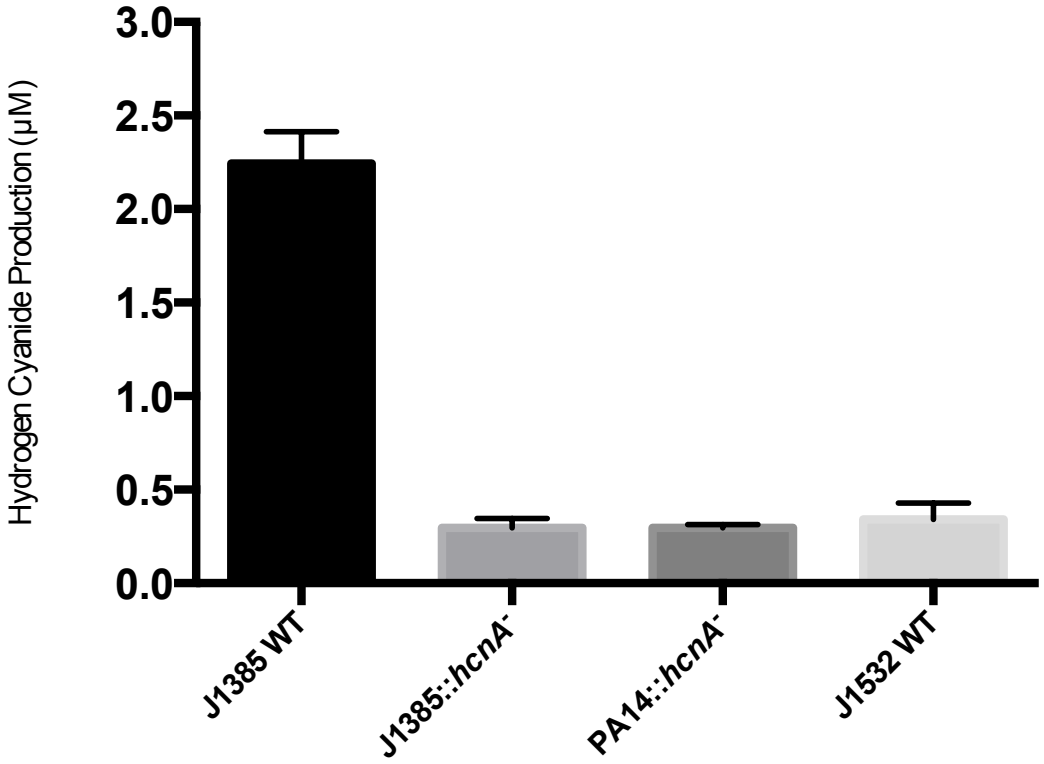


Figure 5.14 Graph showing HCN production of J1385 WT versus J1385::*hcnA*$^{-}$. PA14::*hcnA*$^{-}$ and J1532 are included for reference purposes. HCN by J1385 can be reduced to levels observed in J1532 by silencing *hcnA* by transposon insertion.

As such, the new mutant J1385::*hcnA*$^{-}$ strain was been confirmed to be a satisfactory imitation of the deletion observed in J1532 using both genotypic and phenotypic tests and as a result was suitable to perform a *G. mellonella* infection assay. As can be seen in Figure 5.15, the hypothesis that reduced HCN production will attenuate virulence in J1385 has been disproved. In fact, in both *hcnA* mutants it was observed that reduced HCN production resulted in increased virulence of both strains compared to their respective WT. Conversely, using *Caenorrhabditis elegans* as an infection model, Gallagher and Manoil (2001) showed that cyanide alone is sufficient for killing and that virulence was attenuated in a variety of *hcn* mutants. There are several possible explanations for this result. Gallagher and Manoil (2001) used strains less virulent than PA14 and J1385 strains that were used used in our study. Therefore, silencing HCN production in our strains might have a lesser effect on the arsenal of virulence factors and subsequent killing effect compared to PA01 and other strains used in the mentioned study. A second explanation could be that reduced HCN production may free up metabolic precursors for the production of other potent virulence factors. For example, the amino acid glycine is a major precursor in the formation of HCN (Castric, 1977). *P. aeruginosa* can convert glycine into a variety of other secondary metabolites including the highly toxic pyocyanin. Therefore, the abundance of glycine as a result of reduced HCN production in our strains may have resulted in an increase in the production of other virulence factors such as pyocyanin and may account for the increased virulence observed in our mutants. Figure 5.16 highlights a potential third explanation for the result seen as the viable cell count of bacteria recovered from homogenized larvae infected by J1385 is significantly less than those recovered from PA14 and the two mutant strains. However, this explanation can be discounted as HCN production is independent of bacterial load (Ryall *et al*, 2008). Viable cell counts of J1532 were not possible due to its mucoid phenotype; colonies appeared extremely large and merged with adjacent colonies.

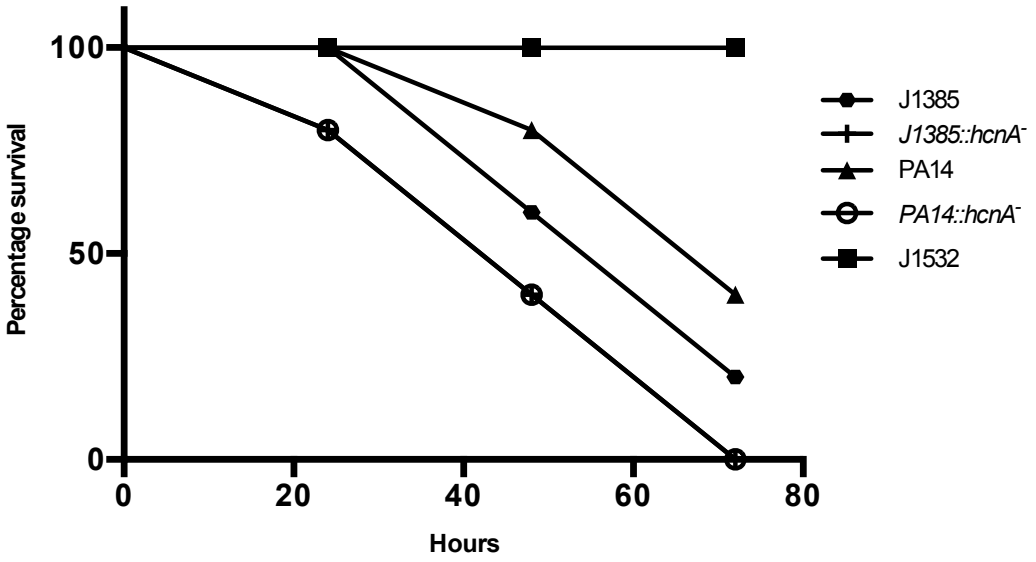


Figure 5.15 Survival graph showing the effect of transposon-mediated silencing of *hcnA* in J1385 and PA14 as well as their WT counterparts and J1532 as a control. As expected and in-keeping with previous results in the laboratory (Amy Ford, unpublished data), J1385 and PA14 are competent at killing whereas J1532 is completely avirulent. Surprisingly, both *hcnA* mutants of J1385 and PA14 have increased virulence compared to their WT. n=5

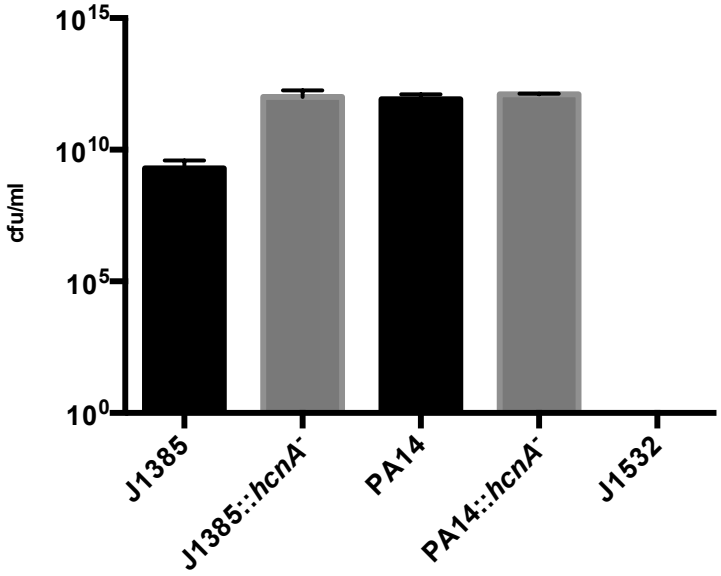


Figure 5.16 Survival graph showing the effect of transposon-mediated silencing of *hcnA* in J1385 and PA14 as well as their WT counterparts and J1532 as a control. As expected and in-keeping with previous results in the laboratory (Amy Ford, unpublished data), J1385 and PA14 are competent at killing whereas J1532 is completely avirulent. Surprisingly, both *hcnA* mutants of J1385 and PA14 have increased virulence compared to their WT. n=5

5.11 Genomic differences between J1385 and J1532 at the single nucleotide level

Previously, this chapter has investigated the major genomic differences between J1385 and J1532. However, as done for the C1426/C1433 pairing, it is essential to take a closer look at their genomes and investigate any minor genomic changes occurring at the single nucleotide level e.g. single nucleotide polymorphisms (SNP) and insertion/deletion polymorphisms (InDel). We hope to identify loss-of-function mutations within this mucoid pairing that may contribute to the altered phenotype of J1532. Over the course of three months during host adaptation, J1532 has acquired 50 non-synonymous SNPs and 13 InDels.

5.11.1 J1532 has acquired 50 non-synonymous SNPs during host adaptation in the CF lung

Over the course of three months in the CF lung, J1385, in response to host conditions, underwent a variety of macro and micro mutations including the acquisition of 50 non-synonymous SNPs during its switch to the mucoid J1532. Table 5.8 outlines the SNPs identified between these strains. As can be seen from this table many SNPs have occurred in unannotated genes with no known function (hypothetical protein). However, several have occurred in a variety of genes whose function has been predicted by either xBASE annotations or based on the putative functions of identical proteins in other strains of *P. aeruginosa.*

Within the CF lung, J1385 will likely encounter a variety of antimicrobial interventions. In response to this the bacterium may adapt and introduce a variety of mutations that will protect the cell from such attacks. For example, we have previously identified a SNP in the *ampR* gene of C1433 that may potentially result in increased antibiotic reduction due to overexpression of RND efflux pumps. J1532 exhibits several examples of this, one of which involves the introduction of a non-synonymous SNP into a gene encoding for

a porin (J1385_0272) (Table 5.8). Porins increase the outer membrane permeability of bacteria allowing a variety of antibioitcs to enter the cell and implement their antimicrobial effects (Hancock and Woodruff, 1988). Therefore, it is possible that a SNP within the gene encoding this porin, J1532 may reduce its outermembrane permeability, thereby increasing its antibiotic resistance. A second example of this phenomenon by J1532 is the introduction of a non-synonymous SNP into the gene encoding for the LpxO2 protein involved in the biosynthesis of an outer membrane lipopolysaccharide. Lipopolysaccharides can be used as targets for a variety of antibiotics such as polymixin and novobiocin (Tamaki, Sato and Matsuhashi, 1971). Therefore, the introduction of a SNP within the gene encoding for this lipopolysaccharide may result in a structural change conferring increased antibiotic resistance.

A non-synonymous SNP was also identified in *pvdL*, a gene coding for a peptide synthase involved in pyoverdin biosynthesis (Table 5.8). Previous work has shown that siderophore production of J1532 is reduced compared to that of J1385. It is therefore hypothesised that this non-synonymous SNP in *pvdL* may contribute to the altered phenotype exhibited by J1532. This is not the only gene associated with iron acquisition within J1532 has undergone the appearance of a non-synonymous SNP. A SNP has also appeared in J1385_4832 that encodes a TonB-like iron transport protein (Takase *et al*, 2000). It is not clear what benefits, if any, J1532 might receive through the reduced iron uptake occurring as a result of the SNPs in these iron acquisition genes. The results of an earlier study investigating the genetic changes of *P. aeruginosa* within a CF lung has also shown mutations within genes involved in this system (Smith *et al*, 2006).

Reference Position	Consensus Position	Reference	Allele Variations	Gene ID	Amino Acid Change	Putative Function
109707	109605	C	T	J1385_0105	p.Ala4Thr	Type VI secretion protein
109715	109613	A	G	J1385_0105	p.Val1Ala	
274237	273941	T	C	J1385_0272	p.Asp63Gly	Porin
457284	456832	G	C	J1385_0459	p.Ser98Thr	Hypothetical protein
583100	582526	C	G	J1385_0580	p.Ala245Gly	Hypothetical protein
594282	593697	T	C	J1385_0595	p.Tyr188His	Acyl-CoA dehydrogenase
934772	933860	T	A	J1385_0961	p.Gln74Leu	Hypothetical protein
934788	933876	G	C	J1385_0961	p.Leu69Val	
976381	975428	G	T	*mpl*	p.Gly224Cys	Ligase
1343367	1342226	G	T	J1385_1384	p.Pro63Thr	Lipoprotein
1468281	1467033	G	C	J1385_1514	p.Val168Leu	Phosphotransferase enzyme I
1675723	1674296	A	C	*phnJ*	p.Asp40Ala	Phosphonate metabolism protein J
1698094	1696649	G	A	J1385_1739	p.Gly295Asp	Glutamate-ammonia ligase
1744831	1743334	C	T	J1385_1782	p.Ser94Phe	Hypothetical protein
2293942	2291873	A	C	J1385_2338	p.Ser364Ala	Hypothetical protein
2331122	2329056	C	A	J1385_2381	p.Gly139Val	Hypothetical protein
2407587	2405441	A	G	J1385_2473	p.Asn12Ser	Hypothetical protein
2736088	2733695	T	C	J1385_2814	p.Ile332Val	MFS transporter
2736096	2733703	G	C	J1385_2814	p.Ala329Gly	
2806355	2803924	C	T	*pvdL*	p.Leu778Phe	Peptide synthase - Pyoverdin synthesis
2820103	2817673	T	G	J1385_2869	p.Lys530Asn	CRISPR associated helicase
2922909	2920373	T	C	J1385_2957	p.Thr50Ala	ABC maltose/ mannitose transporter
2930001	2927456	C	G	J1385_2965	p.Asp130Glu	ABC transporter
2930002	2927457	G	C	J1385_2965	p.Asp131His	
3020867	3018212	G	T	*pslH*	p.Pro20Thr	Glycosyl transferase involved in biofilm production
3304664	3234478	T	C	*pcoA*	p.Ser312Pro	Copper oxidase

3446195	3329565	C	A	J1385_3457	p.Ser188Ile	Azoreductase
3519537	3402872	T	C	J1385_3538	p.Ser447Gly	ABC transporter
3525678	3409013	A	C	J1385_3542	p.Ile20Ser	Hypothetical protein
3623291	3506537	G	A	*nasS*	p.Arg252His	Nitrate transporter
3666714	3549936	T	G	J1385_3680	p.Ser250Ala	Acyl-CoA dehydrogenase
3666779	3550001	T	G	J1385_3680	p.Ser271Arg	
3667235	3550458	CAG	ATG	J1385_3681	p.Gln86Met	Acyl-CoA dehydrogenase
3779224	3662381	T	C	J1385_3806	p.Lys177Arg	Flavin reductase
4093589	3976153	G	C	J1385_4132	p.His237Gln	Hypothetical protein
4196838	4079326	T	C	J1385_4237	p.Val27Ala	Enoyl-CoA hydratase
4196841	4079329	C	G	J1385_4237	p.Ala28Gly	
4479696	4352027	T	G	J1385_4523	p.Val232Gly	Transposase
4518983	4391313	G	T	*lpxO2*	p.Ser294Ile	Lipopolysaccharide biosynthetic protein
4715684	4587886	G	A	*algU*	p.Leu62Phe	Sigma factor
4800002	4672137	A	T	J1385_4832	p.Val119Asp	Energy production protein TonB
4874818	4746832	G	T	J1385_4893	p.Pro21Thr	Hypothetical protein
4983819	4855731	A	C	*murF*	p.Phe72Val	Ligase
4983820	4855732	A	T	*murF*	p.Asp71Glu	Ligase
5053732	4925571	C	G	J1385_5082	p.Gly1197Ala	Hypothetical protein
5333720	5197396	C	G	J1385_5347	p.Ser1116Thr	Hypothetical protein
5483891	5347533	C	G	J1385_5502	p.Arg102Gly	Ferredoxin
5529780	5393420	T	G	J1385_5552	p.Ile376Leu	Hypothetical protein
6296749	6156125	T	G	J1385_6318	p.Asp257Ala	Glutamine synthetase
6396809	6233424	C	G	J1385_6419	p.Arg366Gly	Haemagglutinin

Table 5.8 Table showing the non-synonymous SNPs present in J1532 and the putative function of the genes in which they have occurred. 50 non-synonymous SNPs have occurred in J1532 over the space of three months in the CF lung.

5.11.2 A non-synonymous SNP within the sigma factor *algU* may be responsible for the switch to mucoidy observed in J1532

Perhaps the most interesting and most significant non-synonymous SNP to have occurred between the switch from J1385 to J1532 has appeared within the sigma factor *algU* (Table 5.8). AlgU is an alternative sigma factor that plays a role in global gene regulation involved in the promotion of a variety of effectors including porins/transporters and lipoproteins (Firoved, Boucher and Deretic, 2002). However, it is best known for its role in alginate biosynthesis where it is responsible for initiating the biosynthetic process (Martin, Holloway and Deretic, 1993). As a result, it seems likely that it is essential for the conversion to mucoidy observed in J1532. Therefore, we hypothesise that this SNP may be involved in the over production of alginate responsible for the appearance of the mucoid phenotype.

This SNP (c184t) has resulted in the switch of a leucine to a phenylalanine at residue 62 (L62F) (Figure 5.17). Due to its high degree of homology to sigma factor E of *Escherichia coli* K12, we have used the crystal structure of this protein (Campbell *et al*, 2003) to computationally visualise the positioning of this residue (red) within the protein and predict what, if any, effect this amino acid change might exert upon the protein conformation (Figure 5.17). The affected amino acid, residue 62, is located at the start of region 2.3 of the protein. Since alginate production is not compromised in J1532, compared to J1385, it is not possible for this SNP to have caused a loss-of-function. Instead it is hypothesised that this SNP has lead to constitutive expression of the protein by preventing correct binding of its anti-sigma factor MucA. To the best of our knowledge, no switch to mucoidy in any *P. aeruginosa* strain has ever been attributed to a mutation within the AlgU sigma factor. Therefore, if confirmed, this SNP in AlgU will be the first of its kind reported.

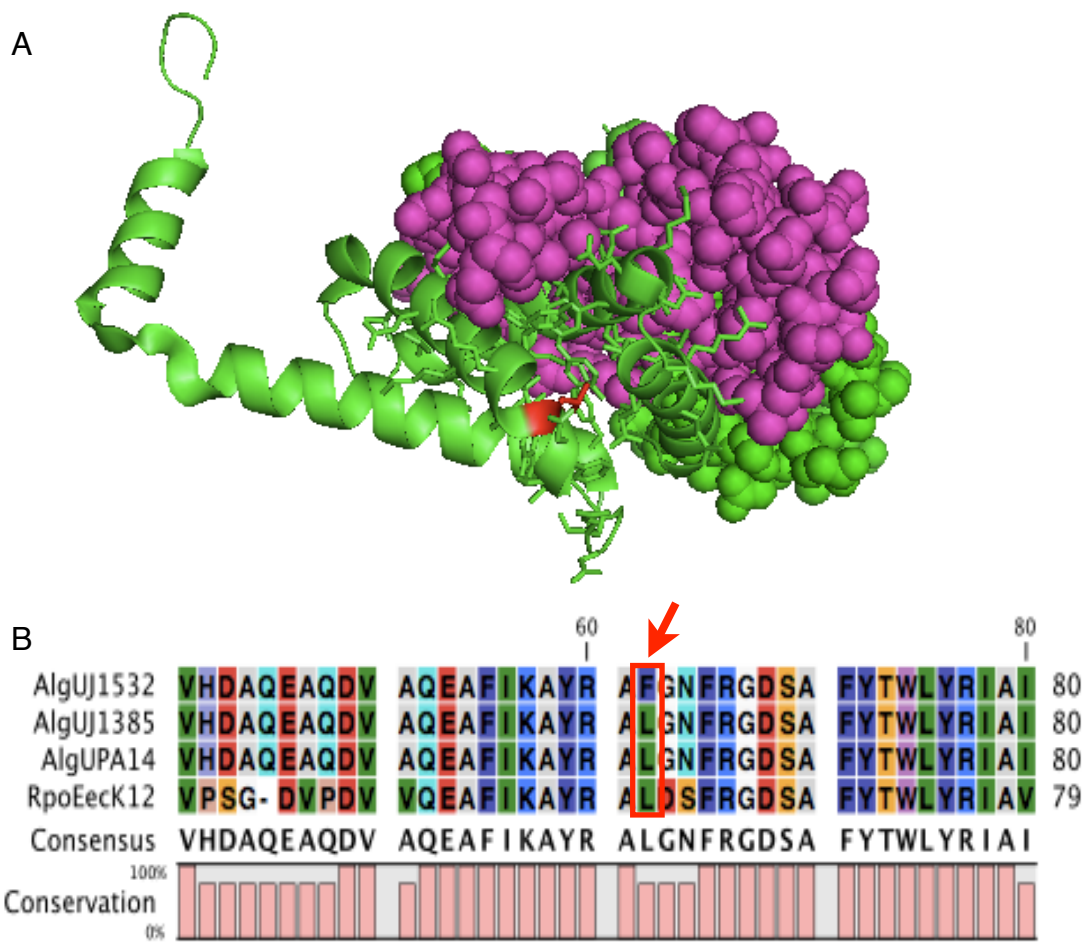


Figure 5.17 **A** PyMol generated crystal structure of sigma factor E from *E. coli* K12 which shows a high degree of homology to sigma factor AlgU. The residue affected by the L62F SNP is coloured red and can be seen located at the start of section 2.3 of the protein. **B** Alignment of J1532 vs J1385, PA14 and *E. coli* K12 showing the L62F amino acid change resulting from the SNP in AlgU.

5.11.3 J1532 has experienced 13 non-synonymous InDel polymorphisms

Over the course of three months within the CF lung, J1532 has acquired 13 InDel polymorphisms, 54% of which are insertion polymorphism with the remainder deletions (Table 5.9). Many of the genes affected by InDels have occurred in genes encoding for a variety of secondary metabolites. These may result in in changes within the metabolic profile of J1532 but without investigation, this cannot be confirmed.

Reference Position	Consensus Position	Reference	Variants	Allele Variations	Gene ID	Amino Acid Change	Putative function
320143	319801	-	1	T	J1385_0315	p.Pro40fs	Zinc dependant protease
444579	444128	C	1	-	J1385_0444	p.Gly123fs	Alpha/beta hydrolase
504069	503582	-	1	G	*mexB*	p.Leu73fs	Multidrug resistance protein B
949924	948994	T	1	-	*ispA*	p.Thr123fs	Geranyltransferase
1675722	1674294	-	1	G	*phnJ*	p.Asp40fs	Phosphonate metabolism protein
1793799	1792234	-	1	T	*rdgC*	p.Pro126fs	Recombination associated protein
3223659	3153718	--	1	CG	*cupA3*	p.Arg850fs	Usher protein
3525071	3408407	C	1	-	J1385_3542	p.Arg222fs	Hypothetical protein
3666975	3550197	-	1	A	J1385_3680	p.Ser337fs	Acyl-CoA dehydrogenase
4715012	4587215	C	1	-	*mucA*	p.Gly81fs	Anti-sigma factor
5085471	4949190	G	1	-	*dppF*	p.Gln160fs	ABC transporter
5552933	5416566	G	1	-	*speA*	p.His571fs	Putrescine biosynthetic protein
5633811	5497428	-	1	A	J1385_5669	p.Gln458fs	ABC transporter

Table 5.9 Table showing the non-synonymous InDels present in J1532 and the putative function of the genes in which they have occurred. 13 non-synonymous InDels have occurred in J1532 over the space of three months in the CF lung.

5.11.4 An InDel present within the anti-sigma factor MucA

One InDel in particular stands out from the list of InDels presented in table 5.9. This InDel occurs in the *mucA* gene responsible for encoding the anti-sigma factor MucA. Inactivation of this *mucA* results in an overproduction of alginate due to uncontrolled *algU* expression (Govan and Deretic, 1996). In fact, 80% of mucoid *P. aeruginosa* strains isolated from the lungs of CF patients exhibit a variety of mutations within *mucA* (Boucher *et al*, 1997) and as such, this gene is regarded to be a "hot-spot" for mucoid-inducing mutations. Supporting this fact, we have previously identified an InDel within this gene in C1433 that is expected to be solely responsible for its altered mucoid phenotype, as mentioned earlier.

The InDel within *mucA* of J1532 is the result of a deletion of a single cytosine nucleotide causing a frameshift at amino acid 81 that results in the introduction of a premature stop codon at amino acid 94 and a truncated protein compared to its normal size of 194 amino acids (Figure 5.18). This significant reduction in protein size is likely to have devastating effects on the functionality of this MucA protein in its anti-sigma factor role. We therefore suggest that this deletion polymorphism will result in uncontrolled expression of the previously mutated AlgU sigma factor, overproduction of alginate and the appearance of the mucoid phenotype exhibited by J1532.

Discovery of this mutation has led us to doubt the significance of the *algU* SNP in the conversion to the mucoid phenotype in J1532. In contrast to this hypothesis, previous studies have identified point mutations affecting this gene that arise following MucA knockout mutations resulting in the appearance of a mucoid phenotype (DeVries and Ohman, 1994 and Schurr *et al*, 1994). It has been suggested that these second-site suppressor mutations are an attempt by the bacterium to autoregulate, promoting the switch back to a non-mucoid phenotype. Therefore, it is predicted that the L62F mutation observed in J1532 AlgU is an attempt by the bacteria to

counteract the metabolite-draining switch to the mucoid phenotype and is no longer expected to have any role in alginate overproduction.

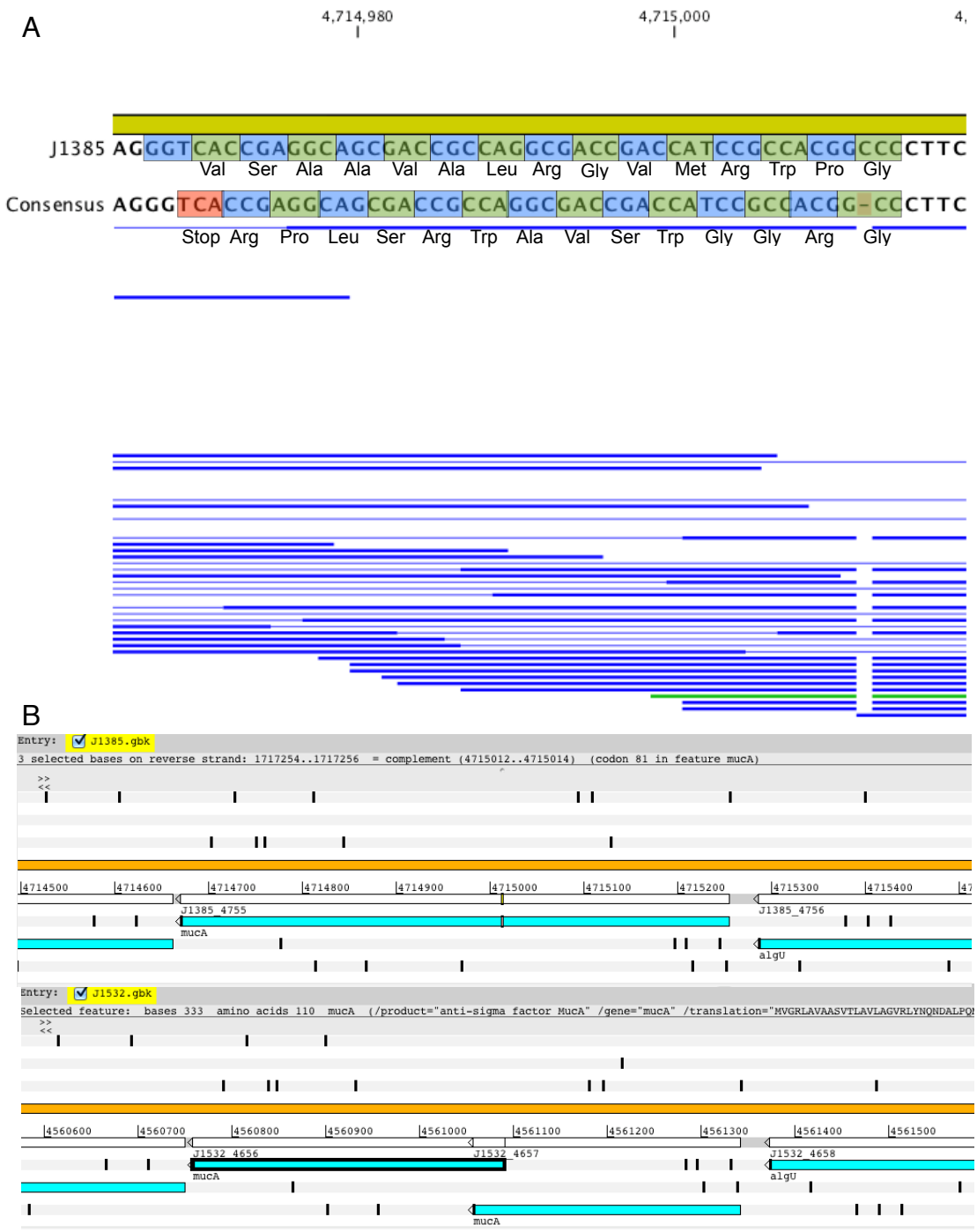


Figure 5.18 **A** CLC Genomics Viewer screen grab showing a deletion of a cytosine within the *mucA* gene in J1532 compared to its progenitor J1385. This results in a frameshift mutation in J1532 and insertion of a premature stop codon (red). **B** Artemis screen grab showing the truncated MucA protein in J1532 compared to J1385 as a result of the single nucleotide deletion in the *mucA* gene.

5.12 Chapter 5 discussion

Throughout this chapter, we have analysed and discussed a variety of macro and micro differences observed between the non-mucoid J1385 and its mucoid descendant strain J1532. We have also compared both of these strains to 53 other fully sequenced *P. aeruginosa* strains from a variety of niches including environmental and clinical isolates. Evidence presented in this chapter demonstrates that this non-mucoid/mucoid pairing has acquired a variety of large genomic regions as a result of horizontal gene transfer that are present in none or very few *P. aeruginosa* strains included in this study. One strain that does possess all acquired genomic regions of J1385 is the hypervirulent PA14. This is to be expected as previously, J1385 and PA14 were shown to share a high degree of similarity (Figures 6.2, 6.3 and 6.4). Two such regions, a 33 kbp transposon and a 42 kbp plasmid, carry genes that may confer resistance to a variety of heavy metals including mercury, zinc, cobalt and cadmium. Within its niche of the CF lung, there is no identifiable benefit that J1385/J1532 might experience as a result of acquiring these genes. Therefore, we expect that these genomic elements will have been acquired during a previous environmental colonisation, a location synonymous with high heavy metal concentrations that may be toxic to cells lacking such resistance mechanisms (Cánovas, Cases and de Lorenzo, 2003).

In contrast to this, we have also identified several horizontally acquired regions that may confer a selective advantage to J1385/J1532 within the CF lung. The first of these is a ~21 kbp region located within the hypervariable pyoverdin locus. Acquisition of this region has resulted in J1385 gaining several extra genes involved in iron sequestration and transport. This includes genes encoding the pyoverdin synthesis proteins PvdJ, PvdD and PvdE as well as a gene encoding for the ferripyoverdin receptor FpvA. This is not the only set of horizontally acquired siderophore biosynthesis and transport genes acquired by J1385/J1532. In addition to the previously acquired ~21 kbp region described, these strains (and only two further *P.*

aeruginosa strains included) have acquired a ~34 kbp bacteriophage, over 50% of the genes of which are involved in siderophore biosynthesis and transport. In this case the siderophore in question may be pyochelin due to the presence of a thioesterase, an enzyme that forms part of the non-ribosomal synthesis of this siderophore (Reimmann *et al*, 2004). Therefore, J1385 and J1532 both have an extra two sets of genes involved in iron acquisition from the bacterias immediate environment. Possessing the genes acquired on these two large horizontally acquired regions may confer a selective advantage to J1385, allowing it to increase iron uptake, enabling it to compete more effectively against any co-colonising bacterium within the lung and establish strain dominance.

The competitive advantage conferred by these regions to J1385 has been further supplemented by the acquisition of the pathogenicity islands PAPI-1 and PAPI-2. These islands contribute to the increase in virulence exhibited by PA14 both individually and synergistically (He *et al*, 2004 and Harrison *et al*, 2010). Possession of the PAPI-1 island results in increased motility due to possession of an extra set of type IVb pili. These islands also contain the *cupD1-5* fimbrial genes and PvrR, a two component response regulator both of which are involved in biofilm formation. The latter of these is also involved in antibiotic resistance (Harrison *et al*, 2010). Increased motility, increased biofilm formation and increased antibiotic resistance are likely to aid J1385 in establishing a successful chronic infection within the lung via protection from host and drug interventions. The second pathogenicity island PAPI-2, carries the genes coding for the potent cytotoxin ExoU and its chaperone SpcU (He *et al*, 2004). ExoU is an effector of the type III secretion system of *P. aeruginosa* and can be injected into a variety of eukaryotic cells where it will cause rapid cell lysis due to its phospholipase A_2 activity (Kurahashi *et al*, 1999 and Philips *et al*, 2003). It can limit the host immune response via infiltration and killing of immune effectors such as phagocytes (Diaz and Hauser, 2010) and is toxic to yeasts such as *S. cerevisiae* (Rabin and Hauser,2003). Therefore, possession of the PAPi-2 island and therefore a

copy of the *exoU* gene and protein might confer a selective advantage to J1385/J1532, allowing it to colonise the CF lung with minimal host immune intervention whilst initiating killing of any co-colonising and competing yeasts within the lung.

Each of these previously described macro differences are present in both J1385 and J1532 and either few or no other *P. aeruginosa* strains included in the study. However, investigating the role of host adaption within the CF lung, we must look specifically at macro and micro differences existing between J1385 and its mucoid descendant J1532 only. Over the course of three months within the CF lung, we have shown that J1532 has lost four large bacteriophages when compared to its progenitor J1385. Two of these, located at 4245 kbp (Figure 5.9) and 5070 kbp (Figure 5.10) appear to be fairly innocuous in the role of altering the phenotype of J1532. Neither contains any genes that seem likely to confer a selective advantage to J1532 in chronically infecting the CF lung. However, the former of these two bacteriophages, at 4245 kbp, shares a high degree of homology to the filamentous phage Pf1 (Hill *et al*, 1999). The genes within this phage are amongst the most highly upregulated genes during biofilm development and as a result may play a role during this event (Webb *et al*, 2003 and Webb, Lau and Kjelleberg, 2004).

A third bacteriophage at 3395 kbp was lost by J1532. Out of the 57 genes encoded by this bacteriophage only one stands out and may have a potential effect on the phenotype of J1532. This is a putative homologue to *algR*, the gene encoding for the broad range transcriptional regulator AlgR. In its role as a transcriptional regulator it controls a variety of different systems including alginate biosynthesis, type IV pili functioning and several other virulence genes (Lizewski *et al*, 2004) one of which is activation of the *fimU-pilVWXY1Y2* operon, the operon that controls twitching motility in *P. aeruginosa*. Therefore, deletion of the *algR* homologue within this large bacteriophage may account for the reduced twitching motility observed in

J1532 compared to J1385 within our lab (Amy Ford, unpublished results). It has also been shown that AlgR is essential for virulence (Lizewski *et al*, 2003) and loss of a putative homologue may therefore contribute to the reduced virulence observed in mucoid CF strains such as J1532 as shown in Figure 5.15 and by Amy Ford (unpublished results). AlgR is also involved in positive regulation of alginate biosynthesis, a fact that is contradictory to the mucoid phenotype displayed by J1532 although this will be discussed later.

The fourth bacteriophage lost by J1532 during host adaptation is located at 3080 kbp, spans ~79 kbp and encodes for 73 proteins. Many genes lost within this bacteriophage, based on the putative functions of their encoded proteins, may contribute to the altered phenotype of J1532 and act as evidence of host adaptation within this CF pairing. Genes lost include several of those involved in the heavy metal resistance of copper. Within the CF lung, J1532 will not encounter toxic concentrations of such heavy metals and therefore there is no benefit to possessing this resistance mechanism. This may be an example of negative selection, removing these genes from its repertoire allowing redirecting of essential efforts. Another gene deleted within this bacteriophage is responsible with encoding the outermembrane porin OpdG. This porin is responsible for transporting amino acids into the cell as well as structurally similar antibiotics such as imipenem and meropenem (Hancock and Brinkman, 2002). Therefore, the loss of OpdG may result in increased resistance to such antibiotics, conferring a selective advantage to J1532, demonstrating a further example of negative selection in host adaptation.

The loss of this bacteriophage also results in the loss of genes encoding proteins normally associated in provoking the host into an increased immune response. This includes a gene encoding for a protease, an enzyme associated with cleaving host proteins. The loss of this gene may account for the reduced protease production previously observed in J1532 compared to J1385 (Table 5.1), a change in phenotype commonly associated with

conversion to mucoidy (Kamath, Kapatral and Chakrabarty, 1998). A second antagonistic protein lost due to deletion of its gene within this bacteriophage is ExoY, an extracellular adenylate cyclase (Yahr *et al*, 1998). ExoY will cause conformation changes in host epithelial cells and conferred increased virulence in an *in vivo* mouse model (Vallis *et al*, 1999, Prasain *et al*, 2009 and Hritonenko *et al*, 2011). The reduced virulence experienced as a result of reduced protease and ExoY expression may contribute to the ability of J1532 to establish a chronic infection within the CF lung due to reduced antagonistic effects on its host and the resulting increased host immune intervention.

We have also identified a variety of micro differences between J1385 and its mucoid descendant J1532 including 50 non-synonymous SNPs and 13 non-synonymous InDels. The majority of SNPs occurring in J1532 have appeared in genes encoding proteins of hypothetical function. However, some have appeared in gene that encode proteins that may contribute to the altered phenotype in J1532. For example, SNPs have been identified in a gene encoding a porin. Altering the amino acid structure of this porin may result in reduced outermembrane permeability and as a result increased antibiotic resistance. This may be particularly useful for J1532, as in its time colonising the CF lung it is likely to experience a variety of antibiotic interventions. Increased antibiotic resistance may also be conferred by a SNP within the gene encoding for the LpxO2 lipopolysaccharide. Such lipopolysaccharides can be used as targets for a variety of antibiotics such a polymixin and novobiocin (Tamaki, Sato and Matsuhashi, 1971). Therefore, altering the lipopolysaccharide LpxO2 structure via introduction of a non-synonymous SNP that potentially may alter the structure recognised by the antibiotic active site, thereby increasing the resistance of J1532. Both of these examples of SNPs that may cause increased antibiotic resistance are potential examples of host adaptation and negative selection within the CF lung allowing J1532 to persist and establish a chronic infection.

Non-synonymous SNPs have also been identified in genes involved in iron sequestration and transport. SNPs have appeared in *pvdL*, the gene encoding a protein involved in siderophore biosynthesis and within a gene encoding for a TonB-like iron transport protein. It is not clear what, if any, benefits J1532 will experience from reduced iron uptake as a result of these SNPs. However, this is not the first time SNPs in the iron sequestration system have been identified (Smith *et al*, 2006)

5.12.1 The appearance of the mucoid phenotype can be attributed to an InDel in its anti-sigma factor MucA

As previously mentioned, the switch to mucoidy in *P. aeruginosa* is associated with the overproduction of alginate due to uncontrolled activity of the sigma factor AlgU (Govan and Deretic, 1996). 80% of the mucoid strains isolated from the lungs of CF patients possess a variety of mutations within the anti-sigma factor MucA (Boucher *et al*, 1997). To the best of our knowledge, no mutations within the sigma factor AlgU has ever been attributed with the switch to mucoidy. As a first, we report the presence of a non-synonymous SNP within this anti-sigma factor resulting in a leucine being replaced by a phenylalanine at residue 62 located at the beginning of section 2.3 of the protein (Campbell *et al*, 2003). As alginate production is not compromised in this strain, this SNP cannot result in attenuation of the sigma factor activity of this protein. Instead, we initially hypothesised that this SNP may result in unregulated activity of AlgU via preventing binding of its anti-sigma factor MucA.

One potential contradiction to this is the fact that a non-synonymous InDel has been detected in *mucA* of the same strain. This deletion polymorphism has resulted in a frameshift mutation causing the insertion of a premature stop codon truncating the protein by over 50% at 94 amino acids. This severely truncated protein will be unable to exert is anti-sigma factor effects on the AlgU protein, allowing uncontrolled activation of the alginate

biosynthesis pathway and appearance of the mucoid phenotype. It has previously been identified that mutations of this kind, resulting in a defective MucA, are associated with a subsequent point mutation in *algU* (DeVries and Ohman, 1994 and Schurr *et al*, 1994). These mutations are most likely an attempt by the bacterium to counteract the metabolite-sapping over production of alginate associated with the mucoid phenotype. Therefore it seems likely that our initial hypothesis that the L62F mutation identified in J1532 AlgU contributes to this switch in phenotype is incorrect and mucoidy can be attributed solely to the truncation of the MucA protein.

Throughout this chapter we have discussed a variety of macro and micro differences that may contribute to the altered phenotype observed in J1532 compared to its progenitor strain J1385. However, our investigations have only just initiated the vast amount of further analysis required to fully map these differences. As can be seen in each table in this the chapter, the vast majority of genes/proteins identified are of unknown functions. It is almost certain that the mutations within or the entire loss of these proteins will contribute to the overall phenotype of J1532. To fully understand the effect of host adaptation between J1385 and J1532, the function of these proteins must be investigated further.

Chapter 6 - Metabolomic analysis reveals a variety of changes in the bacterial metabolome as a result of host adaption within the CF lung

6.1 Introduction

Genomic analysis of non-mucoid/mucoid host adapted *P. aeruginosa* pairs reveals that during host colonisation and infection of the CF lung, strains undergo a variety of changes in their genomes. During previous chapters, we have discussed a variety of macro and micro differences between mucoid strains C1433 and J1532 and their non-mucoid progenitors C1426 and J1385, respectively. As previously suggested, these genomic alterations may contribute to the phenotypic differences observed between pairings that have most likely arisen as a result of host adaption. To do this we performed a large-scale metabolomic analysis using Liquid Chromatography-Mass Spectrometry (LC-MS). Using this method of analysis we planned to map a variety of the genomic changes observed between pairs to specific changes in the metabolome of these strains. For example, it is our aim to link specific changes in the metabolome to specific genotypic alterations previously identified in previous chapters. To perform this, we replicated growth conditions that these strains would experience in the lungs of cystic fibrosis patients. We utilised a modified version of Artificial Sputum Medium (ASM) developed by Kirchner *et al* (2012) and replicated chronic infection by allowing prolonged microaerophilic incubation at 37^{o}C with minimal shaking. A whole culture method of extraction was chosen to include both acquisition of both intracellular and extracellular metabolites. Our analysis included metabolite extraction and subsequent LC-MS analysis of uninoculated ASM culture media to allow for detection of metabolites present within the sterile medium. Also included was the reference strain PA14.

6.2 Metabolite extraction

Following incubation, the first step in any good metabolomic study is the efficient extraction of metabolites from the culture. A variety of simple quenching and extraction methods were available for use and were considered before settling upon a cold (-40^{o}C) methanol method of extraction. This method was chosen based upon the literature in which, two distinct studies compared variety of metabolite extraction methods to detect the most efficient in a global metabolite analysis of another Gram-negative bacterium, *E. coli* (Maharjan and Ferenci, 2003 and Winder *et al*, 2008). In both studies, cold methanol was more effective than any other of the extraction methods tested and were used to produce reproducible and homogenous biological datasets.

Due to the highly variable nature of bacterial growth and despite our best efforts at standardising culture inoculum there was a concern that each of our biological replicates could vary significantly. In order to combat this, we performed metabolite extraction from six cultures, double the amount of replicates required for testing and chose the most homogenous three extractions. Using thin layer chromatography and ninhydrin as a colourimetric substrate to detect amino acids, we visualised our extractions allowing us to make an informed decision when choosing appropriate biological replicates for subsequent LC-MS (Figure 6.1).

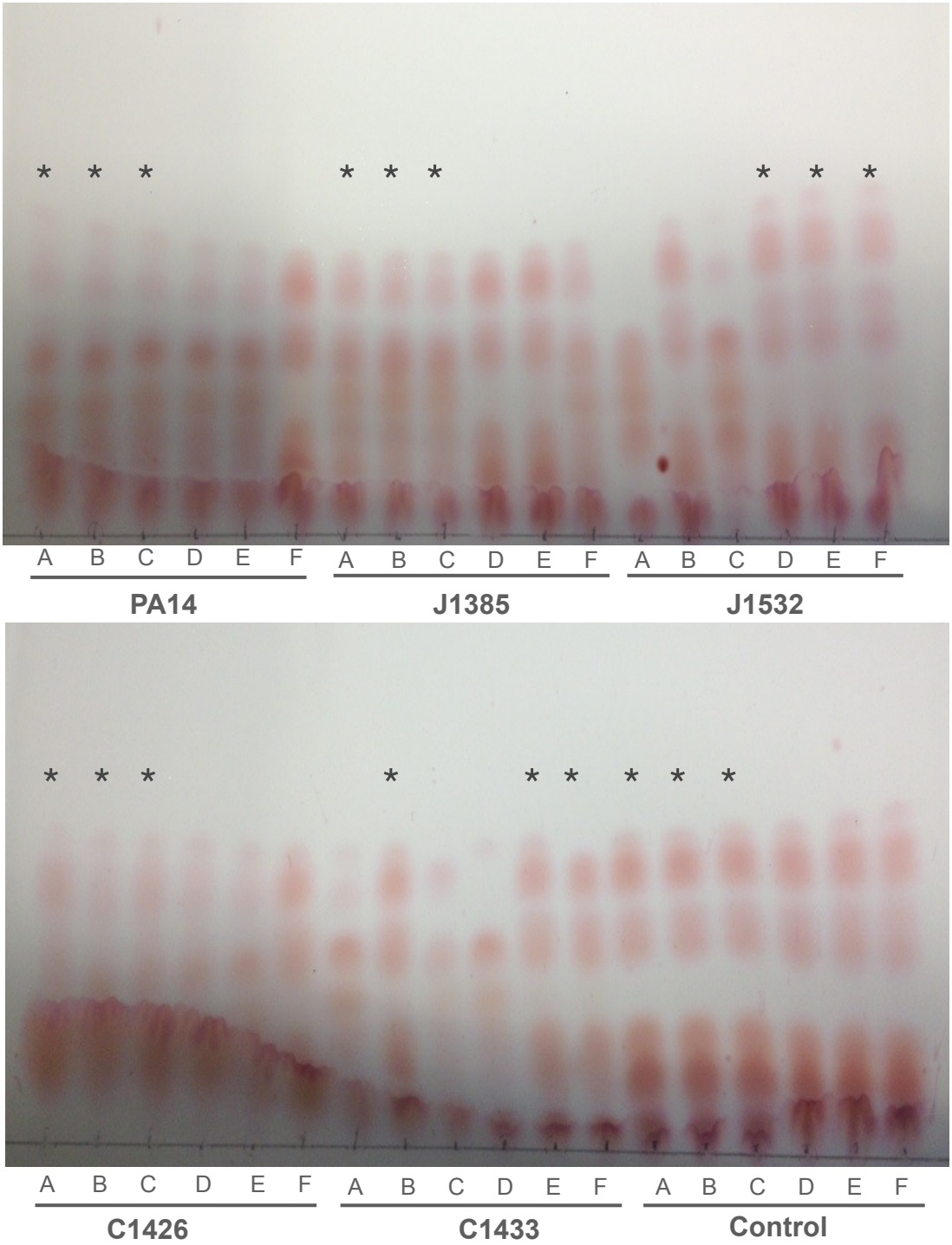


Figure 6.1 TLC chromatograms showing total metabolite extraction as visualised using ninhydrin as a colourimetric substrate for amino acid detection from each strain and uninoculated control included in the study. The six replicates for each strain are labelled A-F and the three replicates chosen for each strain or control, based on homogeneity, to be analysed via LC-MS are indicated by *.

6.3 Data processing

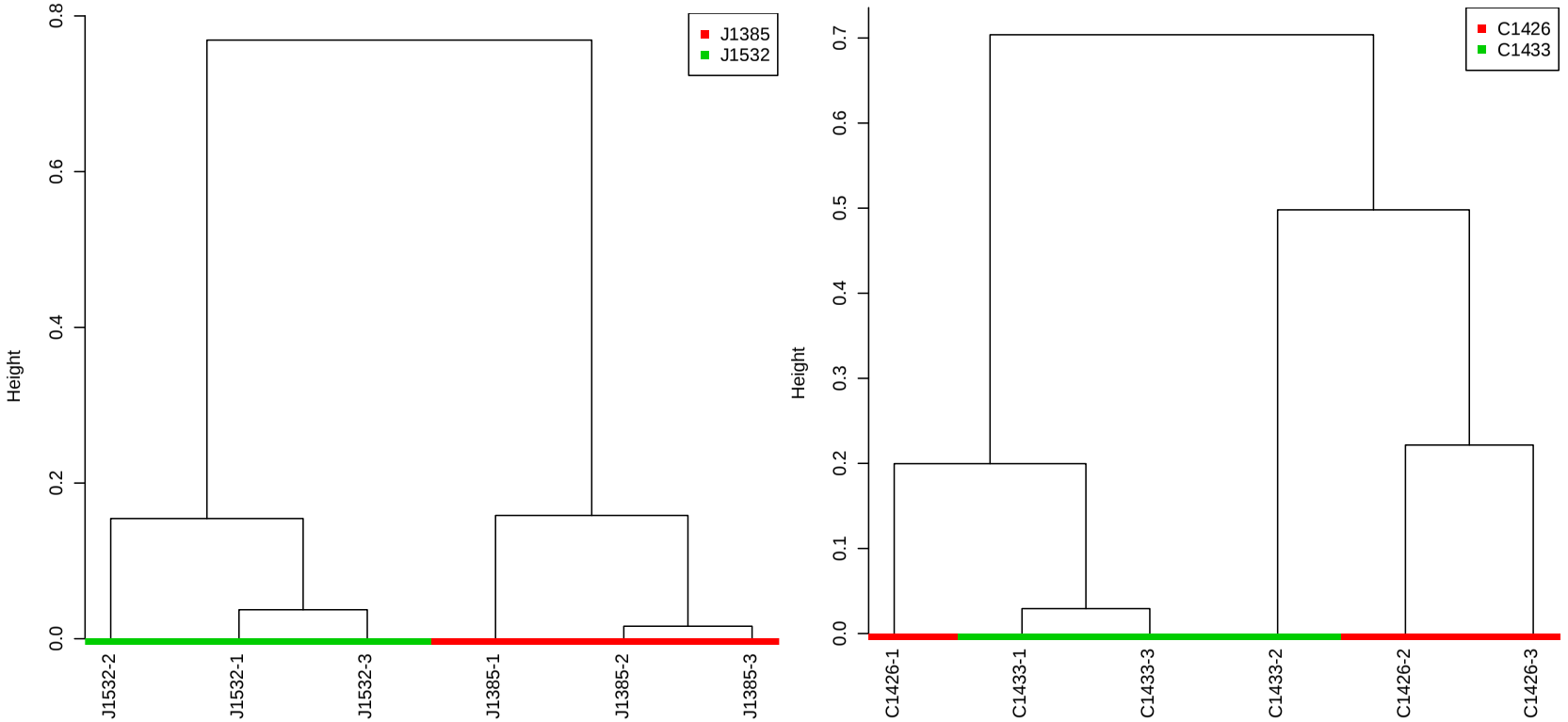
Following LC-MS, there is an enormous amount of data to be analysed. This includes combining positive and negative ion mode spectra and identifying spectra peaks each representing a single metabolite via comparison with a database with known peaks. This initial data analysis, along with LC-MS analysis, was performed within the University of Strathclyde's metabolomics team led by Dr Gavin Blackburn. Following this initial analysis, we received our metabolomics data in a .xlm file format, containing a complete list of peaks (metabolites) for each strain and their respective replicates. This file was visualised and data manipulated using IDEOM, an Excel interface used for analysing LC-MS based metabolomics data (Creek *et al*, 2012). Within this file, a variety of macros are enabled to allow users to identify any significant changes in metabolite concentration/presence between strains and perform basic statistical analysis on this data.

Before commencing upon a significant study of our newly acquired metabolomics data we sought to assess the quality of this data to ensure that any subsequent analysis was not in vain. From the literature, we used an online set of tools designed for analysis of high-throughput metabolomic processing including statistical analysis of LC-MS outputs called MetaboAnalyst 2.0 (Xia *et al*, 2012).

We have utilised the dendrogram drawing function of MetaboAnalyst 2.0 to create images illustrating the clustering of the metabolic profile of each biological replicate for each strain. Analysis of our first non-mucoid/mucoid pairing, J1385 and J1532, has shown that there is a significant degree of homology between the replicates of each strain (Figure 6.2). Although there appears to be one outlier for each of the strains, the apparent distance between these replicates and their respective counterparts is relatively small and it is not expected that these differences will have any impact on subsequent analysis. Figure 6.2 also demonstrates that there is a significant difference in the metabolic profiles between our J strains suggesting that data

obtained may be useful for elucidating any metabolic changes that may occur as a result of host adaptation and accompanying genomic alterations.

In contrast to the dendrogram produced for our J pairing, the tree showing the similarity between replicates for our second non-mucoid/mucoid pairing, C1426 and C1433, demonstrates that these replicates cannot be considered fully homologous. We can see that two out of three replicates for each strain are highly similar, both strains have one significant outlier that actually appears to be more similar to its opposite strain. When performing subsequent analysis into the metabolic profiles of these strains, it may be sensible to consider discounting these outlying replicates to ensure that their variation does not skew potential observations. However, despite this anomaly, as with the J pairing, Figure 6.2 demonstrates that there is a significant difference between the metabolic profiles of C1426 and C1433 and therefore, further analysis may prove worthwhile.



Firgure 6.2 Dendrogram images demonstrating the homology between replicates of each strain and the homology observed between non-mucoid/mucoid pairings.

6.4 Identification of changes to the metabolome in response to host adaptation

Based on IDEOM data, LC-MS analysis on our strain collection has identified approximately 4600 distinct metabolites present in at least one strain included. Of these, only ~2400 have a putative identification. This represents a significant amount of data to trawl through considering there are five samples and an uninoculated control to scrutinise. This vast amount of data can be significantly trimmed by discounting small di-, tri- and tetra-peptides which comprised a large proportion of metabolites detected, approximately 1000, and will have little significance on the findings of our study. In a further attempt to significantly reduce our workload and improve our chances of identifying any significant changes between our non-mucoid/mucoid pairs, we have further utilised MetaboAnalyst 2.0. This tool suite includes fold change analysis wherein, the metabolite profile of the mucoid isolate can be compared that of the non-mucoid strain. Metabolites with significant (>2.0) fold change can easily be visualised graphically (Figure 6.3).

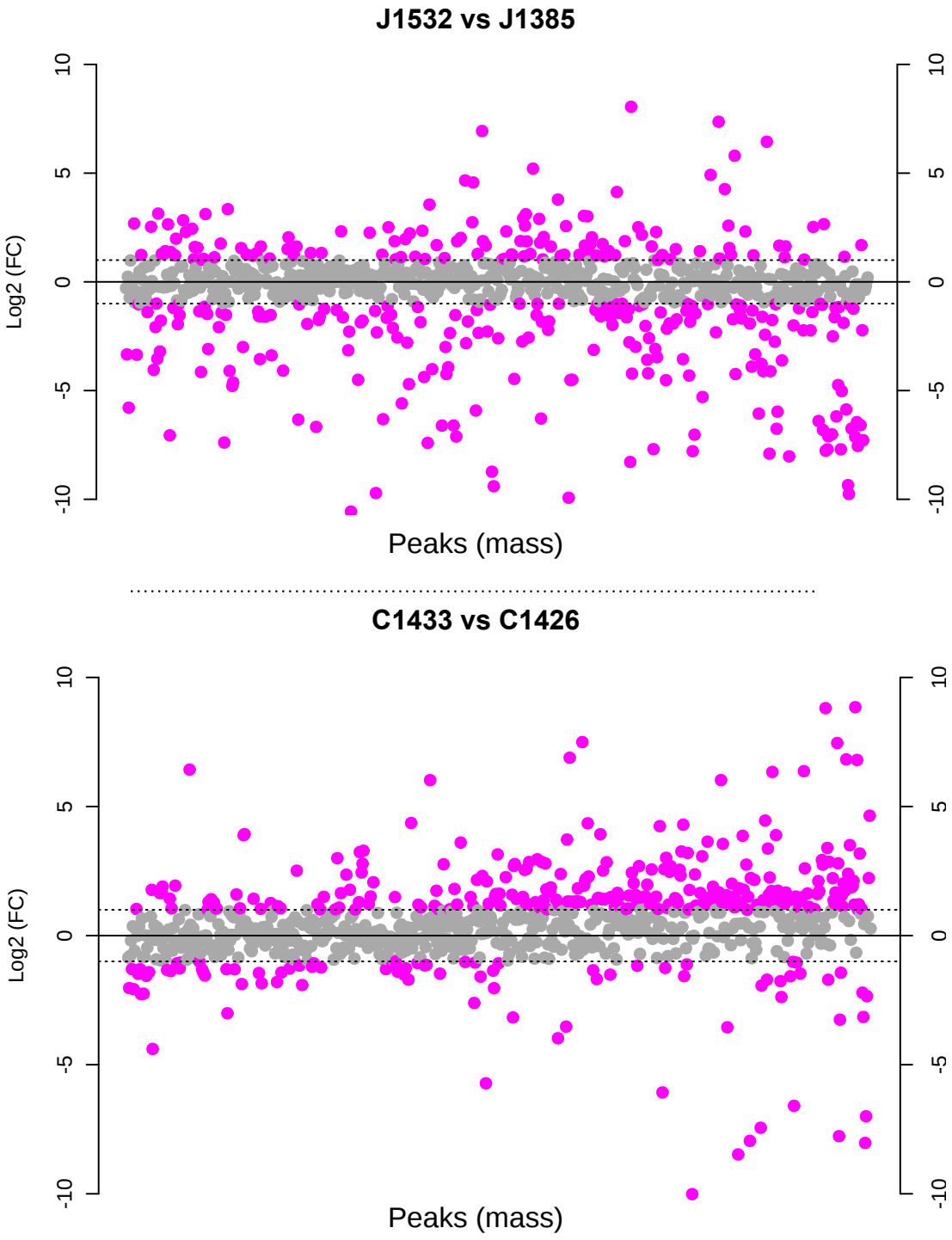

J1532 vs J1385
Log2 (FC)
10
5
0
-5
-10
Peaks (mass)
C1433 vs C1426
Log2 (FC)
10
5
0
-5
-10
Peaks (mass)

Figure 6.3 demonstrates the range of metabolites that are considered either up- or down-regulated when comparing J1532 and C1433 to their non-mucoid counterparts J1385 and C1426, respectively. A very crude analysis, this graphical representation of the fold change between metabolites as generated by MetaboAnalyst 2.0, does not include statistical analysis and therefore, metabolites with no statistical significance may be included. This was evident upon trawling through the 350 metabolites included. Many of the metabolites that were considered later turned out to be nonsignificant (p-value >0.5). There was a need to perform further analysis to exclude these metabolites, establishing a more manageable dataset of significantly up- or down-regulated metabolites to consider. MetaboAnalyst 2.0 has yet another available analysis that allows the user to exclude nonsignificant metabolites.
Using our entire, untrimmed dataset, we identified the 25 most statistically significant altered metabolites identified for each of our strain pairs as displayed using separate heat maps for J and C pairings (Figures 6.4 and 6.5, respectively).

As can be seen from these figures, the majority of metabolites that are significantly altered between the strains of both pairs have no predicted identification. This is obviously a problem when trying to attempt our aim of matching changes in the metabolome to changes in the genome as previously planned. This, however, was to be expected as, as previously stated, a large proportion of metabolites have no putative identification and therefore, to streamline analysis, should be excluded as should small, insignificant di-, tri- and tetra peptides.

6.4.1 Changes in the metabolome of J1532 can be attributed to genomic alterations affecting the stress response sigma factor AlgU and its anti-sigma factor MucA

After excluding unidentified metabolites and small peptides, only two of the top 25 metabolites of the J pairing have been putatively identified and will be considered, 2-Octenoic acid and L-Histidinol. A review of the literature to date reveals that 2-Octenoic acid, or its chemical formula of $C_8H_{14}O_2$ does not share any similarity with any identified metabolite of *P. aeruginosa*. Figure 6.4 shows that L-histidinol is significantly down-regulated in the mucoid J1532. L-histidinol can be utilised by bacteria as a carbon source via oxidation to l-histidinaldehyde then l-histidine via the action of the enzyme histidinol dehydrogenase (HDH) (Barbosa *et al*, 2002). The lower concentration of L-histidinol in the extract of J1532 compared to J1385 suggests that this strain is more efficient at utilising this metabolite from its environment. The question is, can this be attributed to changes in the genome of this strain?

Interestingly, it has previously been shown that the HDH enzyme responsible for utilising L-histidinol in *E. coli* shows a high degree of homology with AlgD of *P. aeruginosa* (Deretic, Gill and Chakrabarty, 1987). It is therefore expected that the increased utilisation of L-histidinol by J1532 is due to either increased activity of AlgD or increased transcritpional activation of the *hisD* gene responsible for HDH production as a result of increased AlgD-targeting effector proteins. This protein is the first effector in the highly complex alginate biosynthesis operon, the operon responsible for the appearance of the mucoid phenotype in *P. aeruginosa* strains, including J1532 (Qui *et al*, 2007). AlgD is directly under the control of the stress-response sigma factor AlgU that, in turn, is controlled by its anti-sigma factor MucA. We have previously demonstrated in Chapter 8 that, as a result of host adaptation, J1532 has acquired changes in *mucA*. Therefore, we hypothesise that this genomic anomaly in J1532 *mucA* may contribute to the increased L-histidinol utilisation observed (Figure 6.4) in this strain compared to its non-mucoid

counterpart J1385 by means of increased activation of the histidinol dehydrogenase HisD or its homologue AlgD.

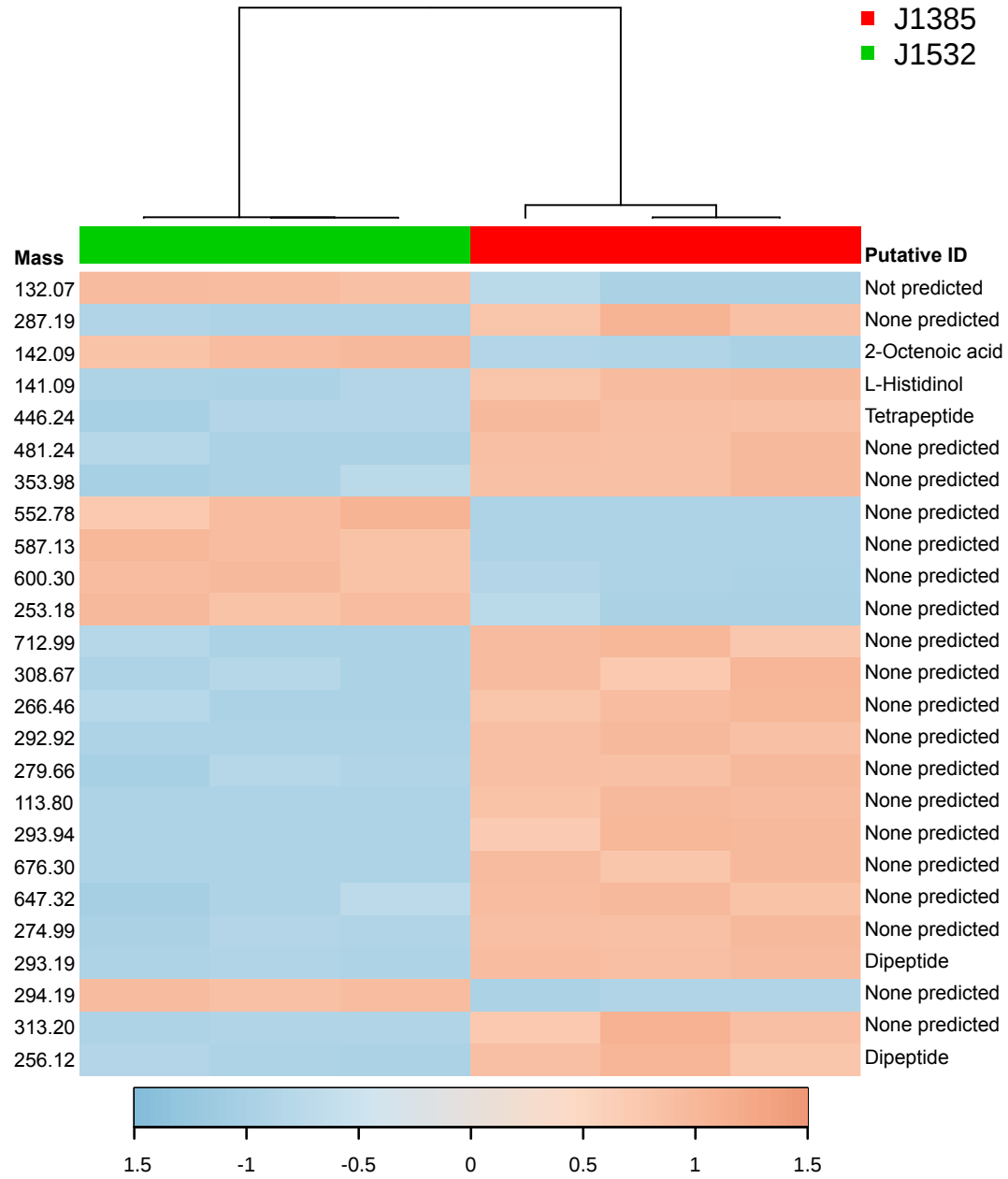


Figure 6.4 Heat map showing the 25 most significantly up- or down-regulated metabolites based on t-test analysis between the non-mucoid J1385 and its mucoid counterpart J1532. The mass for each of these metabolites and their putative identification are listed adjacently. Also included is a dendrogram showing the homogeneity between replicates as previously demonstrated in Figure 6.2.

6.4.2 Changes in the oligosaccharide Δ4,5-unsaturated trigalacturonate may be attributed indirectly to a SNP in *wspF*

Applying similar rules to the C pairing as we did the J pairing, excluding unidentified metabolites and small peptides, leaves us with only five of the top 25 metabolites to consider (Figure 6.5). Of these, Adenosine Diphosphate (ADP), due to its frequent occurrence in metabolism and its essential role in many important cellular pathways mean that mapping any change in this metabolite to any specific genomic change is impossible and was investigated no further.

The first metabolite that is significantly increased in C1433 was Δ4,5-unsaturated trigalacturonate. In fact, this metabolite was not detected in any of the replicates of C1426 but is present in all C1433 replicates (40825 ± 1901 [mean intensity ± standard deviation]) This oligosaccharide, although poorly understood, is catabolised by the enzymatic reaction of PelA (BioCyc, 2014). PelA, an effector of the *pel* operon, is involved in both early and late stages of biofilm formation (Vasseur *et al*, 2004). This led us to try to identify any link between the increased biofilm formation and the increased levels of Δ4,5-unsaturated trigalacturonate that are shared by C1433. The literature revealed no strong direct link between each of these observations. However, as Colvin *et al* (2013) showed, both *pelA* expression and biofilm formation are increased when introducing an in-frame mutation in *wspF*, the negative regulator of the diguanylate cyclase WspR. As has previously been shown in Chapter 4 (Table 4.7), C1433 has acquired a non-synonymous SNP in *wspF*. We, therefore, predict that this SNP may also result in increased *pelA* expression as Colvin *et al* (2013) have previously shown, promoting subsequent accumulation of the oligosaccharide Δ4,5-unsaturated trigalacturonate as observed in our metabolomic analysis (Figure 6.5). Unfortunately, there is no evidence to suggest that increased Δ4,5-unsaturated trigalacturonate concentrations will confer any selective advantage to our strain over its non-mucoid counterpart C1426. Instead, it is expected that the accumulation of this metabolite occurs only as a byproduct

of the increased biofilm formation conferred by the mutation observed in *wspF*.

The remaining three metabolites classed as significant by MetaboAnalyst 2.0 (Figure 6.5) do not appear to have any significance in regards to *P. aeruginosa* infection, at least in their current putative identifications. A search of the literature reveals that neither methyl 9-butylperoxy-10,12-octadecadienoate, varanic acid nor (9Z)-9-Octadecenamide have any isotopes to any *P. aeruginosa* metabolite and a search using their chemical formulas provides no suggestions. This highlights the need to consider all identifications made by IDEOM as strictly putative and without known standards being analysed simultaneously cannot be considered definitive.

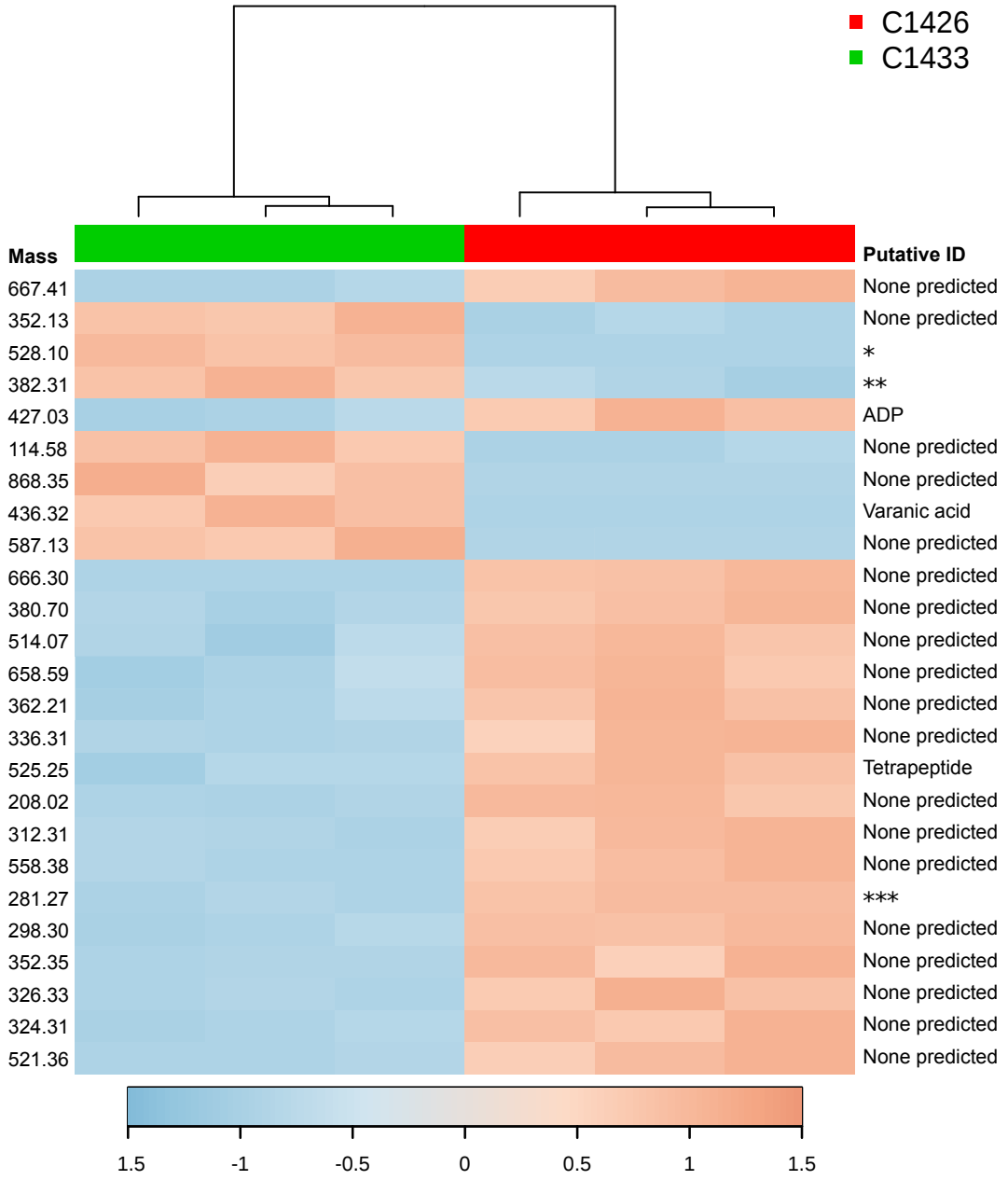


Figure 6.5 Heat map showing the 25 most significantly up- or down-regulated metabolites based on t-test analysis between the non-mucoid C1426 and its mucoid counterpart C1433. The mass for each of these metabolites and their putative identification are listed adjacently. Also included is a dendrogram showing the homogeneity between replicates as previously demonstrated in figure 6.2. Due to the large nature of the nomenclature of some metabolites, their names have been omitted and replaced with *. These are * Δ4,5-unsaturated trigalacturonate, ** methyl 9-butylperoxy-10,12-octadecadienoate and *** (9Z)-9-Octadecenamide.

6.4.3 A reduction in concentration of the osmo-protectant glycine-betaine may be attributed to the truncation of the anti-sigma factor MucA

During infection of the CF lung, *P. aeruginosa* is likely to experience a significant degree of osmotic stress. In response to this, the bacterium produces a variety of osmo-protectant buffers including proline, carnitine, choline and the predominant, glycine-betaine (Fitzsimmons, Hampel and Wargo, 2012). It was shown that for successful infection, *P. aeruginosa* maintains a pool of both glycine-betaine and choline and epletion of glycine-betaine severely inhibits bacterial growth under conditions of high salt concentration (Fitzsimmons, Hampel and Wargo, 2012). Mimicking chronic lung infection, metabolomic data from our strains somewhat contradicts this observation, in that both of our mutated mucoid strains have significantly (p-value <0.05) reduced concentrations of both glycine-betaine and choline (Figure 6.6). This would suggest that both J1532 and C1426 would struggle to establish a successful infection within the CF lung. This was not the case as, as has previously been described, both strains were isolated as the dominant strain from their respective hosts. Our data has revealed that instead of relying upon glycine-betaine and choline as osmo-protectant buffers, our mucoid strains have instead switched osmo-protectant dependence to proline and carnitine. Figure 6.6 demonstrates that J1532 and C1433 have significantly increased the concentrations of each of these alternative buffers to counteract decreased glycine-betaine and choline protection.

We have attempted to map these metabolomic changes to genomic alterations that have occurred as a result to host adaptation as per our overall aim. It has previously been described that for glycine-betaine mediated osmotic tolerance, *P. aeruginosa* must possess a functional copy of the MucA anti-sigma factor. Behrends *et al* (2010) demonstrated that the introduction of a stop codon at position 22 of the *mucA* gene, resulting in a truncated MucA protein, significantly reduces the relative concentration of

glycine-betaine compared to the WT PA01. Therefore, we predict that the mutations that have resulted in truncation of the MucA protein and appearance of the mucoid phenotype in both J1532 and C1433 will also have contributed to the reduced glycine-betaine concentrations exhibited by these strains compared to their non-mucoid counterparts. Unfortunately, despite extensive review of our genomic anomalies and the literature, no link between host adaptation and the altered concentrations of the remaining osmo-protectants, as per Figure 6.6, can be made. It would, however, be very interesting to look for this pattern in further non-mucoid/mucoid, host adapted pairs to see if such changes are synonymous with truncation of the MucA anti-sigma factor.

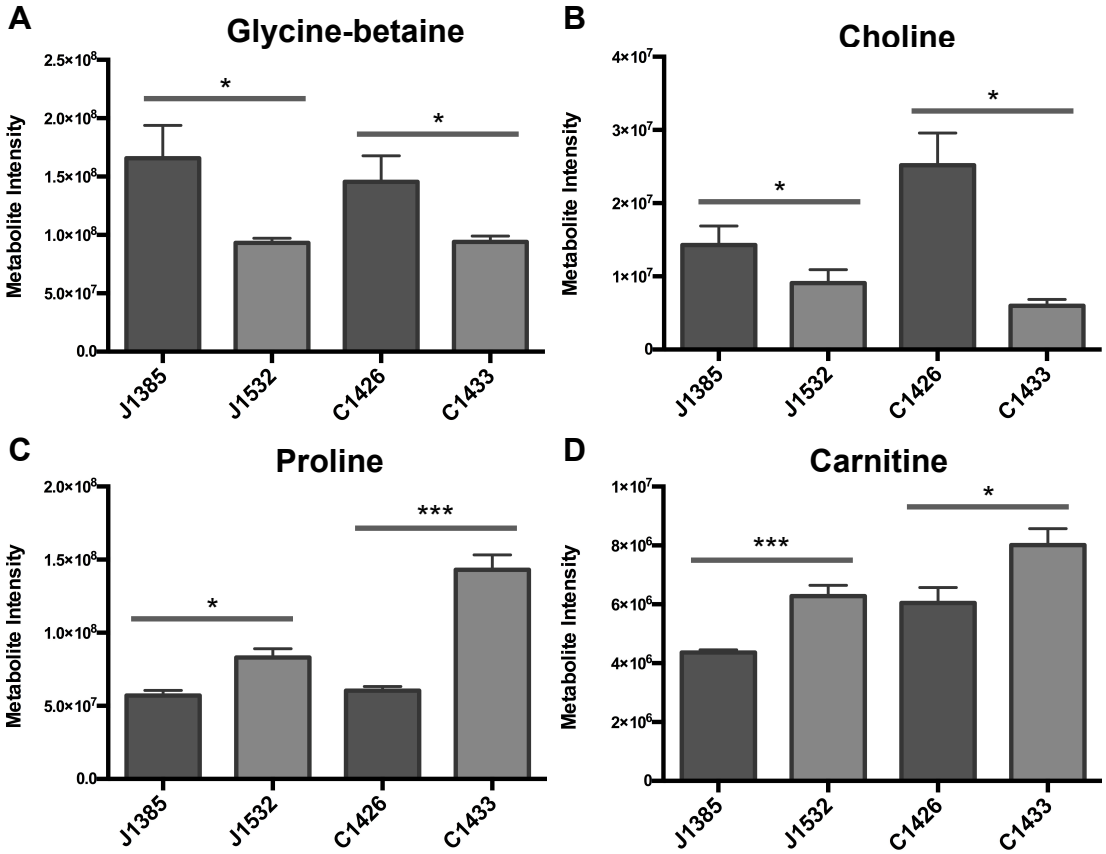


Figure 6.6 Graphs showing altered concentrations of a variety of osmo-protectants utilised by *P. aeruginosa*. Mucoid strains J1532 and C1433, compared to their non-mucoid counter parts J1385 and C1426, respectively, have demonstrated decreased **A** glycine-betaine and **B** choline concentrations and increased **C** proline and **D** carnitine concentrations.

6.4.4 The loss-of-function mutation previously identified in C1433 *ampR* may contribute to the significant reduction in the production of Rhamnolipid-2

Many bacteria, including *P. aeruginosa*, produce a class of biosurfactant glycolipids called rhamnolipids. *P. aeruginosa* has been shown to produce two main rhamnolipids; L-rhamnosyl-L-rhamnosyl-β-hydroxydecanoyl-β-hydroxydecanoate (rhamnolipid-1) and L-rhamnosyl-β-hydroxydecanoyl-β-hydroxydecanoate (rhamnolipid-2). Our metabolomics data has demonstrated that there is a significant, ~25 fold decrease in production of rhamnolipid-2 in our mucoid strain C1433 compared to its non-mucoid counterpart C1426 (Figure 6.7).

The role of these rhamnolipids in bacterial cell functioning is not well understood. However, it has been demonstrated that they are involved in virulence, biofilm dispersal, motility and possess antimicrobial properties (Chrzanowski, Lawniczak and Czaczyk, 2012). Production of both rhamnolipids is under control of the *rhlAB* operon which encodes two sequential rhamnosyltransferases. Transcriptional activation of this operon is dependent upon the presence of a functional RhlR regulatory protein (Ochsner, Fietcher and Reiser, 1994). RhlR is involved in the complex quorum sensing network of *P. aeruginosa*, with transcriptional activation being under the control of *lasR*. Transcriptional activation of *rhlR,* and subsequent rhamnolipid biosynthesis, is also under the control of the global regulator AmpR (Kong *et al*, 2005). In that study it was shown that insertional inactivation of *ampR* resulted in a significant reduction in transcriptional activation of *rhlR* compared to the WT. Our work, via previous genomic analysis of C1426 and C1433, has demonstrated a loss-of-function mutation in the *algR* global regulator via qPCR (Chapter 4). It is expected that, similar to the results observed by Kong *et al* (2005), *RhlR* transcription and subsequent rhamnolipid-2 production, might be reduced as a result of the non-synonymous SNP occurring in C1433 *ampR*.

It is expected that the purpose of host adaptation by *P. aeruginosa* is to confer a selective advantage to the strain allowing it to establish and maintain a successful infection in host. Although not well understood, an argument can be made for this reduction in rhamnolipid-2 production conferring a selective advantage to C1433 over C1426. Rhamnolipids are key virulence determinants of *P. aeruginosa*, having the ability to integrate with and damage host cells. For example, within the CF lung, rhamnolipids can disrupt cell membranes, promoting paracellular invasion and reducing mucociliary clearance (Zulianello *et al*, 2006). Previous work within our laboratory, utilising a *G. mellonella* infection model, has demonstrated that C1433 is significantly less virulent than its counterpart C1426 (Amy Ford, unpublished data). Reduced virulence might confer an advantage to C1433, helping the bacterium avoid host recognition and intervention in its attempt at maintaining a successful infection. Rhamnolipids are involved in promoting swarming motility by *P. aeruginosa* by means of decreasing surface tension (Glick *et al*, 2010). Again, work within our laboratory has demonstrated that compared to its counterpart, C1433 has reduced motility (Amy Ford, unpublished data). The reduced motility as a result of reduced rhamnolipid production may aid in establishment of infection by allowing C1433 to attach to host surfaces, initiating biofilm formation thereby conferring an advantage to the strain. Rhamnolipid production is further involved in biofilm development at the dispersal stage. Fluorescence microscopy was used to demonstrate that immediately prior to dispersal, induction of the *rhlAB* operon occurs (Boles, Thoendel and Singh, 2005). Increased biofilm attachment and persistence via decreased motility and dispersal, respectively, as a result of decreased rhamnolipid production may account for the increased biofilm formation exhibited by C1433 and C1426 as previously demonstrated in Figure 4.1. This increased biofilm formation might confer a selective advantage to C1433 by promoting a more persistent infection via reduced immune and antibiotic intervention.

Therefore, collectively, the altered phenotypes resulting from decreased rhamnolipid production in C1433 including decreased virulence and motility as well as increased biofilm formation, will confer a selective advantage to this strain. Together, these phenotypes should increase persistence via reduced host and medical intervention.

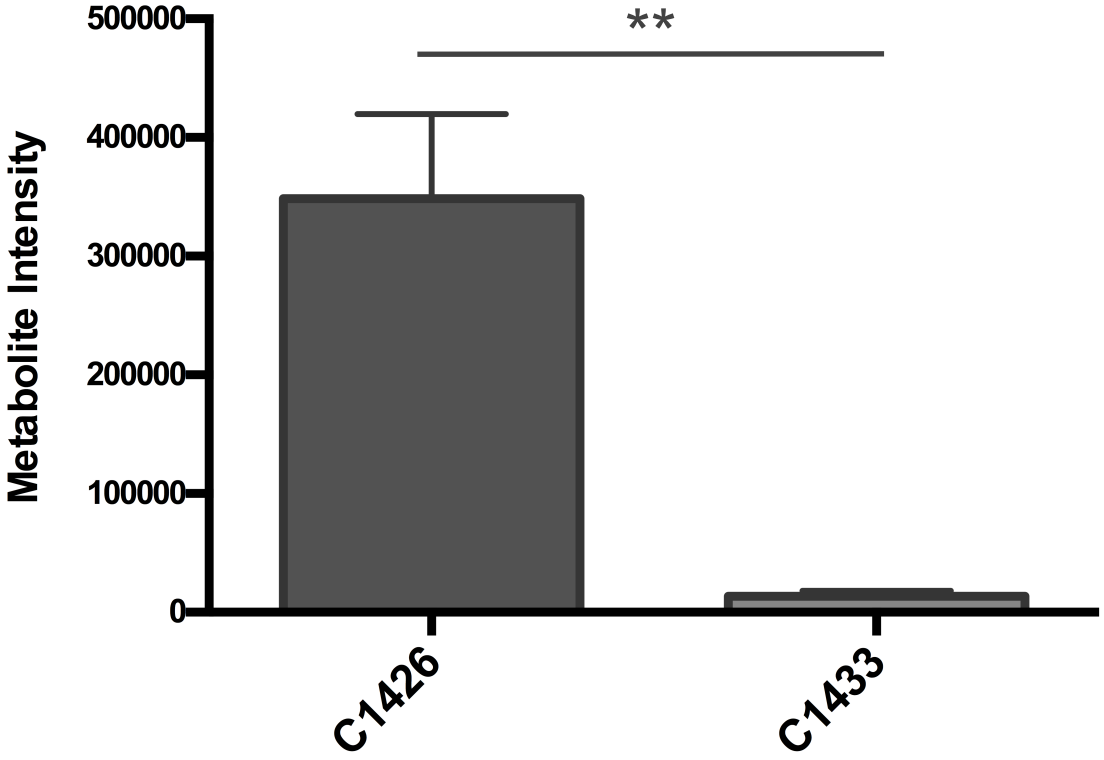


Figure 6.7 Graph showing the difference in rhamnolipid-2 production between the non-mucoid C1426 and its mucoid counterpart C1433. The ~25 fold difference in metabolite intensity may be attributable to a single non-synonymous SNP in the global regulator *ampR*.

6.5 Chapter 6 discussion

Using our non-mucoid/mucoid paired CF isolates J1385/J1532 and C1426/C1433, this chapter has sought to map changes in the metabolome to changes in the genome that have occurred as a result of host adaptation. We have attempted to replicate growth conditions that these strains would experience within the CF lung, utilising prolonged incubation in Artificial Sputum Medium (Kirchner *et al*, 2012). Whole cell extract of these cultures were analysed by LC-MS in an attempt to detect all metabolites present. We included two controls in our LC-MS analysis in the form of a negative blank media control and the reference strain PA14. With hindsight, if our experiment was to be repeated, these controls would be excluded due to their analysis proving insignificant and of little help to our efforts. Instead, we would seek to include a higher number of replicates of our test strains to provided more statistically significant results.

To ensure that metabolite identification and mapping was performed with confidence, two quality control checkpoints were incorporated into our workflow. Prior to LC-MS analysis, TLC was used to choose the three most similar replicates out of six for each strain (Figure 6.1). Post LC-MS, it was demonstrated, using dendrogram drawing (Figure 6.2) that there was a significant degree of homology between replicates of each of our strains. Quality control ensured that subsequent analysis and any potential metabolite heterogeneities between strains should be considered significant. Initial analysis revealed that approximately 4600 distinct metabolites were identified as present in at least one of our strains. Of these, only ~2400 were putatively identified of which, approximately 1000 were insignificant di-, tri- or tetra-peptides and could be excluded. This represents a substantial dataset to be mined through and analyses were sought to highlight only those metabolites showing a degree of significance. Using MetaboAnalyst 2.0, we generated heat maps showing the 25 most statistically significant up- or down-regulated metabolites based on t-test analysis identified between non-mucoid strains J1385 and C1426 and their mucoid counterparts J1532 and

C1433, respectively (Figures 6.4 & 6.5). From this analysis, it was observed that the majority of metabolites identified as significant had no putative identification and therefore had to be discounted. Despite this, we have managed to map several metabolite changes to genomic adaptations previously identified in Chapters 4 and 5.

It was noted that J1532 has experienced reduced L-histidinol production compared to its counterpart J1385. It has previously been shown that the enzyme responsible for the catabolism of this metabolite, histidinol dehydrogenase, shows a high degree of homology to *P. aeruginosa* AlgD (Deretic, Gill and Chakrabarty, 1987). AlgD is directly under control of the stress-response sigma factor AlgU that, in turn is regulated by its anti-sigma factor MucA. An InDel present within *mucA* has led us to predict that the resulting truncated protein is no longer able to exert its inhibiting effect on AlgU, allowing its constitutive expression. Therefore, we hypothesise that the genomic changes that have occurred in *mucA* may lead to the increased L-histidinol utilisation observed in J1532 compared to its non-mucoid counterpart J1385 by means of either increased activation of the histidinol dehydrogenase protein HisD or its homologue AlgD (Figure 6.8).

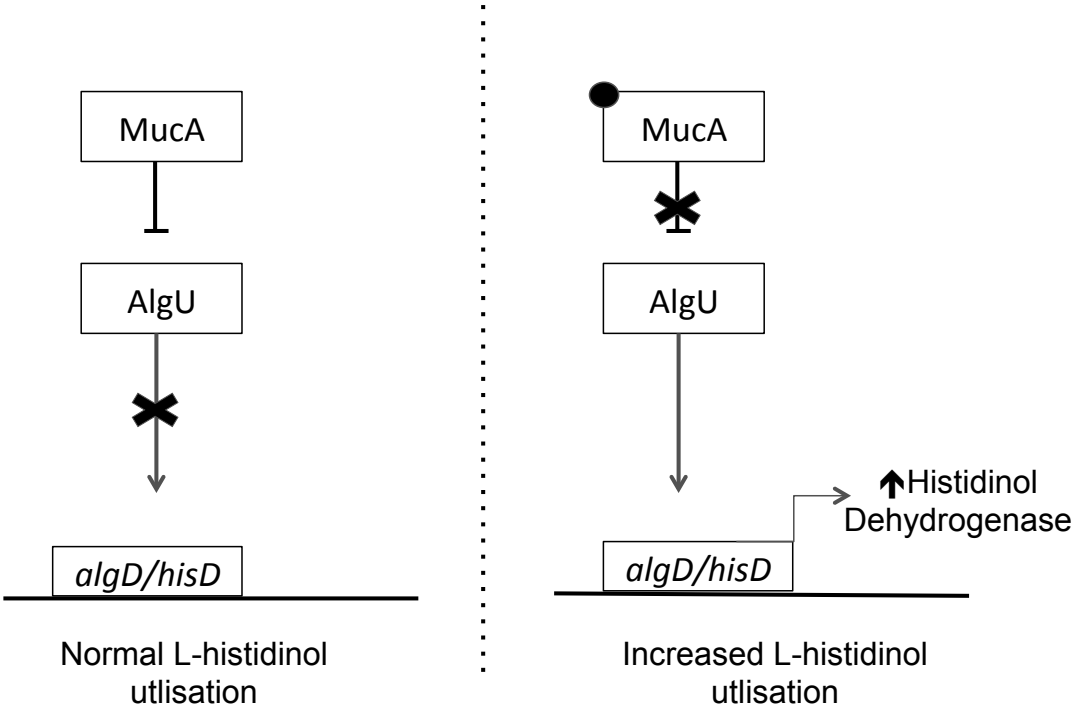


Figure 6.8 In WT *P. aeruginosa*, L-histidinol utilisation is controlled by normal expression of the AlgU sigma factor that is responsible for promoting transcription of AlgD, a homologue of the histidinol dehydrogenase-encodning enzyme HisD. In J1532, as a result of truncation of the anti-sigma factor MucA, AlgU is constitutively expressed resulting in increased AlgD activity. Our hypothesis states that this increase in AlgD results in increased histidinol dehydrogenase activity and therefore, increased utilisation of L-histidinol accounting for the reduced concentration observed in this strain.

We further utilised our heat maps to identify a metabolite present in C1433 but not C1426. This metabolite, the oligosaccharide Δ4,5-unsaturated trigalacturonate, is catabolised by the enzymatic reaction of PelA (BioCyc, 2014). Based on this, we expect that the increase in this metabolite is due to increased transcriptional activation of the *pelA* gene. A review of the literature has revealed that an in-frame mutation in *wspF* is sufficient to induce increased expression of *pelA* as well as increased biofilm formation (Colvin *et al*, 2013). As previously shown in Chapter 4 (Table 4.7), *wspF* of C1433 has acquired a single non-synonymous SNP. We therefore hypothesise that this polymorphism is responsible for the accumulation of Δ4,5-unsaturated trigalacturonate in C1433 as a result of increased *pelA* expression. In time, it would be interesting to test this hypothesis with the use of qPCR to confirm the inactivation of *wspF* and the resulting increased expression of the *pelA* gene. Unfortunately, the literature revealed no evidence to suggest that increased Δ4,5-unsaturated trigalacturonate concentrations will provide any selective advantage to our strain over its predecessor C1426. Instead, it is expected that the accumulation of this metabolite is merely a “side effect” of the increased biofilm formation conferred by the mutation observed in *wspF* (Figure 6.9).

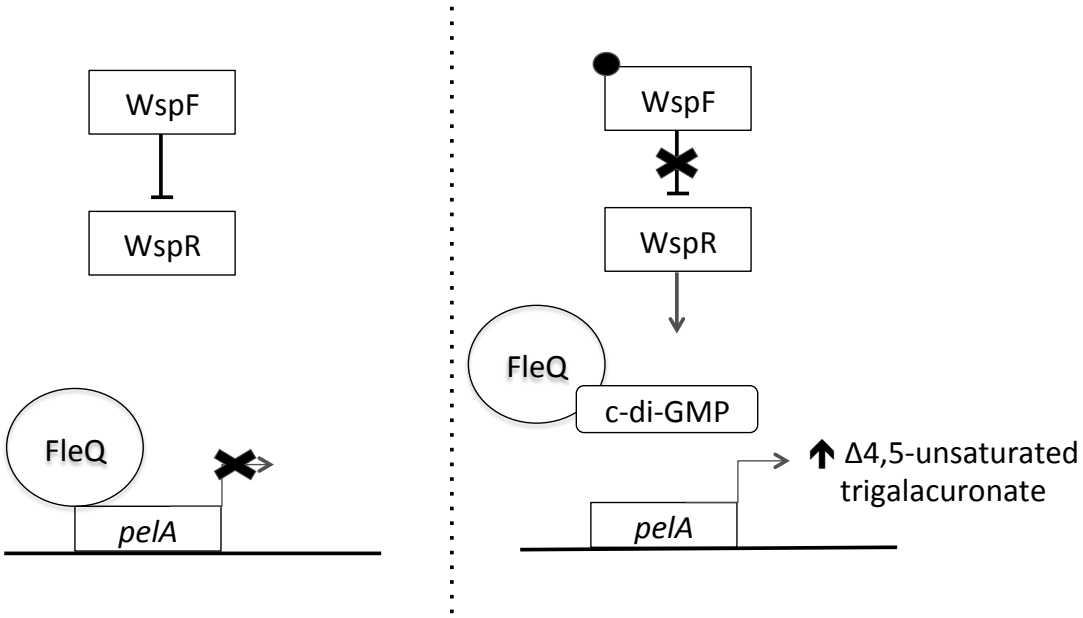


Figure 6.9 In WT *P. aeruginosa*, *pelA* transcription is negatively controlled by FleQ, resulting in normal catabolisation of Δ4,5-unsaturated trigalacturonate. In C1433, we predict that a SNP in *wspF*, the negative regulator of WspR, results in increased c-di-GMP production, subsequent inactivation of FleQ and transcription of *pelA*. The resulting increased PelA is expected to account for the increased Δ4,5-unsaturated trigalacturonate concentration exhibited by C1433.

Our heat maps, due to the high proportion of unidentified metabolites, did not provide much success in our attempt at mapping metabolite change to genomic changes occurring as a result of host adaptation. We switched the focus of our attention to our genomic observations outlined in Chapters 7 and 8. Using this data, based on previous observations in the literature, we have attempted to predict any resulting metabolite changes and link these to changes occurring in the metabolomic profile of our strains. Using this new tactic we have identified a variety of changes in the genome that have potentially resulted in changes in the metabolome. This includes a switch in osmo-protectant buffer preference from glycin-betaine and choline to proline and carnitiane by both J1532 and C1433. No links can be made to map the changes in choline, proline and carnitine to our genomic anomalies. However, it has previously been shown that for sufficient glycine-betaine mediated osmotic tolerance, *P. aeruginosa* must possess a functional copy of the MucA anti-sigma factor (Behrends *et al*, 2010). As both J1532 and C1433 have acquired InDel mutations resulting in truncation of this protein it is postulated that the absence of a functional MucA will contribute to the reduced glycine-betaine concentrations exhibited by these mucoid strains.

Finally, we have also identified that a ~25 fold decrease in the production of the biosurfactant rhamnolipid-2 by C1433 may be accounted for by a single SNP occurring in *ampR*. This hypothesis was formulated based on the requirement of a functional RhlR regulatory protein for rhamnolipid production (Ochsner, Fietcher and Reiser, 1994), the transcription of which is under the control of the global regulator AmpR (Kong *et al*, 2005). This reduction of rhamnolipid-2 production by C1433 will confer a selective advantage to this strain in several ways. As rhamnolipids are key virulence determinants of *P. aeruginosa*, involved in disrupting host cell membranes and reducing mucociliary clearance (Zulianello *et al*, 2006), reduced production, and therefore virulence, will aid the bacterium mask itself from the host preventing immune intervention thereby increasing persistence. Further to this, high concentrations have also been shown to aid bacterial

motility by means of reducing surface tension (Glick *et al*, 2010). This will allow the strain to attach to host surfaces, initiating biofilm formation further aiding strain persistence. Biofilm formation is further increased by rhamnolipid reduction by preventing biofilm dispersal (Boles, Thoendel and Singh, 2005). Each of these altered phenotypes resulting from decreased rhamnolipid have been observed in C1433. For example, this strain is less virulent (Amy Ford, unpublished data), less motile (Amy Ford, unpublished data) and more proficient at biofilm formation (Figure 4.1) compared to its predecessor C1426.

Therefore, we have, to a certain degree, managed to achieve our aim of mapping changes in the metabolome to changes in the genome of our non-mucoid/mucoid pairs. Obviously, these observations must be taken as strictly provisional as no confirmatory analyses have been performed. It would also be desirable to repeat analysis with an increased number of replicates. It is important to remember that metabolite identification must be assumed as putative unless known standards are ran concurrently with the analysis. If these measure are taken we are in no doubt that metabolomics will provide invaluable in attempts at elucidating the effect that genomic alterations have on the phenotype of a particular *P. aeruginosa* strain.

Chapter 7 - General discussion

7.1 Contributing to the genomic understanding of *P. aeruginosa*

There is a real need to develop a greater understanding of the genomic basis of the clinically significant and environmental organism *P. aeruginosa*. At the inception of this project, only four strains of this species had their genomes fully sequenced; PA01 (Stover *et al*, 2000), PA14 (Lee *et al*, 2006), LESB58 (Winstanley *et al*, 2009) and PA7 (Roy *et al*, 2010). We aimed to contribute significantly to this limited understanding via sequencing of a further twelve host adapted and environmental *P. aeruginosa* genomes (Table 3.3). Included are four clinical pairs, three non-mucoid early CF lung isolates and their subsequently isolated mucoid pairs as well as PA17 and its lab evolved small colony variant. Four environmental strains have also been sequenced to provide the project with a significant aspect of strain diversity. Using next-generation sequencing technologies and the latest bioinformatic tools, we have successfully met this aim, publishing the draft sequences of these twelve diverse strains (Stewart *et al*, 2013). This data was subsequently used to map the core genome phylogeny of the *P. aeruginosa* species and to evaluate the role of host adaptation during infections of the cystic fibrosis lung.

7.2 Core genome phylogeny of *P. aeruginosa*

The majority of phylogenetic analyses of *P. aeruginosa*, and bacterial populations in general, currently focus on the widely used Multilocus Sequence Typing technique (Jolley *et al*, 2004). This analysis is extremely limited in its use in predicting the population structure of *P. aeruginosa*. MLST bases its analysis on only ~0.12% of the genome, meaning many significant genomic anomalies may be overlooked. There is a need to develop a more thorough and exact strain typing method. We have devised a new typing method, utilising Panseq (Laing *et al*, 2010) that considers only

the genes present in the core genome of this species, based on all 55 available genomes submitted to GenBank. The core genome was defined as the “collection of all genes that exist in all 55 strains analysed with no exceptions and no less than 90% homogeneity across these strains”. This ensures that no genes that may have been horizontally acquired will be included in the analysis.

Our new, novel method for *P. aeruginosa* typing has demonstrated that 3 distinct groups exist within this species (Figure 3.4). The largest of these groups, Group 1, is comprised of 75% of the strains analysed and contains many notable strains including DK2, PA01, LESB58 and PAK. The second major group, Group 2, contains only 12 strains, approximately 22% of analysed strains, including the highly virulent PA14, a strain featured significantly throughout this study. The third group, Group 3, containing only 2 strains, PA7 and VRFPA01, lies at a striking distance compared to the two major groups. Almost 40% of the total SNPs identified by Panseq are accounted for by the two strains existing within this group, a significant amount demonstrating how outlying these strains truly are. This analysis failed to identify any pattern regarding grouping and the location from which they were identified. This is partly due to the small sample size of only 55 strains. It is important to remember that such a small pool of genomes may not provide a true reflection of the core genome phylogeny of *P. aeruginosa* and any observations made must be considered with caution. However, based on initial results, we believe that this analysis has contributed significantly to the limited understanding of the species structure described in previous studies (Pirnay *et al*, 2009, Kiewitz & Tummler, 2000 and Wiehlmann et al, 2007) and in time will be a very useful method to type and group strains as more genomes become publicly available to the wider *P. aeruginosa* genomics community.

We have used our new data, alongside the traditional MLST technique, to analyse the relationships existing between our twelve sequenced strains. We

have demonstrated that we the non-mucoid/mucoid pairings of C1426/C1433 and J1385/J1532 are true isogenic, host adapted pairs (Figures 3.2, 3.3 and 3.4). Our third non-mucoid/mucoid pair, C763 and C1334, were distinct from one another demonstrating that the switch in strain dominance in the host occurred as a result of strain succession rather than host adaption. Alternatively, it is important to consider that it is possible that both strains may have been present within the patient and culture failed to isolate both strains. Based on these results, we have selected our two isogenic pairings to investigate the role of host adaptation in infections of the CF lung.

7.3 Host adaptation of *P. aeruginosa* strains infecting the CF lung

Host adaptation is a phenomenon employed by bacteria as a mechanism of survival and growth in hostile environments such as the CF lung. Particularly adept at this is *P. aeruginosa*, displaying an ability to hypermutate in response to internal and external challenges such as host immunity and antibiotic intervention (Folkesson *et al*, 2012). This process is essential to the success of the bacterium within the CF lung, resulting in a variety of phenotypic changes that result in the switch from an acute to chronic infection. Included in these changes are deflagination, increased antibiotic sensitivity and perhaps most commonly and significantly, a switch from a non-mucoid to a mucoid phenotype as previously detailed (Hassett *et al*, 2009). As mentioned, we have undertaken a significant study into the effects of host adaptation on the genomic structure of two non-mucoid/mucoid CF isolated pairings. We have attempted to and successfully mapped a variety of phenotypic changes displayed by the late mucoid strains compared to their non-mucoid counterparts to changes occurring in their genomes. To demonstrate this, we have utilised a variety of phenotypic analyses including virulence models, biofilm growth assays, antibiotic resistance assays and a large-scale metabolomic analysis.

7.3.1 The switch to the mucoid phenotype can be attributed to InDel polymorphisms occurring within the anti-sigma factor *mucA* of both C1433 and J1532

The most striking phenotype that has arisen as a result of host adaptation within the CF lung is the switch to a mucoid phenotype experienced by C1433 and J1532 compared to their progenitor strains C1426 and J1385, respectively. Previous studies have revealed that this switch can be attributed to inactivation of the *mucA* anti-sigma factor resulting in uncontrolled AlgU sigma factor activity and overproduction of alginate (Govan and Deretic, 1996). Mutations affecting this gene occur in approximately 80% of mucoid *P. aeruginosa* strains (Boucher *et al*, 1997). Our genomic analysis reveals that both C1433 and J1532 have experienced truncation, and consequently attenuation, of the MucA protein as a result of the acquisition of InDel polymorphisms and introduction of premature stop codons. We therefore hypothesise that truncation and deactivation of this MucA protein in both C1433 and J1532 will result in uncontrolled expression of the AlgU sigma factor, overproduction of alginate and the appearance of the mucoid phenotype.

Interestingly, our analysis has also revealed nucleotide polymorphism have occurred in the *algU* sigma factor gene of both mucoid strains. Polymorphism prediction by CLC Genomics Viewer has surprisingly suggested that an InDel has occurred in *algU* of C1433 resulting in truncation of its protein product. If correct, this InDel would result in attenuation of the AlgU, inhibiting alginate biosynthesis. Obviously, due to the mucoid phenotype of C1433 this cannot be the case as without a functioning AlgU, alginate biosynthesis and the mucoid switch could not occur. Closer inspection reveals that this polymorphism is only present in seven out of twenty-two reads and therefore, this should not be considered a true InDel (Figure 4.13). In J1532, *algU* has experienced a nonsynonymous SNP resulting in a L62F amino acid change. We hypothesised that this SNP resulted in constitutive expression of AlgU resulting in uncontrolled alginate biosythesis and, consequently, the

appearance of the mucoid phenotype. To the best of our knowledge, no mutation of this gene has ever been reported to be sufficient for emergence of the mucoid phenotype. A review of the literature did however reveal that it has previously been noted that point mutations affecting this gene have arisen following MucA knockout mutations resulting in the appearance of a mucoid phenotype (DeVries and Ohman, 1994 and Schurr *et al*, 1994). It was suggested that these second-site suppressor mutations are an attempt by the bacterium to autoregulate, promoting the switch back to a non-mucoid phenotype. Therefore, it is predicted that this mutation may be an attempt by the bacteria to counteract the energy expending switch to the mucoid phenotype and is no longer expected to have any role in alginate overproduction.

Through its role of negatively regulating the sigma factor AlgU, MucA can be considered a global regulator involved in the regulation of a variety of *P. aeruginosa* processes (Jones *et al*, 2010). Using our large-scale metabolomic analysis, we attempted to attribute changes in the metabolome to the polymorphisms identified in the *mucA* genes of C1433 and J1532. In J1532, we have identified that, compared to J1385, L-histidinol is significantly down-regulated (Figure 6.4). This suggests that J1532 is more proficient at utilising this metabolite via increased metabolism by the HDH enzyme. This enzyme, in *E. coli*, shares a high degree of homology with AlgD of *P. aeruginosa* (Deretic, Gill and Chakrabarty, 1987). This protein is the first effector in the alginate biosynthesis process and is directly under the control of the sigma factor AlgU. Therefore, truncation of the anti-sigma factor MucA in J1532 as described may result in uncontrolled expression of AlgU that, in turn, promotes transcription of *algD*. We hypothesise that the increased L-histidinol utilisation observed in this strain occurs as a result of increased activation of the histidinol dehydrogenase homologue AlgD due to inactivation of the MucA global regulator.

Metabolomic analysis also revealed that a switch in osmo-protectant buffers utilised by both C1433 and J1532 has occurred. This includes a significant reduction in the production of the primary buffer glycine-betaine by both mucoid strains compared to their non-mucoid counterparts (Figure 6.6). The literature has revealed that for glycine-betaine mediated osmotic tolerance, *P. aeruginosa* must posses a functional copy of the MucA sigma factor (Behrends *et al*, 2010). As described, both C1433 and J1532 posses only truncated and nonfunctional copies of this protein suggesting that the reduction in glycine-betaine exhibited by both mucoid strains can be attributed to the InDels present in their copies of the *mucA* gene. Findings by Behrends *et al* (2010) have corroborated this hypothesis, demonstrating that the introduction of a stop codon at position 22 of the *mucA* gene, resulting in a truncated MucA protein, will significantly reduce the relative concentration of glycine-betaine compared to the WT PA01.

Therefore, we have potentially identified the causative genomic anomaly responsible for the switch to the mucoid phenotype in both C1433 and J1532. Further to this we have hypothetically identified a variety of phenotypic changes that have arisen as a “side effect” of this switch. Unfortunately without further confirmatory investigations, these observations cannot be considered definite.

7.3.2 C1426 vs. C1433

The first of our isogenic non-mucoid/mucoid pairings, C1426 and C1433, were isolated from the lungs of a CF patient one month apart. Our analysis, involving BLASTn comparison with 53 other *P. aeruginosa* genomes (Table 3.3), has identified a variety of rare large genomic regions that have been horizontally acquired by these strains during colonisation of their previous niches but are present in very few other *P. aeruginosa* genomes. This includes two identified prophages (C1426 Prophage 2 and 3), a large genomic island (C1426 Genomic Island 1) and several small horizontally acquired regions. Phages C1426 Prophage 2 and 3, based on putative function of contained genes, do not appear to confer any competitive advantage to strains acquiring these phages. However, each displays a degree of similarity to phages described in the literature; Prophage 2 occupies a region previously identified as a region of genome plasticity (Klockgether *et al*, 2011) and resembles LESB58 Prophage 5 (Winstanley *et al*, 2009). Prophage 3, also occupies a region of genome plasticity and is 47% homologous to the *Pseudomonas* filamentous phage Pf1 (Hill *et al*, 1991), containing many of the same structural phage genes (Table 4.4) but in its entirety, is unique to strains C1426 and C1433.

Both C1426 and C1433 host a large genomic island in a region normally occupied by the PA14 pathogenicity island PAPI-1. This island, named C1426 Genomic Island 1, shows some degree of homology to PAPI-1, however, this 139 Kbp island is significantly larger and contains a 56 Kbp region unique to our pairing. This region encodes a further 49 genes, including two putative ferripyoverdine receptors allowing increased iron sequestration from their environment, potentially providing them with a competitive edge against other co-colonising bacterial strains. Two operons encoding two, distinct tripartite RND multidrug efflux pumps have also been acquired within this region. Possession of these pumps may contribute to the increased antibiotic profiles exhibited by these strains (Rita Lagido, unpublished results). This, in turn, may confer an increased chance of

successfully infecting the CF lung, a niche commonly saturated with high antibiotic concentrations.

Each of the small regions present in both C1426 and C1433 but very few other strains, all contain a variety of genes with significant putative functions that support the hypothesis that they confer a selective advantage to our strains. For example, a small 3 Kbp region containing an extra set of fimbriae synthesising *cup* genes has been acquired by our pairing as has a second 4.5 Kbp region containing the two-component regulatory system *bfmSR*. Both *cup* and *bfmSR* (Petrova and Sauer, 2009) genes are involved in the process of biofilm formation. It is hypothesised that acquisition of these genes might result in increased biofilm production, conferring an increased ability to persist within a host. This may be the causative factor for the increased biofilm formation observed in these strains compared to the second non-mucoid/mucoid pairing of J1385 and J1532 (Figure 4.1). Further to this, and providing yet another selective advantage, both C1426 and C1433 have acquired a R-type pyocin resembling the bacteriophage P2 (Nakayama *et al*, 2000). Production of such bacteriolytic pyocins may allow strains that possess these genes to outcompete rival infectious bacterium present in the host, increasing the likelihood of establishing strain dominance (Michel-Briand and Baysee, 2002).

The focus of this study was to explore the genomic effects host adaptation had on C1426 during its one month progression to C1433. That means we must look specifically at any macro and micro genomic anomalies existing between these strains. We have identified one bacteriophage, C1426 Prophage 1, that exists in C1426 but not in C1433. It is hypothesised that such a change might have occurred as an attempt by the bacterium to gain a selective advantage, improving the chance of establishing a persistent chronic infection. Demonstrating this, loss of the C1426 Prophage 1 by C1433 has resulted in the loss of a gene encoding for a putative chitinase. These enzymes, responsible for the breakdown of chitin in the cell wall of

fungi (Thompson *et al*, 2001), have previously been shown to be antigenic. This suggests that C1433 has selected against this gene in an attempt to further prevent host immune recognition and intervention, allowing continued infection. Also lost within this bacteriophage is a gene putatively encoding for the alginate biosynthesis regulator AlgR3 (Misra and Chakrabarty, 1990). Although AlgR3 regulates alginate production in a positive manner, loss of this protein will not disadvantage the strain due to the existing overproduction of the mucoid phenotype causing alginate.

Evaluation of micro differences between the non-mucoid C1426 and its mucoid descendent C1433 has revealed that over the course of one month, C1433 has acquired 27 nonsynonymous SNPs and 10 InDels, including one affecting the previously discussed *mucA*. In a similar manner to macro differences observed, we have predicted that the occurrence of these polymorphisms will potentially confer a selective advantage to C1433. Supporting this argument, we have identified a SNP in the global regulator *ampR*. qPCR has revealed that this SNP has resulted in a loss-of-function of the AmpR regulator (Figure 4.11). We subsequently predicted that this loss of a functioning AmpR protein might contribute to the increased biofilm formation and pyocyanin production exhibited by this strain (Figure 4.1 and Table 4.1). Metabolomic analysis has revealed a significant change in rhamnolipid production by C1433, down ~25 fold compared to C1426 that we have hypothesised is attributable to the *ampR* SNP. Rhamolipid production is dependent upon a functioning RhlR regulatory protein (Ochsner, Fietcher and Reiser, 1994) the expression of which is dependent upon the regulatory effects of AmpR (Kong *et al*, 2005). Therefore, the loss-of-function mutation effecting C1433 AmpR might in turn result in decreased expression of *rhlR* and subsequent decrease in rhamnolipid prodcution. Decreased rhamnolipid production as a result of decreased *ampR* expression may confer a variety of advantages to the C1433 including reduced host cell disruption and virulence as demonstrated by *G. mellonella* infection models (Amy Ford, unpublished data). Further to this, decreased rhamnolipid production might contribute to

the reduced motility (Amy Ford, unpublished data) and increased biofilm production (Figure 4.1) demonstrated by C1433 by aiding initial attachment and preventing dispersal (Boles, Thoendel and Singh, 2005).

A SNP was identified in the chemotaxis gene *wspF*. Possible attenuation of WspF resulting from this SNP may contribute further to the reduced motility demonstrated (Amy Ford, unpublished data). Aside from this, this SNP may also be indirectly responsible for the increased production of the oligosaccharide Δ4,5-unsaturated trigalacturonate (Figure 6.5) via increased *pelA* expression. The introduction of an in-frame mutation in *wspF* is sufficient to induce increased expression of *pelA* as well as increased biofilm formation (Colvin *et al*, 2013). It is therefore hypothesised that the SNP existing in *wspF* results in inactivation of this gene and subsequent accumulation of Δ4,5-unsaturated trigalacturonate as demonstrated by metabolomic analysis.

Non-synonymous SNPs have also been identified in three genes involved in pyoverdin synthesis; *pvdL*, C1426_2790 and C1426_2791 (Table 4.7). Pyoverdin is one of two siderophores produced by *P. aeruginosa* that act as powerful iron chelators, sequestering iron from the environment surrounding the bacterium (Meyer *et al*, 1996). Work within our lab (Table 4.1) has shown that compared to C1426, C1433 has a reduced level of siderophore production. We have failed to identify any further mutations affecting other genes involved in the second siderophore pyochelin suggesting that these mutations observed in genes involved in pyoverdin have the sole responsibility for the reduction in siderophore production observed.

7.3.3 J1385 vs. J1532

We have applied a similar investigation to our second isogenic non-mucoid/ mucoid pairing J1385 and J1532 that were isolated from the lungs of one CF patient three months apart. These strains were previously shown to share a high degree of homology to hypervirulent strain PA14 (Figures 3.2, 3.3 and 3.4). In fact, utilising our BLASTn analysis of these strains and the genomes of 53 other publicly available *P. aeruginosa* strains, it is noted that all horizontally acquired regions present within J1385 are also present within PA14, such is the homogeneity existing between these strains. The same cannot be said for J1532, in which, we have identified several large regions that have been deleted in its host adaptation from J1385 in the CF lung. This includes four bacteriophage; J1385 Prophage 1, 2 and 4 as well as the filamentous phage Pf1. The loss of the latter two of these, J1385 Prophage 4 and Pf1, do not appear to confer any selective advantage to J1532. Based on the putative functions of contained genes, it is not expected that their loss will have no significant bearing on the phenotype of the strain.

In contrast to this, loss of J1385 Prophage 1 may provide J1532 with a significant competitive advantage in infecting the CF lung. For example, putative identification of genes lost within this large bacteriophage reveal that the strain may be more resistant to antibiotic intervention through the loss of the outer membrane protein OpdG, a porin shown to be the mode of entry used by antibiotics such as the widely used CF carbapenems imipenem and meropenem (Hancock and Brinkman, 2002). Increased persistence may also be conferred by the loss of Prophage 1 by the reduced virulence, and subsequent provokation of the host immune system, resulting from loss of the ability to produce a putative protease, a potent toxin ExoY and a functioning hydrogen cyanide cluster (*hcnABC*). We theorise that loss of these genes may have contributed to the reduced virulence exhibited by this strain compared to its progenitor J1385 and its homologue PA14, as visualised using a *G. mellonella* infection model (Figure 5.15).

The final bacteriophage lost by J1532, J1385 Prophage 2, during host adaptation may also contribute to the altered phenotype displayed by this strain, possibly conferring a selective advantage. Putative identification of genes within this prophage reveal that J1532 may have lost a second copy of the transcriptional regulator AlgR, a protein involved in the regulation of a variety of different systems including alginate biosynthesis, type IV pili functioning and other virulence genes (Lizewski *et al*, 2004). Through it activation of the twitching motility operon *fimU-pilVWXY1Y2*, loss of AlgR may contribute to the reduced twitching motility exhibited by this strain (Amy Ford, unpublished data). Lizewski *et al* (2003) have demonstrated that AlgR is also essential for virulence of *P. aeruginosa* in a murine model of septicaemia suggesting that the reduced virulence of J1532 (Figure 5.15) may also be partly attributable to the loss of this regulator within the J1385 Prophage 2.

As mentioned, all horizontally transferred regions by J1385 are also present in PA14. We have discussed those that have not occurred in J1532 however, in the interests of investigating the ability of these strains to establish strain dominance in the CF lung, we must also consider and investigate those present in J1532 too. This includes the hypervariable pyoverdin locus, the pathogenicity islands PAPI-1 and PAPI-2 and also the J1385 Prophage 3. The former of these, the pyoverdin locus, is the most divergent alignable locus present in the genome of *P. aeruginosa* strains (Spencer *et al*, 2003). A 21 Kbp region within this 50 Kbp locus present in J1385, J1532 and PA14 and less than 50% of other strains included in this study, encodes an extra set of pyoverdin biosynthesis genes (*pvdJ*, *pvdD* and *pvdE*) and a FpvA ferripyoverdine receptor. This may result in increased iron uptake due to both increased siderophore synthesis and increased siderophore uptake by the receptor. This will provide a selective advantage to J1385 within the CF lung, allowing it to compete against any co-colonising bacterium in iron sequestration. Adding to this, these strains have also acquired a further set of siderophore biosynthesis or transport genes within the J1385 Prophage 3,

conferring yet a further advantage in outcompeting rival bacterium for essential iron. Also present within this bacteriophage is a gene (J1385_4794) that encodes for a putative thioesterase. These enzymes carry out the non-ribosomal synthesis of the siderophore pyochelin (Reimmann *et al*, 2004) further aiding iron sequestration and also exhibit esterase activity potentially contributing to the reduced esterase activity displayed by these strains compared to the reference strain PA01 (Table 5.1). Reduced esterase activity may also confer a selective advantage to these strains by preventing host provokation and immune intervention allowing increased persistence.

J1385 and J1532 both possess the pathogenicity islands PAPI-1 and PAPI-2 previously identified in PA14. These islands work both individually and synergistically in mediating the virulence of PA14 (Harrison *et al*, 2010). Within PAPI-1, host strains will acquire an extra set of type IVb pili, cellular appendages that will aid colonisation of the CF lung via increasing motility and promoting biofilm formation by increasing attachment. Biofilm formation is further promoted by genes encoding for a *cup* fimbrial biosynthesis cluster and PvrR two-component response regulator present within PAPI-1. The latter of these, PvrR might further contribute to the strains success due to its role in increasing the strains antibiotic resistance (Harrison *et al*, 2010). PAPI-2 might also aid hosting strains in their attempts at dominantly colonising the CF lung via possession of a gene putatively encoding for the potent cytotoxin ExoU. ExoU is one of the many effector proteins that is part of the type III secretion repertoire of *P. aeruginosa* that is used to deliver such toxins into eukaryotic cells (Kurahashi *et al*, 1999). ExoU can exert its phospholipase A_2 activity to limit host immune intervention via killing of host immune effectors and also kill other eukaryotic competitors such as *S. cerevisiae*.

Analysis of of strains at the single nucleotide level has revealed that over the course of three months infecting the CF lung, host adaptation has resulted in the appearance of 50 non-synonymous SNPs and 13 InDel polymorphisms.

SNPs have arisen in a variety of genes that may contribute to the increaed antibiotic resistance demonstrated by this strain (Rita Lagido, unpublished results). For example, a SNP has occurred in J1385_0272 that is predicted to code for a porin involved in transportation of antibiotics into the cell. We speculate that such a SNP may result in altered porin confirmation, reduced outermembrane permeability and potentially, increased antibiotic resistance. A SNP has also occurred in the gene coding for the outermembrane lipopolysaccharide biosynthesis protein LpxO2. Antibiotics such as polymixin and novobiocin target such lipopolysaccharide (Tamaki, Sato and Matsuhashi, 1971). It is predicted that the introduction of this SNP may result in a structural change of this protein conferring increased antibiotic resistance.

Table 5.1 demonstrates that compared to its non-mucoid counterpart J1385, J1532 exhibits reduced siderophore production. We predict that this change in production may be partly attributable to two SNPs identified in genes encoding for proteins involved in iron sequestration. For example, a SNP occurring in the pyoverdin peptide synthase encoding gene *pvdL*, potentially resulting in reduced synthesis of the siderophore pyoverdin. A SNP has also appeared in J1385_4832 that encodes a TonB-like iron transport protein (Takase *et al*, 2000). It is not clear what benefits, if any, J1532 will experience by the reduced siderophore production, and subsequent iron acquisition, putatively occurring as a result of the SNPs in these iron acquisition genes. However, results of an earlier study investigating the genetic changes of *P. aeruginosa* within a CF lung has also shown mutations within genes involved in this system (Smith *et al*, 2006).

7.4 Conclusion

We have, to an extent, successfully contributed to the overall understanding of the genomic basis of the clinically and environmentally significant species *P. aeruginosa*. By sequencing twelve diverse strains, we have increased the number of publicly available *P. aeruginosa* strains by approximately 20% taking the current total to 55. We have also developed a more thorough and potentially more accurate method of typing of strains, replacing the very limited MLST format currently used (Stewart *et al*, 2013). Using this Panseq method we have identified that *P. aeruginosa* strains fall into two major groups and a third, very distinct minor group. No obvious patterns have emerged to exist between grouping and the niche from which strains were originally isolated. However, our results must be considered with caution as, as mentioned, only 55 strains have been included in this study. Future analysis would seek to include significantly more *P. aeruginosa* genomes with the aim of providing a more accurate and significant representation of the *P. aeruginosa* lineage. We therefore require the *P. aeruginosa* community, as a whole, to continue to sequence a diverse range of strains, making the data publicly available on GenBank.

Our second aim set out to evaluate the role of host adaptation during infections of the cystic fibrosis lung. Our aforementioned Panseq study identified that in our sequenced strain collection we have two probable near isogenic, host adapted, non-mucoid/mucoid pairings; C1426/C1433 and J1385/J1532. Although unlikely, it must be considered that these strains have arisen due to evolution as a result of host adaptation and may have existed concurrently is a co-infecting manner. This analysis also revealed that a third pairing was not isogenic and the switch in strain dominance can be attribute to strain succession rather than host adaptation. Our subsequent genotypic and phenotypic analysis has successfully attributed a variety of genomic modifications to phenotype changes observed as a result of host adaptation. This includes identifying the genomic anomaly that has led to the mucoid phenotype as an InDel resulting in the truncation of the anti-sigma factor

MucA resulting in uncontrolled aliginate biosynthesis. We have also attributed macro and micro genomic changes to a variety of phenotypic alterations including reduced virulence and motility as well as increased biofilm formation and antibiotic resistance. As a result of the scale of the computational analysis and the diversity of the potential inter strain differences, it has not been possible to analyse every hypothesis made. There is a requirement to develop a significant range of phenotypic tests if we are to gain a greater understanding of the true diversity of the *P. aeruginosa* lineage.

7.4.1 Future work

This project has given rise to many hypotheses, the majority of which remain unconfirmed. There are several lines of further research that have arisen from that could be pursued. It would be beneficial to confirm that the InDel polymorphisms present in both J1532 and C1433 *mucA* genes are sufficient for the appearance of the mucoid phenotype via complementation of the mutant strain. Similarly, all nucleotide polymorphisms and their hypothesised resulting phenotypic changes identified in this study could be investigated further using this technique. It would be interesting to utilise both ChIP-seq and RNA-seq to identify possible targets of the global regulator AmpR that has experienced a non-synonymous SNP in the mucoid C1433.

Several metabolic differences were also observed between our pairs. Unfortunately, each of these observations can only be classed as putative due to potential miscalling of metaboltes. Therefore, it would be beneficial to our study to replicate our metabolomic analysis complemented with a panel of known metabolite standards to aid in our identification.

Bibliography

Aballay and Ausubel. Caenorhabditis elegans as a host for the study of host-pathogen interactions. Current Opinion in Microbiology (2002) vol. 5 (1) pp. 97-101

Agarwal *et al*. Medical significance and management of staphylococcal biofilm. FEMS Immunol Med Microbiol (2010) vol. 58 (2) pp. 147-60

Al Aloul *et al*. Increased morbidity associated with chronic infection by an epidemic *Pseudomonas aeruginosa* strain in CF patients. Thorax (2004) vol. 59 pp. 334-6

Alibaud *et al*. *Pseudomonas aeruginosa* virulence genes identified in a *Dictyostelium* host model. Cellular Microbiology (2008) vol. 10 (3) pp. 729-40

Altschul *et al*. Basic local alignment search tool. J Mol Biol (1990) vol. 215 (3) pp. 403-10

Altschul *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res (1997) vol. 25 (17) pp. 3389-402

Altun and Hall. 2009. Introduction. In WormAtlas. doi:10.3908/wormatlas.1.1
Andersen. Cystic fibrosis of the pancreas and its relation to celiac disease: a clinical and pathological study. Am J Dis Child (1938) vol. 56 pp.344-99

Anderson and Anderson. Proteome and proteomics: new technologies, new concepts, and new words. Electrophoresis (1998) vol. 19 (11) pp. 1853-61

Andrejko and Mizerska-Dudka. Elastase B of *Pseudomonas aeruginosa* stimulates the humoral immune response in the greater wax moth, *Galleria mellonella*. Journal of Invertebrate Pathology (2011) vol. 107 (1) pp. 16-26

Anzai *et al*. Phylogenetic affiliation of the pseudomonads based on 16S rRNA sequence. International Journal of Systematic and Evolutionary Microbiology (2000) vol. 50 (4) pp. 1563-89

Balasubramanian *et al*. The Regulatory Repertoire of *Pseudomonas aeruginosa* AmpC β-Lactamase Regulator AmpR Includes Virulence Genes. PLoS One (2012) vol. 7 (3) e34067

Banin *et al*. Iron and *Pseudomonas aeruginosa* biofilm formation. Proc Natl Acad Sci USA (2005) vol. 102 (31) pp. 11076-81

Banerjee and Stableforth. The treatment of respiratory *Pseudomonas* infection in cystic fibrosis: what drug and which way? Drugs (2000) vol. 60 (5) pp. 1053-64

Barbieri and Sun. *Pseudomonas aeruginosa* ExoS and ExoT. Rev Physiol Biochem Pharmacol (2004) vol. 152 pp. 79-92

Barbosa *et al*. Mechanism of action and NAD+-binding mode revealed by the crystal structure of L-histidinol dehydrogenase. Proc Natl Acad Sci (2002) vol. 99 (4) pp. 1859-64

Barken *et al*. Roles of type IV pili, flagellum-mediated motility and extracellular DNA in the formation of mature multicellular structures in *Pseudomonas aeruginosa* biofilms. Environmental Microbiology (2008) vol. 10 (9) pp. 2331-43

Battle *et al*. Genomic islands of *Pseudomonas aeruginosa*. FEMS Microbiol Lett (2009) vol. 290 (1) pp. 70-8

Behrends *et al*. Metabolic profiling of *Pseudomonas aeruginosa* demonstrates that the anti-sigma factor MucA modulates osmotic stress tolerance. Molecular Biosystems (2010) vol. 6 pp. 562-9

Bennasar *et al*. PseudoMLSA: a database for multigenic sequence analysis of *Pseudomonas* species. BMC microbiology (2010) vol. 10 (1) pp. 118

Benson *et al*. GenBank. Nucleic Acids Research (2008) vol. 36 pp. 25-30

Bentley *et al*. Accurate whole human genome sequencing using reversible terminator chemistry. Nature (2008) vol. 456 (7218) pp. 53-9

Berka and Vasil. Phospholipase C (heat-labile hemolysin) of *Pseudomonas aeruginosa*: purification and preliminary characterization. Journal of bacteriology (1982) vol. 152 (1) pp. 239-45

BioCyc. (2012) *Pseudomonas aeruginosa* PA01 Enzyme: PelA. www.biocyc.org/PAER208964/NEW-IMAGE?type=GENE&object=GCXG-3290 [6 August 2014]

Bleves *et al*. Quorum sensing negatively controls type III secretion regulon expression in *Pseudomonas aeruginosa* PAO1. Journal of bacteriology (2005) vol. 187 (11) pp. 3898-902

Boles, Thoendel and Singh. Rhamnolipids mediate detachment of *Pseudomonas aeruginosa* biofilms. Molecular Microbiology (2005) vol. 57 (5) pp. 1210-23

Boucher *et al*. Mucoid *Pseudomonas aeruginosa* in cystic fibrosis: characterization of *muc* mutations in clinical isolates and analysis of clearance in a mouse model of respiratory infection. Infection and Immunity (1997) vol. 65 (9) pp. 3838-46

Brandel *et al*. Pyochelin, a siderophore of *Pseudomonas aeruginosa*: Physicochemical characterization of the iron(lll), copper(ll) and zinc(ll) complexes. Dalton Transactions (2012) vol. 41 pp. 2820-34

Brenner. The genetics of *Caenorhabditis elegans*. Genetics (1974) vol. 77 (1) pp. 71-94 Brouns *et al*. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science (2008) vol. 321 (5891) pp. 960-4

Browning and Busby. The regulation of bacterial transcription initiation. Nat Rev Micro (2004) vol. 2 (1) pp. 57-65

Bryan *et al*. Overproduction of the CFTR R domain leads to increased levels of asialoGM1 and increased *Pseudomonas aeruginosa* binding by epithelial cells. Am J Respir Cell Mol Biol (1998) vol. 19 pp.269-77

Bryant, Faruqi and Pinney. Analysis of metabolic evolution in bacteria using whole-genome metabolic models. Journal of Computational Biology (2013) vol. 20 (10) pp. 755-64

Buck *et al*. The bacterial enhancer-dependent sigma(54) (sigma(N)) transcription factor. Journal of Bacteriology (2000) vol. 182 (15) pp. 4129-36

Campbell *et al*. Crystal structure of *Escherichia coli* sigmaE with the cytoplasmic domain of its anti-sigma RseA. Molecular Cell (2003) vol. 11 pp.1067-78

Campbell *et al*. Regulation of bacterial RNA polymerase sigma factor activity: a structural perspective. Current Opinion in Microbiology (2008) vol. 11(2) pp. 121-7

Campbell *et al*. Structure of the bacterial RNA polymerase promoter specificity sigma subunit. Mol Cell (2002) vol. 9 (3) pp. 527-39

Cánovas, Cases and de Lorenzo. Heavy metal tolerance and metal homeostasis in *Pseudomonas putida* as revealed by complete genome analysis. Environmental Microbiology (2003) vol. 5 (12) pp. 1242-56

Carter, Chen and Lory. The *Pseudomonas aeruginosa* pathogenicity island PAPI-1 is transferred via a novel type IV pilus. Journal of Bacteriology (2010) vol. 192 (13) pp. 3249-58

Casadevall and Pirofski. Host-pathogen interactions: the attributes of virulence. Journal of Infectious Diseases (2001) vol. 184 (3) pp. 337-44

Ceciliani *et al*. Proteomics in veterinary medicine: applications and trends in disease pathogenesis and diagnostics. Veterinary Pathology (2014) vol. 51 (2) pp. 351-62

Chain *et al*. Penicillin as a chemotherapeutic agent. Lancet (1940) vol. 1 (1172) pp. 226-8 Chaudhuri *et al*. xBASE2: a comprehensive resource for comparative bacterial genomics. Nucleic Acids Research (2008) vol. 36 (Database issue) pp. D543-6

Cheng *et al*. Spread of a beta-lactam-resistant *Pseudomonas aeruginosa* in a cystic fibrosis clinic. Lancet (1996) vol. 348 pp. 639-42

Cheng, Qian and Kirby. Evaluation of the Abbott Real Time CT/NG assay in comparison to the Roche Cobas Amplicor CT/NG Assay. Journal of Clinical Microbiology (2011) vol. 49 (4) pp. 1294-1300

Chmiel and Davis. State of the art: why do the lungs of patients with cystic fibrosis become infected and why can't they clear the infection?. Respir Res (2003) vol. 4 pp. 8

Chrzanowski, Lawniczak and Czaczyk. Why do microorganism produce rhamnolipids?. World Journal of Microbiology & Biotechnology (2012) vol. 28 (2) pp. 401-19

Ciofu *et al*. Genetic adaptation of *Pseudomonas aeruginosa* during chronic lung infection of patients with cystic fibrosis: strong and weak mutators with heterogeneous genetic backgrounds emerge in *mucA* and/or *lasR* mutants. Microbiology (2010) vol. 156 (Pt 4) pp. 1108-19

Colvin *et al*. PelA deacetylase activity is required for Pel polysaccharide synthesis in *Pseudomonas aeruginosa*. Journal of Bacteriology (2013) vol. 195 (10) pp. 2329-39

Cosson *et al*. *Pseudomonas aeruginosa* virulence analyzed in a *Dictyostelium discoideum* host system. Journal of bacteriology (2002) vol. 184 (11) pp. 3027-33

Costerton *et al*. How bacteria stick. Sci Am (1978) vol. 238 (1) pp. 86-95

Costerton, Stewart and Greenberg. Bacterial biofilms: a common cause of persistent infections. Science (1999) vol. 284 (5418) pp. 1318-22

Creek *et al*. IDEOM: An Excel interface for analysis of LC-MS based metabolomics data. Bioinformatics (2012) vol. 28 (7) pp. 1048-9

Cystic Fibrosis Trust. (2009) Antibiotic Treatment for Cystic Fibrosis. www.cysticfibrosis.org.uk/media/82010CD_Antibiotic_treatment_for_CF_May_09.pdf [7 November 2013]

D'Argenio *et al*. Autolysis and Autoaggregation in *Pseudomonas aeruginosa* colony morphology mutants. Journal of Bacteriology (2002) vol. 184 (23) pp. 6481-9

D'Argenio *et al*. *Drosophila* as a model host for *Pseudomonas aeruginosa* infection. Journal of bacteriology (2001) vol. 183 (4) pp. 1466-71

Darling *et al*. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res (2004) vol. 14 (7) pp. 1394-403

Davies, Alton and Bush. Cystic Fibrosis. British Medical Journal (2007) vol. 335 pp. 1255-9

Davies. *Pseudomonas aeruginosa* in cystic fibrosis: pathogenesis and persistence. Paediatr Respir Rev (2002) vol. 3 (2) pp. 128-34

Davis *et al*. Cystic Fibrosis. American Journal of Respiratory and Critical Care Medicine (1996) vol. 154 (5) pp. 1229-56

de Vrankrijker *et al*. Aspergillus fumigatis colonisation in cystic fibrosis: implications for lung infection?. Clinical Microbiology and Infection (2011) vol. 19 (9) pp. 1381-86

Delcher *et al*. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics (2007) vol. 23 (6) pp. 673-9

Deretic *et al*. Gene amplification induces mucoid phenotype in rec-2 *Pseudomonas aeruginosa* exposed to kanamycin. Journal of Bacteriology (1986) vol. 165 (2) pp. 510-6

Deretic, Schurr and Yu. *Pseudomonas aeruginosa*, mucoidy and the chronic infection phenotype in cystic fibrosis. Trends Microbiol (1995) vol. 3 (9) pp. 351-6

DeVries and Ohman. Mucoid-to-nonmucoid conversion in alginate-producing *Pseudomonas aeruginosa* often results from spontaneous

mutations in algT, encoding a putative alternate sigma factor, and shows evidence for autoregulation. Journal of Bacteriology (1994) vol. 176 (21) pp. 6677-87

Deziel *et al*. Initiation of biofilm formation by *Pseudomonas aeruginosa* 57RP correlates with emergence of hyperpiliated and highly adherent phenotypic variant deficient in swimming, swarming and twitching motilities. Journal of Bacteriology (2001) vol. 183 (4) pp. 1195-204

Diaz and Hauser. *Pseudomonas aeruginosa* cytotoxin ExoU is injected into phagocytic cells during acute pneumonia. Infect Immun (2010) vol. 78 (4) pp. 1447-56

Dingemans *et al*. The deletion of TonB-dependent receptor genes is part of the genome reduction process that occurs during adaption of *Pseudomonas aeruginosa* to the cystic fibrosis lung. Pathogens and Disease (2014) vol. 71 pp. 26-38

Doggett *et al*. An atypical *Pseudomonas aeruginosa* associated with cycstic fibrosis of the pancreas. Journal of Pediatrics (1966) vol. 68 pp. 215-21

Dong and Schellhorn. Role of RpoS in virulence of pathogens. Infection and Immunity (2010) vol. 78 (3) pp. 887-97

Donlan and Costerton. Biofilms: survival mechanisms of clinically relevant microorganisms. Clin Microbiol Rev (2002) vol. 15 (2) pp. 167-93

Drenkard. Antimicrobial resistance of *Pseudomonas aeruginosa* biofilms. Microbes and Infections (2003) vol. 5 pp. 1213-9

Drenkard and Ausubel. *Pseudomonas* biofilm formation and antibiotic resistance are linked to phenotypic variation. Nature (2002) vol. 416 pp. 740-43

Edenborough *et al*. Genotyping of *Pseudomonas aeruginosa* in cystic fibrosis suggests need for segregation. Journal of Cystic Fibrosis (2004) vol. 3 pp. 37-44

Einhauer and Jungbauer. The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins. J Biochem Biophys Methods (2001) vol. 49 (1-3) pp. 455-65

El Hage *et al*. Proteolysis of *Pseudomonas* exotoxin A within hepatic endosomes by cathepsins B and D produces fragments displaying in vitro ADP-ribosylating and apoptotic effects. FEBS J (2010) vol. 277 (18) pp. 3735-49

Emerson *et al*. *Pseudomonas aeruginosa* and other predictors of mortality and morbididty in young children with cystic fibrosis. Pediatric Pulmonology (2002) vol. 34 pp. 91-100

Euzéby. List of Bacterial Names with Standing in Nomenclature: a folder available on the Internet. International Journal of Systematic Bacteriology (1997) vol. 47 pp. 590-92.

Ewbank. Tackling both sides of the host-pathogen equation with *Caenorhabditis elegans*. Microbes Infect (2002) vol. 4 (2) pp. 247-56

Farinha *et al*. Localization of the virulence-associated genes *pilA*, *pilR, rpoN, fliA, fliC, ent,* and *fbp* on the physical map of *Pseudomonas aeruginosa* PAO1 by pulsed-field electrophoresis. Infection and Immunity (1993) vol. 61 (4) pp. 1571-5

Feil *et al*. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. Journal of Bacteriology (2004) vol. 186 (5) pp.1518-30

Felsenstein. PHYLIP- Phylogeny Inference Package (Version 3.2). Cladistics (1989) vol. 5 pp.164-166

Feltman *et al*. Prevalence of type III secretion genes in clinical and envirnonmental isolates of *Pseudomonas aeruginosa*. Microbiology (1997) vol. 147 pp. 2659-69

Firoved and Deretic. Microarray analysis of global gene expression in mucoid *Pseudomonas aeruginosa*. Journal of Bacteriology (2003) vol. 185 (3) pp. 1071-81

Firoved, Boucher and Deretic. Global Genomic Analysis of AlgU (σE)-Dependent Promoters (Sigmulon) in *Pseudomonas aeruginosa* and Implications for Inflammatory Processes in Cystic Fibrosis. Journal of Bacteriology (2002) vol. 184 (4) pp. 1057-64

Fitzsimmons, Hampel and Wargo. Cellular choline and glycine betaine pools impact osmoprotection and phospholipase C production in *Pseudomonas aeruginosa*. Journal of Bacteriology (2012) vol. 194 (17) pp. 4718-26

Fleischmann *et al*. Whole-genome random sequencing and assmebly of *Haemophilus influenzae* Rd. Science (1995) vol. 269 (5223) pp. 496-512

Fournier, Drancourt and Raoult. Bacterial genome sequencing and its use in infectious diseases. Lancet Infectious Diseases (2007) vol. 7 pp. 711-23

Gallant *et al*. *Pseudomonas aeruginosa* cystic fibrosis clinical isolates produce exotoxin A with altered ADP-ribosyltransferase activity and cytotoxicity. Microbiology (Reading, England) (2000) vol. 146 (Pt 8) pp. 1891-9

Gilbert. PHYLODENDRON verison 0.8d. Biology Department, Indiana University, USA (Available at: http://iubio.bio.indiana.edu/treeapp/treeprint-form.html)

Gilligan. Microbiology of airway disease in patients with cystic fibrosis. Clin Microbiol Rev (1991) vol. 4 (1) pp. 35-51

Gimenez-Ibanez and Rathjen. The case for the defense: plants versus *Pseudomonas syringae*. Microbes Infect (2010) vol. 12 (6) pp. 428-37

Glick *et al*. Increase in rhamnolipid synthesis under iron-limiting conditions influences surface motility and biofilm formation in *Pseudomonas aeruginosa*. Journal of Bacteriology (2010) vol. 192 (12) pp. 2973-80

Goffeau *et al*. Life with 6000 genes. Science (1996) vol. 274 (5287) pp. 563-7

Gourse *et al*. UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. Mol Microbiol (2000) vol. 37 (4) pp. 687-95

Govan and Deretic. Microbial pathogenesis in cystic fibrosis: mucoid *Pseudomonas aeruginosa* and *Burkholderia cepacia*. Microbiol Rev (1996) vol. 60 (3) pp. 539-74

Guilbault and Kramer. Ultra sensitive, specific method for cyanide using p-nitrobenzaldehyde and o-dinitrobenzene. Analytical Chemistry (1966) vol. 38 pp. 834-6

Guzzo *et al*. Cloning of the *Pseudomonas aeruginosa* alkaline protease gene and secretion of the protease into the medium by *Escherichia coli.* Journal of Bacteriology (1990) vol. 172 (2) pp. 942-8

Haase *et al*. Bacterial conjugation mediated by plasmid RP4: RSF1010 mobilization, donor-specific phage propagation, and pilus production require the same Tra2 core components of a proposed DNA transport complex. Journal of Bacteriology (1995) vol. 177 (16) pp. 4779-91

Hall-Stoodley *et al*. Bacterial biofilms: from the natural environment to infectious diseases. Nat Rev Microbiol (2004) vol. 2 (2) pp. 95-108

Hamlett *et al*. Roles of the Tn21, *merT*, *merP* and *merC* gene products in mercury resistance and mercury binding. Journal of Bacteriology (1992) vol. 174 (20) pp. 6377-85

Hanahan. Studies on transformation of *Escherichia coli* with plasmids. J Mol Biol (1983) vol. 166 (4) pp. 557-80

Hancock and Brinkman. Function of *Pseudomonas* porins in uptake and efflux. Annual Reviews Microbiology (2002) vol. 56 pp. 17-38

Hancock and Woodruff. Roles of porin and beta-lactamse in beta-lactam resistance of *Pseudomonas aeruginosa*. Rev Infect Dis (1988) vol. 10 (4) pp. 770-5

Harrison *et al*. Pathogenicity island PAPI-1 and PAPI-2 contribute individually and synergistically to the virulence of *Pseudomonas aeruginosa* strain PA14. Infection and Immunity (2010) vol. 78 (4) pp. 1437-46

Hassett *et al*. *Pseudomonas aeruginosa* hypoxic or anaerobic biofilm infections within cystic fibrosis airways. Trends in Microbiology (2009) vol. 17 (3) pp. 130-8

Hauser *et al*. Clinical significance of microbial infection and adaption in cystic fibrosis. Clinical Microbiology Reviews (2011) vol. 24 (1) pp. 29-70

He *et al*. The broad host range pathogen *Pseudomonas aeruginosa* strain PA14 carries two pathogenicity islands harboring plant and animal virulence genes. Proc Natl Acad Sci USA (2004) vol. 101 (8) pp. 2530-5

Hendrickson *et al*. Differential roles of the *Pseudomonas aeruginosa* PA14 *rpoN* gene in pathogenicity in plants, nematodes, insects, and mice. Journal of Bacteriology (2001) vol. 183 (24) pp. 7126-34

Hewitt *et al*. Respiratory symptoms in older people and their association with mortality. Thorax (2005) vol. 60 pp. 331-4

Hill *et al*. DNA sequence of the filamentous bacteriophage Pf1. Journal of Molecular Biology (1991) vol. 218 (2) pp. 349-64

Hoang *et al*. Integration-proficient plasmids for *Pseudomonas aeruginosa*: site-specific integration and use for engineering of reporter and expression strains. Plasmid (2000) vol. 43 (1) pp. 59-72

Hritonenko *et al*. Adenylate cyclase activity of *Pseudomonas aeruginosa* ExoY can mediate bleb-niche formation in epithelial cells and

contributes to virulence. Microbial Pathogenesis (2011) vol. 51 (5) pp. 305-12

Hutchison and Gross. Lipopeptide phytotoxins produced by *Pseudomonas syringae pv*. *syringae*: comparison of the biosurfactant and ion channel-forming activities of syringopeptin and syringomycin. Mol Plant Microbe Interact (1997) vol. 10 (3) pp. 347-54

Iglewski *et al*. *Pseudomonas aeruginosa* exoenzyme S: an adenosine diphosphate ribosyltransferase distinct from toxin A. Proc Natl Acad Sci USA (1978) vol. 75 (7) pp. 3211-5

Imperi *et al*. Molecular basis of pyoverdine siderophore recycling in *Pseudomonas aeruginosa*. Proc Natl Acad Sci USA (2009) vol. 106 (48) pp. 20440-5

Jander, Rahme and Ausubel. Positive correlation between virulence of *Pseudomonas aeruginosa* mutants in mice and insects. Journal of Bacteriology (2000) vol. 182 (13) pp. 3843-5

Jolley *et al*. mlstdbNet - distributed multi-locus sequence typing (MLST) databases. BMC Bioinformatics (2004) vol. 5 pp. 86

Jones *et al*. Activation of the *Pseudomonas aeruginosa* AlgU regulon through *mucA* mutation inhibits cyclic AMP/Vfr signaling. Journal of Bacteriology (2010) vol. 192 (21) pp. 5709-17

Jung *et al*. Overexpression of cold shock protein A of *Psychromonas arctica* KOPRI 22215 confers cold-resistance. Protein J (2010) vol. 29 (2) pp. 136-42

Kamath, Kapatral and Chakrabarty. Cellular function of elastase in *Pseudomonas aeruginosa*: role in the cleavage of nucleoside diphosphate kinase and in alginate synthesis. Molecular Microbiology (1998) vol. 30 (5) pp. 933-41

Kaplan. Biofilm dispersal: mechanisms, clinical implications, and potential therapeutic uses. J Dent Res (2010) vol. 89 (3) pp. 205-18

Karimova *et al*. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. Proc Natl Acad Sci USA (1998) vol. 95 (10) pp. 5752-6

Kato, Misra and Chakrabarty. AlgR3, a protein resembling eukaryotic histone H1, regulates alginate synthesis in *Pseudomonas aeruginosa*. Proc Natl Acad Sci USA (1990) vol. 87 pp. 2887-91

Kievit. Quorum sensing in *Pseudomonas aeruginosa* biofilms. Environmental Microbiology (2009) vol. 11 (2) pp. 279-88

Kiewitz and Tummler. Sequence diversity of *Pseudomonas aeruginosa*: impact on population structure and genome evolution. Journal of Bacteriology (2000) vol. 182 (11) pp. 3125-35

Kirchner *et al*. Use of Artificial Sputum Medium to test antibiotic efficacy against *Pseudmonas aeruginosa* in conditions more relevant to the cystic fibrosis lung. Journal of Visualised Experiments (2012) vol. 64 pp. 3857

Klockgether *et al*. *Pseudomonas aeruginosa* genomic structure and diversity. Frontiers in Microbiology (2011) vol. 2 (150) pp. 1-18

Knapp *et al*. Antibiotic resistanc egene abundancies correlate with metal and geochemical conditions in archived Scottish soils. PLoS ONE (2011) vol. 6 (11) e27300

Kong *et al*. *Pseudomonas aeruginosa* AmpR is a global transcriptional factor that regulates expression of AmpC and PoxB β-lactamases, proteases, quorum sensing, and other virulence factors. Antimicrobial Agents and Chemotherapy (2005) vol. 49 (11) pp. 4567-75

Kukavica-Ibrulj *et al*. In vivo growth of *Pseudomonas aeruginosa* strains PA01 and PA14 and the hypervirulent strain LESB58 in a rat model of chronic lung infection. Journal of Bacteriology (2008) vol. 190 (8) pp. 2804-13

Kurahashi *et al*. Pathogenesis of septic shock in *Pseudomonas aeruginosa* pneumonia. J Clin Invest (1999) vol. 104 pp. 743-50

Kurioka and Liu. Effect of the hemolysin of *Pseudomonas aeruginosa* on phosphatides and on phospholipase C activity. Journal of bacteriology (1967) vol. 93 (2) pp. 670-4

Kurtz *et al*. Versatile and open software for comparing large genomes. Genome Biol (2004) vol. 5 (2) pp. R12

Kwon *et al*. *Pseudomonas koreensis sp. nov*., *Pseudomonas umsongensis sp. nov*. and *Pseudomonas jinjuensis sp. nov*., novel species from farm soils in Korea. Int J Syst Evol Microbiol (2003) vol. 53 (Pt 1) pp. 21-7

Lagesen *et al*. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Research (2007) vol. 35 (9) pp. 3100-8

Lamont *et al*. Siderophore-mediated signalling regulates virulence factor production in *Pseudomonas aeruginosa*. Proc Natl Acad Sci (2002) vol. 99 pp. 7072-7

Langille and Brinkman. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. Bioinformatics (2009) vol. 25 (5) pp. 664-5

Larsen *et al*. Multilocus sequence typing of total-genome-sequenced bacteria. Journal of clinical microbiology (2012) vol. 50 (4) pp. 1355-61

Latifi *et al*. A hierarchical quorum-sensing cascade in *Pseudomonas aeruginosa* links the transcriptional activators LasR and RhIR (VsmR) to expression of the stationary-phase sigma factor RpoS. Mol Microbiol (1996) vol. 21 (6) pp. 1137-46

Lau *et al*. The role of pyocyanin in *Pseudomonas aeruginosa* infection. Trends Mol Med (2004) vol. 10 (12) pp. 599-606

Ledgham *et al*. Global regulation in *Pseudomonas aeruginosa*: the regulatory protein AlgR2 (AlgQ) acts as a modulator of quorum sensing. Research in Microbiologoy (2003) vol. 154 (3) pp. 207-13

Lee *et al*. Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial. Genome Biol (2006) vol. 7 (10) pp. R90

Levene *et al*. Zero-mode wavelguides for single-molecule analysis at high concentrations. Science (2003) vol. 299 pp. 682-6

Li *et al*. The Sequence Alignment/Map format and SAMtools. Bioinformatics (2009) vol. 25 (16) pp. 2078-9

Liang *et al*. Identification of a genomic island present in the majority of pathogenic isolates of *Pseudomonas aeruginosa*. Journal of bacteriology (2001) vol. 183 (3) pp. 843-53

Liberati *et al*. An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. Proc Natl Acad Sci USA (2006) vol. 103 (8) pp. 2833-8

Liebert, Hall and Summers. Tansposon Tn21, Flagship of the floating genome. Microbiol Mol BioL Rev (1999) vol. 63 (3) pp. 507-22

Lindeberg. Pseudomonas-Plant Interactions [online] (2010) Available at http://pseudomonas-syringae.org/ [accepted 23rd November 2010]

Liu and Nizet. Color me bad: microbial pigments as virulence factors. Trends in Microbiology (2009) vol. 17 (9) pp. 406-13

Livak and Schmittgen. Analysis of relative expression data using real-time quantitative PCR and the 2(-Delta Delta C (T)) method. Methods (2001) vol. 25 (4) pp. 402-8

Livermore *et al*. Clavulanate and beta-lactamase induction. Journal of Antimicrobial Chemotherapy (1989) vol. 24 pp 23-33

Lizewski *et al*. Identification of AlgR-regulated genes in *Pseudomonas aeruginosa* by use of microarray analysis. Journal of Bacteriology (2004) vol. 186 (17) pp. 5672-84

Lizewski *et al*. The transcriptional regulator AlgR is essential for *Pseudomonas aeruginosa* pathogenesis. Infect Immun (2002) vol. 70 pp. 6083-93

Lowe and Eddy. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res (1997) vol. 25 (5) pp. 955-64

Lyczak, Cannon and Pier. Lung infection associated with cystic fibrosis. Clinical Microbiology Reviews (2002) vol. 15 (2) pp.194-222

Madigan *et al*. Brock Biology of Microorganisms. 12th edn, (2009). San Francisco, USA: Pearson Benjamin Cummings, pp 413-5

Mah *et al*. A genetic basis for *Pseudomonas aeruginosa* biofilm antibiotic resistance. Nature (2003) vol. 426 pp. 306-10

Maharjan and Ferenci. Global metabolite analysis: the influence of extraction methodology on metabolite profiles of *Escherichia coli*. Analytical Biochemistry (2003) vol. 313 (1) pp. 145-54

Mahenthiralingam, Campbell and Speert. Nonmotility and phagocytic resistance of *Psedumonas aeruginosa* isolates from chronically colonized patients with cystic fibrosis. Infection and Immunity (1994) vol. 62 pp. 596-605

Maki *et al*. Ice nucleation induced by *Pseudomonas syringae*. Appl Microbiol (1974) vol. 28 (3) pp. 456-9

Mardis. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet (2008) vol. 9 pp. 387-402

Martin *et al*. Mechanism of conversion to mucoidy in *Pseudomonas aeruginosa* infecting cystic fibrosis patients. Proc Natl Acad Sci USA (1993) vol. 90 (18) pp. 8377-81

Martin, Holloway and Deretic. Characterisation of a locus determining the mucoid status of *Pseudomonas aeruginosa*: AlgU shows sequence similarities with a *Bacillus* sigma factor. Journal of Bacteriology (1993) vol. 175 (4) pp. 1153-64

Mathee *et al*. Dynamics of *Pseudomonas aeruginosa* genome evolution. Proc Natl Acad Sci (2008) vol. 105 (8) pp. 3100-5

Mathee, McPherson and Ohman. Posttranslational control of the *algT* (*algU*)-encoded sigma22 for expression of the alginate regulon in *Pseudomonas aeruginosa* and localization of its antagonist proteins MucA and MucB (AlgN). Journal of Bacteriology (1997) vol. 179 (11) pp. 3711-20

Matsui *et al*. Evidence for periciliary liquid layer depletion, not abnormal ion composition, in the pathogenesis of cystic fibrosis airways disease. Cell (1998) vol. 95 pp. 1005-15

McCallum *et al*. Spread of an epidemic *Pseudomonas aeruginosa* strain from a patient with cystic fibrosis to non-CF relatives. Thorax (2002) vol. 57 pp. 559-60

McCaughey *et al*. Lectin-like bacteriocins from Pseudomonas spp. utilise D-Rhamnose containing lipopolysaccharide as a cellular receptor. PLOS Pathogens (2014) vol.10 (2) e1002898

Medema *et al*. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res (2011) vol. 39 (Web Server issue) pp. W339-46

Medini *et al*. The microbial pan-genome. Curr Opin Genet Dev (2005) vol. 15 (6) pp. 589-94

Meyer *et al*. Pyoverdin is essential for virulence of *Pseudomonas aeruginosa*. Infection and Immunity (1996) vol. 64 (2) pp. 518-23

Meyer, Stintzi and Poole. The ferripyoverdine receptor FpvA of *Pseudomonas aeruginosa* PAO1 recognizes the ferripyoverdines of *P. aeruginosa* PAO1 and *P. fluorescens* ATCC 13525. FEMS Microbiology Letters (1999) vol. 170 (1) pp. 145-50

Meyer. Pyoverdines: pigments, siderophores and potential taxonomic markers of fluorescent *Pseudomonas* species. Arch Microbiol (2000) vol. 174 (3) pp. 135-42

Michel-Briand and Baysse. The pyocins of *Pseudomonas aeruginosa*. Biochimie (2002) vol. 84 (5-6) pp. 499-510

Min Jou *et al*. Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. Nature (1972) vol. 237 (5350) pp. 82-8

Miyata *et al*. Use of the *Galleria mellonella* caterpillar as a model host to study the role of the type lll secretion system in *Pseudomonas aeruginosa* pathogenesis. Infection and Immunity (2003) vol. 71 (5) pp. 2404-13

Mizerska-Dudka and Andrejko. *Galleria mellonella* hemocytes destruction after infection with *Pseudomonas aeruginosa*. Journal of Basic Microbiology (2013) vol. 54 (3) pp. 232-46

Mohan *et al*. Transmission of *Pseudomonas aeruginosa* epidemic strain from a cystic fibrosis patient to a pet cat. Thorax (2008) vol. 63 pp. 839-40

Monson *et al*. The *Pseudomonas aeruginosa* generealized thransducing phage phiPA3 is a new member of the phiKZ-like group of 'jumbo' phages, and infects model laboratory strains and clinical isolates from cystic fibrosis patients. Microbiology (2011) vol. 157 (3) pp. 859-867

Mooij *et al*. Characterization of the integrated filamentous phage Pf5 and its involvement in small-colony formation. Microbiology (Reading, Engl) (2007) vol. 153 (Pt 6) pp. 1790-8

Nickels. Research Summary [online] (2009) Available at http://waksman.rutgers.edu/Waks/Nickels/nickels.html [accepted 26th October 2010]

Nies. The cobalt, zinc and cadmium efflux system CzcABC from *Alcaligenes eutrophus* functions as a cation-proton antiporter in *Escherichia coli*. Journal of Bacteriology (1995) vol. 177 (10) pp. 2707-12

Nishimori *et al*. *Pseudomonas plecoglossicida sp. nov*., the causative agent of bacterial haemorrhagic ascites of ayu, *Plecoglossus altivelis*. Int J Syst Evol Microbiol (2000) vol. 50 Pt 1 pp. 83-9

Nomura *et al*. A bacterial virulence protein suppresses host innate immunity to cause plant disease. Science (2006) vol. 313 (5784) pp. 220-3

Nudler. RNA polymerase active center: the molecular engine of transcription. Annu Rev Biochem (2009) vol. 78 pp. 335-61

O'Callaghan *et al*. Novel method for detection of β-lactamase by using a chromogenic cephalosporin substrate. Antimicrobial Agents and Chemotherapy (1972) vol. 1 (4) pp. 283-8

Ochsner *et al*. Production of *Pseudomonas aeruginosa* rhamnolipid biosurfactants in heterologous hosts. Appl Environ Microbiol (1995) vol. 61 (9) pp. 3503-6

Ochsner, Fietcher and Reiser. Isolation, characterization, and expression in *Escherichia coli* of the *Pseudomonas aeruginosa rhlAB* genes encoding a rhamnosyltransferase involved in rhamnolipid biosurfactant synthesis. The Journal of Biological Chemistry (1994) vol. 269 (31) pp. 19787-95

Ostroff *et al*. Molecular comparison of a nonhemolytic and a hemolytic phospholipase C from *Pseudomonas aeruginosa*. Journal of bacteriology (1990) vol. 172 (10) pp. 5915-23

Palleroni *et al*. Nucleic acid homologies in the genus *Pseudomonas*. Int J Syst Bacteriol (1973) vol. 23 pp (333-339)

Palleroni. Family 1. Pseudomonadaceae Winslow, Broadhurst, Buchanan, Krumwiede, Rogers and Smith, 1917. In Bergey's Manual of Systmematic Bacteriology, 1st edn, vol. 1 (1984). Krieg and Holt (eds). Baltimore, MD, USA: Williams & Wilkins, pp 141-218

Palleroni. The *Pseudomonas* story. Environmental microbiology (2010) vol. 12 (6) pp. 1377-83

Parad *et al*. Pulmonary outcome in cystic fibrosis is influenced primarily by mucoid *Pseudomonas aeruginosa* infection and immune status and only modestly by genotype. Infection and Immunity (1999) vol. 67 pp. 4744-50

Pearson *et al*. Roles of *Pseudomonas aeruginosa las* and *rhl* quorum-sensing systems in control of elastase and rhamnolipid biosynthesis genes. Journal of bacteriology (1997) vol. 179 (18) pp. 5756-67

Pearson *et al*. Structure of the autoinducer required for expression of *Pseudomonas aeruginosa* virulence genes. Proc Natl Acad Sci USA (1994) vol. 91 (1) pp. 197-201

Pedersen *et al*. Role of alginate in infection with mucoid *Pseudomonas aeruginosa* in cystic fibrosis. Thorax (1992) vol. 47 pp. 6-13

Peix *et al*. Historical evolution and current status of the taxonomy of genus *Pseudomonas*. Infect Genet Evol (2009) vol. 9 (6) pp. 1132-47

Perez-Montero *et al*. The embryonic linker Histone H1 variant of *Drosophila*, dBigH1, regulates zygotic genome activation. Developmental Cell (2013) vol. 26 (6) pp. 579-90

Pessi and Haas. Transcriptional control of the hydrogen cyanide biosynthetic genes *hcnABC* by the anaerobic regulator ANR and the quorum-sensing regulators LasR and RhlR in *Pseudomonas aeruginosa*. J Bacteriol (2000) vol. 182 (24) pp. 6940-9

Phillip *et al*. In vivo phospholipase activity of the *Pseudomonas aeruginosa* cytotoxin ExoU and protection of mammalian cells with phospholipase A2 inhibitors. J Biol Chem (2003) vol. 278 pp. 41326-32

Piddock. Multidrug-resistance efflux pumps — not just for resistance. Nature Reviews Microbiology (2006) vol. 4 pp. 629-36

Pier, Grout and Zaidi. Cystic fibrosis transmembrane conductance regulator is an epithelial cell receptor for clearance of *Pseudomonas aeruginosa* from the lung. Proc Natl Acad Sci (1997) vol. 94 pp. 12088-93

Pirnay *et al*. *Pseudomonas aeruginosa* population structure revisited. PLOS One (2009) vol. 4 (11) pp. e7740

Poole *et al*. Multiple antibiotic resistance in *Pseudomonas aeruginosa*: evidence for involvement of an efflux operon. Journal of Bacteriology (1993) vol. 175 (22) pp. 7363-72

Prasain *et al*. Soluble adenylyl cyclase-dependent microtubule disassembly reveals a novel mechanism of endothelial cell retraction. Am J Physiol Lung Cell Mol Physiol (2009) vol. 297 pp. 73-83

Pukatzki *et al*. The human pathogen *Pseudomonas aeruginosa* utilizes conserved virulence pathways to infect the social amoeba Dictyostelium discoideum. Proc Natl Acad Sci USA (2002) vol. 99 (5) pp. 3159-64

Pulcrano *et al*. Different mutations in *mucA* gene of *Pseudomonas aeruginosa* mucoid strains in cystic fibrosis patietns and their effect on *algU* gene expression. New Microbiology (2012) vol. 35 (3) pp. 295-305

Qiu *et al*. Regulated proteolysis controls mucoid conversion in *Pseudomonas aeruginosa*. Proc Natl Acad Sci USA (2007) vol. 104 (19) pp. 8107-12

Qui, Gurkar and Lory. Interstrain transfer of the large pathogenicity island (PAPI-1) of *Pseudomonas aeruginosa.* Proc Natl Acad Sci USA (2006) vol. 103 (52) pp. 19830-5

Rabin and Hauser. *Pseudomonas aeruginosa* ExoU, a toxin transported by the type III secretion system, kills *Saccharomyces cerevisiae*. Infect Immun (2003) vol. 71 (7) pp. 4144-50

Ramarao, Nielsen-Leroux and Lereclus. The insect *Galleria mellonella* as a powerful infection model to investigate bacterial pathogenesis. Journal of Visualized Experiments (2012) vol. 70 e4392

Ramirez *et al*. Metabolomics in toxicology and precinical research. ALTEX (2013) vol. 30 (2) pp. 209-25

Ramsey and Wozniak. A new polysaccharide resembling alginic acid isolated from Pseudomonads. Molecular Microbiology (2005) vol. 56 (2) pp. 309-22

Rau *et al*. Early adaptive developments of *Pseudomonas aeruginosa* after the transition from life in the environment to persistent colonization in the airways of human cystic fibrosis hosts. Environmental microbiology (2010) pp. 1643-58

Reimmann *et al*. PchC thioesterase optimises nonribosomal biosynthesis of the peptide siderophore pyochelin in *Pseudomonas aeruginosa*. Journal of Bacteriology (2004) vol. 186 (19) pp. 6367-73

Reitzer *et al*. Mutations that create new promoters suppress the sigma 54 dependence of *glnA* transcription in *Escherichia coli*. Journal of Bacteriology (1987) vol. 169 (9) pp. 4279-84

Riordan *et al*. Identification of the cystic fibrosis gene: cloning and characterisation of complementary DNA. Science (1989) vol. 245 pp. 1066-73

Rissman *et al*. Reordering contigs of draft genomes using the Mauve Aligner. Bioinformatics (2009) vol. 25 (18) pp. 2071-73

Ronaghi, Uhlen and Nyren. A sequencing method based on real-time pyrophosphate. Science (1998) vol. 281 (5375) pp. 363-5

Rosenstein and Cutting, for the Cystic Fibrosis Foundation Consensus Panel. The diagnosis of cystic fibrosis: a consensus statement. Journal of Paediatrics (1998) vol. 132 pp. 589-95

Roy *et al*. Complete genome sequence of the multiresistant taxonomic outlier *Pseudomonas aeruginosa* PA7. PLoS ONE (2010) vol. 5 (1) pp. e8842

Rusk. Torrent of sequence. Nature Methods (2011) vol. 8 (1) pp. 44

Rutherford *et al*. Artemis: sequence visualization and annotation. Bioinformatics (2000) vol. 16 (10) pp. 944-5

Saiman. Clinical utility of synergy testing for multidrug-resistant Pseudomonas aeruginosa isolated from patients with cystic fibrosis: 'the motion for'. Paediatric Respiratory Review (2007) vol. 8 pp. 249–55.

Sanger *et al*. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci (1997b) vol. 74 (12) pp. 5463-7

Sanger *et al*. Nucleotide sequence of bacteriophage phi X174 DNA. Nature (1997b) vol. 265 (5596) pp. 687-95

Sato *et al*. The mechanism of action of the *Pseudomonas aeruginosa*-encoded type III cytotoxin, ExoU. EMBO J (2003) vol. 22 (12) pp. 2959-69

Schilling *et al*. Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. Journal of Theoretical Biology (2000) vol. 203 pp. 249-83

Schilling *et al*. Genome-scale metabolic model of *Helicobacter pylori* 26695. Journal of Bacteriology (2002) vol. 184 pp. 4582-93

Schmidt, Tummler and Romling. Comparative genome mapping of Pseudomonas aeruginosa PAO with P. aeruginosa C, which belong to a major clone in cystic fibrosis patient and aquatic habitats. Journal of Bacteriology (1996) vol. 178 (1) pp. 85-93

Schurr *et al*. Control of AlgU, a member of the sigma E-like family of stress sigma factors, by the negative regulators MucA and MucB and *Pseudomonas aeruginosa* conversion to mucoidy in cystic fibrosis. Journal of Bacteriology (1996) vol. 178 (16) pp. 4997-5004

Schurr *et al*. Gene cluster controlling conversion to alginate-overproducing phenotype in *Pseudomonas aeruginosa*: functional analysis in a heterologous host and role in the instability of mucoidy. Journal of Bacteriology (1994) vol. 176 (11) pp. 3375-82

Schwyn and Neilands. Universal chemical assay for the detection and determination of siderophores. Analytical Biochemistry (1987) vol. 160 (1) pp. 47-56

Scott and Pitt. Identification and characterization of transmissible *Pseudomonas aeruginosa* strains in cystic fibrosis patients in England and Wales. Medical Microbiology (2004) vol. 53 pp. 609-15

Seng *et al*. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight

mass spectrometry. Clinical Infectious Diseases (2009) vol. 49 (4) pp. 543-51

Severinov and Darst. A mutant RNA polymerase that forms unusual open promoter complexes. Proc Natl Acad Sci USA (1997) vol. 94 (25) pp. 13481-6

Shao *et al*. mGenomeSubtractor: a web-based tool for parallel in silico subtractive hybridization analysis of multiple bacterial genomes. Nucleic Acids Res (2010) vol. 38 (Web Server issue) pp. W194-200

Smith *et al*. Evidence for diversifying selection at the pyoverdine locus of *Pseudomonas aeruginosa*. Journal of Bacteriology (2005) vol. 187 (6) pp. 2138-47

Smith *et al*. Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. Proc Natl Acad Sci USA (2006) vol. 103 (22) pp. 8487-92

Smyth *et al*. Prophylactic antibiotics for cystic fibrosis. Chochrane Database System Review (2003). Issue 3.

Soberón-Chávez *et al*. Production of rhamnolipids by *Pseudomonas aeruginosa*. Appl Microbiol Biotechnol (2005) vol. 68 (6) pp. 718-25

Sousa. Tie me up, tie me down: inhibiting RNA polymerase. Cell (2008) vol. 135 (2) pp. 205-7

Spaink *et al*. Promoters in the nodulation region of the *Rhizobium leguminosarum* Sym plasmid pRLIJI. Plant Molecular Biology (1987) pp. 1-13

Spencer *et al*. Whole genome sequence variation among multiple isolates of *Pseudomonas aeruginosa*. Journal of Bacteriology (2003) vol. 185 pp. 1316-25

Starkey *et al*. *Pseudomonas aeruginosa* rugose small-colony variants have adaptions that likely promote persistence in the cystic fibrosis lung. Journal of Bacteriology (2009) vol. 191 (11) pp. 3492-503

Stewart *et al*. Draft genomes of 12 host-adapted and environmental isolates of *Pseudomonas aeruginosa* and their positions in the core genome phylogeny. Pathogens and Disease (2013) vol. 71 (1) pp. 20-5

Stothard and Wishart. Circular genome visualization and exploration using CGView. Bioinformatics (2005) vol. 21 (4) pp. 537-539

Stover *et al*. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. Nature (2000) vol. 406 (6799) pp. 959-64

Studholme and Dixon. Domain architectures of sigma54-dependent transcriptional activators. J Bacteriol (2003) vol. 185 (6) pp. 1757-67

Tager, Wu and Vermeulen. The effect of chloride concentration on human neutrophil functions: potential relevance to cystic fibrosis. Am J Respir Cell Mol Biol (1998) vol. 19 pp. 643-52

Takase *et al*. Requirement of the *Pseudomonas aeruginosa tonB* gene for high-affinity iron acquisition and infection. Infect Immun (2000) vol. 68 (8) pp. 4498-504

Tamaki, Sato and Matsuhashi. Role of lipopolysaccharides in antibiotic resistance and bacteriophage adsorption of *Escherichia coli* K-12. Journal of Bacteriology (1971) vol. 105 (3) pp. 968-75

Tang *et al*. Role of *Pseudomonas aeruginosa* pili in acute pulmonary infection. Infection and Immunity (1995) vol. 63 (4) pp. 1278-85

Thompson *et al*. Identification and characterization of a chitinase antigen from *Pseudomonas aeruginosa* strain 385. Applied and Environmental Microbiology (2001) vol. 67 (9) pp. 4001-8

Tosi *et al*. Neutrophil elastase cleaves C3bi on opsonized *Pseudomonas* as well as CR1 on neutrophils to create a functionally important opsonin receptor mismatch. J Clin Invest (1990) vol. 86 (1) pp. 300-8

Totten *et al*. The *rpoN* gene product of *Pseudomonas aeruginosa* is required for expression of diverse genes, including the flagellin gene. Journal of bacteriology (1990) vol. 172 (1) pp. 389-96

Usher *et al*. Induction of neutrophil apoptosis by the *Pseudomonas aeruginosa* exotoxin pyocyanin: a potential mechanism of persistent infection. J Immunol (2002) vol. 168 (4) pp. 1861-8

Vallis *et al*. Biological effects of *Pseudomonas aeruginosa* type III-secreted proteins on CHO cells. Infection and Immunity (1999) vol. 67 pp. 2040-44

Vasseur *et al*. The *pel* gene of the *Pseudomonas aeruginosa* PAK strain are involved at early and late stages of biofilm formation. Microbiology (2004) vol. 151 (3) pp. 985-97

Venter *et al*. The sequence of the human genome. Science (2001) vol. 291 (5507) pp. 1304-51

Venturi. Regulation of quorum sensing in *Pseudomonas*. FEMS microbiology reviews (2006) vol. 30 (2) pp. 274-91

Visca *et al*. Pyoverdine siderophores: from biogenesis to biosignificance. Trends Microbiol (2007) vol. 15 (1) pp. 22-30

Walker *et al*. *Pseudomonas aeruginosa*-plant root interactions. Pathogenicity, biofilm formation, and root exudation. Plant Physiol (2004) vol. 134 (1) pp. 320-31

Webb *et al*. Cell death in *Pseudomonas aeruginosa* biofilm development. Journal of Bacteriology (2003) vol. 185 (15) 4585-92

Webb, Lau and Kjelleberg. Bacteriophage and phenotypic variation in *Pseudomonas aeruginosa* biofilm development. Journal of Bacteriology (2004) vol. 186 (23) pp. 8066-73

Whiteley *et al*. Gene expression in *Pseudomonas aeruginosa* biofilms. Nature (2001) vol. 413 (6858) pp. 860-4

Wiehlmann *et al*. Population structure of *Pseudomonas aeruginosa*. Proc Natl Acad Sci (2007) vol. 104 (19) pp. 8101-6

Wiehlmann *et al*. Population structure of *Pseudomonas aeruginosa*. Proc Natl Acad Sci USA (2007) vol. 104 (19) pp. 8101-6

Wilhelm, Tommassen and Jaeger. A novel lipolytic enzyme located in the outer membrane of *Pseudomonas aeruginosa.* Journal of Bacteriology (1999) vol. 181 (22) pp. 6977-6986

Winder *et al*. Global metabolic profiling of *Escherichia coli* cultures: an evaluation of methods for quenching and extraction of intracellular metabolites. Analytical Chemistry (2008) vol. 80 (8) pp. 2939-48

Windsor *et al*. *Pseudomonas* genome database: Improved comparative analysis and population genomics capability for Pseudomonas genomes. Nucleic Acids Res (2011) vol. 39 pp. 596-600

Winkler and Stuckmann. Glycogen, hyaluronate and some other polysaccharides greatly enhance the formation of exolipase by *Serratia marcescens*. Journal of Bacteriology (1979) vol. 138 (3) pp. 663-670

Winsor *et al*. *Pseudomonas* Genome Database: facilitating user-friendly, comprehensive comparisons of microbial genomes. Nucleic Acids Res (2009) vol. 37 (Database issue) pp. D483-8

Winstanley *et al*. Newly introduced genomic prophage islands are critical determinants of *in vivo* competitiveness in the Liverpool Epidemic Strain of *Pseudomonas aeruginosa*. Genome Res (2009) vol. 19 (1) pp. 12-23

Woods *et al*. *In vivo* regulation of virulence in *Pseudomonas aeruginosa* associated with genetic rearrangement. Journal of Infectious Diseases (1991) vol. 163 (1) pp. 143-9

Wozniak *et al*. Alginate is not a significant component of the extracellular polysaccharide matrix of PA14 and PAO1 *Pseudomonas aeruginosa* biofilms. Proc Natl Acad Sci USA (2003) vol. 100 (13) pp. 7907-12

Xia *et al*. MetaboAnalyst 2.0 - a comprehensive server for metabolomic data analysis. Nucleic Acids Research (2012) vol. 40 (W1) pp. W127-33

Xiong and Jayaswal. Molecular characterisation of a chromosomal determinant conferring resistance to zinc and cobalt ions in *Staphylococcus aureus*. Journal of Bacteriology (1998) vol. 180 (16) pp. 4024-9

Yahr *et al*. ExoY, an adenylate cyclase secreted by the *Pseudomonas aeruginosa* type III system. Proc Natl Acad Sci USA (1998) vol. 95 (23) pp. 13899-904

Yanagihara *et al*. Role of elastase in a mouse model of chronic respiratory *Pseudomonas aeruginosa* infection that mimics diffuse panbronchiolitis. Journal of Medical Microbiology (2003) vol. 52 (6) pp. 531-5

Zulianello *et al*. Rhamnolipids are virulence factors that promote early infiltration of primary human airway epithelia by *Pseudomonas aeruginosa*. Infection and Immunity (2006) vol. 74 (6) pp. 3134-47

Pathogens and Disease ISSN 2049-632X

RESEARCH ARTICLE

Draft genomes of 12 host-adapted and environmental isolates of *Pseudomonas aeruginosa* and their positions in the core genome phylogeny

Lewis Stewart[1], Amy Ford[1]*, Vartul Sangal[1], Julie Jeukens[2], Brian Boyle[2], Iréna Kukavica-Ibrulj[2], Shabhonam Caim[3], Lisa Crossman[3], Paul A. Hoskisson[1], Roger Levesque[2] & Nicholas P. Tucker[1]

1 Strathclyde Institute of Pharmacy and Biomedical Science, University of Strathclyde, Glasgow, UK
2 Institut de Biologie Intégrative et des Systèmes, Université Laval, Québec City, QC, Canada
3 The Genome Analysis Centre, Norwich Research Park, Norwich, UK

This shortomics study describes the genomes of 12 isolates of *Pseudomonas aeruginosa* and highlights the genome diversity of this organism using non-mucoid/mucoid pairs from a limited number of isolates.

Keywords
Pseudomonas aeruginosa; cystic fibrosis; genomics; core genome; phylogeny; host adaptation.

Correspondence
Nicholas P. Tucker, Strathclyde Institute of Pharmacy and Biomedical Science, University of Strathclyde, 161 Cathedral Street, Glasgow G4 0RE, UK.
Tel.: (44) 141 548 2861
fax: (44) 141 552 2562
e-mail: nick.tucker@strath.ac.uk

***Present address**
Amy Ford, Centre for Infection and Immunity, Health Sciences Building, 97 Lisburn Road, Belfast, BT9 7BL, UK

Received 10 July 2013; revised 11 October 2013; accepted 13 October 2013.

doi:10.1111/2049-632X.12107

Editor: David Rasko

Abstract

Pseudomonas aeruginosa is a Gram-negative opportunistic pathogen particularly associated with the inherited disease cystic fibrosis (CF). *Pseudomonas aeruginosa* is well known to have a large and adaptable genome that enables it to colonise a wide range of ecological niches. Here, we have used a comparative genomics approach to identify changes that occur during infection of the CF lung. We used the mucoid phenotype as an obvious marker of host adaptation and compared these genomes to analyse SNPs, indels and islands within near-isogenic pairs. To commence the correction of the natural bias towards clinical isolates in genomics studies and to widen our understanding of the genomic diversity of *P. aeruginosa*, we included four environmental isolates in our analysis. Our data suggest that genome plasticity plays an important role in chronic infection and that the strains sequenced in this study are representative of the two major phylogenetic groups as determined by core genome SNP analysis.

Introduction

It is well established that the genome of *Pseudomonas aeruginosa* has many regions that exhibit genome plasticity (Lee *et al.*, 2006; Klockgether *et al.*, 2011; Bezuidt *et al.*, 2013). The comparatively large genome of *P. aeruginosa* provides significant potential for strain diversity and differing clinical outcomes (Lee *et al.*, 2006; Fothergill *et al.*, 2010; Klockgether *et al.*, 2011; Silby *et al.*, 2011; Bezuidt *et al.*, 2013; Gellatly & Hancock, 2013). Several studies have quantified genomic changes in both a linear and contemporary manner during the infection of individual patients with cystic fibrosis (CF) and this has provided strong evidence of host adaptation during these chronic infections (Terry *et al.*, 1992; Darling & Evans, 2003; Smith *et al.*, 2006; Winstanley *et al.*, 2009; Cramer *et al.*, 2011; Chung *et al.*, 2012; Workentine *et al.*, 2013). Genomic variation is known to arise through a range of mechanisms including the mobilisation of prophages and loss of function mutations in DNA mismatch repair gene such as *mutS* (Lee *et al.*, 2006; Mena *et al.*, 2008; Winstanley *et al.*, 2009; Ciofu *et al.*, 2010; Oliver & Mena, 2010; Klockgether *et al.*, 2011; Chung *et al.*, 2012; Bezuidt *et al.*, 2013). Genome sequence polymorphism and phage mobilisation of *P. aeruginosa* in the CF

lung has been suggested to be a potential complication during strain typing (Fothergill *et al.*, 2010; Silby *et al.*, 2011; Gellatly & Hancock, 2013). Whole-genome sequencing provides fine-scale analysis with which to evaluate these changes, in addition to allowing the comparison of known virulence factors. This study is motivated by the need for a better understanding of genomic diversity in *P. aeruginosa* as well as the more practical application of producing a panel of well characterised strains for antibiotic drug discovery.

Provisional comparative genomic analyses of the strains sequenced in this work

In this study, we sequenced 12 strains using an Illumina 50-bp paired-end approach with an average insert size of 150 bp. Sequencing was carried out at the GenePool (University of Edinburgh) using a single lane on an Illumina GAII instrument. The resulting data were analysed using CLC Genomics Workbench to generate nonscaffolded assemblies. These strains include six nonmucoid/mucoid pairs from three patients with CF, four environmental isolates and a laboratory evolved small colony variant (SCV) variant and its isogenic progenitor strain originally isolated from a patient with CF. The six nonmucoid/mucoid pairs were originally isolated by the UK CF Microbiology Consortium based at the University of Edinburgh and include strains J1385 and J1532, which have been used in previous studies (Terry *et al.*, 1992; Darling & Evans, 2003; Smith *et al.*, 2006; Winstanley *et al.*, 2009; Cramer *et al.*, 2011; Chung *et al.*, 2012; Workentine *et al.*, 2013). Variations in the representation of each strain in the barcoded pool lead to a range of read depths. As a consequence, the number of contigs in the final assemblies ranged between 172 and 1009. Unannotated contigs have been deposited in GenBank under the Bioproject designation PRJNA200798, and the relevant accession numbers are indicated in Table 1 along with the *P. aeruginosa* genomes currently available from either GenBank or from the Pseudomonas Genome Database (Winsor *et al.*, 2010). Contigs were then reordered using Mauve and concatenated prior to annotation with RAST and xBASE (Aziz *et al.*, 2008; Chaudhuri *et al.*, 2008). To retain contig boundary information in resulting annotated GenBank files, we read the reordered contigs into the draft genomes using Artemis (Carver *et al.*, 2008). This approach allowed us to rapidly generate an annotated list of SNPs and InDels by mapping the reads of mucoid strains to their nonmucoid progenitor using Maq and CLC Genomics Workbench using a minimum SNP percentage cut-off of 85%. In summary, we identified 207 nonsynonymous SNPs in J1532 and 82 in C1433 when compared to J1385 and C1426, respectively. Work is currently underway to determine whether any of these nonsynonymous SNPs affect the phenotypes of these mucoid strains relative to their nonmucoid progenitor. As expected, all three mucoid strains in our study contained frameshift mutations in the *mucA* gene encoding the antisigma factor responsible for sequestering AlgU away from its target promoters, including the alginate biosynthesis operon (Fyfe & Govan, 1980; Martin *et al.*, 1993). Strains J1385 & J1532, C1426 & C1433 and C763 & C1334 are paired nonmucoid/mucoid isolates from three separate patients. BLAST and SNP analysis revealed that pairs J1385 & J1532 and C1426 & C1433 can be considered to be derived from the same strain, whereas C763 & C1334 are not, suggesting either a multistrain infection or strain succession event in the latter case. Within the near-isogenic mucoid pair J1385 and J1532, it is interesting to note that the mucoid strain has a smaller genome of than its nonmucoid progenitor. There is significant evidence of genome shrinkage during bacterial adaptation to a host (Toft & Andersson, 2010), and this is a possible scenario in the case of J1532 that has a significantly smaller genome than its nonmucoid progenitor, J1385. Phylogenetic analysis revealed that J1385 and J1532 are of the PA14 strain type. Notable islands and phages missing from J1532 compared to J1385 include the filamentous phage PF1 and a large deletion adjacent to the *exoY* locus. Moreover, J1385 is the only strain in our data set to encode PF1. ExoY is an adenylate cyclase-type effector protein that is secreted by the type three secretion system (T3SS) of *P. aeruginosa*. During the course of our analysis, we noted that strain PA14 has a frameshift in the *exoY* gene that is also present in strain J1385, consistent with our observation that these strains are genomically similar. The environmental strain E2UoS possesses a different frameshift mutation in *exoY* that has been previously identified in other strains (Ajayi *et al.*, 2003). Strain J1532 lacks the *exoY* gene completely owing to a 65-Kbp deletion relative to its putative parent, J1385. None of the other T3SS *exo* effectors were altered in this way in any other strain. The genome of strain C763 is amongst the largest *P. aeruginosa* genome currently reported at over 7 Mbp.

Core genome SNP analysis reveals three distinct phylogenetic groups of *P. aeruginosa*

We used PANSEQ (Laing *et al.*, 2010) to analyse the core genomes of 55 *P. aeruginosa* strains, including the 12 genomes sequenced in this work using a core genome threshold of 55 and a sequence identity of 90%. These constraints represent the strictest definition of the core genome because each sequence must be present in every strain. The advantage of this method is that it separates only the most conserved portions of the genome such that the resulting phylogenies are not influenced by horizontal gene transfer events in the accessory genome. PANSEQ identified the SNPs in the core genomes of all 55 strains, allowing phylogenetic analysis using SPLITSTREE (version 4.13.1; Kloepper & Huson, 2008). By analysing the SNPs in the core genome, it was possible to confirm that C1426 and C1433 as well as J1385 and J1532 are indeed near-isogenic pairs. The two environmental strains isolated from Mount St. Helens (MSH3 and MSH10) used in the original PA14 genome study are effectively the same strain with only very minor differences.

Of all the 55 strains analysed in this work using PANSEQ 2.0, only one isolate (VRFPA01) was phylogenetically

Table 1 *Pseudomonas aeruginosa* genomes and corresponding accession numbers used in this study

Strain identifier	Comments/Origin	Phylogenetic group	Accession	Reference
J1385	CF, progenitor of J1532	2	ASRB00000000	This study
J1532	CF, mucoid	2	ASQZ00000000	This study
C1426	CF, progenitor of C1433	1	ASRD00000000	This study
C1433	CF, mucoid	1	ASRC00000000	This study
PA17	CF, progenitor of PA17SCV	1	ASQY00000000	This study
PA17SCV	Lab evolved from CF Isolate	1	ASRA00000000	This study
C763	CF, from same patient as C1334	1	ASQS00000000	This study
C1334	CF, mucoid, from same patient as C763	1	ASQT00000000	This study
E2UOS	Soil	2	ASQV00000000	This study
PA62	Soil	1	ASQW00000000	This study
MSH3	Environmental, Mount St. Helens	1	ASQU00000000	This study
MSH10 UoS	Environmental, Mount St. Helens	1	ASQX00000000	This study
2192	CF	1	AAKW00000000	Mathee *et al.* (2008)
39016	Keratitis	2	AEEX01000000	Stewart *et al.* (2011)
138244	Sputum, non-CF	1	AEVV00000000	Soares-Castro *et al.* (2011)
152504	Sputum, non-CF	2	AEVW00000000	Soares-Castro *et al.* (2011)
18A	CF	1	CAQZ00000000	Unpublished data
AES-1R	Australia, CF	1	AFNF00000000	Naughton *et al.* (2011)
AH16	China, sputum, non-CF	1	ALJH00000000	Wu *et al.* (2012)
ATCC 14886	Soil	1	AKZD00000000	Chugani *et al.* (2012)
ATCC 25324	Glass-crusher air plate	1	AKZE00000000	Unpublished data
ATCC 700888	Industrial water system	2	AKZF00000000	Chugani *et al.* (2012)
B136.33	Diarrhoea	2	CP004061	Unpublished data
C3719	Manchester Ep Strain	1	AAKV00000000	Mathee *et al.* (2008)
CI27	Sputum, CF	2	AKZG00000000	Chugani *et al.* (2012)
CIG1	Sputum, CF	1	AKBD00000000	Chugani *et al.* (2012)
DK2	CF, Denmark	1	NC_018080.1	Rau *et al.* (2012)
DK2_CF510	CF, Denmark	1	AJHI00000000	Rau *et al.* (2012)
DQ8	Crude oil, China	1	ALIO00000000	Gai *et al.* (2012)
E2UoW	Plant, USA	1	AKZH00000000	Chugani *et al.* (2012)
LCT-PA102	Blood	1	AJKG00000000	Fang *et al.* (2012)
LESB58	Liverpool Epidemic Strain	1	NC_011770.1	Winstanley *et al.* (2009)
M18	Plant isolate, China	1	CP002496	Wu *et al.* (2011)
MRW44.1	Mutator lineage 44	1	ALBW00000000	Weigand & Sundin (2012)
MSH-10	Environmental	1	ASWW00000000	Broad Institute, Unpublished data
N002	Crude oil, India	1	ALBV00000000	Roy *et al.* (2013)
NCGM2.S1	Urinary tract infection	2	AP012280	Miyoshi-Akiyama *et al.* (2011)
NCMG1179	Respiratory, Japan	1	BADP00000000	Tada *et al.* (2011)
PA01	Burn	1	NC_002516.2	Stover *et al.* (2000)
PA14	Plant, USA	2	NC_008463.1	Lee *et al.* (2006)
PA19BR	Polymixin resistant, Brazil	1	AFXJ00000000	Boyle *et al.* (2012)
PA21_ST175	Blood, MDR	1	AOIH00000000	Viedma *et al.* (2013)
PA213BR	Polymixin resistant, Brazil	1	AFXK00000000	Boyle *et al.* (2012)
PA45	Blood, Italy	1	APMD00000000	Segata *et al.* (2013)
PA7	Nonrespiratory clinical isolate	7	NC_009656.1	Roy *et al.* (2010)
PA9BR	Brazil, non-CF	1	AFXI00000000	Boyle *et al.* (2012)
PAb1	Frostbite	2	ABKZ00000000	Salzberg *et al.* (2008)
PABL056	Blood, USA	2	ALPS00000000	Ozer *et al.* (2012)
PACS2	CF	1	NZ_AAQW00000000	Mathee *et al.* (2008)
PAK	PAK	1	ASWU00000000	Broad Institute, Unpublished data
PAO579	Mucoid	1	ALOF00000000	Withers *et al.* (2012)
SJTD-1	Crude oil, China	1	AKCM00000000	Liu *et al.* (2012)
VRFPA01	Blood, India	7	AOBK00000000	Malathi & Madhavan, unpublished data
VRFPA02	Eye, India	1	AQHM00000000	Malathi & Madhavan, unpublished data
XMG	Soil, China	1	AJXX00000000	Gao *et al.* (2012)

The first 12 strains are newly released as part of the present work.

related to PA7, a known taxonomic outlier (Roy *et al.*, 2010). VRFPA01 is a blood isolate from India and to the best of our knowledge is the first PA7-like isolate to be sequenced since the original Argentinian strain (J. Malathi & H. N. Madhavan, pers. commun.). Analysis of core genome in all 55 strains using PANSEQ identified 118934 SNP positions. When PA7 and VRFPA01 were excluded from the analysis, this number fell to 73206, demonstrating how distantly related these isolates are from the rest of *P. aeruginosa* lineage. Perhaps the most striking observation from this analysis can be seen in the UPGMA phylogenetic tree presented in Fig. 1, which identifies two major subgroups of *P. aeruginosa*. Distances were calculated using the uncorrected P method, and all gap sites were removed prior to the analysis. It is clear from this core genome phylogeny that *P. aeruginosa* can be broadly divided into two major subgroups with the rarer PA7 containing group as a third outlier. We carried out 1000 bootstraps on the core SNP tree and the major splits that separate the key groups all returned scores of 100. However, bootstrap values dropped significantly in the centre of group 1, in which a roughly equal distance separates most branches. One explanation for these lower bootstrap scores is recombination within the core genome, which has been proposed in MLST-based studies (Kiewitz & Tümmler, 2000; Maatallah *et al.*, 2011). When the phylogeny is viewed as a neighbour network, it is clear that the large number of parallel paths in the hub of group 1 is in agreement with this assertion (data not shown). For the sake of clarity here, we have arbitrarily named the two major groups one and two (Fig. 1) and the PA7 and VRFPA01 outliers group three. Group one is clearly the larger of the two main groups and contains notable strains such as DK2, PA01, LESB58 and PAK. Group two is less well populated than group 1, consisting of only 12 strains including PA14. From whole genome BLAST and Maq reference mapping analyses, we previously noted that the nonmucoid/mucoid pair J1385 and J1532 sequenced in this study are very similar to PA14 and this is clearly supported by core phylogeny presented here (Fig. 1a). Isolates from all sources, including CF sputum, are well represented in both groups except for the hydrocarbon remediation strains, which are all located in group 1. At the present time, it is

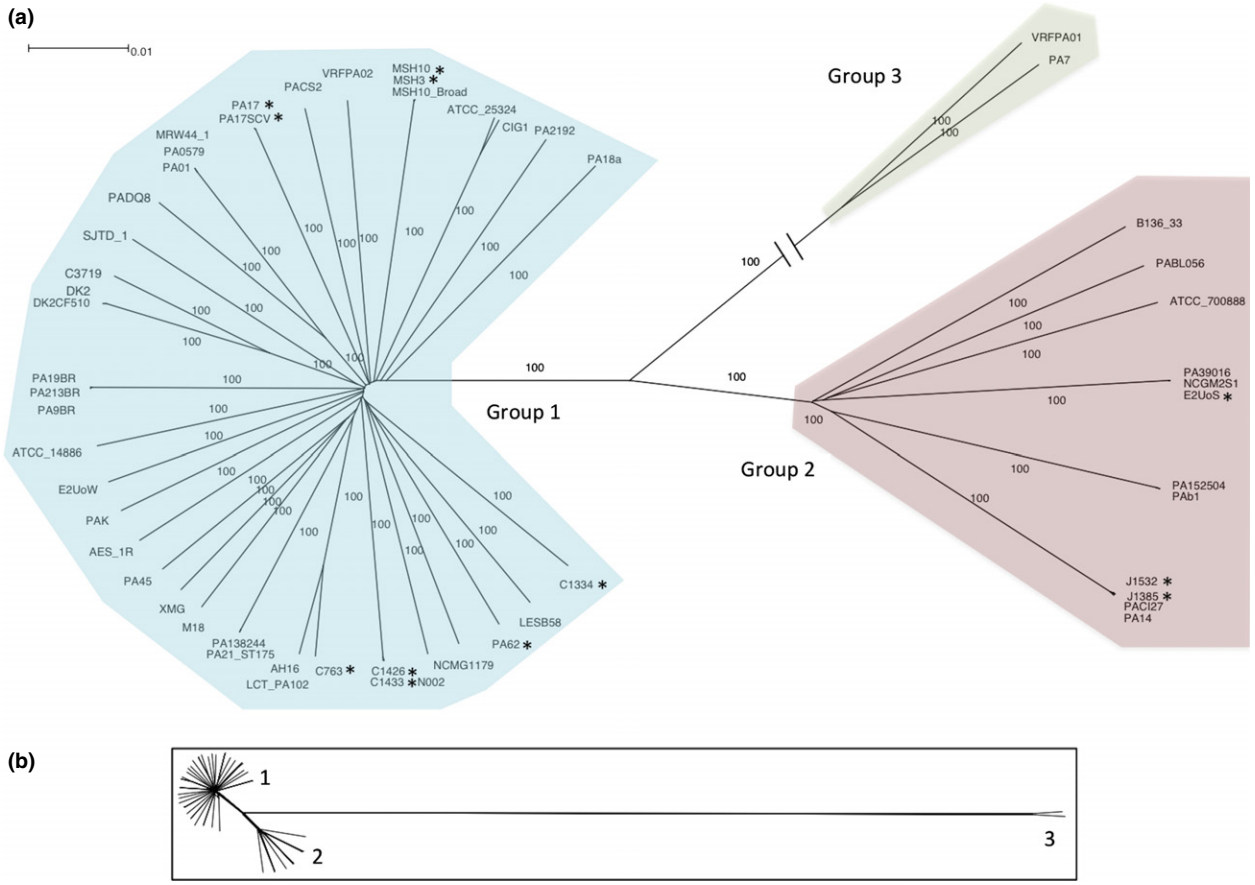


Fig. 1 (a) Unrooted UPGMA tree of 55 *Pseudomonas aeruginosa* genomes based on SNPs within the core genome as defined by PANSEQ with a core genome threshold of 55 and required sequence identity of 90%. The three groups are indicated in blue (group 1), pink (group 2) and green (group 3). Strains sequenced as part of this study are indicated with an asterisk. The tree was prepared in SPLITSTREE 4 and was bootstrapped with 1000 replicates. Because group three is so distant from groups one and two, it has been truncated for clarity. (b) An overview of the tree in (a) provides an impression of the true distances.

unclear whether group one strains are naturally more abundant than group two or whether a bias has been introduced during the selection of strains for the various sequencing projects currently underway. Given that phylogenetic analysis of 55 genomes has produced 35 distinct branches, it seems likely that more extensive survey sequencing in the future will populate these branches. More detailed sequencing is currently underway to improve the assemblies of the nonmucoid CF isolates C1426 and J1385.

Acknowledgements

This work was supported by a starter grant from Strathclyde Institute of Pharmacy and Biomedical Science at the University of Strathclyde to N.P.T. R.C.L. is funded by Cystic Fibrosis Canada and J.J. holds a Cystic Fibrosis Canada postdoctoral research fellowship. The collaboration between Strathclyde and Université Laval, Integrative and systems biology Institute (IBIS) was supported by travel grants from The Society for General Microbiology and the Scottish Universities Life Science Alliance to N.P.T. The authors are grateful to The Genepool at the University of Edinburgh for providing the Illumina data.

References

Ajayi T, Allmond LR, Sawa T & Wiener-Kronish JP (2003) Single-nucleotide-polymorphism mapping of the *Pseudomonas aeruginosa* type III secretion toxins for development of a diagnostic multiplex PCR system. *J Clin Microbiol* 41: 3526–3531.

Aziz RK, Bartels D, Best AA *et al.* (2008) The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9: 75.

Bezuidt OK, Klockgether J, Elsen S, Attree I, Davenport CF & Tümmler B (2013) Intraclonal genome diversity of *Pseudomonas aeruginosa* clones CHA and TB. *BMC Genomics* 14: 416.

Boyle B, Fernandez L, Laroche J, Kukavica-Ibrulj I, Mendes CMF, Hancock RW & Levesque RC (2012) Complete genome sequences of three *Pseudomonas aeruginosa* isolates with phenotypes of polymyxin B adaptation and inducible resistance. *J Bacteriol* 194: 529–530.

Carver T, Berriman M, Tivey A, Patel C, Böhme U, Barrell BG, Parkhill J & Rajandream M-A (2008) Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* 24: 2672–2676.

Chaudhuri RR, Loman NJ, Snyder LAS, Bailey CM, Stekel DJ & Pallen MJ (2008) xBASE2: a comprehensive resource for comparative bacterial genomics. *Nucleic Acids Res* 36: D543–D546.

Chugani S, Kim BS, Phattarasukol S, Brittnacher MJ, Choi SH, Harwood CS & Greenberg EP (2012) Strain-dependent diversity in the *Pseudomonas aeruginosa* quorum-sensing regulon. *P Natl Acad Sci USA* 109: E2823–E2831.

Chung JCS, Becq J, Fraser L, Schulz-Trieglaff O, Bond NJ, Foweraker J, Bruce KD, Smith GP & Welch M (2012) Genomic variation among contemporary *Pseudomonas aeruginosa* isolates from chronically infected cystic fibrosis patients. *J Bacteriol* 194: 4857–4866.

Ciofu O, Mandsberg LF, Bjarnsholt T, Wassermann T & Høiby N (2010) Genetic adaptation of *Pseudomonas aeruginosa* during chronic lung infection of patients with cystic fibrosis: strong and weak mutators with heterogeneous genetic backgrounds emerge in mucA and/or lasR mutants. *Microbiology* 156: 1108–1119.

Cramer N, Klockgether J, Wrasman K, Schmidt M, Davenport CF & Tümmler B (2011) Microevolution of the major common *Pseudomonas aeruginosa* clones C and PA14 in cystic fibrosis lungs. *Environ Microbiol* 13: 1690–1704.

Darling KEA & Evans TJ (2003) Effects of nitric oxide on *Pseudomonas aeruginosa* infection of epithelial cells from a human respiratory cell line derived from a patient with cystic fibrosis. *Infect Immun* 71: 2341–2349.

Fang X *et al.* (2012) Draft genome sequence of *Pseudomonas aeruginosa* strain ATCC 27853. *J Bacteriol* 194: 3755.

Fothergill JL, White J, Foweraker JE, Walshaw MJ, Ledson MJ, Mahenthiralingam E & Winstanley C (2010) Impact of *Pseudomonas aeruginosa* genomic instability on the application of typing methods for chronic cystic fibrosis infections. *J Clin Microbiol* 48: 2053–2059.

Fyfe JA & Govan JR (1980) Alginate synthesis in mucoid *Pseudomonas aeruginosa*: a chromosomal locus involved in control. *J Gen Microbiol* 119: 443–450.

Gai Z, Zhang Z, Wang X, Tao F, Tang H & Xu P (2012) Genome sequence of *Pseudomonas aeruginosa* DQ8, an efficient degrader of n-alkanes and polycyclic aromatic hydrocarbons. *J Bacteriol* 194: 6304–6305.

Gao C *et al.* (2012) Genome sequence of the lactate-utilizing *Pseudomonas aeruginosa* strain XMG. *J Bacteriol* 194: 4751–4752.

Gellatly SL & Hancock REW (2013) *Pseudomonas aeruginosa*: new insights into pathogenesis and host defenses. *Pathog Dis* 67: 159–173.

Kiewitz C & Tümmler B (2000) Sequence diversity of *Pseudomonas aeruginosa*: impact on population structure and genome evolution. *J Bacteriol* 182: 3125–3135.

Klockgether J, Cramer N, Wiehlmann L, Davenport CF & Tümmler B (2011) *Pseudomonas aeruginosa* genomic structure and diversity. *Front Microbiol* 2: 150.

Kloepper TH & Huson DH (2008) Drawing explicit phylogenetic networks and their integration into SPLITSTREE. *BMC Evol Biol* 8: 22.

Laing C, Buchanan C, Taboada EN, Zhang Y, Kropinski A, Villegas A, Thomas JE & Gannon VPJ (2010) Pan-genome sequence analysis using PANSEQ: an online tool for the rapid analysis of core and accessory genomic regions. *BMC Bioinformatics* 11: 461.

Lee DG *et al.* (2006) Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial. *Genome Biol* 7: R90.

Liu H, Liang R, Tao F, Ma C, Liu Y, Liu X & Liu J (2012) Genome sequence of *Pseudomonas aeruginosa* strain SJTD-1, a bacterium capable of degrading long-chain alkanes and crude oil. *J Bacteriol* 194: 4783–4784.

Maatallah M *et al.* (2011) Population structure of *Pseudomonas aeruginosa* from five Mediterranean countries: evidence for frequent recombination and epidemic occurrence of CC235. *PLoS One* 6: e25617.

Martin DW, Holloway BW & Deretic V (1993) Characterization of a locus determining the mucoid status of *Pseudomonas aeruginosa*: AlgU shows sequence similarities with a Bacillus sigma factor. *J Bacteriol* 175: 1153–1164.

Mathee K *et al.* (2008) Dynamics of *Pseudomonas aeruginosa* genome evolution. *P Natl Acad Sci USA* 105: 3100–3105.

Mena A, Smith EE, Burns JL, Speert DP, Moskowitz SM, Perez JL & Oliver A (2008) Genetic adaptation of *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients is catalyzed by hypermutation. *J Bacteriol* 190: 7910–7917.

Miyoshi-Akiyama T, Kuwahara T, Tada T, Kitao T & Kirikae T (2011) Complete genome sequence of highly multidrug-resistant *Pseudomonas aeruginosa* NCGM2.S1, a representative strain of a cluster endemic to Japan. *J Bacteriol* 193: 7010.

Naughton S *et al.* (2011) *Pseudomonas aeruginosa* AES-1 exhibits increased virulence gene expression during chronic infection of cystic fibrosis lung. *PLoS One* 6: e24526.

Oliver A & Mena A (2010) Bacterial hypermutation in cystic fibrosis, not only for antibiotic resistance. *Clin Microbiol Infect* 16: 798–808.

Ozer EA, Allen JP & Hauser AR (2012) Draft genome sequence of the *Pseudomonas aeruginosa* bloodstream isolate PABL056. *J Bacteriol* 194: 5999.

Rau MH, Marvig RL, Ehrlich GD, Molin S & Jelsbak L (2012) Deletion and acquisition of genomic content during early stage adaptation of *Pseudomonas aeruginosa* to a human host environment. *Environ Microbiol* 14: 2200–2211.

Roy PH *et al.* (2010) Complete genome sequence of the multiresistant taxonomic outlier *Pseudomonas aeruginosa* PA7. *PLoS One* 5: e8842.

Roy AS, Baruah R, Gogoi D, Borah M, Singh AK & Deka Boruah HP (2013) Draft Genome Sequence of *Pseudomonas aeruginosa* Strain N002, Isolated from Crude Oil-Contaminated Soil from Geleky, Assam, India. *Genome Announc* 1(1): e00104–12.

Salzberg SL, Sommer DD, Puiu D & Lee VT (2008) Gene-boosted assembly of a novel bacterial genome from very short reads. *PLoS Comput Biol* 4: e1000186.

Segata N, Ballarini A & Jousson O (2013) Genome sequence of *Pseudomonas aeruginosa* PA45, a highly virulent strain isolated from a patient with bloodstream infection. *Genome Announc* 1(3): e00289–13.

Silby MW, Winstanley C, Godfrey SAC, Levy SB & Jackson RW (2011) *Pseudomonas* genomes: diverse and adaptable. *FEMS Microbiol Rev* 35: 652–680.

Smith EE *et al.* (2006) Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *P Natl Acad Sci USA* 103: 8487–8492.

Soares-Castro P, Marques D, Demyanchuk S, Faustino A & Santos PM (2011) Draft genome sequences of two *Pseudomonas aeruginosa* clinical isolates with different antibiotic susceptibilities. *J Bacteriol* 193: 5573.

Stewart RMK *et al.* (2011) Genetic characterization indicates that a specific subpopulation of *Pseudomonas aeruginosa* is associated with keratitis infections. *J Clin Microbiol* 49: 993–1003.

Stover CK *et al.* (2000) Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* 406: 959–964.

Tada T, Kitao T, Miyoshi-Akiyama T & Kirikae T (2011) Genome sequence of multidrug-resistant *Pseudomonas aeruginosa* NCGM1179. *J Bacteriol* 193: 6397.

Terry JM, Piña SE & Mattingly SJ (1992) Role of energy metabolism in conversion of nonmucoid *Pseudomonas aeruginosa* to the mucoid phenotype. *Infect Immun* 60: 1329–1335.

Toft C & Andersson SGE (2010) Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat Rev Genet* 11: 465–475.

Viedma E, Juan C, Otero JR, Oliver A & Chaves F (2013) Draft genome sequence of VIM-2-producing multidrug-resistant *Pseudomonas aeruginosa* ST175, an epidemic high-risk clone. *Genome Announc* 1: e0011213.

Weigand MR & Sundin GW (2012) General and inducible hypermutation facilitate parallel adaptation in *Pseudomonas aeruginosa* despite divergent mutation spectra. *P Natl Acad Sci USA* 109: 13680–13685.

Winsor GL, Lam DKW, Fleming L, Lo R, Whiteside MD, Yu NY, Hancock REW & Brinkman FSL (2010) *Pseudomonas* Genome Database: improved comparative analysis and population genomics capability for Pseudomonas genomes. *Nucleic Acids Res* 39: D596–D600.

Winstanley C *et al.* (2009) Newly introduced genomic prophage islands are critical determinants of *in vivo* competitiveness in the Liverpool Epidemic Strain of *Pseudomonas aeruginosa*. *Genome Res* 19: 12–23.

Withers TR, Johnson SL & Yu HD (2012) Draft genome sequence for *Pseudomonas aeruginosa* strain PAO579, a mucoid derivative of PAO381. *J Bacteriol* 194: 6617.

Workentine ML, Sibley CD, Glezerson B, Purighalla S, Norgaard-Gron JC, Parkins MD, Rabin HR & Surette MG (2013) Phenotypic heterogeneity of *Pseudomonas aeruginosa* populations in a cystic fibrosis patient. *PLoS One* 8: e60225.

Wu D-Q, Ye J, Ou H-Y, Wei X, Huang X, He Y-W & Xu Y (2011) Genomic analysis and temperature-dependent transcriptome profiles of the rhizosphere originating strain *Pseudomonas aeruginosa* M18. *BMC Genomics* 12: 438.

Wu D-Q, Cheng H, Wang C, Zhang C, Wang Y, Shao J & Duan Q (2012) Genome sequence of *Pseudomonas aeruginosa* strain AH16, isolated from a patient with chronic pneumonia in China. *J Bacteriol* 194: 5976–5977.

