Application of Whole Genome Sequencing to the Study of *Pseudomonas aeruginosa*

Hardeep Naghra

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

September 2015

For Masi and Massar

Declaration

Unless otherwise acknowledged, the work presented in this thesis is my own. No part has been submitted for another degree at the University of Nottingham or any other institute of learning.

Hardeep Naghra

September 2015

Abstract

The reduction in cost and increase in throughput of whole genome sequencing (WGS) technologies, and the advent of benchtop WGS instruments such as the Illumina MiSeq, means that WGS is no longer restricted to large genome centres and consortia. The number of microbial genomes in public repositories is ever increasing due to the availability of WGS technologies to research groups, with individual genera having their own dedicated genome databases. *Pseudomonas aeruginosa* is an opportunistic pathogen and a major cause of healthcare associated infection in immunocompromised individuals and cystic fibrosis (CF) patients. The first complete genome sequence of a *P. aeruginosa* strain was that of the most commonly studied laboratory strain *P. aeruginosa* PAO1, sequenced in 2000. It was found in a study, published in 2010, that there were differences between the chromosomal sequences of two isolates of *P. aeruginosa* PAO1 originating from two different public strain collections, as well as differences compared to the reference sequence. It was therefore proposed that *P. aeruginosa* PAO1 exists as variable "sublines" in strain collections, whose differences included a 2.2 Mb chromosomal inversion and a prophage insertion compared to the reference sequence, as well as single nucleotide polymorphisms (SNPs) and short insertions and deletions (INDELs) which are unique to individual sublines. Since the current genomic reference sequence is based on one of these variable sublines, this study aims to deduce the sequence of the original *P. aeruginosa* Strain 1 (PAO) from which all the PAO1 sublines are

derived, using sequence information from several laboratory sublines and from *P. aeruginosa* strains PAO2 and PAO3, which were directly derived from the PAO progenitor strain. This could be used as a more universal reference which is representative of a range of PAO1 sublines, to which they are all more closely related.

Although most genetic studies of *P. aeruginosa* are carried out in PAO1, this strain is in no way the archetype for this diverse species. As whole genomes of other *P. aeruginosa* strains became available, comparative genomics revealed that PAO1 shared only 80% of its genome with other strains. *P. aeruginosa* is a highly ubiquitous organism which is able to adapt to a variety of environmental niches, as well as to human and animal hosts. This is thought to be related to the large genome size and genetic complexity of the organism, with 10% of its genes devoted to regulatory functions. Mutations in quorum sensing (QS) genes have been reported in multiple studies of host adapted *P. aeruginosa* isolates. QS is the mechanism by which a bacterium adapts from the lifestyle of an individual cell to a multicellular community via the regulation of specific target genes, and has been shown to regulate key virulence factors. A study of 49 clinical isolates, collected from various wound sites from distinct inpatients and outpatients at the Queen's Medical Centre in Nottingham within a few days of one another, revealed that this set of isolates displayed a diverse range of QS phenotypes. Two of these isolates were identified as producing high levels of *N*-acylhomoserine lactone (AHL) QS signal molecules (QSSMs) and low levels of alkyl quinolone (AQ) QSSMs, which had not previously been found in clinical isolates of *P. aeruginosa*, at the time of the study. WGS of these 49 clinical isolates was carried out in this study, where the aim was to; (1) determine the genomic diversity of these isolates, and relate this to previously sequenced *P. aeruginosa* strains, and (2) analyse the key QS genotypes of these isolates, and attempt to relate these to the phenotypes

observed, in particular for the two isolates which were AHL-proficient and AQ-deficient.

Acknowledgements

I'd like to begin by thanking my supervisors Richard James, Stephan Heeb, Mathew Diggle and David Turner, as well as Paul Williams, for providing guidance, useful suggestions and encouraging me to go to conferences. This work wouldn't have been possible without the James Tudor Foundation, whose funding also allowed me to go to meetings, conferences and external research visits.

Although I was inwardly reluctant to return to the "wet lab" after two years away from that environment, guidance from Stephan Heeb made this a much smoother transition. I'd like to thank Manuel and Matt for imparting their wisdom, Pantelitsa for being an absolute pleasure to supervise, and all other members of the Pseudomonas lab for helping me out when I was looking a bit lost.

I'd like to thank Chan Kok Gan at the University of Malaya, Kuala Lumpur, for hosting me on my short visit to his lab. I am grateful to Huey Jia and Corrinne for their help and patience, and would like to extend a special thank you to Robson, Yan Lue and Tek Min for taking us to so many awesome places to eat and generally ferrying us around. On the same strand, I'd like to thank Roger, Irena, Luca, Brian and Jerome at Université Laval, Quebec for acting as very welcoming hosts, and for teaching me a lot about handling large datasets.

I would like to thank all friends, old and new, who have made my time in Nottingham not only memorable, but also life changing. Thank you

for all the coffees, lunches, Johnson Arms Fridays, parties, dinners and conversations (silly and serious). I would also like to take this opportunity to thank all those involved in getting me to this stage; from my secondary school Biology and Chemistry teachers who first captured my interest, to my Form Tutor for encouraging me to aim high when applying for University, to friends and tutors from my Bachelors and Masters courses, and finally former colleagues in Team 133 at the Sanger Institute. I wouldn't be where I am now without your help and support!

I'd like to end by thanking my family for their constant stream of unconditional love and support, as well as my newly acquired in-laws for theirs too. And lastly, a massive show of appreciation to Bas for keeping me sane and well fed for the last four years, even when I was a complete bear; a keeper you are :)

Contents

List of Figures

1

2

List of Tables

Abbreviations

Table 1: *Abbreviations*

Abbreviation	Full term
°C	Degrees Celsius
% (w/v)	Percentage weight per volume
3-oxo-C_{12}-HSL	*N*-3-oxo-dodecanoyl-L-homoserine lactone
AHL	*N*-acylhomoserine lactone
AIDS	Acquired immune deficiency syndrome
AQ	2-alkyl-4(1*H*)-quinolones
Amp	Ampicillin
bp	base pair
C_4-HSL	*N*-butanoyl-L-homoserine lactone
CF	Cystic Fibrosis
CLED	Cystein lactose electrolyte deficient
Cm	Chloramphenicol
DNA	Deoxyribonucleic acid
EPS	Extracellular polysaccharide
HGT	Horizontal gene transfer
HHQ	2-heptyl-4-hydroxyquinoline
HQNO	2-heptyl-4-hydroxyquinoline *N*-oxide
ICE	Integrative conjugative element
IGP	Indole-3-glycerol phosphate
INDEL	Insertion/deletion
IS	Insertion sequence
kb	Kilobase
LES	Liverpool epidemic strain
LPS	Lipopolysaccharide
OMF	Outer membrane factor
ORF	Open reading frame
Mb	Megabase
MFP	Membrane fusion protein
MGE	Mobile genetic element
mRNA	Messenger RNA
mvfR	Also known as *pqsR*
NGS	Next generation sequencing
O-Ag	O antigen
PAO1-ATCC	PAO1-ATCC 15692
PAO1-DSM	PAO1-DSMZ 1707

Continued on next page

Table 1 – *Continued from previous page*

Abbreviation	Full term
PIA	Pseudomonas Isolation Agar
PQS	*Pseudomonas* quinolone signal (2-heptyl-3-hydroxy-4(1H)-quinolone)
pqsR	Also known as *mvfR*
QS	Quorum sensing
QSSM	Quorum sensing signal molecule
RGP	Region of genome plasticity
RND	Resistance nodulation division
RNA	Ribonucleic acid
rrn	Ribosomal operon
rRNA	Ribosomal RNA
SNP	Single nucleotide polymorphism
Tet	Tetracycline
WGS	Whole genome sequencing

Chapter 1

General Introduction

1.1 Whole genome sequencing

1.1.1 Ten years of whole genome sequencing

Since the landmark sequencing of the whole genome of *Haemophilus influenza* in 1995 (Fleischmann *et al.*, 1995) and the completion of the Human Genome Project in 2004 (International Human Genome Sequencing Consortium, 2004), both of which were completed using Sanger sequencing (Sanger *et al.*, 1977) and required a large amount of time and resources, whole genome sequencing (WGS) has developed significantly. The development and commercialisation of the so-called next generation sequencing (NGS) technologies has meant that it is possible to sequence whole genomes faster and at a reduced cost when compared with traditional Sanger sequencing technology. The main improvements of these technologies compared to Sanger sequencing are the preparation of NGS libraries in a cell free system rather than relying on bacterial cloning of DNA fragments, the increase in the numbers of sequencing reactions run in parallel from hundreds to millions, and the direct detection of sequencing output using the optical detection of incorporated nucleotides, with base calling occurring in parallel within the sequencing platforms (Grada and Weinbrecht, 2013).

Between 2005 and 2010 there were three major NGS technologies which dominated the market; the Roche 454 pyrosequencing method, the Illumina sequencing platform and the Sequencing by Oligo Ligation Detection (SOLiD) technology by Applied Biosystems (now Life Technologies). Although the Illumina and the SOLiD sequencers generated a larger number of reads than the 454 sequencers, the reads produced were very short in comparison, namely only 35 bp in length (van Dijk *et al.*, 2014). The Illumina and Roche 454 technologies underwent significant improvements in read length and throughput over time, and eventually the focus turned to more portable NGS instruments.

Between 2010 and 2011, three benchtop high throughput sequencing instruments were released; the 454 GS Junior form Roche, the Personal Genome Machine (PGM) from Ion Torrent and the MiSeq from Illumina. The 454 GS Junior is a lower throughput of the previous high throughput 454 GS FLX machine, which still uses the pyrosequencing approach (Ronaghi *et al.*, 1996) of its predecessor but with lower set up and running costs. The Ion Torrent PGM differs in its use of semiconductor technology to detect hydrogen ions released during base incorporation by DNA polymerase, rather than the optical detection of fluorescence upon base incorporation like previous instruments (Rothberg *et al.*, 2011). The Illumina MiSeq is based on the existing Illumina sequencing by synthesis chemistry (Bentley *et al.*, 2008), however it uses a smaller flow cell than its larger predecessor (the Illumina HiSeq), which means it has faster microfluidics and a reduced imaging time.

The PacBio RS, released commercially in 2011, uses a single molecule real time (SMRT) sequencing approach, where a single DNA template is used and fluorescently labelled bases are incorporated by DNA polymerase as the template is read (Eid *et al.*, 2009). This technology is also able to detect modified nucleotides and so can be used for methylation studies as well as WGS (Flusberg *et al.*, 2010). The current instrument, the PacBio RS II, is able to generate significantly longer reads than previous WGS instruments. The long reads generated by the PacBio instrument have been reported to have a high error rate, however correcting errors using short reads from either the Illumina MiSeq or the PacBio itself has been shown to yield high quality bacterial genomes (Bashir *et al.*, 2012; Liao *et al.*, 2015).

The first nanopore sequencer, the MinION, was released to early access customers in 2014 by Oxford Nanopore technologies. The device contains membrane-embedded pores, which are placed over an electrical detection

grid. As single DNA molecules pass through the pores, alterations in ionic current are measured. These changes in ionic current differ per nucleotide, and so the sequence of DNA molecules can be inferred (Clarke *et al.*, 2009). The idea of nanopores was first proposed by Kasianowicz *et al.* (1996), however numerous challenges needed to be overcome before the technology could be used to reliably read nucleotide sequences. The MinION is the smallest WGS instrument to date; quoted as being "roughly the size and shape of a packet of chewing gum", and is powered via a USB 3.0 connection (Hayden, 2014). Initial sequencing runs on the MinION yielded mean read lengths of ~6 kb (Quick *et al.*, 2014).

WGS data in this thesis were generated using the Illumina MiSeq, and so the Illumina sequencing technology and associated biases will be discussed in further detail.

1.1.2 Illumina sequencing

Illumina WGS platforms use a similar method to Sanger chain termination process, with the primary difference being the use of fluorescently labelled terminator nucleotides which allow the reversal of the chain termination process. Template DNA is fragmented, which is followed by the ligation of specific universal adapters in the library preparation step prior to sequencing. These DNA fragments are hybridised to a glass surface, called a flow cell, on which amplification and sequencing take place. Bridge amplification, which consists of a surface PCR using universal primers, generates localised clusters of identical DNA fragments in the flow cell. The first stage of sequencing is the addition of a primer which hybridises to the ligated adapters on each fragment. It is after this that the sequencing chemistry takes place. Fluorescently labelled nucleotide terminators complementary to the base of the DNA fragment are incorporated using DNA polymerase, with a single nucleotide being incorporated to all fragments on the flow cell. The in-

corporated base in each cluster is identified by imaging, which is followed by the cleavage of the dye and the terminating groups of all fragments to prepare for the next cycle of nucleotide incorporation and imaging (Bentley *et al.*, 2008). These stages are repeated over a number of cycles, where the exact number of cycles used determines the read length. The images of the individual clusters, which correspond to a single read, at each cycle are converted into base calls using the Illumina base calling algorithm.

Illumina sequencing chemistry is prone to errors, such as phasing, where DNA polymerase fails to incorporate a base in a given fragment in a given cycle and pre phasing, where multiple bases are synthesised in a single cycle. As well as this, the signal intensity of clusters is known to decay as cycles progress, which leads to an increase in base calling errors towards the end of reads (Fuller *et al.*, 2009). The Illumina base calling algorithm is able to correct phasing and pre phasing, as well as assigning a base quality score, in the form of a Phred score (Ewing and Green, 1998), to each base call in a read (Ledergerber and Dessimoz, 2011). The Phred score is therefore an indicator of base calling accuracy at a given base in a read (Table 1.1).

Table 1.1: *The relationship between Illumina base quality score and base call accuracy. Table produced using data from Illumina; http://www.illumina.com.*

Quality score	Probability of incorrect base call	Inferred base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%

1.1.3 Optical mapping

Optical restriction mapping was first reported by Schwartz *et al.* (1993) and was commercialised by OpGen, whose Argus optical mapping system became adopted by research institutes in 2011. An overview of the technique is shown in Figure 1.1. Upon extraction, DNA is fragmented and single molecules are flowed through microfludic channels on a positively

charged glass surface, where they are stretched and immobilised. DNA is then digested with restriction enzymes, and since it is immobilised, the fragment order is maintained. The digested fragments are then stained with a fluorescent dye, and then imaged. The length of each digested fragment is proportional to its fluorescence intensity, which means that single molecules can be overlapped to generate a chromosomal optical map (Mendelowitz and Pop, 2014). Optical mapping is often combined with short read WGS to generate more contiguous genome assemblies (Latreille *et al.*, 2007; Riley *et al.*, 2011; Tsai *et al.*, 2013; Reid *et al.*, 2014; Howe and Wood, 2015) and has been used to detect large scale chromosomal rearrangements in bacteria (Raeside *et al.*, 2014; Sabirova *et al.*, 2014).

1.2 Bacterial Genomics

Traditionally, whole genome shotgun of bacterial genomes was carried out to generate reference sequences for species which were commonly studied in the laboratory (Blattner *et al.*, 1997; Cole *et al.*, 1998; Kunst *et al.*, 1997; Stover *et al.*, 2000; Parkhill *et al.*, 2001). With the advent of high throughput sequencing, it became more common to sequence multiple strains of the same bacterial species and compare their genomes (Kos *et al.*, 2012). As throughput of WGS technologies increased, thereby reducing running costs, WGS was used to track hospital (Koser *et al.*, 2012; Snitkin *et al.*, 2012), community (Harris *et al.*, 2010; Gardy *et al.*, 2011) and global (Mutreja *et al.*, 2011) outbreaks of bacterial pathogens, as well as to resequence the genomes of laboratory strains and mutants (Klockgether *et al.*, 2010; McEvoy *et al.*, 2013).

WGS was initially restricted to microbes which could be grown in pure culture, however the development of culture-independent methods, as well as the steady increase in throughput of WGS technologies, has led to the possibility of sequencing microbial communities directly from environmen-

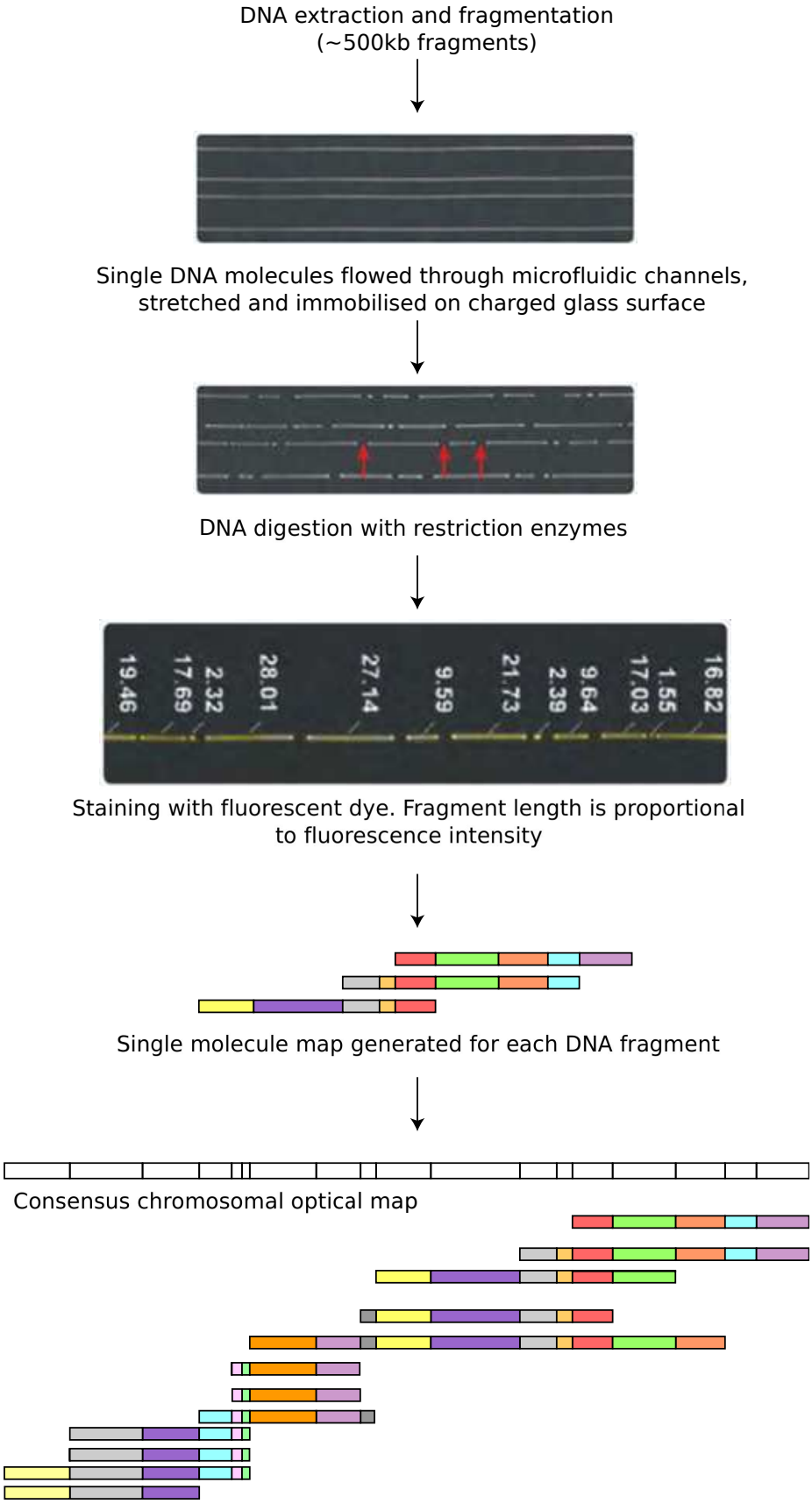

Figure 1.1: Overview of optical restriction mapping. *Argus MapCard images from Opgen Ltd (http://www.opgen.com)*

tal samples; known as metagenomics (Tringe and Rubin, 2005). It is not yet possible to generate complete genomes of all individual strains in a microbial community, with metagenome assemblies usually consisting of multiple contiguous sequences, or "contigs". There are, however, multiple methods to "bin" these contigs into species level groups to generate draft metagenome sequences (Wang *et al.*, 2012; Albertsen *et al.*, 2013; Alneberg *et al.*, 2014; Nielsen *et al.*, 2014). Metagenomics has been applied to the study of the microbial communities of wounds (Scales and Huffnagle, 2013), the human gut, both healthy and during disease, (Qin *et al.*, 2010, 2014) and permafrost (Mackelprang *et al.*, 2011), among others. However, a recent study showed that contaminants from laboratory reagents can skew results of metagenomics studies of low biomass samples, and highlighted the importance of sequencing negative controls (Salter *et al.*, 2014).

1.2.1 Bacterial comparative genomics

There are currently two commonly used strategies employed to analyse whole genome sequencing datasets generated by the "next generation" methods. The first of these, read alignment, relies on the availability of a complete reference sequence closely related to the bacterium which has been sequenced. There is a plethora of read alignment software to choose from for this type of analysis (Li and Homer, 2010). The choice of read aligner used in a study is dependent on multiple factors, including the nature of the genome being studied (for example, if it is repeat-rich), the computational resources available, the WGS technology used and the research questions which need to be answered. For the most part, read alignment is used to identify single nucleotide polymorphisms (SNPs) and short insertions/deletions (INDELs) in the isolate whose genome has been sequenced, compared to the reference sequence.

The second common strategy for analysis of WGS data is *de novo*

genome assembly, which is often employed where no reference sequence is available. *De novo* genome assembly involves the merging and ordering of reads, which are sampled from a set of larger sequences, namely the genomic DNA of the organism being sequenced, in order to reconstruct the genomic sequence. Prior to long read technologies such as PacBio SMRT sequencing, short read technologies were used to generate *de novo* assemblies. Usually, a genome assembly consists of a set of contigs which can be ordered and oriented into "scaffolds", where contigs are known to be joined by distance information but the sequence of the contig cannot be extended, for example due to the presence of large repetitive regions such as ribosomal operons. As with read alignment, there is an extensive set of genome assembly software available (Earl *et al.*, 2011). The PacBio RS II has been reported to generate reads of up to 10-15 kb in length (Biosciences, 2014), and improvements to and methods to correct the error rate associated with previous versions of the technology mean that complete bacterial genomes can be sequenced using this long read technology (Liao *et al.*, 2015). However, the PacBio RS II has a high cost per base compared to platforms such as the Illumina MiSeq (Quail *et al.*, 2012), and so is not routinely used to sequence bacterial genomes. Once a bacterial genome has been assembled, annotation of genes, rRNAs and tRNAs, as well as functional domains and other genetic features is carried out for use in downstream genomic analyses.

1.2.2 WGS in diagnostic microbiology

In theory, WGS of microbial pathogens in hospitals would enable physicians to treat patients with the appropriate antibiotics given the repertoire of resistance genes carried by the causative agent of an infection. The Oxford Nanopore MinION, the smallest sequencer to date, has been used to detect antibiotic resistance genes in isolates of meticillin resistant *Staphylococcus aureus* (MRSA) (Judge *et al.*, 2015). However, although

antibiotic resistance genes were identified in this study, the data generated was not accurate enough, due to the high error rate of MinION reads, to detect point mutations associated with antibiotic resistance in genes in the chromosome of the organism. In order for WGS technologies to be adopted by the NHS for routine use, they must be affordable and generate accurate results rapidly; in addition to the requirement for automated sample preparation (Peacock, 2014). Although there is currently no WGS instrument which fits these requirements, the MinION is thought to display the most potential, given ongoing improvements in accuracy and yield (Judge *et al.*, 2015).

1.2.3 WGS of laboratory strains

Sabirova *et al.* (2014) carried out WGS and optical mapping study of twelve *S. aureus* laboratory strains whose whole genomes had previously been sequenced, and are commonly used as reference sequences. It was found that one of the lab strains, *S. aureus* FPR3757, contained a 13.3 kb genomic deletion compared to the reference sequence for this strain, which corresponded to an integrative conjugative element (ICE). Optical mapping of clones of *Escherichia coli* from a long term evolution experiment revealed a total of 110 rearrangement events had occurred after 25 years, which equated to 40,000 generations (Raeside *et al.*, 2014). Most rearrangements were found to involve recombination between insertion sequence (IS) elements and repeated sequences such as rRNA-encoding operons. Large deletions were the most common type of rearrangement in this study with 30% of involving the deletion of prophage DNA in the chromosome. Just under half of the chromosomal inversions in the *E. coli* populations were found to include more than a quarter of the chromosome. It has previously been found that the commonly used laboratory strain of *E. coli*, K12, has accumulated mutations over the years (Hobman *et al.*, 2007), and this observation has also been

found to extend to *Pseudomonas aeruginosa*. WGS of two *P. aeruginosa* PAO1 laboratory strains, available from public culture collections, not only revealed the presence of SNPs, INDELs and a 2.2 Mb inversion compared to the reference sequence, based on a PAO1 laboratory isolate from the University of Washington (PAO1-UW), but also differences between the two laboratory strains whose genomes were sequenced (Klockgether *et al.*, 2010).

1.3 *Pseudomonas aeruginosa*

P. aeruginosa is a Gram-negative rod shaped bacterium, which was first isolated in pure culture in 1882 by Carle Gessard and is distinguished by its characteristic blue/green colouration in culture as a result of the production of the phenazine compound pyocyanin (Gessard, 1984). *P. aeruginosa* is a ubiquitous organism which has been isolated from a diverse range of environments, such as soil, water, humans, plants and animals (Green *et al.*, 1974; Kiss *et al.*, 1985; Winstanley *et al.*, 2009; Wise, 2012; Haenni *et al.*, 2015). The first complete genome sequence of *P. aeruginosa* was that of PAO1, which was found to have a large genome size and genetic complexity (Stover *et al.*, 2000). These features are thought to reflect evolutionary adaptations which allow the bacterium to survive in a wide range of ecological niches.

P. aeruginosa is a common cause of bacteraemia in hospitals; in 2013 *Pseudomonas* was the eighth most frequently reported genus in both monomicrobial and polymicrobial bacteraemic episodes (PHE, 2014). In hospitals, *P. aeruginosa* has been isolated from water sources such as sinks and taps, catheters, respirators and ventilators (Jefferies *et al.*, 2012; Hooton *et al.*, 2010; Cobben *et al.*, 1996; Crouch Brewer *et al.*, 1996). *P. aeruginosa* also occurs in community reservoirs such as swimming pools, bathroom water, hot tubs and contact lens solution (Barben *et al.*, 2005; Crnich *et al.*,

2003; Fleiszig *et al.*, 1992). *P. aeruginosa* is not usually part of the normal microbial flora in humans; in healthy individuals it colonises the oropharynx of up to 6% of individuals and the faeces of 3-24% of individuals. However, in hospitals up to 50% of patients are colonised with the organism, and so it appears that reservoirs in the hospital have the potential to act as sources of *P. aeruginosa* infection in susceptible individuals. This has been seen in the recent outbreak of *P. aeruginosa* in a neonatal ward in Belfast, where the source of infection was determined to be contaminated wash basin taps (RQIA, 2012).

P. aeruginosa is an opportunistic pathogen which primarily causes infection in immunocompromised patients, such as patients undergoing chemotherapy, acquired immune deficiency syndrome (AIDS) patients and patients with Cystic Fibrosis (CF), as well as patients with burn wounds (Davies, 2002; Shepp *et al.*, 1994; Ohmagari *et al.*, 2005; Lipovy *et al.*, 2010).
P. aeruginosa is intrinsically resistant to antibiotics, and this, in combination with the immunocompromised status of patients, makes these infections particularly difficult to treat.

1.4 *P. aeruginosa* virulence factors

P. aeruginosa produces a number of extracellular virulence factors after colonisation, which can lead to tissue damage, invasion of the bloodstream and propagation of the bacteria (Van Delden and Iglewski, 1998). These include pyocyanin, rhamnolipids, exotoxins secreted by the Type III secretion system, elastases, proteases, siderophores and hydrogen cyanide. *P. aeruginosa* also has a number of cell-associated virulence factors, such as lipopolysaccharides (LPS), type IV pili, flagella and adhesins. *P. aeruginosa* is able to attach to solid surfaces and form organised biofilm communities, which is important for establishing chronic infections, and for persistence of the infection in otherwise hostile environments (Costerton *et al.*, 1999).

Further information relating to some virulence factors of *P. aeruginosa* and their properties are given in Table 1.2, and virulence factors which are relevant to the work in this thesis are described in further detail in the remainder of this section.

Table 1.2: *P. aeruginosa virulence factors and their properties*

Virulence factor	**Properties**	**References**
Alginate	Thought to increase resistance of bacterial cells to host immune system and antibiotics.	(Pedersen *et al.*, 1992)
Alkaline protease	Encoded by *aprA*. Prevents the activation of complement, part of the innate immune response of the host.	(Laarman *et al.*, 2012)
Exotoxin A	Inactivates elongation factor 2 in host cells, an essential factor for protein synthesis. Secreted by Type III secretion system.	(Pavlovskis *et al.*, 1978)
Exotoxin U	Involved in the initiation of a host inflammatory response by activating production of the chemokine interleukin-8, resulting in increased epithelial permeability of thecryz1984 bacterium, facilitating its spread. Secreted by Type III secretion system.	(Cuzick *et al.*, 2006)
Exoenzyme S	Enables bacteria to evade host immune system by acting as an antiphagocytic factor. Secreted by Type III secretion system.	(Frithz-Lindsten *et al.*, 1997)
Exoenzyme T	Inactivates host cell proteins involved in maintaining the actin cytoskeleton. Secreted by Type III secretion system.	(Sundin *et al.*, 2004)
Exoenzyme Y	Involved in the breakdown of the microtubule of host cells, leading to tissue damage. Also thought to chronically impair tissue integrity following damage. Secreted by Type III secretion system.	(Stevens *et al.*, 2014)
HasAp	Protein with high affinity binding for heme (hemophore). Allows for survival of *P. aeruginosa* in low iron conditions, i.e. at the early stages of infection.	(Alontaga *et al.*, 2009)
Hydrogen cyanide (HCN)	The molecule which gives *P. aeruginosa* cultures their distinct almond-like smell, HCN inhibits aerobic respiration in host cells.	(Gallagher and Manoil, 2001)
LasB elastase	Degrades and inactivates host tissues and immune system components; including collagen, fibrin, elastin and immunoglobulin.	(Rust *et al.*, 1996)
LasA protease	Endopeptidase which can lyse *Staphylococcus aureus* cells by cleaving peptide bonds in the peptidoglycan of the cells. Degrades elastin by cleaving peptide bonds between glycine residues.	(Kessler *et al.*, 1993)
Lipopolysaccharide (LPS)	Maintains cell wall structure of bacterial cell, as well an mediating interactions with the neighbouring environment via the LPS O antigen.	(Cryz *et al.*, 1984)
Phosopholipase C	Hemolysin, damages host tissue	(Berka and Vasil, 1982)
Protease IV	Serine protease shown to degrade fibrinogen, plasmin, plamsinogen and capable of degrading corneal epithelium.	(Engel *et al.*, 1998a,b)

Continued on next page

Table 1.2 – *Continued from previous page*

Virulence factor	Properties	References
Pyocyanin	Siderophore (iron chelator), which induces the formation of reactive oxygen species (ROS) in the host	(Mahajan-Miklos *et al.*, 1999)
Pyoverdine	Siderophore (iron chelator), which induces the formation of ROS in the host, as well as regulating the production of exotoxin A and proteases in the bacterial cell.	(Becerra *et al.*, 2001; Lamont *et al.*, 2002)
Rhamnolipids	A biosurfactant which is required for bacterial cells to infiltrate the respiratory epithelium during initial stages of an infection.	(Zulianello *et al.*, 2006)
Type IV pili	Surface adhesin, involved in attachment, biofilm formation and twitching motility. In recent studies have been shown to promote the induction of virulence factors when cells encounter surfaces, such as host tissues.	(Persat *et al.*, 2015)

1.4.1 Lipopolysaccharide

Lipopolysaccharide (LPS) molecules are found in the outer membrane of *P. aeruginosa*, where they not only maintain the structure of the outer membrane, but also mediate interactions with the neighbouring environment. The LPS molecule can be divided into three parts; the hydrophobic lipid A region, which takes the place of phospholipids in the outer cell membrane, the central oligosaccharide region, and the O antigen (O-Ag), a repeating polysaccharide which is exposed on the outer surface of the bacterial cell. LPS can occur in smooth or rough forms; where the O-Ag is attached to the core polysaccharide, which is in turn attached to lipid A in the smooth form, with the rough form lacking the O-Ag of the smooth form (Pier, 2007). The O-Ag has been shown to be essential for virulence (Cryz *et al.*, 1984; Tang *et al.*, 1996; Augustin *et al.*, 2007), and is responsible for serogroup specificity in *P. aeruginosa* (Lam *et al.*, 2011). LPS induces an inflammatory response in host cells, enhancing host defence against bacterial infection. However, LPS can cause excessive stimulation of the immune system, which can lead to sepsis and, in some cases, death (Van Amersfoort *et al.*, 2003).

Most strains of *P. aeruginosa* produce two different forms of O-Ag, the

A band O-Ag, which is formed from a homopolymer of D-rhamnose, and the B band O-Ag, a heteropolymer with repeats units of between three and five individual sugars (Rocchetta *et al.*, 1999). It is the composition of the O-Ag units which affects the serotype of *P. aeruginosa* (Knirel *et al.*, 2006). In *P. aeruginosa* PAO1, which is serotype O5, the A band O-Ag is encoded by the gene products of PA5447 (*wbpZ*) to PA5454 (*rmd*) and the B band O-Ag is encoded by the gene products of PA3141 (*wbpM*) to PA3160 (*wzz*) (Rocchetta *et al.*, 1999).

1.5 Quorum Sensing

In nature, bacteria are able to form multicellecular communities, allowing them to exploit nutrient resources which are otherwise restricted to individual cells, and adapt to and modify their environment (Shapiro, 1998). The mechanism by which a bacterium adapts from the lifestyle of an individual cell to a multicellular community, via the regulation of specific target genes, is termed quorum sensing (QS) (Fuqua *et al.*, 1994). This involves the production of diffusible, low molecular weight QS signal molecules (QSSMs) by specific synthase proteins, which in turn activate specific target genes (Camara *et al.*, 2002).

QSSMs can diffuse into and out of bacterial cells. The regulation of target genes by QS is dependent on the density of the bacterial population; as the cell density increases so does the level of extracellular QSSMs. Once a target threshold has been reached the QSSMs bind to their cognate receptor protein, usually a transcriptional regulator. The QSSM-receptor complex is then able to regulate the transcription of QS controlled genes, which vary between bacterial species (Williams *et al.*, 2007). An example of a basic receptor/synthase QS system in Gram negative bacteria is outlined in Figure 1.2.

QS is thought to regulate over 300 genes in *P. aeruginosa*, which consti-

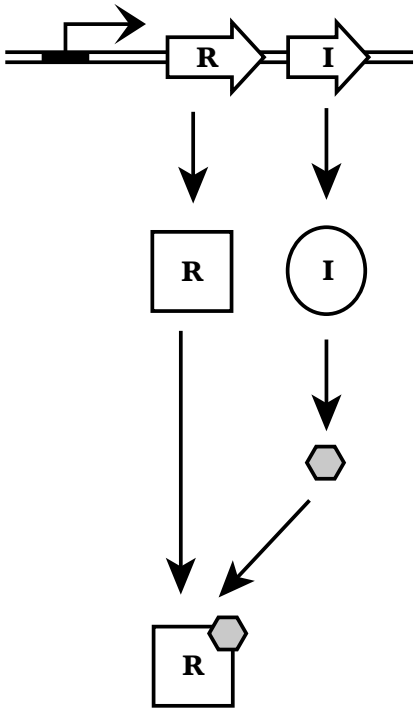

Figure 1.2: Example of a basic QS system in Gram negative bacteria. *The inducer synthase protein (I) synthesises an inducer molecule (hexagon), which binds to the receptor protein (R). This bound complex regulates the transcription of QS-controlled genes. This all occurs once bacterial density and inducer molecule have reached a certain threshold. Genes are represented as open arrows, the receptor protein by a square, the synthase by a circle and the signal molecule by a filled grey hexagon.*

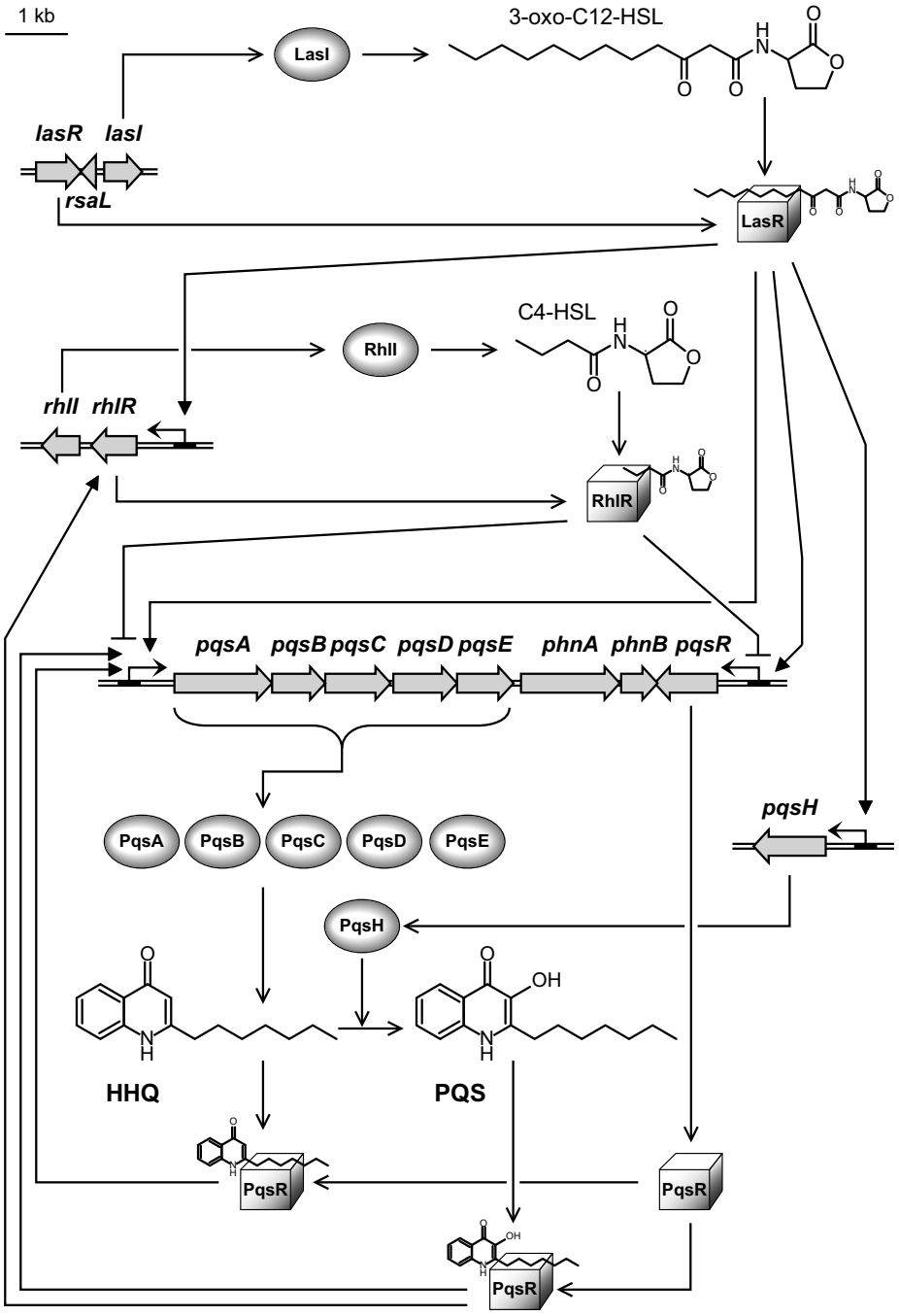

Figure 1.3: Overview of the major QS systems in P. aeruginosa. *3-oxo-C_{12-HSL} is produced via LasI, and binds to LasR. This receptor/signal complex positively regulates the transcription of pqsR, pqsABCDE, pqsH and rhlRI. C_4-HSL is produced via RhlI, and binds to RhlR, which negatively regulates the transcription of pqsR and pqsABCDE. The PqsABCDE proteins synthesise HHQ from anthranilate (produced by PhnAB). HHQ is converted to PQS by PqsH. Either HHQ or PQS can bind to PqsR and positively regulate the transcription of pqsABCDE and the rhl QS system. Enzymes are represented by circles and transcriptional regulators by cubes. Arrows represent positive regulation, and blunt lines negative regulation. Adapted from Figure 5 in (Heeb* et al.*, 2011), with permission from the lead author.*

tutes around 6% of the genome (Whiteley *et al.*, 1999; Schuster *et al.*, 2003; Wagner *et al.*, 2003), and has been shown to be a contributing factor to virulence in this species (Rumbaugh *et al.*, 1999; Hentzer *et al.*, 2003; Kohler *et al.*, 2010). The major QS systems of *P. aeruginosa* are the *las* and *rhl* QS systems, whose QSSMs are *N*-acylhomoserine lactones (AHLs), and the *pqs* QS system, whose QSSMs are alkyl quinolones (AQs). The *las* and *rhl* QS systems have been shown to activate transcription of virulence genes, including those which encode LasB elastase, LasA protease, protease IV, alkaline protease, exotoxin A, pyocyanin and siderophores(Gambello and Iglewski, 1991; Toder *et al.*, 1991; Gambello *et al.*, 1993; Schuster *et al.*, 2003; Schuster and Greenberg, 2007; Conibear *et al.*, 2012). These three systems are hierarchical, where the *las* system positively regulates both the *rhl* and *pqs* systems (Pesci *et al.*, 1997; Wade *et al.*, 2005). In addition, the *rhl* system negatively regulates the *pqs* system (Wade *et al.*, 2005) by inducing the expression of an alternative *pqsA* mRNA isoform with a reduced translation efficiency (Brouwer *et al.*, 2014) and the *pqs* system positively regulates the *rhl* system (McKnight *et al.*, 2000). An outline of these QS systems of *P. aeruginosa* is shown in Figure 1.3. However despite this, a functional *pqs* QS system has been shown in the absence of a functional *las* QS system. In one study a *lasR* mutant was shown to produce PQS in the late stationary phase (Diggle *et al.*, 2003), and other studies the expression of *pqsR* (also known as *mvfR*) and *pqsH* (Xiao *et al.*, 2006) and the *rhl* QS system (Dekimpe and Deziel, 2009) was shown to be delayed but not abolished during the growth of a *lasR* mutant. This indicates the presence of additional positive regulator(s) of the *pqs* QS system; which a couple of recent studies have proposed (Liang *et al.*, 2009; Tipton *et al.*, 2015).

The signal molecule *N*-3-oxo-dodecanoyl-L-homoserine lactone (3-oxo-C_{12}-HSL) is produced by the QSSM synthase LasI, which is encoded by the *lasI* gene (Gould *et al.*, 2004). 3-oxo-C_{12}-HSL binds to the transcriptional

regulator LasR, which is encoded by the *lasR* gene (Gambello and Iglewski, 1991; Bottomley *et al.*, 2007). LasR has been found to be involved in the production of virulence factors such as elastase and pyoverdin (Passador *et al.*, 1993; Stintzi *et al.*, 1998). The LasR/3-oxo-C_{12}-HSL complex binds to the promoters of *rhlIR*, *pqsABCDE*, *pqsR* and *pqsH*, where it positively regulates the transcription of these genes (Latifi *et al.*, 1996; Wade *et al.*, 2005), as well as upregulating its own production by binding to the *lasI* promoter (Seed *et al.*, 1995).

The AHL synthase RhlI, encoded by the *rhlI* gene, produces the QSSM *N*-butanoyl-L-homoserine lactone (C_4-HSL) (Pearson *et al.*, 1995; Winson *et al.*, 1995; Raychaudhuri *et al.*, 2005). The C_4-HSL molecule binds to the N-terminal domain of the transcriptional regulator RhlR, encoded by the *rhlR* gene (Lamb *et al.*, 2003). RhlR has been found to increase the production of rhamnolipid biosurfactants in *P. aeruginosa* (Ochsner and Reiser, 1995). The RhlR/C_4-HSL complex binds to the promoters of *pqsABCDE* and *pqsR*, where it represses the transcription of these genes (Wade *et al.*, 2005), as well as upregulating its own production by binding to the *rhlI* promoter, in a similar fashion to the *las* QS system.

The AQ QSSM of the *pqs* QS system of *P. aeruginosa* is 2-heptyl-4-quinolone (HHQ), which was widely accepted to be produced by the products of genes *pqsABCD* of the *pqsABCDE* operon (Gallagher *et al.*, 2002), and subsequently converted to 2-heptyl-3-hydroxy-4(1*H*)-quinolone (the *Pseudomonas* quinolone signal; PQS) by PqsH (Schertzer *et al.*, 2010). The role of *pqsE* was unknown until a recent study revealed its role as a thioesterase in HHQ synthesis, and that PqsE deletion mutants from previous studies were still able to synthesise HHQ as the thioesterase role of PqsE is taken over to some degree by the broad specificity thioesterase TesB, among others (Drees and Fetzner, 2015). Both PQS and HHQ are able to bind to PqsR to form a signal molecule/regulator complex (Xiao

et al., 2006), which acts to positively regulate the expression of *pqsA* in an autoregulatory feedback loop (Wade *et al.*, 2005), as well as *rhlR* (McKnight *et al.*, 2000). The *pqs* QS system regulates the production of virulence factors such as pyocyanin, elastase and rhamnolipids (Pesci *et al.*, 1999).

A number of functions outside of QS have been described for the molecule PQS in *P. aeruginosa*. PQS has been shown to have the ability to chelate iron (Bredenbruch *et al.*, 2006), and is thought to act as an iron trap at the cell surface in *P. aeruginosa*, potentially facilitating siderophore-mediated iron transport (Diggle *et al.*, 2007). The role of PQS in the promotion of autolysis has also been described (D'Argenio *et al.*, 2002), with PQS thought to act as an endogenous stress factor in *P. aeruginosa*, increasing the susceptibility of the cell to oxidative stress and antibiotics (Haussler and Becker, 2008). Autolysis of bacterial cells leads to the release of intracellular contents such as nutrients, which can be used by surviving cells. PQS has been shown to interact with LPS in the outer membrane of *P. aeruginosa* to induce the formation of membrane vesicles (Mashburn-Warren *et al.*, 2008), which transport hydrophobic molecules, including PQS itself (Mashburn and Whiteley, 2005), between cells.

These three QS systems of *P. aeruginosa* have been shown to be influenced by a number of external regulators, a few of which are summarised in Table 1.3. Although the QS system of *P. aeruginosa* is the most well characterised, studies in other *Pseudomonas* species indicate that the complexity of QS regulation in *P. aeruginosa* extends to other members of the genus (Zhang and Pierson, 2001; Steidle *et al.*, 2002; Yan *et al.*, 2009).

Although traditionally bacterial infections are treated with antibiotics that inhibit bacterial growth, bacteria are able to develop resistance to antimicrobials relatively quickly, which is currently a major healthcare threat (McKenna, 2013). Antimicrobial resistance is a problem in CF patients, who are subject to chronic respiratory infections, often led by *P. aeruginosa*,

Regulator	Mode of regulation	Function	References
GacA	Overexpression of GacA positively regulates *las* and *rhl* QS systems	Part of GacA/S two component system which influences production of pyocyanin, cyanide and lipase	(Reimmann *et al.*, 1997)
RelA	Positively regulates *las* and *rhl* QS	Has an effect on elastase and pyocyanin production, although primary role is in stress response to amino acid starvation	(van Delden *et al.*, 2001) (Erickson *et al.*, 2004)
RsaL	Binds to *lasI* promoter, inhibiting transcription	Reduces pyocyanin and hydrogen cyanide production	(Rampioni *et al.*, 2006) (Rampioni *et al.*, 2007)
RsmA	Negative regulator of *las* and *rhl* QS	Small RNA which affects protease, elastase and pyocyanin levels	(Pessi *et al.*, 2001)
QscR	Homologue of *lasR*, although doesn't have a cognate synthase ("orphan"). Initially thought to negatively regulate *lasI* expression, however later found to have its own regulon	Affects pyocyanin production	(Chugani *et al.*, 2001) (Lequette *et al.*, 2006)
Vfr	Positively regulates *lasR* expression	Influences protease, exotoxin A production and motility	(West *et al.*, 1994), (Albus *et al.*, 1997), (Beatson *et al.*, 2002)
VsqR	Truncated orphan LasR homologue. Positively regulates *lasI* expression	Affects protease and pyocyanin production	(Juhas *et al.*, 2004) (Juhas *et al.*, 2005)

Table 1.3: *Additional regulators of P. aeruginosa QS*

which are not cleared despite intensive treatment with antimicrobials (Conway *et al.*, 2003). In the CF lung, the persistence of chronic *P. aeruginosa* infection has been shown to be due to alginate producing strains (termed mucoid) which form biofilms in the lung (Ohman and Chakrabarty, 1982; Singh *et al.*, 2000). Biofilms are also a contributing factor to severity of wound infections (Watters *et al.*, 2013). A biofilm is an organised community of bacterial cells, which excrete an extracellular polysaccharide (EPS) matrix, allowing them to attach to solid surfaces. The EPS also forms a closed environment for bacterial cells, by keeping in nutrients and protecting cells from changes in the environment (Sutherland, 2001) as well as antibiotics (Stewart and Costerton, 2001).

A number of studies have shown that QS, and virulence factors under the control of QS, play an important role in the formation of biofilms. The biofilm matrix genes of *P. aeruginosa* have been shown to be under the control of QS (Sakuragi and Kolter, 2007), and virulence factors such as rhamnolipids, flagella and type IV pili have been shown to have an impact on biofilms and their formation (O'Toole and Kolter, 1998; Davey *et al.*, 2003). A number of studies have focused on the search for inhibitors of QS and biofilm formation, as a potential novel therapeutic for *P. aeruginosa* infection, in an era where antibiotic resistance is a major cause for concern (Hentzer *et al.*, 2002; Brackman *et al.*, 2011; O'Loughlin *et al.*, 2013; Christiaen *et al.*, 2014).

1.6 Antibiotic resistance in *P. aeruginosa*

P. aeruginosa is intrinsically resistant to a broad spectrum of antibiotics due to the low permeability of the outer membrane (Angus *et al.*, 1982). In addition to this, a number of efflux pumps are encoded in the chromosome to transport antibiotics out of the cell. In general, the multidrug efflux systems of *P. aeruginosa* consist of three major components, an energy de-

Table 1.4: *Multidrug efflux pumps of P. aeruginosa. Table adapted from (Poole, 2001). Substrate abbreviations: AG, aminoglycosides, BL, β-lactams, FQ, fluoroquinolones, CM, chloramphenicol, TC, tetracycline, NV, novobiocin, TP, trimethorpim, SM, sulphonamides, ML, macrolides, ER, erythromycin, EB, ethidium bromide, AC, acriflavine, CV, crystal violet, SDS, sodium dodecyl sulphate, AH, aromatic hydrocarbons, HL, homoserine lactones, CL, cerulenin, TL, thiolactomycin, IR, irgasan, TS, triclosan.*

Efflux pump	Regulatory gene(s)	Substrates	Accession number(s)
mexAB-oprM	*mexR*	BL,FQ,CM,TC,NV,TP, SM,ML,EB,AC,CV,SDS, AH,HL,CL,TL,IR,TS	L11616, L23839
mexCD-oprJ	*nfxB*	BL,FQ,CM,TC,NV,TP, ML,CV,EB,AC,SDS,AH, CL,TS	U57969
mexEF-oprN	*mexT*, *(mexS)*	FQ,CM,TP,AH,TS	AJ007825, X99514
mexXY-oprM	*mexZ* (*ampR*)	FQ,AG,TC,ER	AB015853, AF147719

pendent efflux pump of the Resistance-Nodulation-Division (RND) family in the inner membrane, an outer membrane porin (often called outer membrane factor, OMF) and a periplasmic membrane fusion protein (MFP), which joins the two together (Poole, 2001). This is an efficient system of resistance, given that molecules in the cytoplasm, cytoplasmic membrane, or the periplasm can be exported. Efflux pumps, coupled with the reduced influx of drugs by the outer membrane permeability barrier, provide high levels of antibiotic resistance (Nikaido, 1998).

Four antibiotic efflux systems have been described in *P. aeruginosa*; *mexAB-oprM*, *mexCD-oprJ*, *mexEF-oprN* and *mexXY-oprM*. These pumps export a number of dyes, detergents, inhibitors, organic solvents and homoserine lactones, in addition to antibiotics (Poole, 2001). The regulatory genes and substrates of each efflux pump are outlined in Table 1.4.

As well as reducing the influx and increasing the efflux of molecules, *P. aeruginosa* is also able to inactivate antibiotics by modifying their structure. This is achieved using enzymes such as beta lactamase, which is encoded by the inducible *ampC* gene (Lodge *et al.*, 1990). Wild type

P. aeruginosa produce beta lactamase at a low basal level, which is induced to higher levels by the presence of β-lactams (Giwercman *et al.*, 1991; Bagge *et al.*, 2004). Expression of *ampC* is also controlled by the regulatory gene *ampR*, which represses transcription of *ampC* in wild type cells (Kong *et al.*, 2005). The repression of *ampC* by *ampR* can be disabled by the binding of cell wall degradation products, whose production is increased by the action of beta lactam antibiotics, to the AmpR protein, which leads to the transcription of *ampC*.

Chloramphenicol (Cm) is an inhibitor of protein synthesis, and acts by binding to the 50s ribosomal subunit and preventing peptide chain elongation (Wisseman *et al.*, 1954). The efflux pumps of *P. aeruginosa* (Table 1.4) confer resistance to Cm, and the expression of the MexEF-OprN efflux pump has been shown to be induced by the presence of Cm (Fetar *et al.*, 2011). Chloramphenicol acetyltransferase (CAT), encoded by the constitutively expressed *cat* gene, inactivates Cm by acetylation of the 3-hydroxyl group of the molecule (Shaw, 1967). However, whether or not CAT is a major determinant of chloramphenicol resistance in *P. aeruginosa* remains unclear (White *et al.*, 1999; Wang and Liu, 2004).

1.6.1 Regulation of the MexEF-OprN efflux pump

The global regulator *mexT* has been shown to enhance expression of the MexEF-OprN efflux pump, which is inactive in wild type cells, and decrease expression of the carbapenem-permeable porin OprD (Kohler *et al.*, 1999). The overexpression of this efflux pump and the repression of the porin leads to what is described as an *nfxC* mutant phenotype, whereby cells display resistance to imipenem and chloramphencol, and hyper susceptibility to β-lactams and aminoglycosides (Fukuda *et al.*, 1990).

Maseda *et al* (2000) compared the nucleotide sequences of the *mexT* gene in different PAO1 laboratory strains and their derivative *nfxC* mutants, and

found the sequence of *mexT* to be identical in the wild type and the mutant in a number of the strains studied. Therefore it was proposed that one or more mutations in a suppressor, termed *mexS*, lead to the *nfxC* mutant phenotype in some wild type PAO1 laboratory strains.

Transposon mutagenesis of the clinical isolate *P. aeruginosa* K2153 revealed that disruption of the gene PA2491 increased the production of MexEF-OprN and decreased production of OprD, while levels of MexT remained constant (Sobel *et al.*, 2005). The mutant also displayed resistance to chloramphenicol, quinolones and imipenem, much like previously described *nfxC* mutants (Fukuda *et al.*, 1990). All spontaneous multidrug resistant mutants of this clinical isolate which displayed the *nfxC* phenotype carried mutations in PA2491, which was associated with increased expression of *mexEF-oprN* and reduced OprD production. Gene PA2491 was therefore put forward as a possible *mexS* candidate.

nfxC mutants of three wild type *P. aeruginosa* laboratory strains and three clinical isolates, all with mutations in *mexS* and with a wild type version of the *mexT* gene, were transformed with a plasmid carrying the wild type *mexS* gene (Uwate *et al.*, 2013). In all three laboratory strains this led to an increase in the production of MexEF-OprN, which is unexpected given the previously hypothesised role of *mexS* as a negative regulator of *mexT* (Sobel *et al.*, 2005). In two of the three clinical isolates, the transformation of the wild type *mexS* gene led to an increased amount of MexEF-OprN, however one of the isolates displayed reduction in the amount of the efflux pump, much like the laboratory strains studied. Therefore it was concluded that there are two routes of MexT-mediated regulation of the expression of *mexEF-oprN*; a MexS mediated pathway and a MexS bypassed pathway. MexS did, however, appear to be regulating the production of MexAB-OprM and virulence factors in all six of the strains studied.

1.7 *P. aeruginosa* genomics

As of August 2015, 13 complete genome sequences of *P. aeruginosa* isolates are available at the Pseudomonas Genome Database (Winsor *et al.*, 2011). The first *P. aeruginosa* genome to be sequenced was that of the most commonly used laboratory strain PAO1 by Stover *et al.* (2000). Other isolates such as PA14 and LESB58 are also often used in laboratory studies into this opportunistic pathogen (He *et al.*, 2004; Winstanley *et al.*, 2009). As more and more whole genome sequences become available, it is becoming more and more evident that *P. aeruginosa* has a large genome size compared to other bacteria, reportedly between 5.2 and 7 Mb (Wiehlmann *et al.*, 2007), and the *P. aeruginosa* genome has been found to have a mosaic structure, consisting of regions conserved between most genome sequences, interrupted by regions which are unique to particular strains, the core and the accessory genome, respectively (Klockgether *et al.*, 2011).

The core genome is defined as regions of the genome which are highly conserved in all members of a species and contains conserved genes, usually with housekeeping functions, whereas the accessory genome refers to regions which are missing in some members of the species. The pangenome is the sum of the core genome and the accessory genome. Some accessory elements are unique to particular isolates; these strain-specific segments have been termed regions of genome plasticity (RGP) (Mathee *et al.*, 2008), and many of these are mobile genetic elements (MGEs) acquired by horizontal gene transfer (HGT). As more *P. aeruginosa* genomes are sequenced, the core genome is redefined and reduces in size, and is expected to reach an eventual plateau (Valot *et al.*, 2015), while it is expected that the size of the *P. aeruginosa* accessory genome, and therefore also the pangenome, will continue to increase (Pohl *et al.*, 2014).

1.7.1 Mobile Genetic Elements

MGEs make up a large proportion of the accessory genome of *P. aeruginosa*, and appear to be acquired by horizontal gene transfer (HGT) (Mathee *et al.*, 2008). In prokaryotes, HGT in the acquisition of genes by a bacterial cell by either transduction, DNA transfer via a bacteriophage, conjugation, DNA transfer via cell to cell contact, and potentially also transformation, the transfer of naked DNA, although there are very few reports of this in nature (Hupkova *et al.*, 1994; Domingues *et al.*, 2012). Details of MGEs commonly found in *P. aeruginosa* and relevant to this thesis are provided in the remainder of this section.

Integrative conjugative elements

Integrative conjugative elements (ICEs), also known as conjugative transposons, are elements found in both Gram positive and Gram negative bacteria which encode the genes required to control their excision from the host bacterial chromosome, their cell to cell transfer by conjugation and their integration into the chromosome of the host (Burrus *et al.*, 2002). Once integrated, ICEs are able to replicate as part of the host chromosome. Excision of ICEs from the chromosome is induced by specific environmental conditions for different ICEs, upon which ICEs circularise, undergo rolling circle replication and are transferred to new host cells by their self-encoded conjugation machinery. The ICE in the recipient cell becomes integrated in the recipient cell chromosome, and the ICE remaining in the host cell becomes reintegrated into the chromosome of the host cell (Wozniak and Waldor, 2010). A number of *Pseudomonas* pathogenicity islands (PaPIs) are thought to have evolved from ICEs, some of which have lost their mobility genes, rendering them unable to be excised from the chromosome (Kung *et al.*, 2010).

Transposable elements

Transposable elements are DNA sequences which can move from one location to another in the chromosome (known as translocation), as well as between bacterial cells via conjugation or transduction (Fedoroff, 2012). These were first discovered in maize by McClintock (1950). The simplest transposons are insertion sequences (IS), which usually consist of one or two open reading frames which encode a transposase. The majority of ISs have short terminal inverted repeats, which are recognised and cleaved by the transposase in order to facilitate translocation (Mahillon and Chandler, 1998). Larger tansposons found in *P. aeruginosa* encode functions other than those needed for translocation, such as resistance to antibiotics and heavy metals (Iyobe *et al.*, 1981; Sinclair and Holloway, 1982; Stokes *et al.*, 2007).

Bacteriophages

Two major types of bacterial viruses, bacteriophages or simply phages, are lytic and temperate. Temperate phages undergo one of two types of life cycles upon infecting the bacterial host cell: the lytic cycle or the lysogenic cycle (Figure 1.4). In the lysogenic replication cycle, the bacteriophage can become integrated into the bacterial chromosome by recombination, and replicates with the bacterial chromosome as long as the phage-encoded repressor protein, which prevents the synthesis of phage proteins required for the lytic cycle, is expressed. Integrated bacteriophages in bacterial chromosomes are known as prophages. If the repressor protein becomes inactivated, or its synthesis is stopped, the phage genome is excised from the bacterial chromosome by a phage-encoded enzyme and the lytic cycle begins, usually culminating in the lysis of the host bacterial cell. This, however, is a rare occurrence in temperate phage, typically occurring in 1 in 10,000 cell divisions. Some temperate phages are able to replicate autonomously in the

bacterial cell without integrating into the bacterial chromosome, existing in the cell as plasmids. Lytic phage only replicate via the lytic cycle. It is because of the phage-encoded repressor protein that host bacterial cells are immune to infection by bacteriophage which are already incorporated in their chromosome, as the phage repressor protein prevents the expression of any incoming genes of the same phage, as well as the genes of the integrated prophage (Madigan and Martinko, 2006).

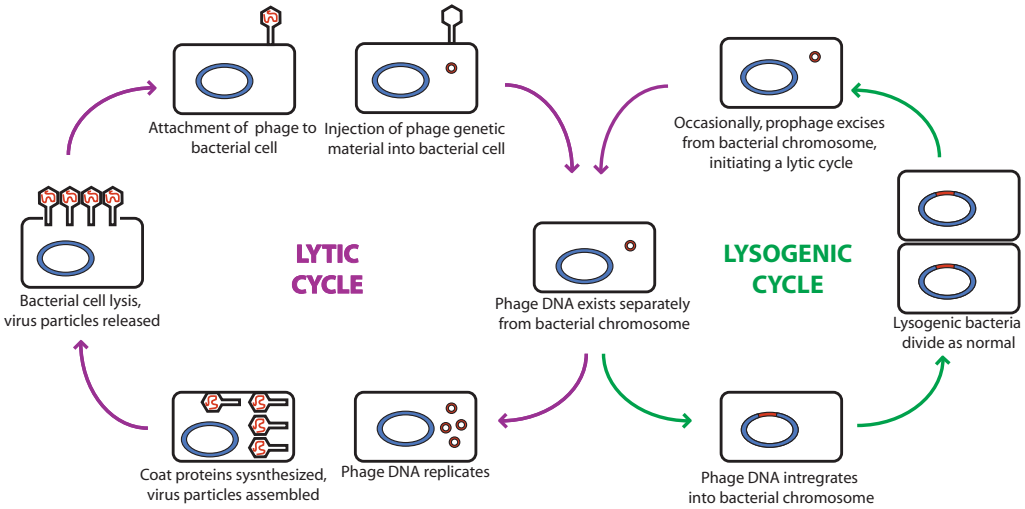

Figure 1.4: Overview of the most common bacteriophage replication cycles and mechanisms.

Upon completion of the whole genome sequence of *P. aeruginosa* PAO1, two regions of the chromosome (PA0616-PA0648 and PA0715-PA0728) were predicted to be probable prophage regions (Stover *et al.*, 2000). The first region, PA0616-PA0648, has been determined to encode narrow-spectrum bacteriocins; compounds which inhibit the growth of closely related bacterial species (Baba and Schneewind, 1998). There genes were specifically found to encode R- and F-type pyocins, whose structure resemble bacteriophage tails (Waite and Curtis, 2009). The latter region, PA0715-PA0728, encodes the filamentous Pf1-like phage Pf4, which is thought to be involved in the development of biofilms (Webb *et al.*, 2003). Prophages have been shown to increase the competitiveness of strain LESB58 (Winstanley *et al.*, 2009), an isolate which was shown to spread in a childrens CF unit in Liver-

pool (Cheng *et al.*, 1996). Prophages of LESB58 are also thought to regulate *P. aeruginosa* population density during chronic lung infection by retaining lytic activity (James *et al.*, 2015).

1.7.2 Large genomic rearrangements

Large genomic rearrangements such as large inversions and deletions have been previously reported in *P. aeruginosa* clinical strains, and are thought to contribute to a method of rapid adaptation to changing environments by *P. aeruginosa* (Kresse *et al.*, 2003; Smith *et al.*, 2006; Rau *et al.*, 2012). This fits with the generally accepted idea that bacteria undergo reductive evolution, where non-essential functions are lost over time (Andersson and Kurland, 1998). However, given that *P. aeruginosa* isolates are found in a variety of different environments, and have been found to acquire genetic elements as well as lose them, therefore displaying a high level of genomic diversity (Mathee *et al.*, 2008), it is unlikely that the genomic reduction seen in host adapted isolates is the norm. In a study on the effect of bacteriophage infection on *P. aeruginosa* it was suggested that clones with large deletions occur naturally in populations at a very low frequency rather than spontaneously arising as a response to a change in conditions (Le *et al.*, 2014). In this study phage infection was found to select for phage-resistant mutants, which had a large deletion containing a gene required for phage adsorption.

1.7.3 Genomic studies of *P. aeruginosa*

There have been a multitude of genomics studies carrying out WGS of a group of related *P. aeruginosa* isolates, with many focusing on mechanisms by which *P. aeruginosa* adapts to the CF lung. Jeukens *et al.* (2014) carried out a genomic analysis of a number of Liverpool epidemic strain (LES)-like isolates from the UK and Canada. Isolates from the two countries

could not be distinguished both phylogenetically and phenotypically in this study. Isolates were found to have distinct mutations in regulatory genes, and therefore it was concluded that this may be the cause of phenotypic diversity in LES-like isolates, which could be a mechanism for adaptation to the CF lung environment. A deletion of LES prophage was noted in one strain. Prophages in LESB58 have previously been shown to increase the competitiveness of the strain in chronic lung infections (Winstanley *et al.*, 2009), however it may be that accessory genome regions such as prophages are not essential as a lung infection evolves, given that prophage deletions have been seen in other studies of CF isolates (Fothergill *et al.*, 2010; Workentine and Surette, 2011). A study of 360 LES isolates from nine chronically infected patients revealed the coexistence of LES lineages in seven of the nine patients, although whether the patients were infected with multiple lineages or whether this was a result of multiple transmissions was unclear. Homologous recombination was observed between divergent lineages, which could be a mechanism of host adaptation by *P. aeruginosa* during chronic infection (Williams *et al.*, 2015).

Darch *et al.* (2015) looked at the genomes of multiple *P. aeruginosa* clones from a single sputum sample from a CF patient. A large number of homoplasic SNPs were found to be involved in adaptive evolution in this dataset, and it was found that these occurred as a result of recombination. In fact, in this dataset genetic variation was more likely to have occurred as a result of recombination than by spontaneous mutation. The results of this study suggested that recombination is an important factor for generating genetic diversity in the CF lung, at least in the short term. This fit previous observations by Rau *et al.* (2012) that homologous recombination plays an important role during early adaptation of *P. aeruginosa* to the CF lung; although in this study it was found to lead to the deletion of genomic content rather than SNPs.

Stewart *et al.* (2014) carried out WGS of a diverse set of 12 *P. aeruginosa* isolates, which included CF lung isolates which had adapted to the host, as well as four environmental isolates, and phylogenetic analysis found that these isolates formed three distinct and diverse phylogenetic groups. Prior to the (Stewart *et al.*, 2014) study, much of the focus of WGS studies of *P. aeruginosa* had been on CF isolates, and since *P. aeruginosa* is a highly ubiquitous organism this study highlighted the need to study the genomes of strains from other environments, in order to get a full idea of *P. aeruginosa* genomic diversity.

1.8 General experimental aims

It has been proposed that *P. aeruginosa* PAO1 exists as genetically variable sublines in strain collections and laboratories, and the current reference sequence is thought to correspond to a distinct subline with its own genetic peculiarities.

Chapter 2 aims to fully characterise the chromosomal differences between a set of commonly studied laboratory sublines of *P. aeruginosa* PAO1, as well as two strains, PAO2 and PAO3, which were directly derived from the original *P. aeruginosa* Strain 1 (PAO). The extent to which these vary from the PAO1 reference sequence will be examined. The final aim of this chapter is to deduce the sequence of *P. aeruginosa* PAO, from which all the PAO1 sublines are derived, using sequence information from these laboratory sublines and strains PAO2 and PAO3. This could be used as a more universal reference sequence which is representative of a range of PAO1 sublines, to which they are all more closely related.

Although PAO1 is the most commonly studied strain in laboratories, it corresponds to a single wound isolate within a diverse species which is able to

adapt to and survive in a range of ecological niches, including human hosts. Previous WGS studies of *P. aeruginosa* have focused on isolates from CF infections, however given that *P. aeruginosa* is able to adapt to a variety of environmental niches, the characteristics of isolates from different types of infection, and therefore different environments, may not be identical. Therefore the need to study a variety of different strains of *P. aeruginosa* from a range of environments is evident.

Chapter 3 aims to determine the genomic diversity of a set of 49 clinical isolates from distinct patients and wound types, in relation to previously sequenced *P. aeruginosa* strains. The key QS genotypes of these isolates will be analysed in order to determine any correlations with the diverse QS phenotypes previously observed; in particular for the two isolates which were previously found to be AHL-proficient and AQ-deficient.

Chapter 2

Determination of the ancestral genomic sequence of *Pseudomonas aeruginosa* PAO

2.1 Introduction

The most commonly used strain in the study of *Pseudomonas aeruginosa* is *P. aeruginosa* Strain 1c (Holloway, 1969), also referred to as strain PAO1. PAO1 is a spontaneous chloramphenicol (Cm) resistant mutant, a phenotype attributed to locus *chl-2*, of the prototroph *P. aeruginosa* Strain 1 (PAO) which was isolated by Bruce Holloway from an infected wound in Melbourne in 1954 (Holloway, 1955; Clarke and Richmond, 1975). Since then, *P. aeruginosa* PAO1 was propagated and distributed to laboratories and strain collections around the world, while the original stocks of strains *P. aeruginosa* PAO and PAO1 slowly became overlooked and unused to the point that they may no longer exist today (Figure 2.1).

Holloway and Jennings (1958) first discovered the fertility factor FP2 in *P. aeruginosa* PAT (then referred to as Strain L), which was comparable to the F factor of *E. coli*. Loutit and Marinus (1968) mapped the first ten genes on the chromosome of *P. aeruginosa* PAO using interrupted mating. A genetic map of the *P. aeruginosa* PAO chromosome was generated in 1979 (Holloway *et al.*, 1979), and with the discovery and use of additional fertility factors, the chromosome was determined to be circular in 1981 (Royle *et al.*, 1981). An initial physical map of the *P. aeruginosa* PAO1-DSMZ 1707 (PAO1-DSM) subline was constructed in 1989 using pulsed field gel electrophoresis (PFGE) (Romling *et al.*, 1989), and was later combined with information from the genetic map (Ratnaningsih *et al.*, 1990). The complete genome of the PAO1 subline from the University of Washington (PAO1-UW) was sequenced in 2000 (Stover *et al.*, 2000) and is currently available at the Pseudomonas Genome Database (Winsor *et al.*, 2011), where peer-reviewed annotations are continually updated. *P. aeruginosa* PAO1-UW was found to contain a 2.2 Mb inversion with respect to the physical map of PAO1-DSM (Stover *et al.*, 2000).

Since the sequencing of *P. aeruginosa* PAO1-UW (Stover *et al.*, 2000)

the whole genomes of a number of additional isolates of *P. aeruginosa* have been sequenced (Genbank Genomes: *P. aeruginosa*), due to the decreasing cost and increasing availability of high throughput genome sequencing. *P. aeruginosa* has been found to consist of a conserved core genome interrupted by strain-specific regions of genome plasticity (RGP) which make up the variable accessory genome (Mathee *et al.*, 2008).

Whole genome sequencing and PFGE of PAO1 sublines MPAO1, used for the construction of a transposon mutant library (Jacobs *et al.*, 2003), and PAO1-DSM, upon which the physical map is based, revealed chromosomal differences between these two PAO1 sublines and the genome sequence of PAO1-UW (Klockgether *et al.*, 2010). PAO1-DSM was found not to have the 2.2 Mb inversion found in PAO1-UW. Additionally, both MPAO1 and PAO1-DSM were found to contain a previously unreported 12 kb prophage-like sequence, designated RGP42, as well as a number of single nucleotide polymorphisms (SNPs) and small insertions and deletions (INDELs) compared to the PAO1-UW genome sequence (Klockgether *et al.*, 2010).

Luong *et al.* (2014) reported that PAO1 sublines from various laboratories display a specific phenotype, termed the P2 phenotype, whose characteristics include altered pyocyanin production, high swarming motility, low resistance to Cm and increased killing of *Caenorhabditis elegans*. The P2 phenotype was also observed in several mutants from the *P. aeruginosa* MPAO1 transposon mutant library. Cloning of the wild type gene back into one of these mutants had no effect on the P2 phenotype, indicating the ancestral strain of the MPAO1 transposon mutant library also displayed this phenotype. The P2 phenotype was thought to be caused by variations in the *mexT* and *mexF* genes of these PAO1 sublines.

In order to further characterise the genetic diversity of commonly published PAO1 sublines from various laboratories, whole genome sequencing and optical mapping of a range of these was carried out. Comparison of the

results obtained with the published sequence of PAO1-UW (Stover *et al.*, 2000) revealed genetic diversity in the form of SNPs and short INDELs as well as large chromosomal deletions and insertions. This level of genetic variation has been shown previously in *P. aeruginosa* clinical strains and is thought to be a means by which *P. aeruginosa* can expand its environmental niches (Mathee *et al.*, 2008), but this is the first report of large scale genetic variation in sublines of the same strain from different laboratories.

Unlike *P. aeruginosa* PAO1, strains PAO2 and PAO3 have not been as highly distributed and propagated in laboratories over the years. These strains are therefore thought to be less variable and more similar to the original *P. aeruginosa* PAO isolate from which they are derived, as serine (PAO2, *ser-3*) and tryptophan (PAO3, *trp-54*) auxotrophs (Figure 2.1), generated by mild manganous chloride mutagenesis (Holloway, 1955). The chromosomes of these strains were therefore analysed as well as the PAO1 sublines, and this data used to deduce the chromosomal sequence of the ancestral *P. aeruginosa* PAO. This could be utilised as a more universal reference sequence for researchers working with *P. aeruginosa* PAO1, to which all the sublines are more closely related.

Figure 2.1: Hierarchy of the isolate PAO, strains PAO1, PAO2 and PAO3, and the PAO1 sublines. *Strains PAO1, PAO2 and PAO3 are all derived from the original isolate, P. aeruginosa PAO (Holloway, 1955). PAO1 currently exists as a number of genetically variable sublines. This nomenclature of isolates, strains and sublines is used throughout the thesis. Cm = chloramphenicol.*

2.2 Materials and Methods

2.2.1 Bacterial strains, plasmids and growth conditions

Strains and plasmids used in this study are listed in Table 2.1. Unless otherwise stated, for all strains and sublines, a 1 μl loopful of glycerol stock was streaked on an LB agar purity plate, which was grown overnight at 37 °C. Batch cultures were made from a pool of colonies on the plate, as opposed to from a single colony, if the plate was not comtaminated with other bacterial species. This was carried out in order to characterise the global genome of the stock instead of that of an individual clone. A 1 μl loopful of bacterial cells were cultured at 37 °C in 10 ml of liquid lysogeny broth (LB) in a shaking incubator at 200 rpm, to allow for sufficient aeration of the bacterial cells during growth. Glycerol stocks were stored at -80 °C, and kept on dry ice when removed from the freezer for use.

Table 2.1: *Bacterial strains and plasmids used in this study*

Strain	Description	Source
PAO1-ATCC	*P. aeruginosa* PAO1 (Schroeter) Migula, ATCC 15692. First mentioned in (Holloway, 1969)	ATCC Strain Collection (http://http://www.atcc.org), purchased September 2014
PAO1-M	University of Manchester laboratory subline	University of Manchester, via A. McBain
PAO1-L	Lausanne laboratory subline	University of Lausanne, via D. Haas
PAO1-N	University of Nottingham laboratory subline	via P. Gilbert, obtained before 1992
PAO1-DSM	PAO1-DSMZ 1707 subline, used to construct physical map (Romling *et al.*, 1989)	German Collection of Microorganisms and Cell Cultures (DSMZ)
MPAO1	PAO1 subline used to construct transposon library (Jacobs *et al.*, 2003)	University of Rochester Medical Centre, NY, via B. Iglewski
PAO1-W	PAO1 subline from University of Warsaw	M. Bagdasarian, via V. Shingler
PAO1-E	University of Tartu laboratory subline	Pseudomonas Genetic Stock Centre (http://www.pseudomonas.med.ecu.edu) via M. Kivisaar, purchased 2004
PAO2	*ser-3* auxotroph of PAO	B. Holloway, via D. Haas
PAO3	*trp-54* auxotroph of PAO	B. Holloway, via D. Haas
PAO2-AUX	PAO2 restored to prototroph	P. Papakyriacou, 2014
PAO3-AUX	PAO3 restored to prototroph	P. Papakyriacou, 2014
P. aeruginosa Nottingham K13	Nottingham clinical isolate	Zoie Aiken, Clinical Microbiology, Queen's Medical Centre
pME6032	Cloning and expression vector derived from pVS1, containing the Ptac promoter (Tet^R)	(Heeb *et al.*, 2000)
pGEM-T Easy Vector II	Multi-copy cloning vector linearised with *Eco*RV with a thymine overhang, within *lacZ*(Amp^R)	Promega, UK
E. coli JM109	High efficiency competent cells	Promega, UK
E. coli S17-1 *λpir*	Tp^R Sm^R *recA thi pro hsdR*17 (rK^-mK^+) RP4-2 [Tc::Mu Km::Tn7] *λpir*	(de Lorenzo *et al.*, 1993)

2.2.2 M9 minimal medium for *Pseudmonas*

M9 minimal medium consisted of a 250 ml basal salt solution (monobasic potassium phosphate; 4.4 x10^{-2} M, dibasic sodium phosphate dihydrate; 8.4 x10^{-2} M, sodium chloride; 1.7 x10^{-2} M) and 250 ml Agar Oxoid L11 No1;

2.2 % (Thermo Fisher Scientific Inc., Loughborough, UK), supplemented with 1.9 x10^{-2} M) ammonium chloride, 10^{-4} M calcium chloride, 10^{-3} M magnesium sulphate and 10^{-2} M glucose. Minimal medium was prepared using dH2O and autoclaved at 121 °C for 20 minutes at a pressure of 15 pounds per square inch (p.s.i.). Unless otherwise stated, reagents for the minimal medium were purchased from Sigma Aldrich (Dorset, UK).

2.2.3 Genomic DNA extraction

Genomic DNA was isolated from bacterial cultures using the GenElute Bacterial Genomic DNA Kit according to the manufacturer's instructions (Sigma Aldrich, Dorset, UK), unless otherwise described. The DNA concentration and the ratios of absorbance at 260/280 nm and 260/230 nm were measured to assess the purity of each sample using the NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific Inc.) and DNA concentration was further verified using the Qubit 2.0 fluorometer (Invitrogen, Life Technologies, Paisley, UK). Chromosomal DNA quality was assessed by running agarose gel electrophoresis on a 0.8% (w/v) agarose gel in 1x Tris-Acetate-EDTA buffer (TAE). DNA was visualised on the gel using SYBR-gafe DNA gel stain following manufacturers instructions (Thermo Fisher, UK).

2.2.4 Whole genome sequencing

Whole genome sequencing and optical mapping were carried out for all PAO1 sublines, and strains PAO2 and PAO3. Paired end sequencing libraries were prepared using the Illumina Nextera kit according to the manufacturer's protocols (Illumina, San Diego, CA) [1] and whole genome sequencing was performed on the Illumina MiSeq platform.

[1] Library preparation of all but one Illumina MiSeq library was carried out by Huey Jia Cheng, University of Malaya, Kuala Lumpur.

2.2.5 Modification of the PAO1-UW reference genome sequence

The PAO1-UW reference genome (Genbank Accession Number: NC_002516) was modified using scripts inversion.pl and insertion.pl (Sections 5.3 and 5.2 in Appendix), to include the 2.2 Mb inversion between ribosomal operons *rrnA* and *rrnB* (between genome positions 727,255 and 4,788,573) and the RGP42 insertion after tRNA-Met (PA4673.1, genome position 5,242,102), once both of these features had been confirmed by optical mapping (Section 2.2.11) in all studied strains. (Genbank accessions for reference sequences can be found in Table 5.1). Genome annotations were transferred from the PAO1-UW reference sequence and the RGP42 sequence to the modified sequence using the Rapid Annotation Transfer Tool (Otto *et al.*, 2011).

2.2.6 Analysis of Sequence Read Archive data

The Illumina reads for PAO1-DSM and MPAO1, as published by Klockgether *et al.* (2010), were downloaded from the Sequence Read Archive (SRA). Both datasets consisted of 35 bp single end Illumina reads.

Reads from both datasets were combined and mapped to the PAO1-UW reference genome (Genbank Accession Number: NC_002516) with the ELAND read aligner, using parameters as described in (Klockgether *et al.*, 2010). The resulting alignment file was used to separate the Illumina reads for each subline into two groups; reads which map to the PAO1-UW reference sequence, and reads which do not map to the reference. These two groups of reads were then mapped separately to the RGP42 sequence (Genbank Accession Number: GQ141978), the PAO1-UW reference sequence, and the modified PAO1-UW reference sequence, using Smalt version 0.7.0.1 (Ponstingl, 2012) with a kmer of 12 and a step size of 2.

2.2.7 Read mapping, realignment and variant calling

Whole genome sequence datasets were subjected to adapter trimming using Trimmomatic version 0.30 (Bolger *et al.*, 2014), as well as hard trimming at the 5' end using the script fastq_trim_5.pl (Section 5.4 in Appendix). Trimmed reads were aligned to the modified PAO1-UW sequence using Smalt version 0.7.0.1 (Ponstingl, 2012) with a kmer of 12 and a step size of 2. The alignment files were sorted, and PCR duplicates removed using Samtools (Li *et al.*, 2009). Details of the percentage of reads mapped, and the reference genome coverage for each dataset can be found in Table 2.3. INDEL realignment was carried out with the Genome Analysis Toolkit, (McKenna *et al.*, 2010), and SNPs and INDELs determined using Samtools mpileup, using varying base quality thresholds (parameter -Q). SNPs were filtered to exclude those with low coverage (<5x) and those which were not supported by at least two reads on each strand using scripts filter_vcf.pl and filter_vcf_dp4.pl (Sections 5.5 and 5.6 in Appendix). SNPdat (Doran and Creevey, 2013) was used to characterise SNPs in genes, and the resulting amino acid changes. Regions with zero read coverage were detected using the genomeCoverageBed functionality of bedtools with the -bga option (version 2.17.0, Quinlan and Hall, 2010).

2.2.8 *De novo* assembly

De novo assembly was carried out using Velvet version 1.2.07 (Zerbino and Birney, 2008) with a kmer of 121 with no scaffolding, and the expected coverage flag set to auto.

2.2.9 Nucleotide BLAST

Nucleotide BLAST searches of novel sequences were carried out at the NCBI BLAST server (Johnson *et al.*, 2008) against the nucleotide sequence database (nt). Distance trees of BLAST results were calculated at the NCBI

web server, using the Fast Minimum Evolution tree method (Desper and Gascuel, 2004).

2.2.10 Phage genome annotation

Initial phage genome annotation was carried out using the PHAST web-server (Zhou *et al.*, 2011). Gene product annotations were assigned based on inspection of NCBI BLASTn results and Pfam (http://pfam.xfam.org/) results of each predicted coding region. Gene models were modified according to these results where appropriate. Additional coding sequences were assigned using Artemis open reading frame (ORF) finder and subsequent inspection of NCBI blastn and Pfam results.

2.2.11 Optical Mapping

High molecular weight (HMW) DNA was isolated from a single colony using the HMW DNA Isolation Kit (Opgen, Inc., Gaithersburg, MD) according to the manufacturer's instructions. The quality of HMW DNA was assessed using QCard (Opgen, Inc., Gaithersburg, MD) and the concentration of the HMW DNA was adjusted using dilution buffer. DNA molecules were stretched and linearised on a charged glass surface using a channel forming device (CFD). The restriction endonuclease *Bam*HI was added into a MapCard and incubated in a MapCard Processor to allow digestion of the DNA molecules on the glass surface. The fragmented DNA was stained with DNA intercalating fluorescent dye and the DNA molecules were imaged under a high resolution microscope. DNA molecules were assembled into a whole genome map using the OpGen Genome Builder software.[2] Optical maps were aligned to PAO1-UW (GenBank accession no.: NC_002516) using the OpGen MapSolver software and manually inspected to locate regions of dif-

[2]Optical map preparation and the building of the whole genome map was carried out by Corrinne Lee, University of Malaya, Kuala Lumpur.

ference larger than 2 kb, since fragments shorter than 1 kb or two adjacent fragments that are both smaller than 1 kb cannot efficiently be detected by optical mapping, and are therefore underrepresented in the dataset.

2.2.12 Polymerase Chain Reaction

Polymerase chain reactions (PCRs) were performed using GoTaq polymerase (Promega, Wisconsin, USA). The PCR mix was made up according to the manufacturers instructions.

Primers were designed using the Primer3 version 0.4.0 web interface (http://bioinfo.ut.ee/primer3-0.4.0/) and were synthesised by Sigma Aldrich (Dorset, UK). Primer sequences and cycling conditions used can be found in Section 2.2.13, Table 2.2.

PCR purification was performed using the QIAquick PCR purification kit (QIAGEN, Hilden, Germany) following the supplier's instructions. PCR products were visualised on a 1% w/v agarose gel, as described in Section2.2.3.

2.2.13 Primers and Cycling Conditions

Primer sequences and cycling conditions used in this thesis are shown in Table 2.2.

Table 2.2: *Oligonucleotides and cycling conditions used in this study. Sequences are in 5' to 3' order. Artificial restriction sites are underlined. All oligonucleotides were obtained from Sigma-Aldrich, UK. Symbol ' = minute.*

Oligonucleotide	Description	Cycling conditions
PA0316FW	TATCAATTGATGAGCAAGACCTCTCTCGAC, forward, *serA*, *Mfe*I	2' at 95 °C; 30 cycles: 1' at 95 °C, 30 sec at 64 °C, 45 sec at 72 °C; 5' at 72 °C
PA0316RV	TATATCGATTCAGAACAGCACGCGGCTA, reverse, *serA*, *Cla*I	
PA0651FW	TATGAATTCATGAGTGTGCCGACGGTT, forward, *trpC*, *Eco*RI	2' at 95 °C; 30 cycles: 1' at 95 °C, 30 sec at 65 °C, 45 sec at 72 °C; 5' at 72 °C
PA0651RV	ACCTGTGCCAGCAACCGGACGCCA, reverse, *trpC*	
RGP42_1F	CGCAGATAGCGATCACACAT, forward, RGP42 fragment 1	2' at 95 °C; 30 cycles: 1' at 95 °C, 30 sec at 60 °C, 45 sec at 72 °C; 5' at 72 °C
RGP42_1R	TCACTGCGAGACCAACCA, reverse, RGP42 fragment 1	
RGP42_2F	TGTCGACCTTGATGTTGTGG, forward, RGP42 fragment 2	2' at 95 °C; 30 cycles: 1' at 95 °C, 30 sec at 60 °C, 45 sec at 72 °C; 5' at 72 °C
RGP42_2R	CGCTATGAGGACTGGAAACAG, reverse, RGP42 fragment 2	
RGP42_3F	TTGGTGTCTTGGGTCTTGC, forward, RGP42 fragment 3	2' at 95 °C; 30 cycles: 1' at 95 °C, 30 sec at 60 °C, 45 sec at 72 °C; 5' at 72 °C
RGP42_3R	GAACTGATCGACAGTCTTGGC, reverse, RGP42 fragment 3	
RGP42_4F	CCGGGGACTTGTCTTGTTT, forward, RGP42 fragment 4	2' at 95 °C; 30 cycles: 1' at 95 °C, 30 sec at 60 °C, 45 sec at 72 °C; 5' at 72 °C
RGP42_4R	TCGGAACAGCTTCACCGTT, reverse, RGP42 fragment 4	
RGP42_5F	TTCAGGGTCTTAGCGAGCAT, forward, RGP42 fragment 5	2' at 95 °C; 30 cycles: 1' at 95 °C, 30 sec at 60 °C, 45 sec at 72 °C; 5' at 72 °C
RGP42_5R	TCACAGGTGCTGTGGAGAAG, reverse, RGP42 fragment 5	
RGP42_6F	ATGGCGTCATCCCAAGT,forward, RGP42 fragment 6	2' at 95 °C; 30 cycles: 1' at 95 °C, 30 sec at 60 °C, 45 sec at 72 °C; 5' at 72 °C
RGP42_6R	ACTCGCCCGTGGTTACCT, reverse, RGP42 fragment 6	
PAO1_DSF	TCTCCGCTAGATACCCCCTCA, forward, upstream of PAO1 58.6 kb deletion	2' at 95 °C; 30 cycles: 1' at 95 °C, 30 sec at 64 °C, 45 sec at 72 °C; 5' at 72 °C
PAO1_DSR	TACCGGCCAGGTAATCGT, reverse, downstream of PAO1 58.6 kb deletion	
PAO1_DWR	TACCGATCCAGACGTCATGG, reverse, within PAO1 58.6 kb deletion	2' at 95 °C; 30 cycles: 1' at 95 °C, 30 sec at 60 °C, 1' at 72 °C; 5' at 72 °C
mexT_FW	TCTGAACCTGCTGATCGTGT, forward, *mexT*	2' at 95 °C; 30 cycles: 1' at 95 °C, 30 sec at 63 °C, 2' at 72 °C; 5' at 72 °C
mexT_RV	CCGATGAACATGCTGATCCG, reverse, *mexT*	
PAO1MIN_FW	GCTCCAGTTGCCTGAGAAAC, forward primer within PAO1-M insertion	2' at 95 °C; 35 cycles: 1' at 95 °C, 30 sec at 60 °C, 1' at 72 °C; 5' at 72 °C
PAO1MIN_RV	GTTGAACGCCTCGTATGGTT, reverse, within PAO1-M insertion	

2.2.14 Random amplified polymorphic DNA analysis

For each strain or subline analysed, multiple colonies were sampled from each LB agar purity plate for the RAPD analysis. DNA extraction was carried out using Chelex-100 Resin (BioRad 143-2832, Hertfordshire, UK). A 5% solution of Chelex-100 in molecular grade water (Sigma Aldrich W4502, Dorset, UK) was used for the extraction, as described in (de Lamballerie *et al.*, 1992). A 1 μl loopful of the multiple colonies was suspended in 100 μl of 5% Chelex solution. After mixing with a vortex and pulse centrifugation, the suspension was incubated at 56 °C for 15 minutes. Following further mixing with a vortex and pulse centrifugation, the suspension was then incubated for a further 8 minutes at 99 °C, and then centrifuged at 13,000 rpm in a microcentrifuge (Heraeus Pico Biofuge) for two minutes. The supernatant was extracted and used for PCR.

PCR amplification was carried out using primers D10514 and 272 for *P. aeruginosa* (Vogel *et al.*, 1999; Mahenthiralingam *et al.*, 1996), and Ready-To-Go RAPD Analysis Beads (GE Healthcare 27-9500-01, Buckinghamshire, UK). 0.25 μl of primer (100mM), 22.75 μl of molecular grade water (Sigma Aldrich W4502, Dorset, UK) and 2 μl of DNA were added to each Ready-To-Go RAPD bead.

Forty-five PCR cycles were run as previously described for *P. aeruginosa* (Vogel *et al.*, 1999). Agarose gel electrophoresis was carried out using a 1.5% agarose gel in 1x Tris-Borate EDTA (TBE) buffer. The gel was stained with ethidium bromide and visualised using a UV transilluminator.

2.2.15 Auxotroph Phenotype Determination and Complementation of Prototroph Phenotype

The laboratory work in this section was carried out under my supervision by Pantelitsa Papakyriacou as part of her MSc Research Project at the University of Nottingham.

Strains PAO1-L, PAO2 and PAO3 were grown overnight on M9 minimal medium agar and on M9 minimal medium agar supplemented with serine or tryptophan at concentrations of 3 mM and 9 mM in order to confirm the auxotrophies of PAO2 and PAO3.

Primer sequences, which incorporate restriction enzyme recognition sequences, as well as cycling conditions for each primer set, are listed in Table 2.2. Ligations were carried out using T4 DNA Ligase (Promega, UK) as per manufacturer's instructions.

PA0316 was cloned into the shuttle expression vector pME6032 (Heeb *et al.*, 2000) in two steps. TA cloning was carried out into *E. coli* S17 using the pGEM-T Easy Vector System II (Promega, UK) according to manufacturers instructions with the following adjustment; SOC medium was replaced with lysogeny broth (LB). Positive clones were verified using Sanger sequencing. PA0316 from the TA cloning vector was subcloned into pME6032. Wild type PA0651 from PAO1-L was cloned directly into the pME6032 shuttle vector (Heeb *et al.*, 2000). Competent cells were transformed by electroporation, as previously described (Green *et al.*, 2012).

Transformation into *P. aeruginosa* was carried out using standard molecular biology methods (Sambrook and Russell, 2001). Transformed cells were subsequently grown on M9 minimal media agar to test for restoration of the prototroph phenotype.

2.2.16 Bacteriophage plaque assay

PAO1-L and PAO1-M cultures were grown as described in Section 2.2.1. One of the PAO1-M cultures was centrifuged at 6000 rpm for 10 minutes in a Beckman Coulter Allegra 64R centrifuge. The supernatant of the centrifuged PAO1-M culture was filter sterilised using a 0.2 μm pore filter. Five rounds of 1 in 10 serial dilution were carried out on the PAO1-M supernatant.

The bacteriophage plaque assay was carried out using the filter sterilised PAO1-M supernatant at varying dilutions, with cells from PAO1-L, to test for lytic phage activity, and PAO1-M, the negative control. 3 ml of 0.7% (w/v) soft top agar was seeded with 10 μl of supernatant and 30 μl of bacterial cells and grown over a layer of 1.5% (w/v) LB agar.

Phage were picked from the plaques using a toothpick. These were then subjected to a colony PCR using primers specific to the phage sequence (Table 2.2). PCR purification was carried out using the GENECLEAN Turbo Kit (MP Biomedicals, Santa Ana, USA) according to manufacturer's instructions, followed by Sanger sequencing of the PCR product.

2.2.17 GC Skew Analysis

The cumulative GC skew across the chromosome of PAO1-UW and the modified PAO1-UW sequence were calculated using GenSkew (http://genskew.csb.univie.ac.at/) with a window size of 500.

2.3 Results

2.3.1 Data from the Sequence Read Archive (SRA) for sublines MPAO1 and PAO1-DSM

The quality of the Illumina data for sublines PAO1-DSM and MPAO1, obtained from (Klockgether *et al.*, 2010), was inspected using FastQC version 0.10.0 (Babraham Institute, unpublished) with varying kmer sizes[3] A kmer corresponds to all possible subsequences of length k present in a sequence. The per base sequence quality plots, which show the average Phred quality score of each base across the length of all Illumina reads, for the two sublines are shown in Figure 2.2.

[3]Raw fastq files for PAO1-DSM and MPAO1 from the (Klockgether *et al.*, 2010) study were provided by the authors.

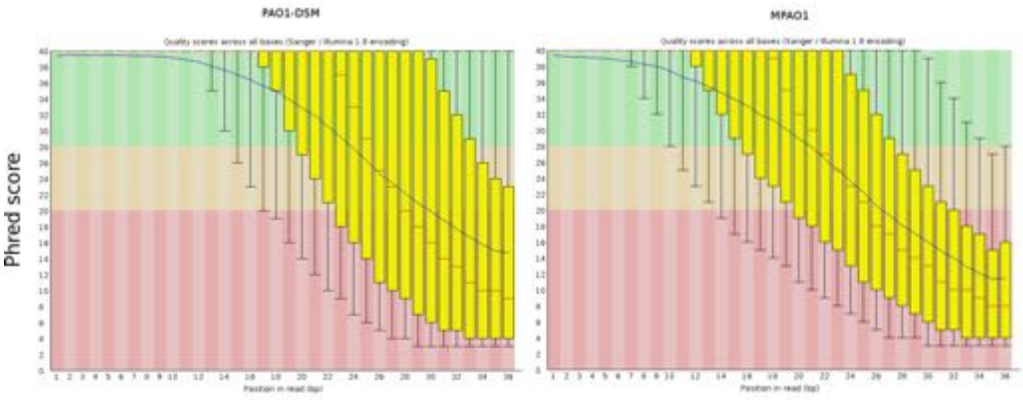

Figure 2.2: Average quality (Phred) score across the length of the single end SRA Illumina reads for PAO1-DSM and MPAO1. *Quality scores are given as Illumina Q score, where Q10 is equal to 90% base call accuracy Q20 is equal to 99% base call accuracy, and Q30 is equal to 99.9% base call accuracy. Quality plots were generated using FastQC version 0.10.0 (Babraham Institute, unpublished)*

There is a considerable drop in quality around halfway along the 35 bp reads in these two Illumina sequencing runs; after base 25 the average base quality drops below Q20 (99% accuracy). The overall percentage of bases with a quality score greater than or equal to Q20 was 76% for PAO1-DSM and 68.9% for MPAO1. Given the short read length, quality trimming would result in very short reads of $\leq$ 35 bp. In order to have uniform read lengths and overall data quality for all datasets, the decision was taken to resequence the MPAO1 and PAO1-DSM sublines.

2.3.2 Quality control of whole genome sequencing reads

Read quality of Nextera reads was assessed using FastQC. The plots for read quality and sequence content for the forward and reverse reads of MPAO1 are shown in Figure 2.3. From the quality score plots, it is evident that the reverse reads have a greater variation in quality score towards the 3' end of the reads compared to the forward reads. This is reflected in the sequence content plot (Figure 2.4) for the reverse read, by a change in sequence composition at the 3' end of reverse reads, which is not seen to the same extent in the forward reads.

By default, FastQC displays overrepresented kmers of length 5 in the input dataset, the results for which are shown in Figure 2.5. From this graph, it is clear that there are several overrepresented 5mers (all possible subsequences of length 5 present in a sequence) at the 5' end of all forward and reverse reads. The sequences of these 5mers correspond to the sequences of the Nextera primers used. There also appears to be an enrichment of poly-A sequences at the 3' end of both forward and reverse reads and an enrichment of poly-T sequences in the reverse read only. To investigate this further, the FastQC analysis was repeated with a kmer value of 10 (Figure 2.6). In the forward reads, there is an increase in the enrichment of 10mers which correspond to the Nextera adapter sequence. This is likely to be due to a combination of long read length and short fragment size. Similar FastQC results were obtained for all sequenced strains and sublines.

In light of these findings, adapter trimming was carried out on all datasets, as well as a hard trim at the 5' end of reads, in order to remove primer sequences.

2.3.3 Read mapping

The read mapping statistics for each strain and subline are shown in Table 2.3. In general, >99% of reads for each strain/subline mapped to the reference sequence, apart from strain PAO2, where only 90.06% of reads mapped. Unmapped reads from PAO2 were found to correspond to reads from a novel prophage sequence, found in PAO1-M (section 2.3.11).

Strains PAO1-M, PAO2 and PAO3 were initially sequenced at the University of Nottingham, prior to being sequenced at the University of Malaya, along with the other PAO1 sublines shown in Table 2.1. *De novo* assemblies were carried out using WGS data for each isolate, from both universities. The PAO1-M insertion sequence was used as a query in a BLAST search of all thirteen de novo assemblies, and according to the results of this BLAST

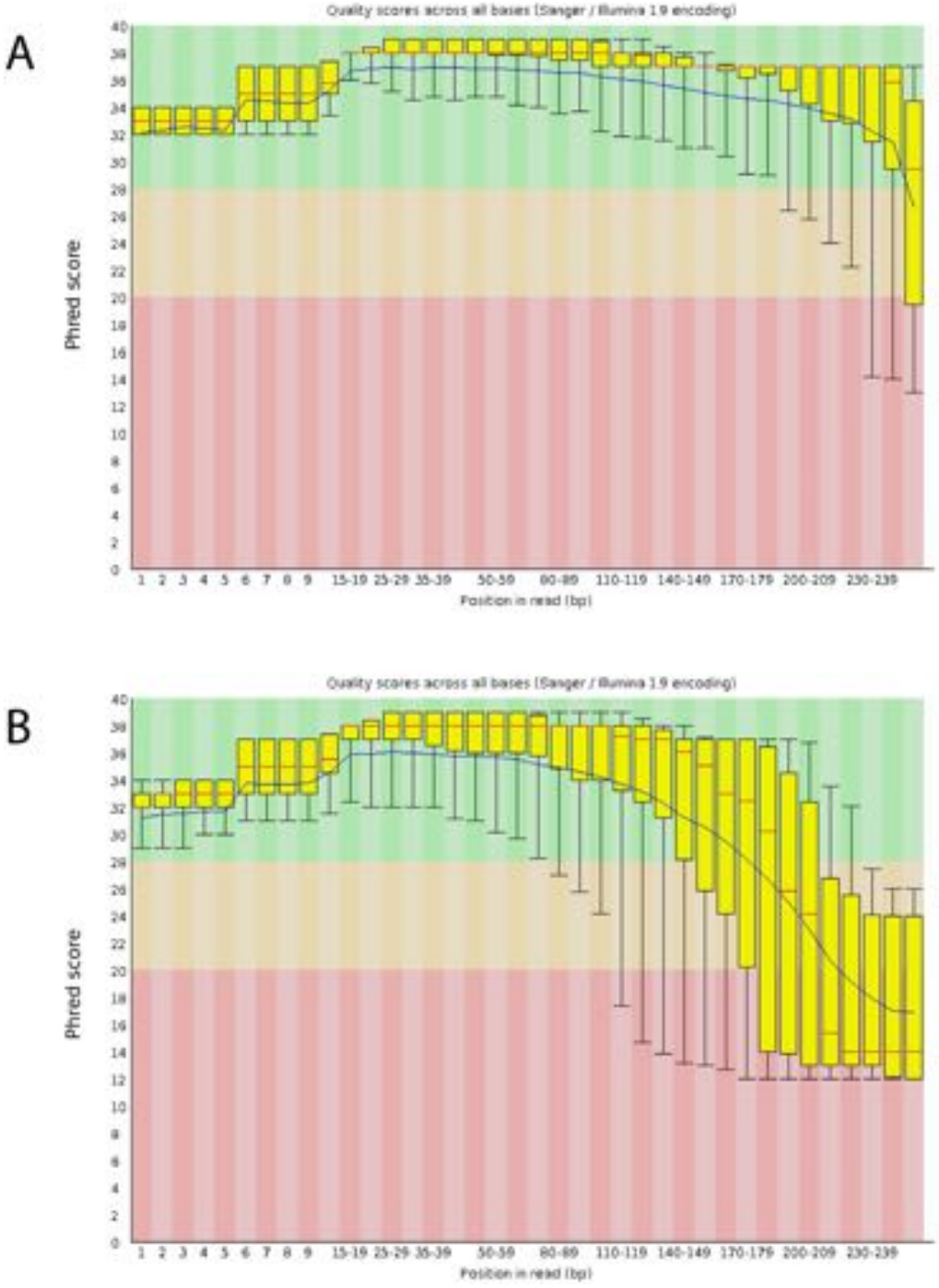

***Figure 2.3: FastQC quality plots for raw reads of subline MPAO1. A**: Quality scores across all bases, forward reads. **B**: Quality scores across all bases, reverse reads.*

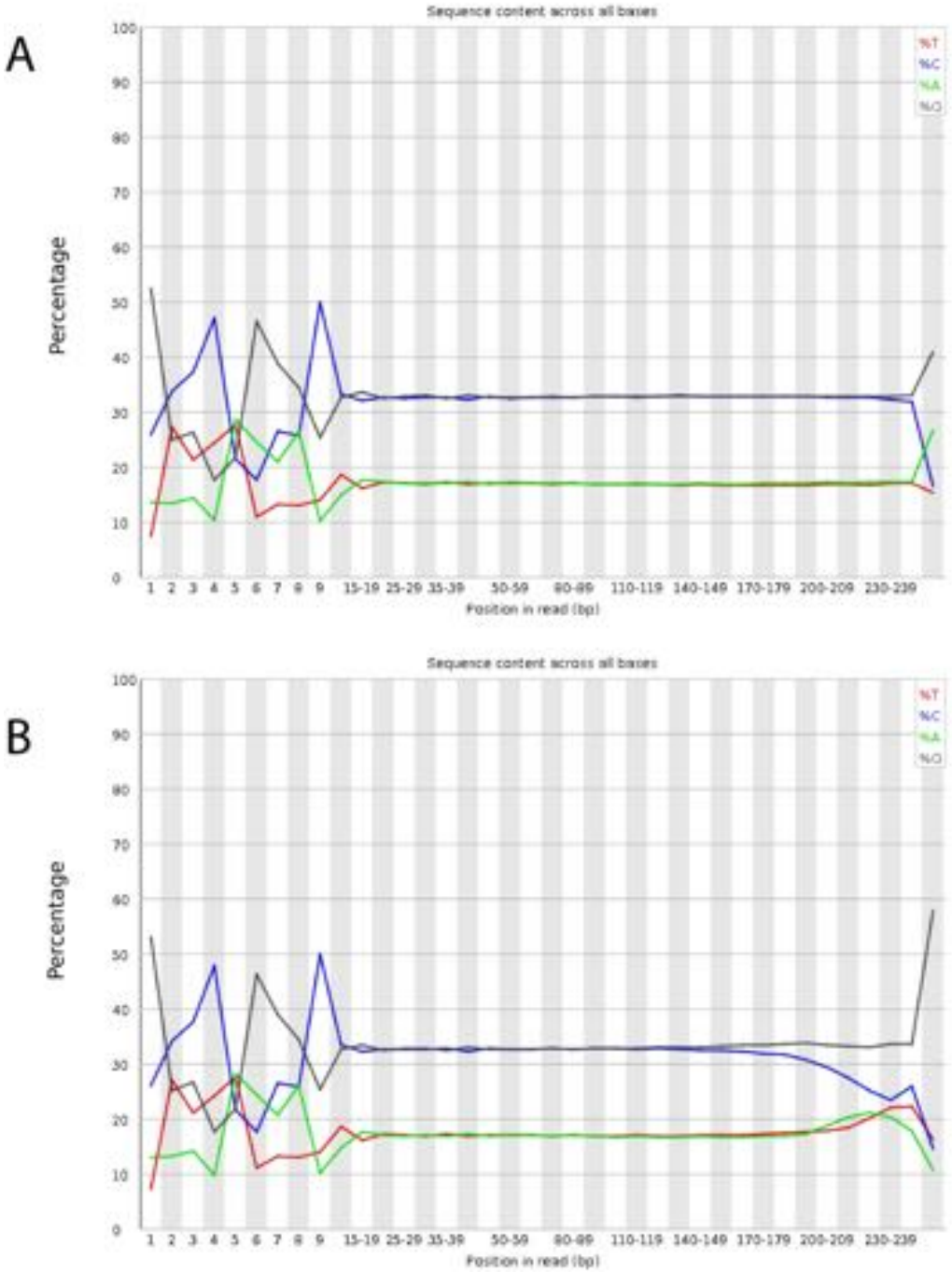

Figure 2.4: FastQC percentage sequence content plots for raw reads of subline MPAO1. ***A****: Sequence content across all bases, forward reads.* ***B****: Sequence content across all bases, reverse reads.*

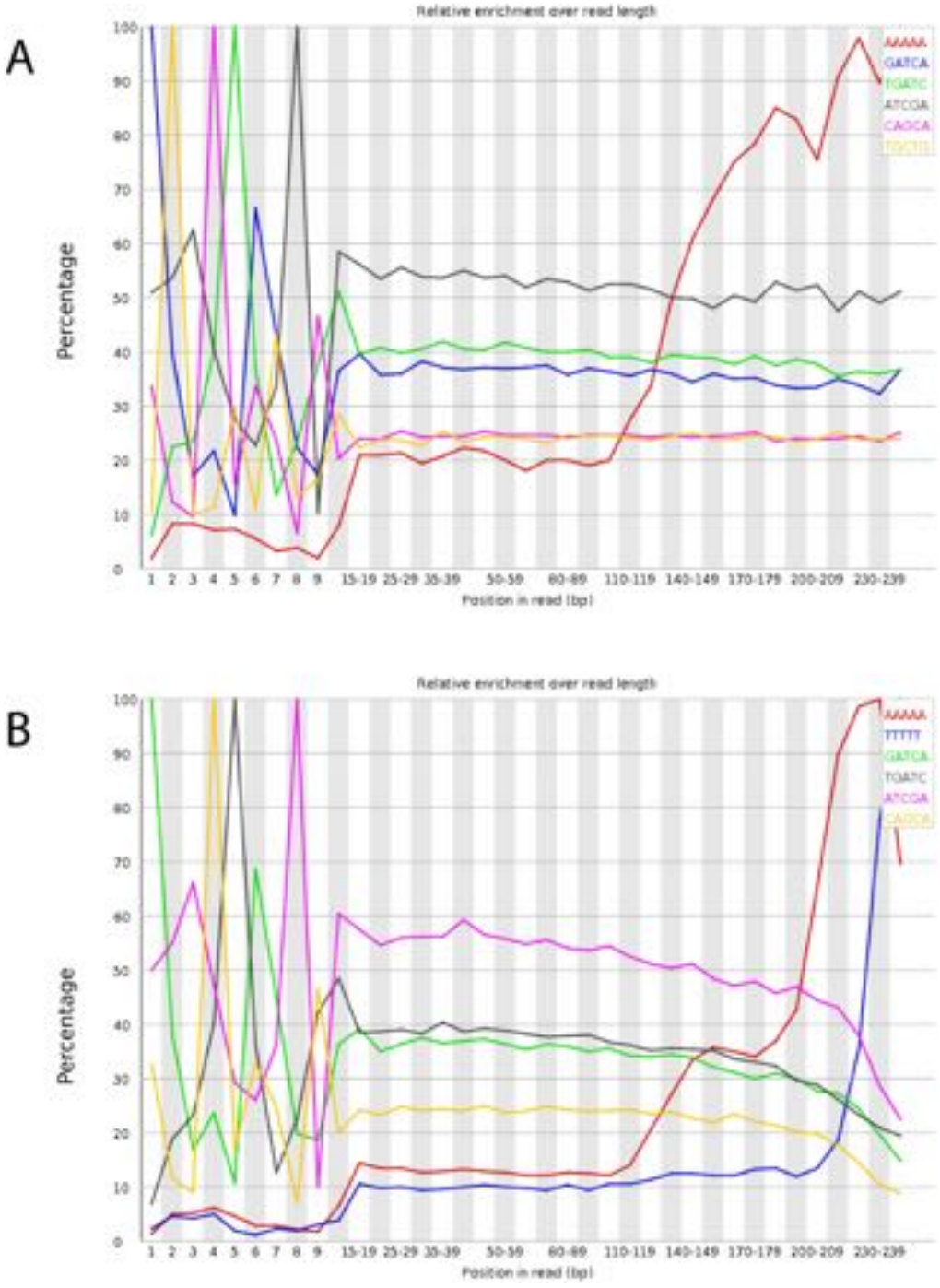

Figure 2.5: FastQC percentage kmer content plots for raw reads of subline MPAO1, where kmer=5. *Graphs show relative enrichment of 5mers over the read position in* ***A****: forward reads and* ***B****: reverse reads.*

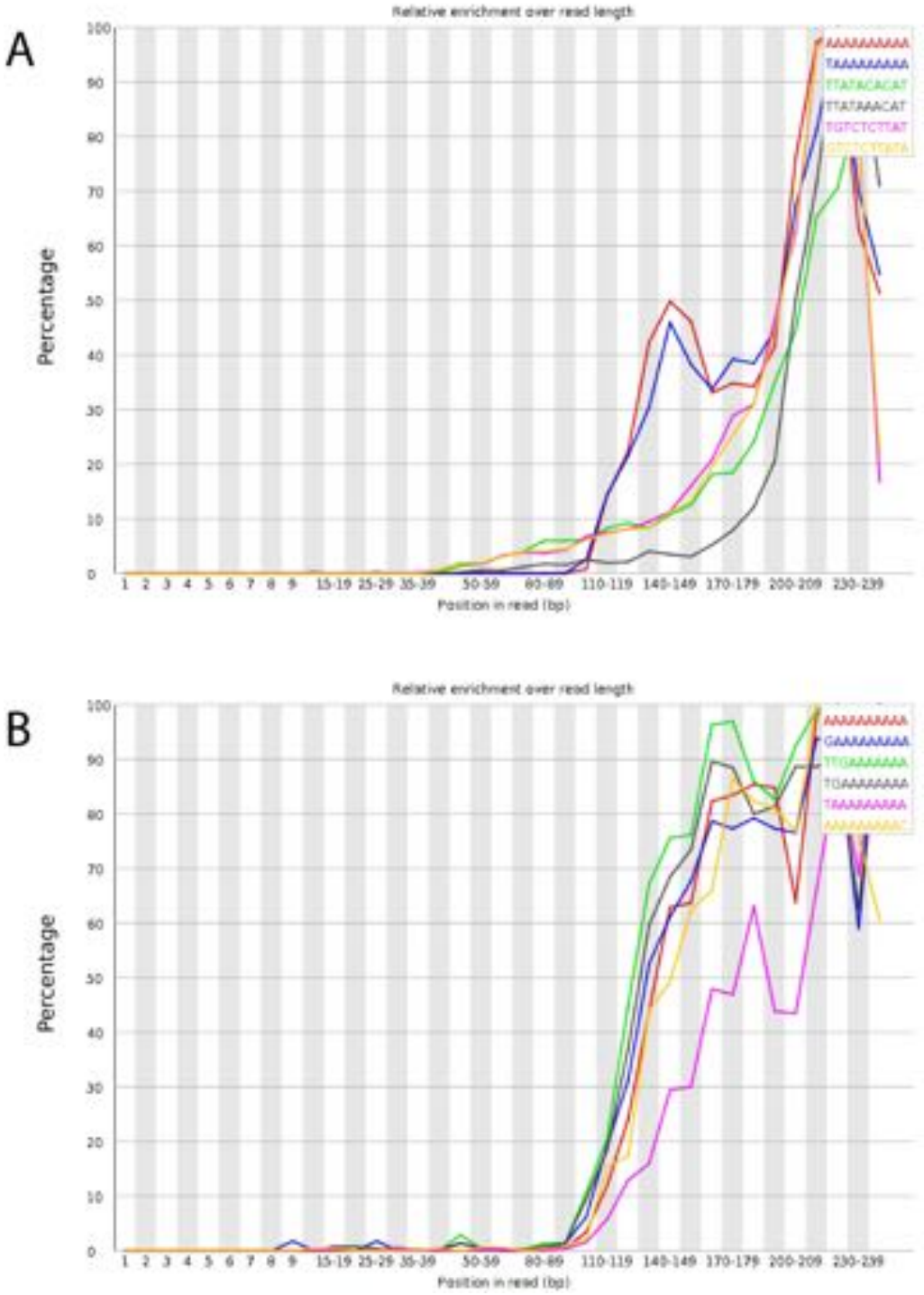

Figure 2.6: FastQC percentage kmer content plots for raw reads of subline MPAO1, where kmer=10. *Graphs show relative enrichment of 10mers over the read position in* ***A****: forward reads and* ***B****: reverse reads.*

Table 2.3: *Read mapping statistics for each strain/subline.*

Strain/subline or strain	Percentage reads mapped	Percentage genome coverage
PAO1-ATCC	99.93	99.97
PAO1-M	99.38	99.99
PAO1-L	99.95	99.99
PAO1-N	99.93	99.13
PAO1-DSM	99.87	100
MPAO1	99.97	99.98
PAO1-W	99.96	99.98
PAO1-E	99.77	99.12
PAO2	90.06	100
PAO3	99.87	100

search the PAO1-M insertion sequence was present in PAO1-M sequenced at both universities, and in PAO2 sequenced at the University if Malaya, but not in PAO2 sequenced at the University of Nottingham. This indicates that the PAO2 sample squenced at the University of Malaya became contaminated with the PAO1-M phage at some point during DNA extraction or sample preparation. The PAO1-M insertion was not found in the optical map of PAO2, and was only found in the optical map of PAO1-M.

Sequencing reads of all strains/sublines covered >99% of the modified reference sequence, with all strains/sublines showing complete coverage of the RGP42 region (Klockgether *et al.*, 2010). Optical mapping results were in agreement with the whole genome sequencing results, indicating the presence of the RGP42 in all strains and sublines (Figure 2.7).

Bedtools was used to detect areas of the modified PAO1-UW genome where no sequencing reads mapped. These regions were inspected visually using Artemis to determine if they corresponded to putative deletions in each sequenced strain and subline. The determined putative deletions are shown in Table 2.4.

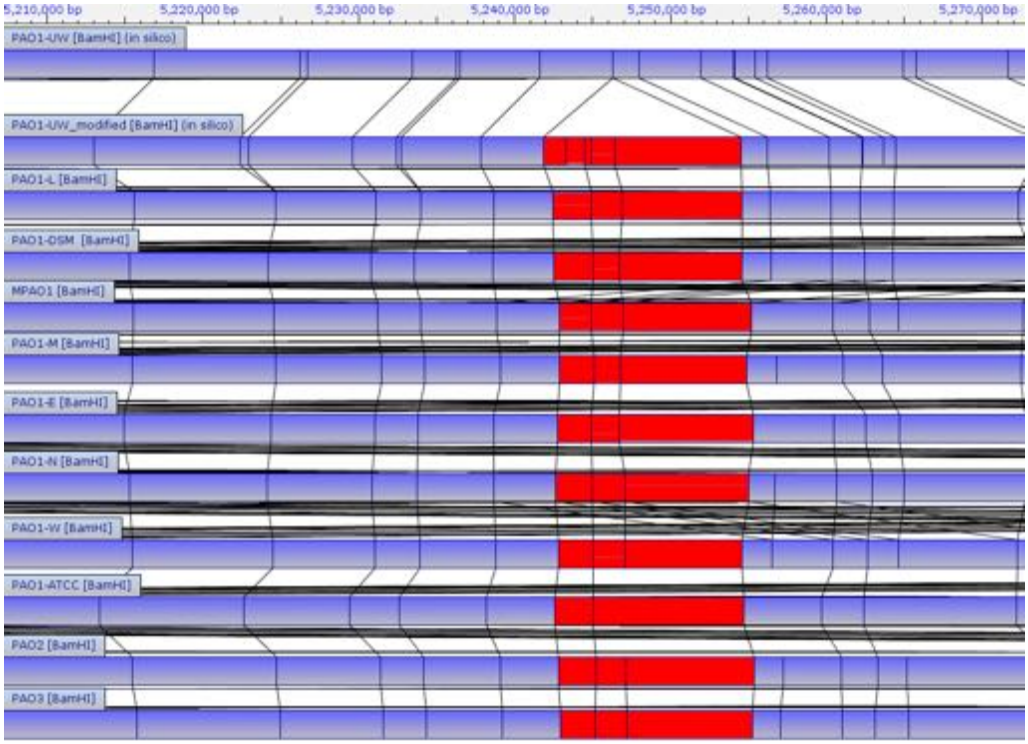

Figure 2.7: Optical mapping confirms the presence of the RGP42 insertion sequence in all studied strains and sublines. *The RGP42 sequence is highlighted in red. Optical maps for all studied strains and sublines, as well as in silico optical maps for PAO1-UW and the modified version of PAO1-UW were compared using the OpGen MapSolver software.*

Table 2.4: Possible deletions in all sequenced strains and sublines. *Deletions were indicated by Bedtools and subsequent inspection of whole genome sequencing data in Artemis. Genomic coordinates are with reference to the modified PAO1-UW sequence, and correspond to the largest putative deletion size of all sublines/strains in which it is contained. Functional annotations were obtained from the Pseudomonas database (Winsor* et al.*, 2011).*

Genomic coordinates	Putative deletion size	Strain/subline(s)	Genes Involved	Functional annotation
169,284	1 bp	All	PA0149	probable sigma-70 factor, ECF subfamily
355,708 - 355,761	54 bp	PAO2	PA0316, *serA*	D-3-phosphoglycerate dehydrogenase
411,126	1 bp	All	Upstream of PA0367	probable transcriptional regulator
3,927,909 - 3,928,016	108 bp	PAO2	PA1458, *cheA*	probable two-component sensor
4,075,203	1 bp	All	PA1327	probable protease
4,240,059	1 bp	All	PA1174, *napA*	periplasmic nitrate reductase protein
4,734,861 - 4,789,127	54,267 bp	PAO1-N, PAO1-E	PA0707 - PA0668.4	See Table 2.13
4,789,045 - 4,789,122	78 bp	All	PA0668.4, *rrnA*	See Table 2.13
4,837,060 - 4,837,063	4 bp	PAO1-W	PA4309, *pctA*	chemotactic transducer
5,054,947	1 bp	PAO1-N, PAO1-E	PA4514, *piuA*	probable outer membrane receptor for iron transport
5,095,403 - 5,095,525	123 bp	PAO3	PA4547, *pilR*	two-component response regulator
5,246,891 - 5,246,893	3 bp	All	PA4673.6, *coaA* (RGP42)	phage coat protein A
5,265,759 - 5,266,752	994 bp	PAO1-M, MPAO1, PAO1-N, PAO1-W, PAO1-E, PAO1-L	PA4684-PA4685	hypothetical proteins

2.3.4 Initial variant calling reveals a high proportion of SNPs occur in the RGP42 insertion sequence

Initial variant calling was carried out using mpileup base quality thresholds of Q10, Q15 and Q20, where the variant caller skips all bases in reads with a base call error probability greater than or equal to 0.1, 0.03 or 0.01, respectively (Ewing and Green, 1998). SNPs were filtered as described in Section 2.2.7. A large proportion of SNPs for each strain/subline at each base quality threshold were found to be located within RGP42 (Figure 2.8).

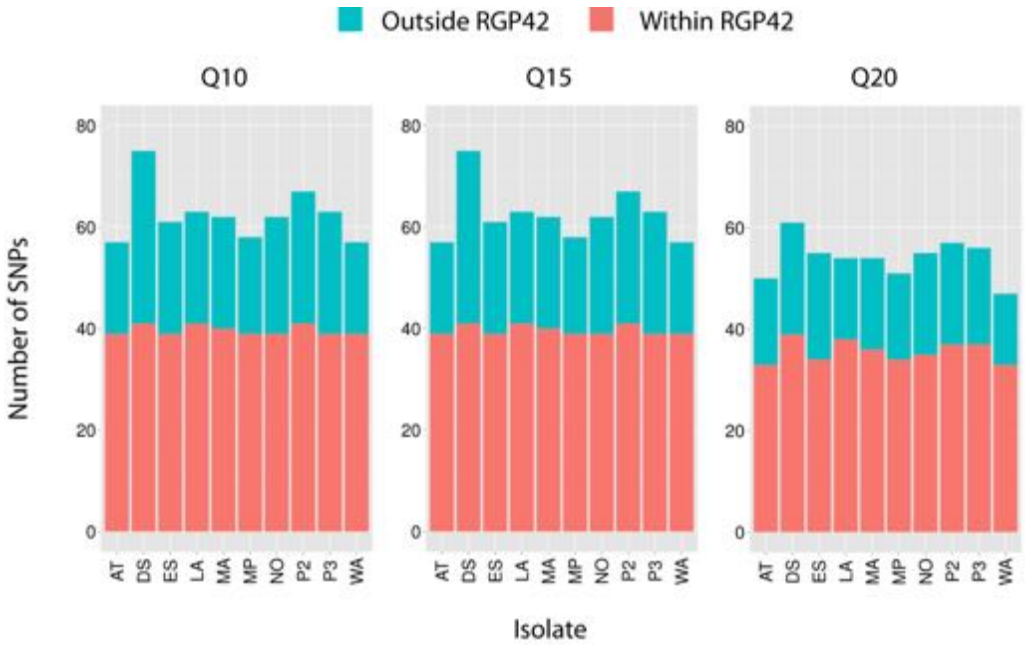

Figure 2.8: A large proportion of total SNPs for each strain and subline are located within RGP42. *Number of SNPs within RGP42 (shown in pink) and in the remainder of the genome, or outside RGP42, (shown in blue) for each strain/subline at base quality thresholds Q10, Q15 and Q20. AT=PAO1-ATCC, DS=PAO1-DSM, ES=PAO1-E, LA=PAO1-L, MA=PAO1-M, MP=MPAO1, NO=PAO1-N, P2=PAO2, P3=PAO3, WA=PAO1-W.*

It is evident that a similar number of SNPs are found within the RGP42 island between strains/sublines at base quality thresholds Q10 and Q15, but that there is a decrease in the number of SNPs in RGP42 at Q20 (Figure 2.8). The number of SNPs in the remainder of the genome (outside RGP42) in each strain/subline follows the same pattern, with less SNPs outside of RGP42 at base quality threshold Q20 compared to Q10 and Q15.

However, SNPs within RGP42 outnumber those outside of RGP42 for all strains/sublines and base quality thresholds.

The RGP42 island was first described by Klockgether *et al.* (2010), and contains a prophage-like region that shares synteny and sequence identity with the region between PA0727 and PA0717 in the PAO1-UW chromosome. This region in PAO1-UW corresponds to the genomic island RGP5 (Figure 2.9).

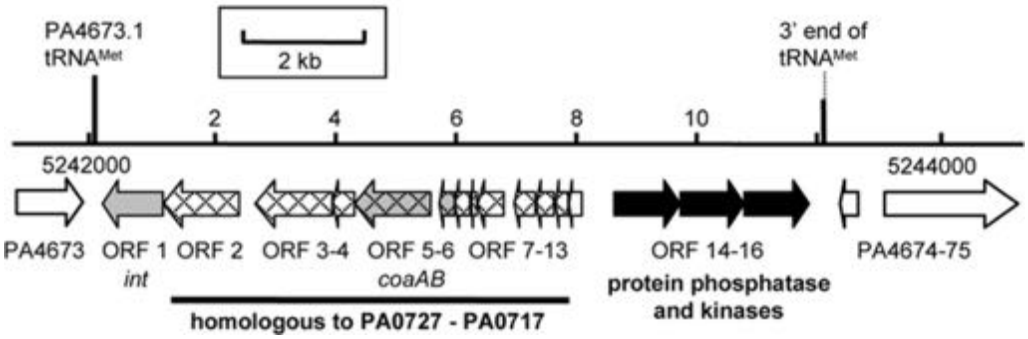

Figure 2.9: Gene map of the 12 kb RGP42 insertion and flanking region in the PAO1-UW chromosome (adapted from (Klockgether* et al., *2010)). *Homologs of open reading frames (ORFs) PA0717-PA0727 in the PAO1-UW region, RGP5, are marked with a criss cross pattern. ORFs outside of the RGP42 insertion are shown in white. Genes encoding phage proteins are coloured in grey (integrase, int; coat proteins, coaAB. ORFs encoding serine/threonine phosphotransferases, which are missing in PAO1-UW, are shown in black.*

It was observed that each sequenced strain/subline had a similar number of SNPs in RGP42 at base quality thresholds Q10 and Q15 (Figure 2.8). These SNPs were found to be located between the second and twelfth open reading frame (ORF) of RGP42, which strikingly corresponds to the region which is homologous to RGP5. For each base quality threshold, SNPs were located in the same position in each strain/subline (Figure 2.10).

Figure 2.10: Location of SNPs (base quality threshold Q10) in RGP42 in the sequenced strains and sublines. *Open reading frames are shown in light blue, and the annotation of the RGP42 island is shown in yellow. The tRNA-Met PA4673.1 is shown in pink, and the corresponding phage integration attachment site (attP) is shown in dark blue. SNPs are shown in red, and each row of SNPs corresponds to a single strain/subline (labelled). The region homologous to RGP5 is marked. Coordinates are with respect to the PAO1-UW reference sequence with the inversion and RGP42 insertion, as described in Section 2.2.5.*

In the original study carried out by (Klockgether *et al.*, 2010), the RGP42 region was assembled using reads which did not map to the PAO1-UW genome sequence. This assembly was then extended using reads which did map to the PAO1-UW genome sequence. In order to investigate if the high SNP density in RGP42 was an artefact of the genome assembly method, the reads from the original dataset of (Klockgether *et al.*, 2010), which were publicly available at the Sequence Read Archive (SRA), were mapped to both the PAO1-UW genome sequence and the RGP42 sequence, as described in Section 2.2.6.

After aligning the reads of the original dataset to the PAO1-UW sequence, the dataset was divided into reads which mapped to the reference sequence and reads which did not map. These two subsets of data were then mapped separately to the RGP42 sequence (Figure 2.11).

Reads which did not map to PAO1-UW were found to map across the entire length of the RGP42 sequence, although at a lower read coverage between ORFs 2 and 12. Reads which mapped to the PAO1-UW reference sequence were found to map to the RGP42 sequence with greater coverage than reads which did not map to the reference between the second and twelfth ORF of RGP42. This region is homologous to RGP5, and is where SNPs were found in the sequenced strains and sublines. These reads were found to map to the RGP5 region of PAO1-UW (Figure 2.12A), specifically at the region between genes PA0717 and PA0727 (Figure 2.12B). This indicates that the SNPs in the sequenced strains and sublines are likely to correspond to the actual sequence of RGP42, and the reference base at the SNP positions is likely to correspond to the sequence within the RGP5 duplication.

Alignment and comparison of the RGP42 and RGP5 sequences revealed that the reference bases of RGP42 at SNP positions were identical to the corresponding base in the alignment in RGP5. SNPs which were called in

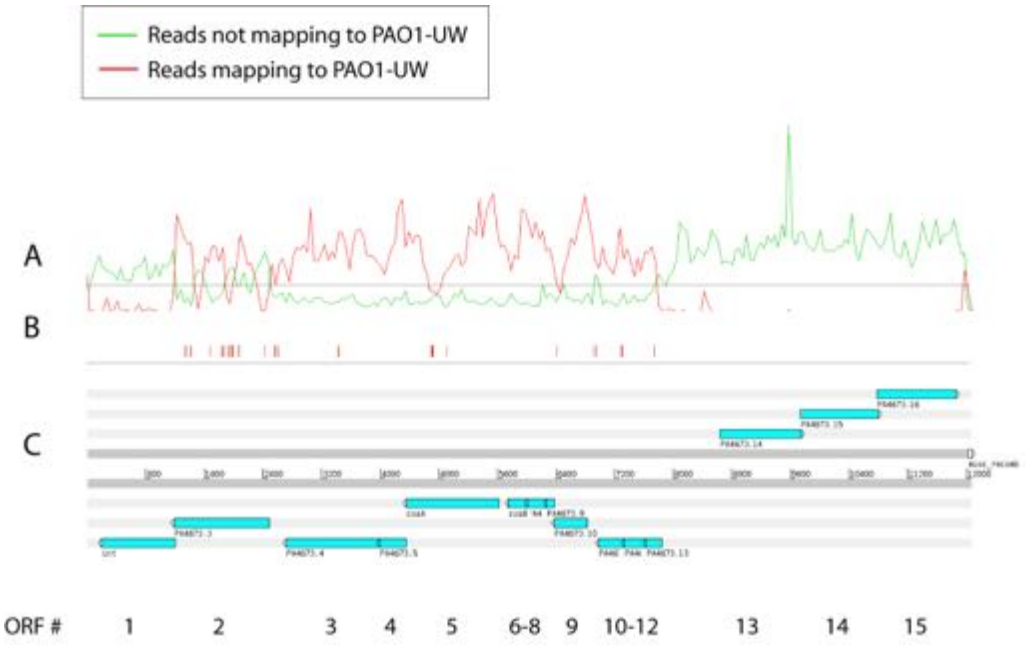

Figure 2.11: Mapping of MPAO1 and PAO1-DSM reads from (Klockgether* et al.*, 2010) dataset to the RGP42 sequence. ***A****: Coverage plot of mapped reads. Coverage of reads which mapped to the PAO1-UW reference sequence are shown in red, and coverage of reads which did not map to the PAO1-UW reference sequence are shown in green.* ***B****: SNPs in RGP42 in all strains and sublines sequenced in this study.* ***C****: Coding sequences in the RGP42 sequence (cyan blue).*

all sequenced strains and sublines by the variant caller, of which there were 39 SNP loci, were confirmed by PCR amplification and subsequent Sanger sequencing of corresponding regions in PAO1-DSM and MPAO1 (Table 2.5). In all cases the SNPs in RGP42 were confirmed by PCR, indicating that these correspond to the correct RGP42 sequence.

There were an additional three SNP locations where variants were not called in all strains/sublines, although manual inspection of the reads mapping to these regions in Artemis seemed to indicate a SNP in all strains/-sublines at these three positions. These SNPs were confirmed to exist in PAO1-DSM and MPAO1 by PCR and subsequent Sanger sequencing (Table 2.6). A further three SNP positions were confirmed to exist in PAO1-DSM and MPAO1 by PCR, and were not called by the variant caller in any of the sequenced strains or sublines (Table 2.7).

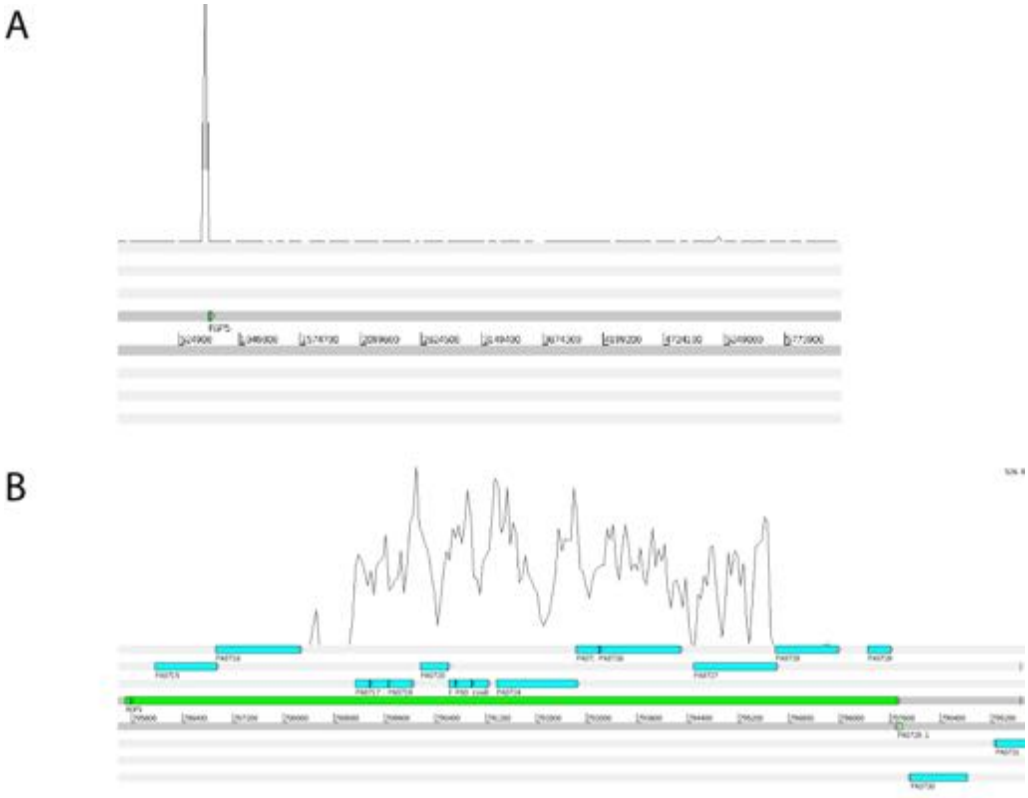

Figure 2.12: Mapping position of reads which mapped to both PAO1-UW and RGP42 in the PAO1-UW sequence. *A: Reads which map to both PAO1-UW and RGP42 were found to map to a single region of the PAO1-UW sequence. B: Zoomed in view of this region reveals that these reads map to the RGP5 region (genes PA0717 - PA0727), which is homologous to the region between the second and twelfth ORF of RGP42.*

Table 2.5: *PCR confirmed discrepancies in RGP42.*

Position in modified PAO1-UW sequence	Position in RGP42 sequence (GQ141978)	Ref. Base	Corresponding base in RGP5	PCR confirmed discrepancies
5243522	1420	A	G	G
5243698	1596	A	A	G
5243713	1611	A	A	G
5243779	1677	G	G	A
5243782	1680	T	T	C
5243787	1685	G	G	T
5244022	1920	T	T	C
5244025	1923	C	C	T
5244175	2073	G	G	A
5244202	2100	A	A	G
5244208	2106	T	T	C
5244265	2163	T	T	C
5244277	2175	A	A	G
5244313	2211	G	G	A
5244328	2226	G	G	A
5244403	2301	C	C	G
5244406	2304	G	G	C
5244408	2306	G	G	A
5244409	2307	T	T	C
5244454	2352	A	A	G
5244755	2653	T	T	A
5244875	2773	C	C	A
5244888	2786	G	G	A
5244899	2797	G	G	C
5244905	2803	A	A	G
5244921	2819	A	A	G
5245736	3634	G	C	T
5246959	4858	C	C	T
5246964	4863	G	G	A
5246969	4868	T	T	C
5247141	5040	A	A	G
5247159	5058	T	T	G
5248597	6496	C	C	G
5248598	6497	A	A	G
5248606	6505	G	G	A
5249103	7002	A	A	G
5249123	7022	T	T	G
5249468	7367	A	A	G
5249897	7796	T	T	G

Table 2.6: *SNPs in RGP42 where variants were not called in all sequenced strains and sublines.*

Position in modified PAO1-UW sequence	Position in RGP42 sequence (GQ141978)	Ref. Base	Corresponding base in RGP5	PCR confirmed SNP	Sublines and strains in which variant called
5243773	1671	G	G	A	PAO1-L, PAO2
5243774	1672	A	A	T	PAO1-L, PAO2
5247153	5052	A	A	G	PAO1-M

Table 2.7: *SNPs in RGP42 where variants were not called in any sequenced strain or subline, but were confirmed by PCR in PAO1-L.*

Position in modified PAO1-UW sequence	Position in RGP42 sequence (GQ141978)	Ref. Base	Corresponding base in RGP5	PCR confirmed SNP
5246890	4788	G	G	A
5247154	5053	T	T	C
5247156	5055	C	C	T

PCR amplification and Sanger sequencing of the region corresponding to the deletion in RGP42, as indicated by Bedtools (Table 2.4), was carried out. The deletion was determined to be located at positions 5,246,891 - 5,246,893 and was found to be preceded by a SNP at position 5,246,890.

2.3.5 Analysis of variants outside of RGP42

Upon inspection of the results of variant calling using quality score cut offs Q10, Q15 and Q20, it was found that a number of SNPs were missing in the Q20 dataset compared to Q10 and Q15 datasets. This was found to include the three SNPs in RGP42 which were not called in all datasets, but were confirmed to be true SNPs in PAO1-DSM and MPAO1 (Table 2.6). To avoid the risk of false negative SNP calls, SNPs called using the quality cut off Q10 were compared to SNPs called using the quality cut offs Q13 and Q15. A large number of SNPs were found in all three datasets, and so were not affected by the quality cut off used. Manual inspection of read alignments with Artemis supported the presence of this subset of SNPs, which consisted of 72 SNP loci, of which 39 were found in RGP42 (Table 2.5). The data for the 33 SNP loci outside of RGP42 is shown in Table 2.8.

Twelve out of the 33 SNP loci outside of RGP42 corresponded to SNPs in all ten sequenced strains and sublines, including PAO2 and PAO3. These either correspond to misassemblies in the PAO1-UW reference sequence, or to SNPs unique to PAO1-UW. In these cases is it likely that the SNPs called in the ten strains/sublines correspond to the sequence of the ancestral *P. aeruginosa* PAO. Fourteen of the SNP loci correspond to SNPs unique to individual strains/sublines. In the case of PAO1 sublines, these SNPs are likely to have occurred after the PAO1 strain was distributed. In the case of strains PAO2 and PAO3, these SNPs are likely to have occurred during the process of manganese mutagenesis. Although there are a small number of SNPs which are unique to particular strains/sublines, a large proportion

(11/14) occur within protein coding sequences and lead to an amino acid change or a premature stop codon.

A further seven SNP loci corresponded to SNPs which were shared by two or more strains/sublines, but were not conserved across all sequenced strains and sublines. Four of these SNP loci consist of SNPs shared by PAO1-N and PAO1-E. The remaining three SNP loci can be divided into two loci where SNPs were shared between PAO1-DSM and PAO2, and an additional locus where a SNP is shared between PAO1-DSM, PAO2 and PAO3. At all three of these SNP loci, PAO1-DSM appears to consist of a mixed population of wild type and mutant.

Table 2.8: SNPs in strains/sublines found using mpileup quality cut offs Q10, Q13 and Q15. *The presence of a SNP is denoted by the ● symbol. Functional annotations were obtained from the Pseudomonas database (Winsor* et al., *2011).*

Genomic position in PAO-consensus	Reference base in PAO-consensus	SNP base in PAO-consensus	Genomic position in PAO1-UW	Reference base in PAO1-UW	SNP base in PAO1-UW	Reference codon	Mutated codon	Amino acid change	Gene ID	Gene Name	PAO1-DSM	PAO1-L	PAO1-M	PAO2	PAO3	MPAO1	PAO1-N	PAO1-W	PAO1-E	PAO1-ATCC	Functional annotation
183697	T	G	183697	T	G	TGT	TGG	C310W	PA0159	-	●	●	●	●	●	●	●	●	●	●	probable transcriptional regulator
587916	A	G	587916	A	G	-	-	-	-	-								●			-
705310	G	A	705310	G	A	GAG	AAG	E61K	PA0651	*trpC*											indole-3-glycerol-phosphate synthase
875031	C	G	4640797	G	C	CTG	GTG	L195V	PA4147	*acoR*	●	●		●		●	●	●		●	transcriptional regulator
1171562	T	C	4344266	A	G	CTT	CTC	L190L	PA3877	*narK1*	●	●		●	●	●	●	●	●	●	nitrite extrusion protein 1
1303627	T	C	4212201	A	G	CAC	CGC	H636R	PA3760	-	●	●		●	●	●	●	●	●	●	N-Acetyl-D-Glucosamine phosphotransferase system transporter
1505695	C	T	4010133	G	A	AGC	AGT	S65S	PA3578	-							●				-
2178520	G	A	3337308	C	T	GAA	AAA	E129K	PA2978	*ptpA*					●						conserved hypothetical protein
2704842	G	C	2810986	C	G	TAC	TAG	Y326*	PA2494	*mexF*									●		RND multidrug efflux transporter
2707503	G	A	2808325	C	T	GCC	GTC	A286V	PA2492	*mexT*						●	●				transcriptional regulator
2707846	A	T	2807982	T	A	TTC	ATC	F172I	PA2492	*mexT*			●	●	●	●	●	●	●	●	transcriptional regulator
2709203	G	A	2806625	C	T	GAC	AAC	D249N	PA2491	*mexS*	●	●	●	●	●	●	●	●	●	●	-
2846653	C	G	2669175	G	C	CCG	GCG	P819A	PA2400	*pvdJ*	●	●	●	●	●	●	●	●	●	●	-
3276281	A	C	2239547	T	G	-	-	-	-	-	●	●	●	●	●	●					-
3281490	A	G	2234338	T	C	CAC	CAC	Y87H	PA2042	-											probable transporter (membrane subunit)
3513875	C	T	2001953	G	A	GTC	GTT	V547V	PA1843	*metH*							●		●		methionine synthase
3799387	C	G	1716441	G	C	GCG	GCC	A211A	PA1576			●		●							probable 3-hydroxyisobutyrate dehydrogenase
3816223	C	T	1699605	G	A	CGG	TGG	R115W	PA1561	*aer*	●	●									aerotaxis receptor
3926390	C	G	1589438	G	C	GGC	GCC	G34A	PA1459	-	●	●	●	●	●	●	●	●	●	●	probable methyltransferase
4108613	G	A	1407215	C	T	GCG	GTG	A70V	PA1296	-	●	●									probable 2-hydroxyacid dehydrogenase
4729181	T	C	786647	A	G	AAA	GAA	K227E	PA0715	-				●	●	●					hypothetical protein
4869855	T	G	4869855	T	G	GAA	GAC	E158D	PA4341	-	●	●	●	●	●	●	●	●	●	●	probable transcriptional regulator
4895782	T	G	4895782	T	G	TAC	GAC	Y442D	PA4367	*bifA*							●	●			-
4992704	G	A	4992704	G	A	-	-	-	-	-					●	●					-
5036891	A	C	5036891	A	C	-	-	-	-	-	●	●	●	●	●	●	●	●	●	●	-
5041200	G	C	5041200	G	C	GAG	CAG	E266Q	PA4502	-	●	●									probable binding protein component of ABC transporter
5069340	G	T	5069340	G	T	AGC	AGA	S64R	PA4525	*pilA*			●								type 4 fimbrial precursor
5069342	T	C	5069342	T	C	AGC	GGC	S64G	PA4525	*pilA*							●		●		type 4 fimbrial precursor
5688553	C	A	5676487	C	A	GAG	TAG	E458*	PA5040	*pilQ*										●	Type 4 fimbrial biogenesis outer membrane protein precursor
5815280	G	A	5803214	G	A	GCC	ACC	A132T	PA5155	-	●	●									amino acid ABC transporter
6091288	A	G	6079222	A	G	CTA	CTG	L393L	PA5399	*dcgB*	●	●	●	●	●	●	●	●	●	●	Dimethylglycine catabolism
6110847	G	C	6098781	G	C	GGG	GGC	G586G	PA5418	*soxA*	●	●	●	●	●	●	●	●	●	●	sarcosine oxidase alpha subunit
6127521	T	G	6115455	T	G	AAA	AAC	K286N	PA5434	*mtr*	●	●	●	●	●	●	●	●	●	●	tryptophan permease

SNP calls at 23 loci were not identical at all three base quality cut offs (Q10, Q13 and Q15). Three of these were in RGP42 and are shown in Table 2.6. The remaining 20 are shown in Table 2.9. Five of these 20 SNP loci (721611-721740 in PAO-consensus, Table 2.9) were located upstream of the ribosomal operon *rrnB* in the modified PAO1-UW sequence, in a region which is >99.5% identical in all four ribosomal operons in PAO1. Due to the repetitive nature of this region, there is a chance that these SNPs are miscalled due to reads carrying sequence variants which are actually derived from one of the other three ribosomal operons. Alternatively, the SNPs at this region could be a signature of recombination at this repetitive region. The SNPs were concatenated and an attempt was made at a phylogenetic reconstruction, however due to the small number of differences between sublines a reliable phylogeny could not be obtained.

Table 2.9: SNPs in strains/sublines which were found either with mpileup quality cut off Q10, Q13 or Q15. *The presence of a SNP is denoted by the ● symbol. Columns with alternative symbols correspond to "heterozygous" variant calls predicted by Samtools mpileup: ① = C/A, ② = T/G. Functional annotations were obtained from the Pseudomonas database (Winsor* et al., *2011).*

Genomic position in PAO-consensus	Reference base in PAO-consensus	SNP base in PAO-consensus	Genomic position in PAO1-UW	Reference base in PAO1-UW	SNP base in PAO1-UW	Reference codon	Mutated codon	Amino acid change	Gene ID	Gene Name	PAO1-DSM	PAO1-L	PAO1-M	PAO2	PAO3	MPAO1	PAO1-N	PAO1-W	PAO1-E	PAO1-ATCC	Functional annotation
64783	A	G	64783	A	G	-	-	-	-	-	●			●	●						-
64793	A	G	64793	A	G	-	-	-	-	-											-
413850	T	C	413850	T	C	-	-	-	-	-	●	①	●	●		●	●	●	●	②	-
721611	C	T	721611	C	T	-	-	-	-	-	●	●	●	●	●	●	●	●	●	●	-
721622	C	T	721622	C	T	-	-	-	-	-	●	●	●	●	●	●	●		●	●	-
721718	A	G	721718	A	G	-	-	-	-	-	●	●	●	●	●	●	●		●	●	-
721725	C	T	721725	C	T	-	-	-	-	-		●		●	●	●	●				-
721740	C	T	721740	C	T	-	-	-	-	-		●		●	●	●	●				-
1066972	C	G	4448856	G	C	-	-	-	-	-	●	●		●							-
1066973	G	C	4448855	C	G	-	-	-	-	-	●	●		●							-
2708163	C	G	2807665	G	C	-	CGT	R66P	PA2492	*mexT*										●	transcriptional regulator
3659765	T	C	1856063	A	G	-	ATC	I68V	PA1710	*exsC*	●										exoenzyme synthesis protein S precursor
3659770	T	C	1856058	A	G	-	GAG	E66G	PA1710	*exsC*				●							exoenzyme synthesis protein S precursor
3957011	C	A	1558817	G	T	-	CGG	R216L	PA1430	*lasR*	●		●								transcriptional regulator
4924552	C	G	4924552	C	G	-	GTG	V178L	PA4394	*yggB*	●		●								conserved hypothetical protein
4924553	G	C	4924553	G	C	-	CCC	P177P	PA4394	*yggB*	●		●								conserved hypothetical protein
5036302	A	C	5036302	A	C	-	ACC	T20P	PA4499	-	●										-
5595558	T	G	5583492	T	G	-	ACC	T52P	PA4973	*thiC*					●						thiamin biosynthesis protein
5755527	C	G	5743461	C	G	-	ACG	T431T	PA5100	*hutU*	●	●						●			urocanase
5755528	G	C	5743462	G	C	-	ACG	T431R	PA5100	*hutU*	●	●						●			urocanase

Six of the 20 SNP loci were not present when cut offs Q13 or Q15 were applied, however from a visual inspection of the mapped reads these appear to be SNPs common to all sequenced strains and sublines (1066972-3, 4924552-3, 5755527-8 in PAO-consensus, Table 2.9). Visual inspection of the mapped reads for a further five of the 20 SNP loci revealed that these SNPs occur in regions where there are a high density of single base changes in read sequences compared to the reference (64783, 64793, 3659765, 3659770, 5595558 in PAO-consensus, Table 2.9). It is unclear if these correspond to real SNPs or whether they are due to errors in the Illumina reads.

Three of the 20 SNP loci correspond to SNPs which were not called when the Q10 or Q13 cut off was used, but were called with the Q15 cut off (2708163, 3957011, 5036302 in PAO-consensus, Table 2.9). Visual inspection of read alignment at all three of these SNPs indicates that these are all low frequency SNPs in their respective strains/sublines. One of these corresponds to a mutation in *lasR* in PAO1-DSM, which has been previously reported (Klockgether *et al.*, 2010). The final SNP locus corresponds to a SNP which was called in all strains/sublines apart from PAO3 (413850 in PAO-consensus, Table 2.9). Visual inspection of the mapped reads show a low read coverage at this region in PAO3, however it looks like this region does correspond to a SNP in PAO3. These cases remain to be resolved by PCR.

2.3.6 INDEL analysis

Short insertions and deletions (INDELs) were called using the Genome Analysis Toolkit, following INDEL realignment, using the modified PAO1-UW sequence as the reference. A total of 54 INDEL positions were called in the sequenced strains/sublines, with 36 of these corresponding to INDELs conserved between all strains/sublines (Table 2.10). Given the conservation of these INDELs between all sequenced PAO1 sublines as well as strains

PAO2 and PAO3, these INDELs are likely to correspond to the sequence of the ancestral *P. aeruginosa* PAO. Eight of these INDEL positions correspond to single base pair deletions, of which four were also predicted by the Bedtools analysis of regions of zero read coverage (Table 2.4).

18 of the 54 INDELs were not conserved between all sequenced strains and sublines (Table 2.11). Two of these INDELs (4775408 and 4788571) were conserved in all sequenced strains/sublines except PAO1-N and PAO1-E, as they were located within the 58.6 kb deletion shared between these two strains (described in further detail in Section 2.3.10). One INDEL (2708112) was predicted to be an 8 bp deletion in *mexT*, which was conserved in all strains and sublines apart from PAO1-L. Visual inspection of the mapped reads at this position appears to be in agreement with this result.

Table 2.10: INDELs conserved across all 10 sequenced strains and sublines*. Single base pair deletions also predicted by Bedtools are marked with a *. Functional annotations were obtained from the Pseudomonas database (Winsor* et al.*, 2011).*

INDEL type	Genomic position in PAO-consensus	Reference base in PAO-consensus	Indel in PAO-consensus	Genomic position in PAO1-UW	Reference base in PAO1-UW	Indel in PAO1-UW	Gene ID	Gene Name	Functional annotation
Deletion*	169283	C	-	169283	C	-	PA0149	-	
Deletion*	411125	A	-	411125	A	-	-	-	
Insertion	667028	G	GC	667028	G	GC	-	-	
Insertion	727254	T	TCC	727254	T	TCC	PA0668.5	rRNA	5S ribosomal RNA
Insertion	4539470	G	GC	976358	C	CG	PA4059	-	hypothetical protein
Insertion	3919509	G	GC	1596319	C	CG	-	-	
Insertion	3083197	C	CG	2432631	G	GC	PA2727	-	hypothetical protein
Insertion	3016846	A	AC	2498982	T	TG	PA2668	-	hypothetical protein
Insertion	2753523	T	TC	2762305	A	AG	PA2452	-	hypothetical protein
Insertion	2532047	T	TC	2983781	A	AG	-	-	
Deletion	2356685	T	-	3159143	A	-	PA2141	-	hypothetical protein
Insertion	2355772	C	CG	3160056	G	GC	PA2139	-	hypothetical protein
Insertion	2342112	C	CT	3173716	G	GA	-	-	
Insertion	2239556	C	CG	3276272	G	GC	-	-	
Insertion	2195463	T	TC	3320365	A	AG	-	-	
Insertion	2186928	G	GC	3328900	C	CG	-	-	
Insertion	2169351	A	AG	3346477	T	TC	-	-	
Insertion	1835046	T	TC	3680782	A	AG	PA1685	*masA*	enolase-phosphatase E-1
Insertion	1467485	A	AG	4048343	T	TC	-	-	
Insertion	1467484	G	GG	4048344	C	CC	-	-	
Insertion	1467483	C	CG	4048345	G	GC	-	-	
Insertion	1445358	C	CG	4070470	G	GC	-	-	
Deletion*	1440626	G	-	4075202	C	-	PA1327	-	probable protease
Deletion*	1275769	T	-	4240059	A	-	PA1174	*napA*	periplasmic nitrate reductase
Insertion	1215660	A	AG	4300168	T	TC	PA1122	-	putative peptide deformylase
Insertion	1116214	T	TC	4399614	A	AG	PA1029	-	hypothetical protein
Insertion	891101	G	GC	4624727	C	CG	-	-	
Insertion	816533	G	GC	4699295	C	CG	PA0748	-	probable transcriptional regulator
Insertion	4888194	A	AG	4888194	A	AG	PA4360	-	hypothetical protein
Insertion	5033101	G	GC	5033101	G	GC	PA4496	-	ABC transporter binding protein
Deletion	5071543	A	-	5071543	A	-	-	-	
Insertion	5230036	C	CA	5242102	C	CA	-	-	
Deletion	5234824	T	-	5246890	T	-	PA4666	*hemA*	glutamyl-tRNA reductase
Deletion	5242101	C	-	5254167	C	-	PA4673.1	-	tRNA-Met
Insertion	5472415	C	CG	5484481	C	CG	-	-	
Insertion	5655220	C	CCGG	5667286	C	CCGG	PA5024	*ytnM*	hypothetical protein

Table 2.11: *INDELs which are not conserved across all 10 sequenced strains and sublines. The presence of an INDEL is indicated by the ● symbol. Functional annotations were obtained from the Pseudomonas database (Winsor* et al., *2011).*

INDEL type	Genomic position in PAO-consensus	Reference base in PAO-consensus	Indel in PAO-consensus	Genomic position in PAO1-UW	Reference base in PAO1-UW	Indel in PAO1-UW	Gene ID	Gene Name	PAO1-DSM	PAO1-L	PAO1-M	PAO2	PAO3	MPAO1	PAO1-N	PAO1-W	PAO1-E	PAO1-ATCC	Functional annotation
Insertion	417866	G	GCCGAGC	417866	G	GCCGAGC	PA0373	*ftsY*					●		●				signal recognition particle receptor
Insertion	4539469	C	CC	976359	G	GG	PA4059	-											hypothetical protein
Deletion	4443210	G	-	1072618	C	-	PA3963	*yiiP*				●							probable transporter
Deletion	4276418	A	-	1239410	T	-	PA3820	*secF*				●							secretion protein
Deletion	2974786	G	-	2541042	C	-	PA2630	*ycfD*				●							conserved hypothetical protein
Insertion	2807991	G	GCTC	2707837	C	CGAG	PA2492	*mexT*			●	●		●	●			●	transcriptional regulator
Deletion	2807716	T	-	2708112	A	-	PA2492	*mexT*	●		●	●		●	●		●	●	transcriptional regulator
Insertion	2342111	T	TT	3173717	A	AA	-	-											
Insertion	2169349	G	GG	3346479	C	CC	-	-	●			●							
Insertion	2002203	A	AC	3513625	T	TG	PA1843	*metH*				●							methionine synthase
Deletion	1893934	C	-	3621894	G	-	PA1753	-				●							conserved hypothetical protein
Insertion	1818461	G	GA	3697367	C	CT	PA1668	*dotU2*				●			●				
Deletion	780976	G	-	4734852	C	-	-	-									●		
Insertion	740420	T	TC	4775408	A	AG	PA0683	*hxcY*	●	●		●		●		●		●	
Deletion	727257	T	-	4788571	A	-	-	-	●	●		●		●		●		●	
Insertion	5035080	C	CCGGTG	5035080	C	CCGGTG	PA4498	*mdpA*							●		●		metallo dipeptidase
Deletion	5054946	G	-	5054946	G	-	PA4514	*puuA*							●		●		probable OM receptor for Fe transport
Insertion	5824960	C	CT	5837026	C	CT	PA5173	*arcC*			●								carbamate kinase

Three INDELs were found to be conserved between PAO1-N and PAO1-E (Table 2.11). One of these (4734852) is a predicted 1 bp deletion located at the base upstream of the start of the 58.6 kb deletion in PAO1-N and PAO1-E (see Section 2.3.10), and upon observation of the mapped reads, appears to be a false positive called only in reads which span the deletion in the two sublines. The presence of the remaining two INDELs conserved in PAO1-N and PAO1-E (5035080 and 5054946) is indicated by the visual inspection of mapped reads. One INDEL was found to be common to PAO1-DSM and PAO2 (3346479), however upon inspection of mapped reads this 1 bp insertion appears to be conserved in all strains and sublines. The remaining INDELs are unique to individual strains/sublines, of which all but one occur within protein coding sequences.

2.3.7 PAO2 and PAO3 are auxotroph mutants of *P. aeruginosa* PAO

Strains PAO2 and PAO3 are thought to correspond to the serine (ser-3) and tryptophan (trp-54) auxotrophs, respectively, which were originally derived from *P. aeruginosa* PAO, as described in (Holloway, 1955; Holloway and Rolfe, 1964; Isaac and Holloway, 1968; Haas and Holloway, 1976). Whole genome sequencing revealed the presence of a deletion in the gene PA0316 in PAO2, and a non synonymous SNP in gene PA0651 in PAO3. Neither of these genetic changes were found in other chromosomal sequences, or in any other *Pseudomonas* isolate in Genbank (data not shown).

Strains PAO2 and PAO3 were determined to be unable to grow on M9 minimal medium (MM), but grew successfully on rich medium. PAO2 was able to grow on M9 MM supplemented with 3 mM and 9 mM serine (Figure 2.13A), and PAO3 was able to grow on M9 MM supplemented with 3 mM tryptophan (Figure 2.13B). PAO3 did not grow in the presence of 9 mM tryptophan, however neither did PAO1-L, which contains the wild type form

of PA0651. Therefore it appears that a high concentration of tryptophan inhibits a vital metabolic process in *P. aeruginosa* in MM. From these results it can be confirmed that PAO2 is a serine auxotroph and PAO3 is a tryptophan auxotroph.

PA0316 corresponds to the gene *serA*, which potentially encodes D-3-phosphoglycerate dehydrogenase. This enzyme catalyses the conversion of phospho-D-glycerate to 3-phosphonooxypyruvate, in a reaction which utilises the coenzyme nicotinamide adenine dinucleotide (NAD) (Tobey and Grant, 1986). The gene is part of the glycine, serine and threonine KEGG matabolic pathway, and is specifically within the pathway for the biosynthesis of serine (Figure 2.16) Initial whole genome sequence analysis indicated the presence of a 54 nucleotide deletion in gene PA0316 (Table 2.4).

The *serA* gene was amplified by PCR from chromosomal DNA of strain PAO2 and subline PAO1-L and sequenced using Sanger sequencing[4]. The sequence of the *serA* gene in PAO1-L was found to be 100% identical to the *serA* gene in PAO1-UW, while the *serA* gene in PAO2 was found to harbour a 69 bp deletion, which encompassed the putative 54 bp deletion indicated by whole genome sequencing. The deletion in *serA* in PAO2 leads to an in frame deletion causing the loss of thirteen amino acids in the SerA protein (Figure 2.14A).

In order to predict if the deletion in *serA* in PAO2 could affect protein function, the amino acid sequence of PA0316 in PAO1-UW was used a query for a BLAST search against the Protein Data Bank (http://www.rcsb.org, (Rose *et al.*, 2015)). Of the number of hits with percentage identity higher than 55%, the structure of the active form of D-3-phosphoglycerate dehydrogenase was selected (PDB identifier: 1YBA). The sequence of this protein from *E. coli* shared 61% sequence identity with the protein sequence of PA0316 from PAO1-UW. The alignment of these sequences can be seen in

[4]PCR of *serA* in PAO2 and PAO1-L was carried out by Pantelitsa Papakyriacou, University of Nottingham as part of her MSc research project

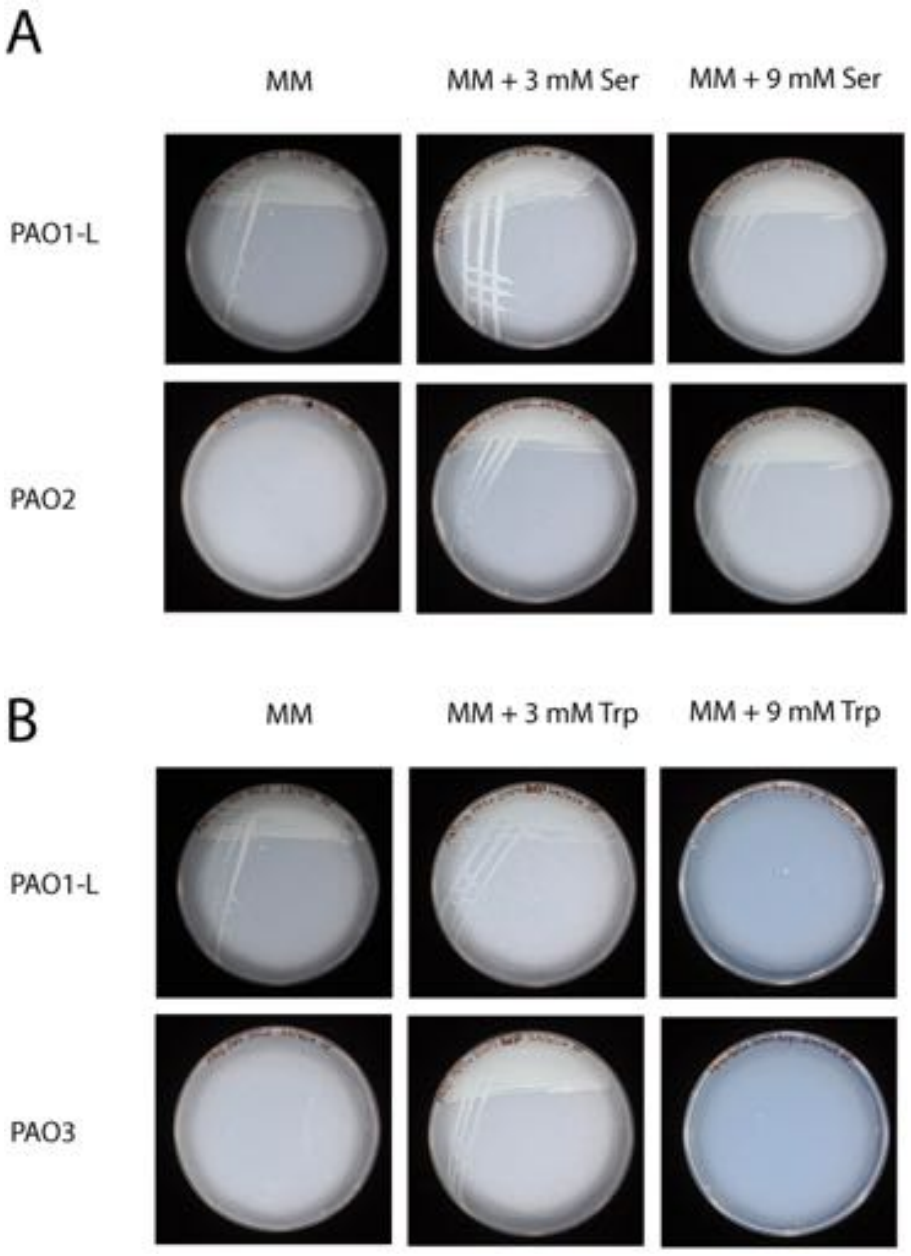

Figure 2.13: Confirmation of auxotroph phenotype of strains PAO2 and PAO3. A: *Comparison of growth of PAO1-L and PAO2 on minimal media (MM) and minimal media supplemented with 3 mM serine (MM + 3mM Ser) and 9mM serine (MM + 9mM Ser). B: Comparison of growth of PAO1-L and PAO3 on MM and minimal media supplemented with 3 mM tryptophan (MM + 3mM Trp) and 9mM tryptophan (MM + 9mM Trp).*

Figure 2.14A.

The residues corresponding to the PAO2 deletion in D-3-phosphoglycerate dehydrogenase in *E. coli* were mapped to the three dimensional protein structure, the result for which can be seen in Figure 2.14B. The PAO2 deletion residues map to part of a beta sheet, an alpha helix and a loop in the Rossmann fold of the protein. It's likely that deletion of this part of the Rossmann fold would disrupt the structure of the protein.

PA0651 corresponds to gene *trpC*, which encodes the protein indole-3-glycerol-phosphate synthase. This enzyme catalyses the conversion of 1-(o-carboxyphenylamino) 1-deoxyribulose 5-phosphate (CdRP) to indole-3-glycerol phosphate (IGP), the fifth reaction in tryptophan biosynthesis, and involves the decarboxylation and dehydration of CdRP, which is the enzyme substrate (Gerth *et al.*, 2012). The gene is part of the phenylalanine, tyrosine and tryptophan biosynthesis KEGG matabolic pathway, and is specifically within the pathway for the biosynthesis of tryptophan (Figure 2.17). Whole genome sequence analysis revealed the presence of a non synonymous SNP in the *trpC* gene in PAO3 at position 705,310 with reference to the modified PAO1-UW sequence, which would lead to the mutation E61K in the protein sequence. The SNP in *trpC* in PAO3 was confirmed by PCR amplification of PA0651 from PAO3 and PAO1-L chromosomal DNA and subsequent Sanger sequencing [5]. The sequence of the *trpC* gene in PAO1-L was found to be 100% identical to the *trpC* gene in PAO1-UW. The alignment of the amino acid sequences of *trpC* from PAO1-UW and PAO3 is displayed in Figure 2.15A.

In order to predict if the SNP in *trpC* in PAO3 could affect protein function, the amino acid sequence of PAO651 from PAO1-UW was used as a query for a BLAST search against the Protein Data Bank (http://rcsb.org, (Rose *et al.*, 2015)). The resulting protein structure with the highest per-

[5]PCR of *trpC* in PAO3 and PAO1-L was carried out by Pantelitsa Papakyriacou, University of Nottingham as part of her MSc research project

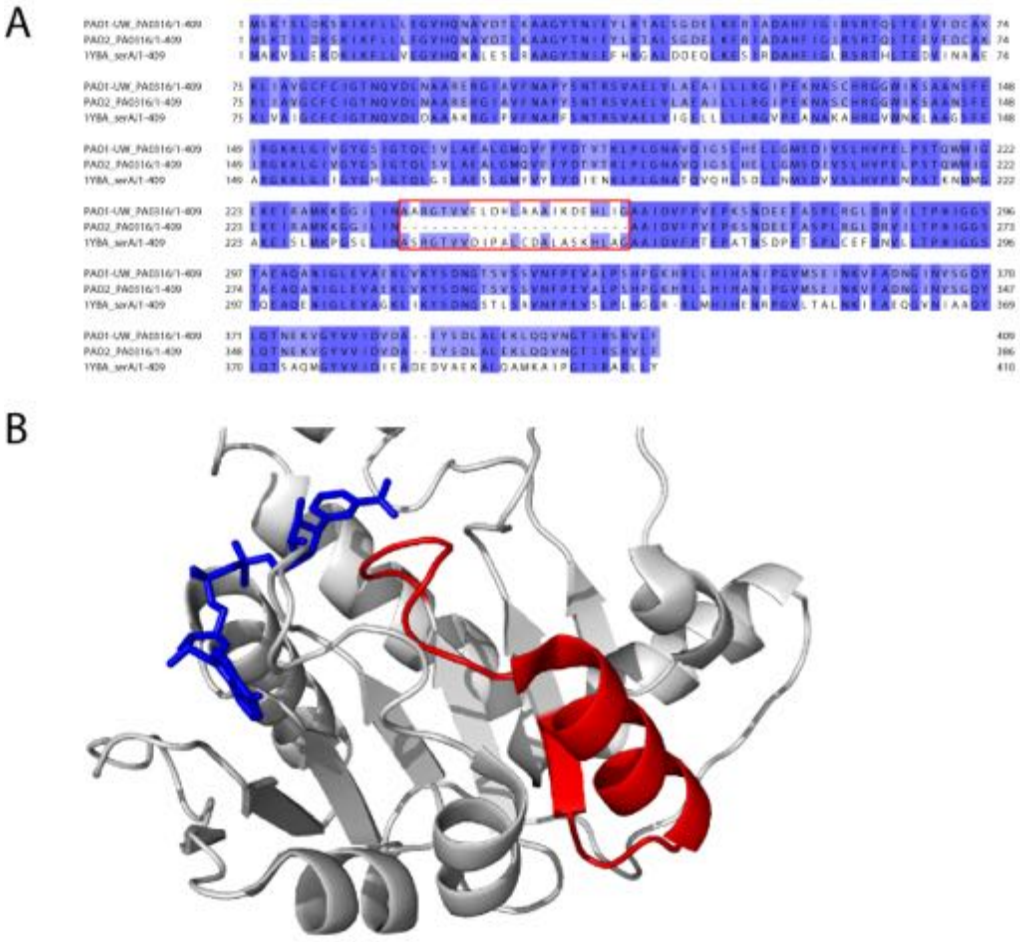

Figure 2.14: Location of the PAO2 deletion in PA0316 in the protein sequence and three dimensional protein structure. A: *Multiple sequence alignment of PA0316 protein sequence in PAO1-UW (wild type), PAO2 (auxotroph) and E. coli (PDB accession number: 1YBA). The alignment is coloured according to percentage sequence identity. The region of the alignment corresponding to the 13 amino acid deletion in PA0316 in PAO2 is marked in red.* ***B:*** *Residues in the protein structure of wild type PA0316 in E. coli (1YBA) contained within the PAO2 deletion. The region of the protein which is deleted in PAO2 is shown in red, NAD is shown in blue.*

centage sequence identity to the sequence from PAO3, as well as a crystal structure with bound ligand, was determined to be the crystal structure of indole-3-glycerol phosphate synthase from *Mycobacterium tuberculosis* (PDB identifier: 3T44). The protein from *M. tuberculosis* shared 40% sequence identity with *trpC* from PAO1-UW. The alignment of these sequences can be seen in Figure 2.15A.

The residue at which there is a SNP in *trpC* in PAO3 was mapped onto the three dimensional protein structure (Figure 2.15B), and was found to be interacting with the co-crystallised product, IGP, via a hydrogen bond in the crystallised protein from *M. tuberculosis*.

In order to verify that the SNP in PA0615 in PAO3 and the deletion in PA0316 in PAO2 lead to the auxotroph phenotypes, these mutants were transformed with the expression vector pME6032 (Heeb *et al.*, 2000), into which the wild type versions of each corresponding gene (from PAO1-L) were cloned. The transformed strains were then grown on minimal media to test for the restoration of the prototroph phenotype [6]. The expression of the wild type gene in each auxotroph mutant led to the restoration of the prototroph phenotype, meaning that the deletion in *serA* in PAO2 and the SNP in *trpC* in PAO3 are the causes of the serine and tryptophan auxotrophy phenotypes, respectively (Figure 2.18).

2.3.8 Random amplified polymorphic DNA analysis

Random amplified polymorphic DNA analysis (RAPD) was carried out on nine of the ten sequenced *P. aeruginosa* strains and sublines 2.12 using two different primers, to see if the genomic changes in these strains and sublines affect the observed banding pattern. The clinical strain for which the RAPD profile is known, *P. aeruginosa* Nottingham K13, was used as a

[6] Cloning and transformation of PA0316 and PA0651 into PAO2 and PAO3, respectively, was carried out by Pantelitsa Papakyriacou, University of Nottingham as part of her MSc research project

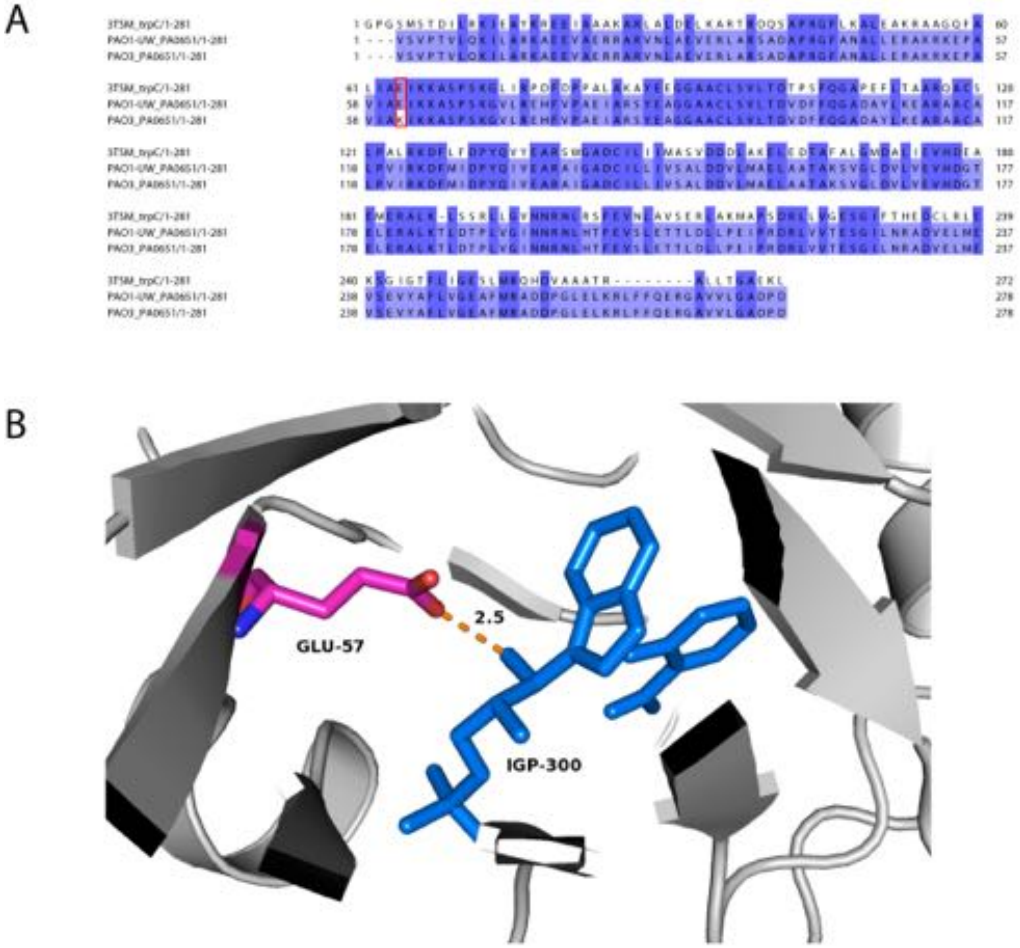

Figure 2.15: Location of the PAO3 deletion in PA0651 in the protein sequence and three dimensional protein structure. A: *Multiple sequence alignment of PA0651 protein sequence in PAO1-UW (wild type) and in PAO3 (auxotroph). The alignment is coloured according to percentage sequence identity. The region of the alignment corresponding to the amino acid change in PA0651 (E61K) in PAO3 is marked in red.* ***B:*** *The interaction between glutamine 61 of indole-3-glycerol-phosphate synthase in PAO3 (which corresponds to glutamine 57 in the equivalent M. tuberculosis protein, PDB accession number: 3T44) and the co-crystallised product, indole-3-glycerol phosphate (IGP). Glutamine 57 is shown in pink, and IGP molecule is shown in blue. The hydrogen bond between glutamine 57 and IGP is shown by an orange dashed line.*

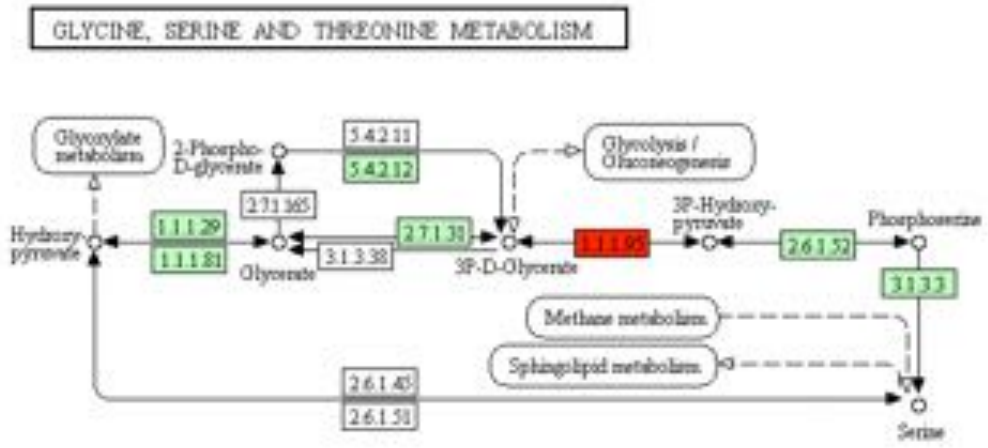

Figure 2.16: Part of KEGG metabolic pathway involving gene PA0316. *Gene PA0316 is shown in red in the serine metabolism part of the glycine, serine and threonine metaboilsm pathway (KEGG reference: pae00260)*

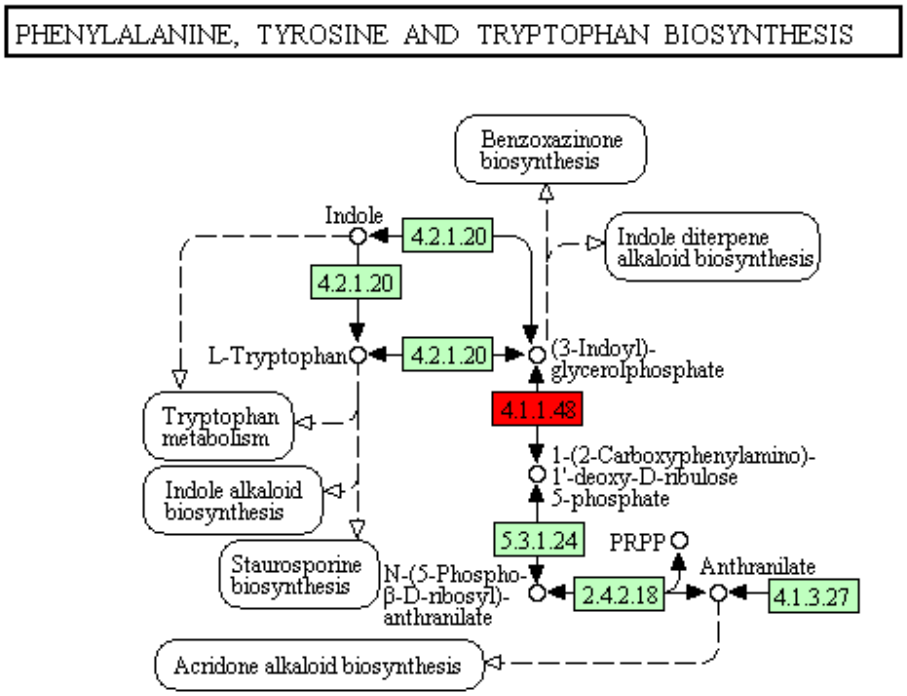

Figure 2.17: Part of KEGG metabolic pathway involving gene PA0651. *Gene PA0651 is shown in red in the tryptophan metabolism part of the phenylalanine, tyrosine and tryptophan biosynthesis pathway (KEGG reference: pae00400)*

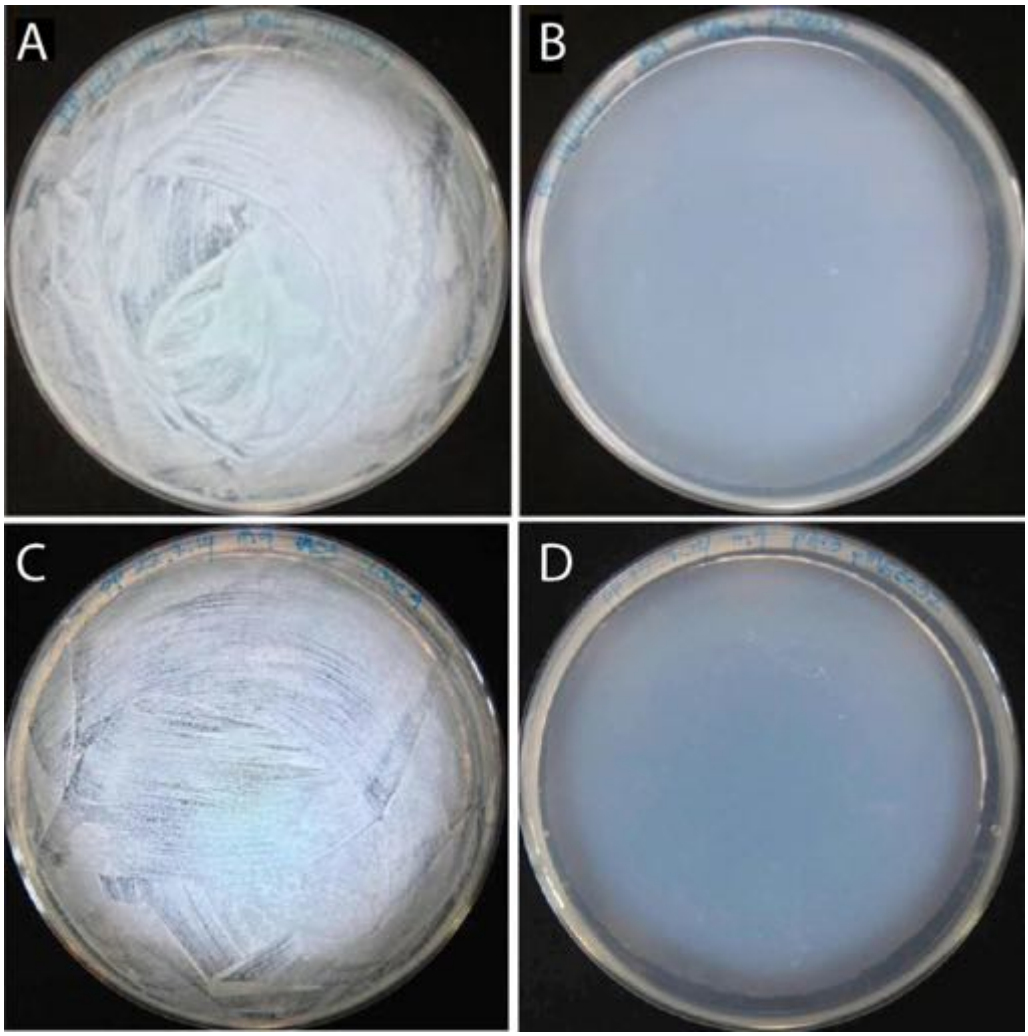

Figure 2.18: A deletion in PA0316 (serA) and a mutation in PA0651 (trpC) cause serine and tryptophan auxotrophy in strains PAO2 and PAO3, respectively. *Results of transformation of PAO2 ($serA^-$) with plasmid vector (pME6032) containing wild type serA from PAO1-L and PAO3 ($trpC^-$) with plasmid vector containing wild type trpC from PAO1-L. All strains are grown on M9 minimal media agar.* ***A****: PAO2 transformed with vector containing wild type serA,* ***B****: PAO2 transformed with empty vector,* ***C****: PAO3 transformed with vector containing wild type trpC,* ***D****: PAO3 transformed with empty vector.*

Table 2.12: *Strains and sublines used in RAPD analysis*

Sample Number	Sample ID
1	PAO1-N
2	PAO1-L
3	MPAO1
4	PAO1-DSM
5	PAO1-M
6	PAO1-E
7	PAO1-W
8	PAO2
9	PAO3
10	K13 (Positive control)

positive control. All nine strains/sublines show identical banding patterns with each set of primers (Figure 2.19), but show a different banding pattern from the control, clinical isolate *P. aeruginosa* Nottingham K13[7]. This method therefore does not discriminate the PAO1 sublines.

2.3.9 PAO1-UW contains a unique 2.2 Mb inversion

It has been previously reported by (Stover *et al.*, 2000) that PAO1-UW contains a 2.2 Mb inversion between the almost identical ∼5.2 kb ribosomal operons *rrnA* and *rrnB* compared to the physical map of PAO1, as well as PAO1-DSM. The inversion has been visualised previously using pulsed field gel electrophoresis in PAO1-DSM (Klockgether *et al.*, 2010), where it was found that this PAO1 subline does not contain the inversion found in PAO1-UW. A large chromosomal inversion was also previously reported in the sequence of PA14 compared to that of PAO1-UW (Lee *et al.*, 2006). Optical mapping was used in this study to see if the inversion was a unique feature of PAO1-UW, or whether large inversions are a regular occurrence in commonly used laboratory sublines of PAO1, and to determine if this feature was likely to occur in the ancestral *P. aeruginosa* PAO sequence.

[7]Strain obtained from Zoe Aiken, Clinical Microbiology, Queen's Medical Centre, Nottingham

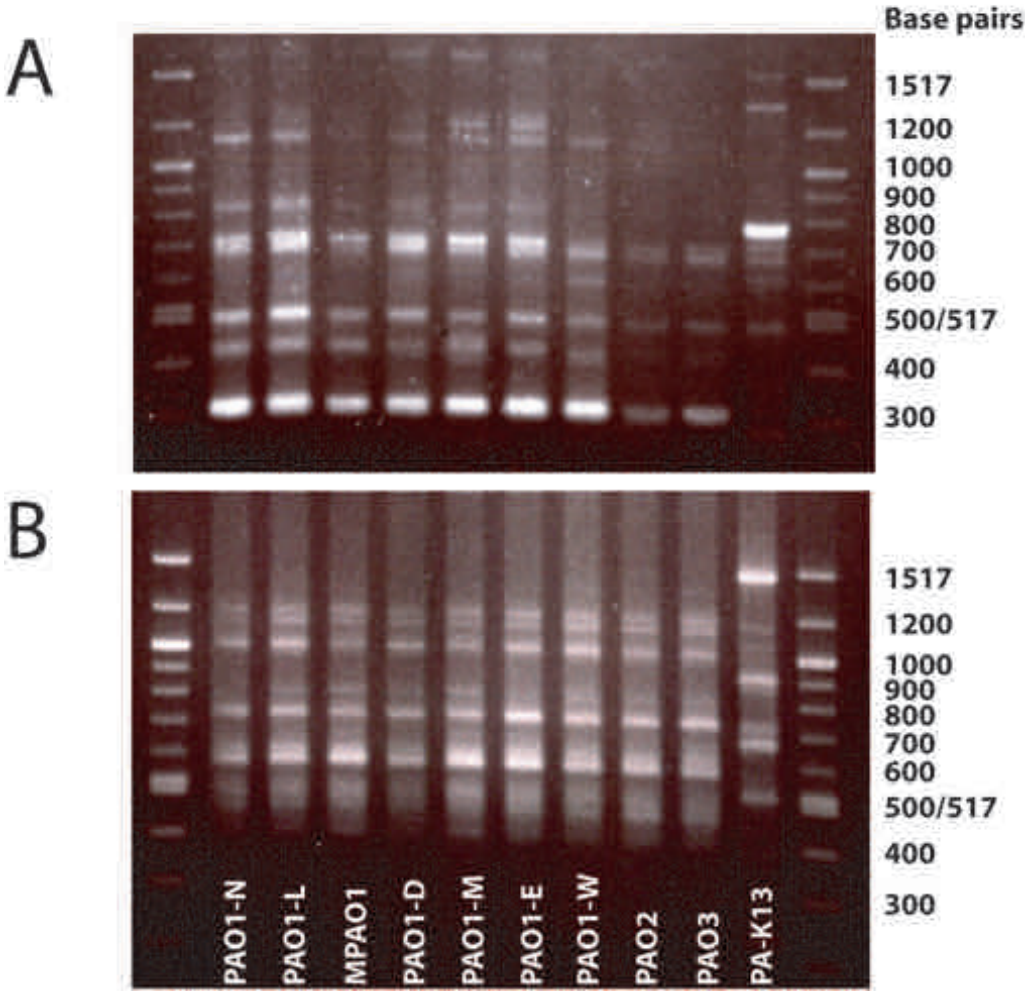

Figure 2.19: ***Gel electrophoresis of RAPD fragments with P. aeruginosa primer A: D1054 and B: 272***

Optical mapping revealed that all ten strains and sublines, including PAO2 and PAO3 did not contain the 2.2 Mb inversion found in the PAO1-UW reference sequence (Figure 2.20A). The inversion between PA14 and PAO1-UW was found to correspond to the 2.2 Mb inversion described by Stover *et al.* (2000), and the PAO1 isolates in this study were found to be more similar to PA14 than to PAO1-UW with respect to the inversion. The inversion in PAO1-UW is asymmetrical with respect to the origin of replication, resulting in it no longer being opposite to the *dif* replication termination sequence (Rocha and Danchin, 2003; Mathee *et al.*, 2008) in the chromosome (Figure 2.20B), as is seen in previously sequenced *P. aeruginosa* chromosome sequences (Rau *et al.*, 2012; Winstanley *et al.*, 2009; Lee *et al.*, 2006; Roy *et al.*, 2010). This leads to a change in the GC skew of the PAO1-UW chromosome compared to the chromosome of strains and sublines without the inversion (Figure 2.21), which is expected given the change of position of the origin of replication with respect to the replication termination point (Grigoriev, 1998).

2.3.10 PAO1-N and PAO1-E share an identical 58.6 kb deletion

Upon examining genomic regions where there were no reads mapping in each sequenced strain and subline, it was determined that there is a large putative deletion in PAO1-N and PAO1-E which is not found in any of the other PAO1 sublines in this study. According to whole genome sequencing, the putative deletion occurs between genes PA0707 and PA0669, and contains 39 ORFs, including that of the transcriptional regulator *toxR* and the Cm resistance gene *cat* (Figure 2.22A). *toxR* is part of the *toxA*/*toxR* toxin-antitoxin system, where *toxA* encodes exotoxin A and *toxR* regulates the expression of *toxA*.

This deletion was also confirmed by whole genome sequencing of the

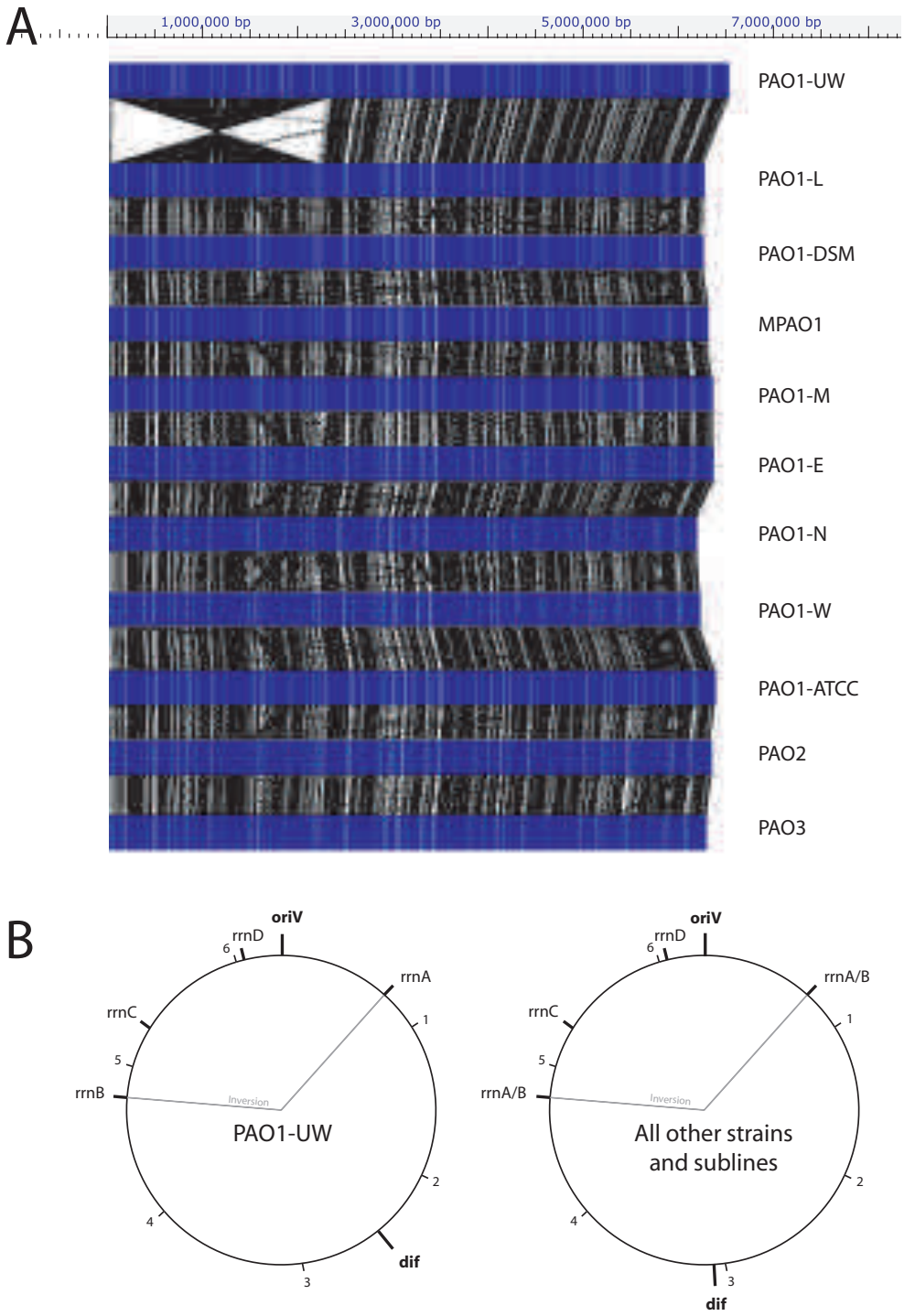

Figure 2.20: The 2.2 Mb inversion is a unique feature of the PAO1-UW sequence. A: *Alignment of the in silico optical map of PAO1-UW reference sequence (Genbank Accession Number: NC_002516) with the optical maps of all strains and sublines. Optical maps are shown in blue, and the alignment of restriction enzyme cut sites between optical maps are shown in black.* ***B:*** *Schematic of genomic orientation in PAO1-UW and strains/sublines without the inversion. oriV = origin of replication, dif = dif termination sequence, rrnA-D = ribosomal operons A-D. The coordinates of the inversion are shown in grey. The marked scale is in Mb.*

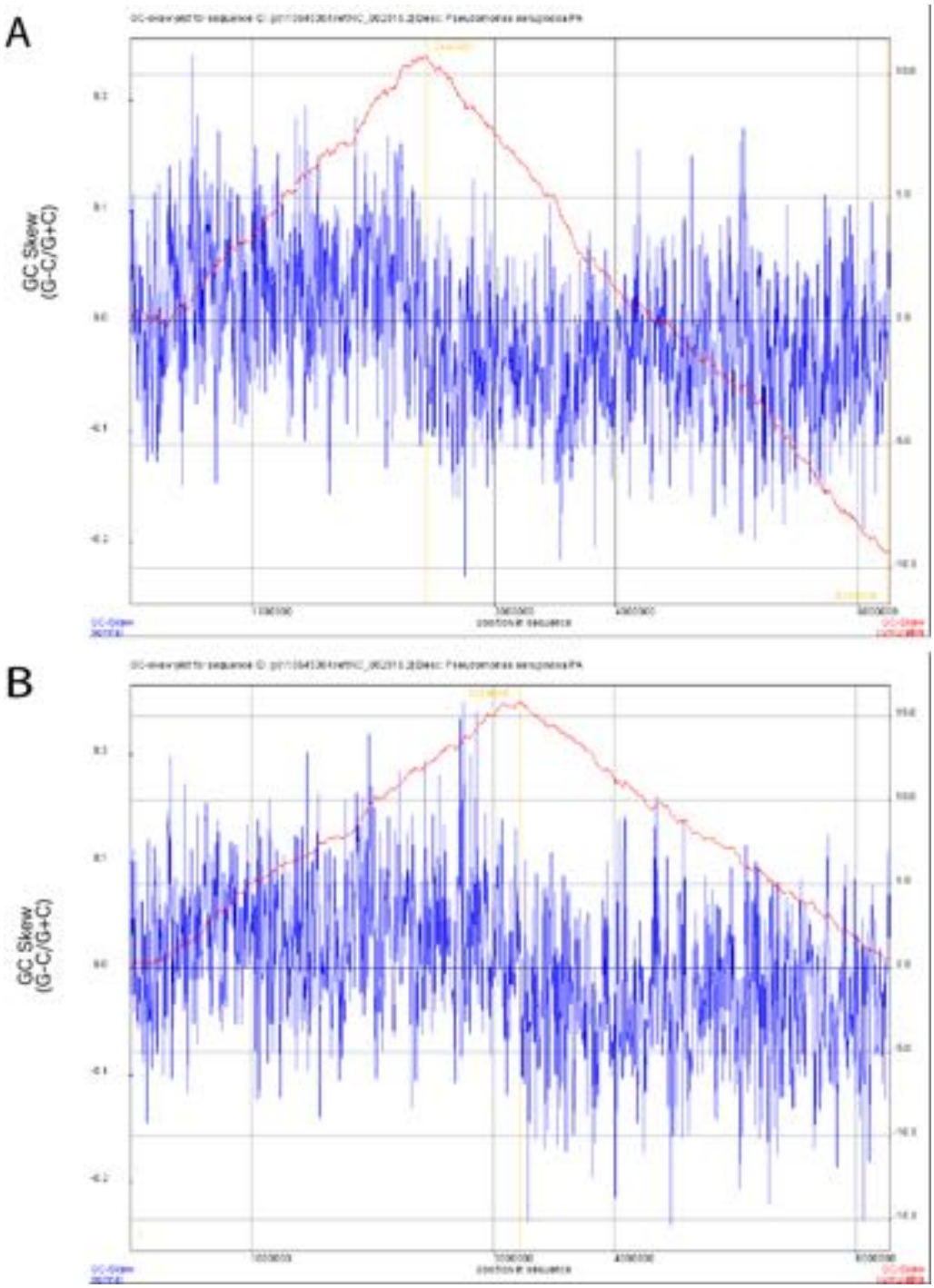

Figure 2.21: Analysis of GC skew of PAO1-UW and that of other PAO1 sublines. *GC skew plots for* ***A****: PAO1-UW and* ***B****: modified PAO1-UW sequence are shown. The GC skew plot is shown in blue and the cumulative GC skew plot is shown in red. The predicted origin and termination points of replication are shown in yellow. The origin of replication for the modified PAO1-UW sequence was predicted to occur at position 1 in the sequence. The GC skew is calculated using the formula (G-C/G+C).*

PAO1-N derived mutant PAN067, the first *P. aeruginosa* construct published at the University of Nottingham (Latifi *et al.*, 1995). Optical mapping confirmed a deletion in this region in both PAO1-N and PAO1-E compared to the reference sequence (Figure 2.22B).

In order to see how conserved the deleted region is in *Pseudomonas* species, a NCBI nucleotide BLAST search was carried out against the nucleotide sequence database nt. The resulting distance tree of BLAST hits is shown in Figure 2.23. All hits with sequence similarity covering >20% of the query sequence corresponded to *Pseudomonas* species. All hits with >=73% query coverage were strains of *P. aeruginosa*, which could be further divided into those with 73-78% query coverage of the deleted sequence, which included strains PA7 and PA14, and those with 92-99% query coverage, which included the Liverpool Epidemic Strain (LES) strains and the Danish CF isolate DK2. Hits from four other *Pseudomonas* species are shown in Figure 2.23, all of which share >= 80% sequence identity with the deletion region over 29-35% of the sequence. The deleted region was checked for phage sequences at the Phast server, which did not detect any intact, incomplete or questionable phage sequences. Based on the BLAST and Phast results it appears that the deletion is a unique feature of *Pseudomonas* species, and the deleted region shows varying conservation across *P. aeruginosa* strains.

PCR was carried out using primers which were designed to span the putative deletion region, in order to determine its exact location (PAO1_DSF, PAO1_DSR and PAO1_DWR in Table 2.2). In concordance with the whole genome sequencing data, a PCR product was not obtained in PAO1-L, as it would be >58 kb in size, and identical PCR products were obtained for PAO1-N and PAO1-E (Figure 2.24A). Based on sequence analysis of the PCR products from PAO1-N and PAO1-E it was determined that both of these sublines contain an identical 58.6 kb deletion between genes PA0707, which encodes the gene *toxR*, and PA0668.1, a 16S ribosomal RNA gene

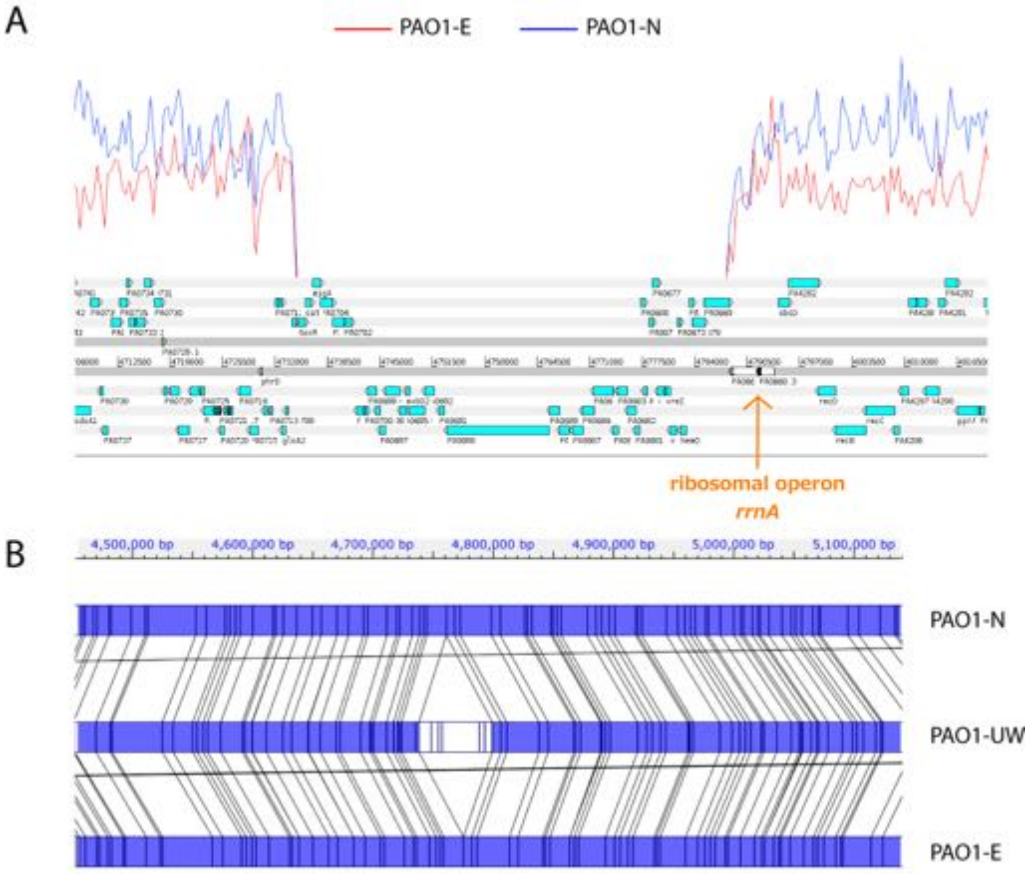

Figure 2.22: Whole genome sequence and optical mapping indicate the presence of a large deletion in PAO1-E and PAO1-N. A: *Mapping of whole genome sequencing reads from PAO1-N (blue) and PAO1-E (red) to the modified PAO1-UW reference sequence reveal a large area where no reads map, corresponding to a putative deletion in these sublines. Coding sequences are shown in cyan blue, and the ribosomal operon rrnA is shown in white, and are labelled in orange.* ***B:*** *Comparison of optical maps of PAO1-E and PAO1-N with the in silico optical map of the modified PAO1-UW sequence reveal a large deletion, at the same location as that determined by whole genome sequencing. Maps of PAO1-E and PAO1-N are shown in blue, and PAO1-UW in red. The lines between the optical maps correspond to alignments between restriction sites. The white region in PAO1-UW corresponds to the region which is not found in both PAO1-N and PAO1-E, which corresponds to the deletion. The scale corresponds to the genomic coordinates of the modified PAO1-UW sequence.*

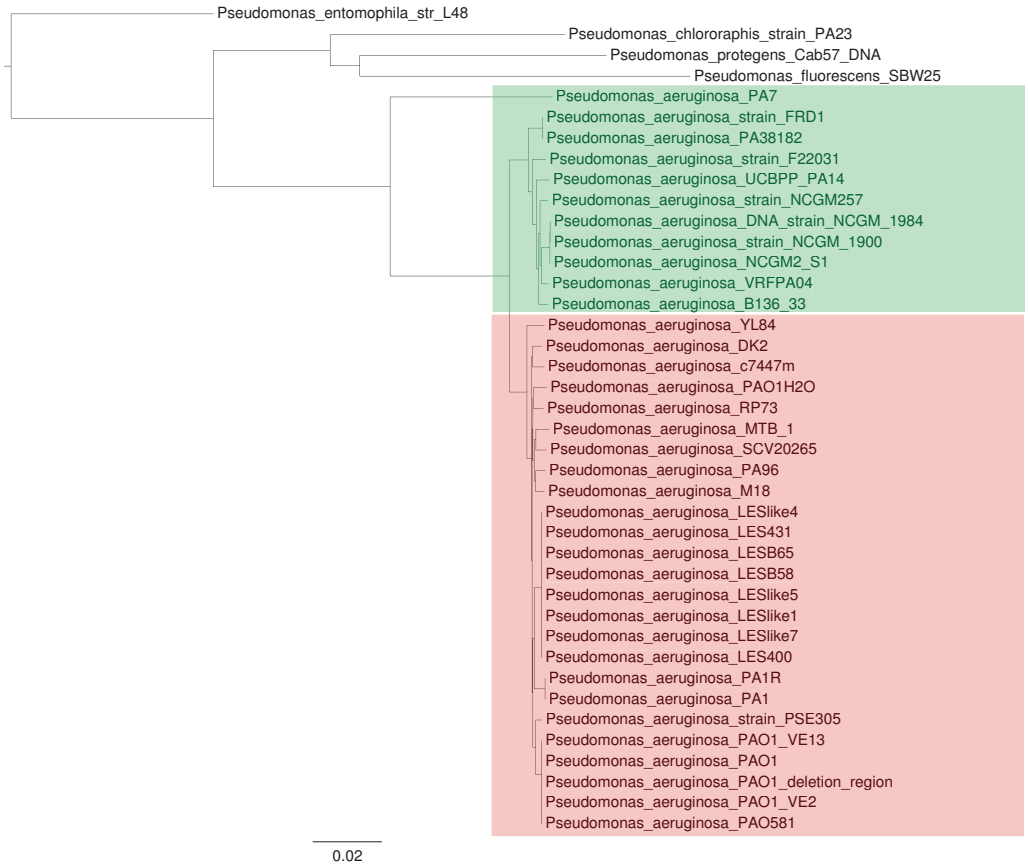

Figure 2.23: Unrooted distance tree of BLAST hits to the 58.6 kb sequence deleted in PAO1-N and PAO1-E. *P. aeruginosa strains with 99-100% identity over 98-100% of the query sequence are highlighted in pink. Strains with 92-99% identity over 73-78% of the query sequence are highlighted in green.*

(Figure 2.24B).

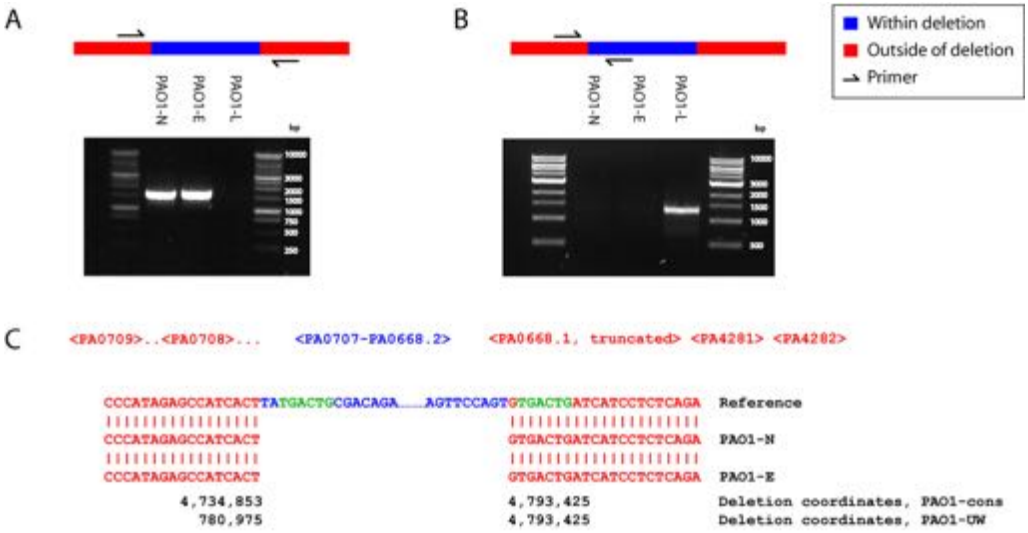

Figure 2.24: PCR confirmation of the 58.6 kb deletion in PAO1-E and PAO1-N. *Visualisation of PCR products by gel electrophoresis, using* ***A****: primers designed to span the deletion, reveals a product in PAO1-N and PAO1-E but not in PAO1-L.* ***B****: Forward primer outside the deletion, reverse primer within the deletion, reveals a product in PAO1-L but not in PAO1-N and PAO1-E.* ***C****: Sanger sequencing of the PCR products spanning the deletion from PAO1-N and PAO1-E confirms that both of these sublines share an identical 58.6 kb deletion between genes PA0707-PA0669.1. A 6 bp repeat sequence present at the start and end of the deletion is marked in green.*

Inspection of the sequence at either side of the deletion revealed a 6 bp repeat sequence (TGACTG) at the start and at the end of the deletion (Figure 2.24). This short sequence occurs multiple times in the genome sequence, including seven additional times within the deleted region itself. However, the placement of this sequence before and after the deletion, with only one copy remaining once the deletion occurred could be indicative of a deletion caused by homologous recombination. A NCBI nucleotide BLAST of the sequence flanking the deletion was carried out to see if the deletion occurs in any other organism. It was found that there are no other sequenced isolates with the same deletion, so it appears to be a unique feature of PAO1-N and PAO1-E.

Table 2.13: Genes within the 58.6 kb deletion in PAO1-N and PAO1-E. *Annotations are taken from the Pseudomonas Genome Database, apart from those marked with an asterisk, where product name annotations are taken from (Llamas* et al.*, 2009).*

Locus tag	Feature type	Name	Product name
PA0668.1	rRNA		16S ribosomal RNA
PA0668.2	tRNA		tRNA-Ile
PA0668.3	tRNA		tRNA-Ala
PA0668.4	rRNA		23S ribosomal RNA
PA0668.5	rRNA		5S ribosomal RNA
PA0669	CDS		probable DNA polymerase alpha chain
PA0670	CDS		hypothetical protein
PA0671	CDS		hypothetical protein
PA0672	CDS	*hemO*	heme oxygenase
PA0673	CDS		hypothetical protein
PA0674	CDS	*vreA*	ECF sigma factor, transcriptional regulator
PA0675	CDS	*vreI*	ECF sigma factor, transcriptional regulator
PA0676	CDS	*vreR*	sigma factor regulator, VreR
PA0677*	CDS	*hxcW*	putative pseudopilin
PA0678*	CDS	*hxcU*	putative pseudopilin
PA0679*	CDS		probable hcx type II secretion system protein
PA0680*	CDS	*hxcV*	putative pseudopilin
PA0681*	CDS	*hxcT*	pseudolipin
PA0682*	CDS	*hxcX*	atypical pseudolipin
PA0683*	CDS		probable hcx type II secretion system protein
PA0684*	CDS		probable hcx type II secretion system protein
PA0685*	CDS		probable hcx type II secretion system protein
PA0686*	CDS		probable hcx type II secretion system protein
PA0687*	CDS		probable hcx type II secretion system protein
PA0688*	CDS		alkaline phosphatase
PA0689	CDS		hypothetical protein
PA0690*	CDS	*pdtA*	putative two partner secretion system
PA0691*	CDS	*phdA*	putative transposase
PA0692*	CDS	*pdtB*	putative two partner secretion system
PA0693	CDS	*exbB2*	transport protein ExbB2
PA0694	CDS	*exbD2*	transport protein ExbD2
PA0695	CDS		hypothetical protein
PA0696	CDS		hypothetical protein
PA0697	CDS		hypothetical protein
PA0698	CDS		hypothetical protein
PA0699	CDS		probable peptidyl-prolyl cis-trans isomerase, PpiC-type
PA0700	CDS		hypothetical protein
PA0701	CDS		probable transcriptional regulator
PA0702	CDS		hypothetical protein
PA0703	CDS		probable major facilitator superfamily (MFS) transporter
PA0704	CDS		probable amidase
PA0705	CDS	*migA*	alpha-1,6-rhamnosyltransferase MigA
PA0706	CDS	*cat*	chloramphenicol acetyltransferase
PA0707	CDS	*toxR*	transcriptional regulator ToxR

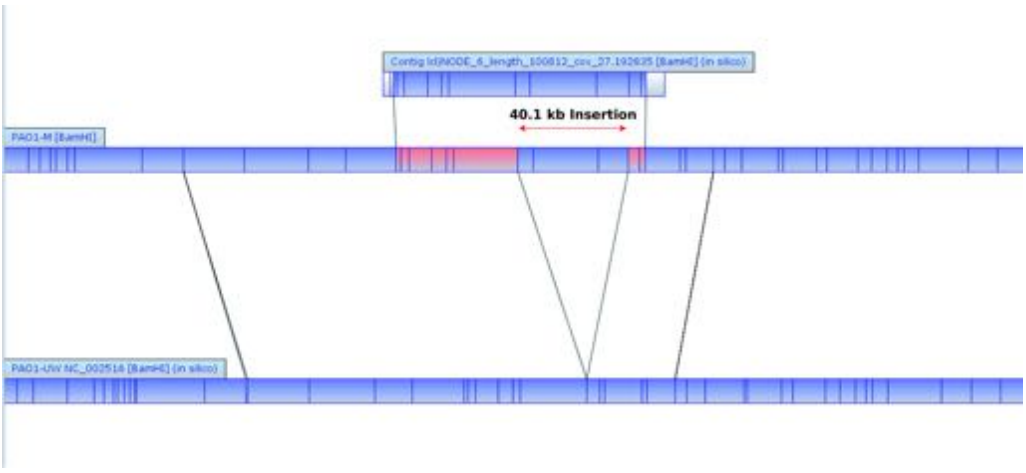

Figure 2.25: Optical mapping of PAO1-M indicates the presence of an insertion compared to PAO1-UW. *The bottom track shows the in silico optical map of PAO1-UW (NC_002516), the middle track shows the optical map of PAO1-M, and the top track shows the in silico optical map of the de novo assembly contig which maps to the insertion region in PAO1-M. The insertion is indicated by the red arrow.*

2.3.11 Novel prophage sequence in PAO1-M

Optical mapping of *P. aeruginosa* PAO1-M revealed a 40.1 kb insertion, which was found to be unique to this subline. Alignment of the contigs generated from *de novo* assembly of PAO1-M to the optical map revealed a contig of length 100 kb which spanned the insertion (Figure 2.25). The sequence of this contig was compared with the genomic sequence of PAO1-UW in order to pinpoint the location of the insertion.

The 40.1 kb insertion in PAO1-M was found to occur within gene PA3620, which encodes the DNA mismatch repair protein MutS (Figure 2.26). The insertion in PAO1-M occurs after the 40th nucleotide of the PA3620 coding sequence, and results in a truncated form of PA3620 at the 5' end of the insertion due to the introduction of a premature stop codon. This sequence has been designated PA3620.1 (Figure 2.26A). The sequence at the 3' end of the insertion is almost identical to that of the complete PA3620, with the exception of the first few amino acids. This sequence has been designated PA3620.2 (Figure 2.26A).

A nucleotide BLAST search search of the 40,144 bp insertion sequence

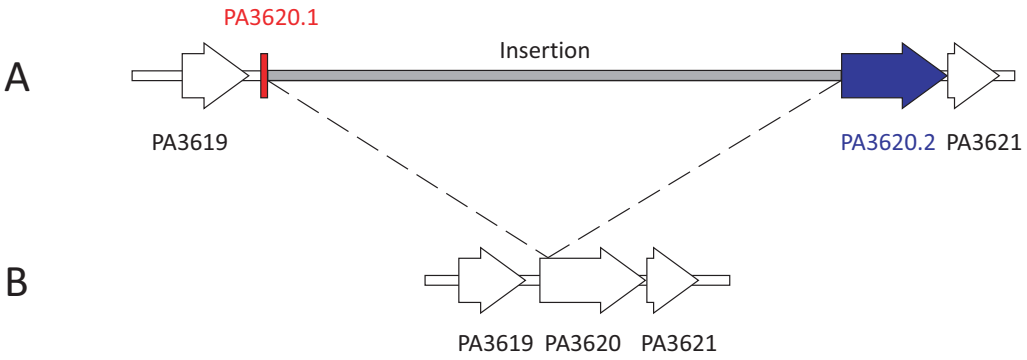

Figure 2.26: Insertion in PAO1-M occurs within PA3620, mutS. *Schematic (not to scale) of genomic region between PA3619 and PA3621 in* ***A****: PAO1-M,* ***B****: All other PAO1 sublines. PA3620.1 is marked in red, and PA3620.2 is marked in blue. The insertion is shown in grey, and the point of insertion in PA3620 is indicated by the dashed lines.*

was carried out against the nucleotide collection database at NCBI (http://blast.ncbi.nlm.nih.gov/). The resulting distance tree of hits is shown in Figure 2.27, and includes all hits which shared >80% sequence identity over >10% of the insertion (query) sequence.

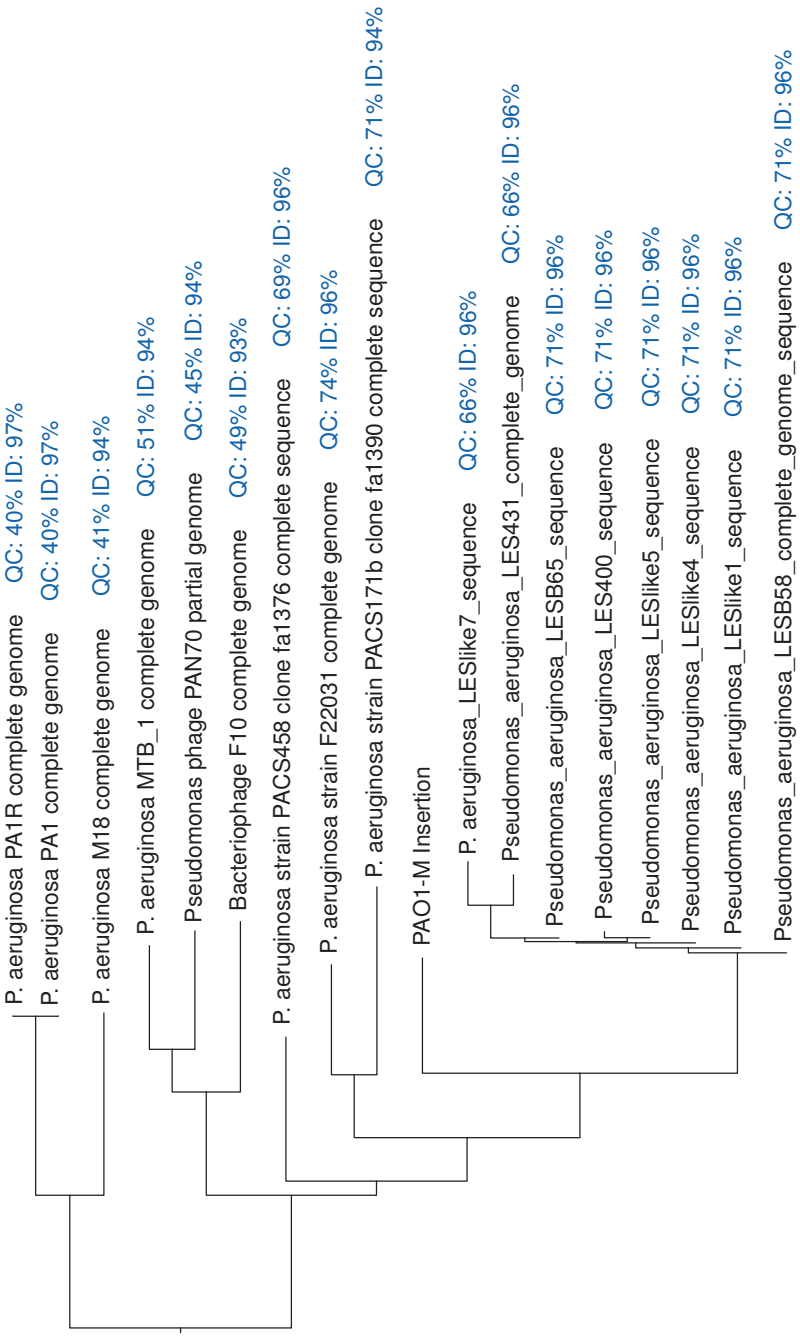

Figure 2.27: Unrooted distance tree of BLAST hits to the 40.1 kb insertion sequence in PAO1-M. *Percentage query coverage (QC) and percentage sequence identity (ID) of hits are shown next to the strain names in blue.*

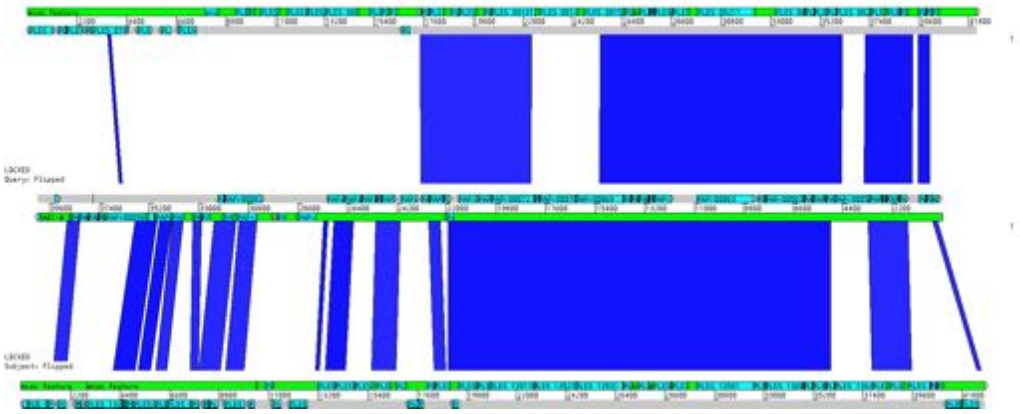

Figure 2.28: Comparison of prophages 2 and 3 from LESB58 with the PAO1-M insertion. *Top track: LESB58 prophage 2. Middle track: PAO1-M insertion. Lower track: LESB58 prophage 3. Blue regions between tracks indicate sequence similarity according to nucmer.*

The PAO1-M insertion shares >94% sequence identity over >66% of its sequence with multiple LES-like sequences and two *P. aeruginosa* cystic fibrosis PACS clones. The insertion sequence shares 94% sequence identity over 51% of its sequence with *P. aeruginosa* MTB-1, which is an environmental isolate from Japan, as well as sharing 93% sequence identity with Bacteriophage F10, which covers 49% of the insertion sequence.

The sequence of the contig containing the PAO1-M insertion was compared to the sequence of *P. aeruginosa* LESB58 using nucmer (Kurtz *et al.*, 2004) and visualised using Artemis Comparison Tool (Carver *et al.*, 2005). The region surrounding the insertion was found to occur in the LESB58 sequence as a single contiguous region with 99% sequence identity to the PAO1-UW sequence. The point of the insertion mapped to the *mutS* gene in LESB58.

The PAO1-M insertion was found to share sequence identity with two regions of the LESB58 sequence, which correspond to sequences of LES prophage 2 and prophage 3 (Figure 2.28). The PAO1-M insertion shares 89% - 94% nucleotide sequence identity with prophage 2, and 88%-96% identity with prophage 3.

It is evident from Figure 2.28 that a larger proportion of the PAO1-

M insertion sequence shares sequence identity with prophage 3 than with prophage 2; 65.39% of the PAO1-M sequence shares nucmer sequence identity hits with prophage 2, compared to 45.86% with prophage 2. This includes a region of prophage 3 which is annotated as "Duplication 2a". This is duplication which also occurs in LESB58 prophage 5, where it is annotated as "Duplication 2b", and according to a nucleotide BLAST comparison the two sequences share 99% identity. Apart from the duplication, and a small region of a hypothetical protein (PLES_13711 in prophage 3 and PLES_25651 in prophage 5), the two prophages share no sequence identity. Prophage 5 only shares sequence similarity with PAO1-M at the region corresponding to duplication 2b. Therefore the PAO1-M insertion sequence is most similiar to that of LES prophage 3.

Annotation of this putative phage sequence was carried out using the Phast server (Zhou *et al.*, 2011), followed by manual curation of gene models. In addition to this, open reading frames above 100 nucleotides were considered and manually annotated using NCBI BLAST and Pfam searches. The phage sequence was predicted to contain 61 open reading frames, the putative products of which are shown in Table 2.14.

In order to determine whether this prophage is lytic against other PAO1 sublines, serial dilutions of supernatant from PAO1-M were grown overnight at 37 °C with bacterial cells from PAO1-L and PAO1-M (Figure 2.29). From concentration $10^{-}3$ onwards, plaques were observed when PAO1-M supernatant was grown with PAO1-L cells. This is indicative of a bacteriophage in PAO1-M killing PAO1-L bacterial cells. No such plaques were seen in the control, where PAO1-M bacteriophage were grown with PAO1-M cells.

Table 2.14: Putative functions of predicted genes in the PAO1-M insertion sequence. *The predicted gene product, and the number of genes with this annotation in PAO1-M are shown.*

Putative gene product	Number of genes with this annotation
Hypothetical protein	44
Clp protease	1
DNA-binding protein Roi	1
Holin	1
Integrase	1
LuxR family transcriptional regulator	1
NinG family protein	1
NinG protein	1
Phage DNA packaging protein	1
Phage portal protein	1
Phage terminase, large subunit	1
Cro family transcriptional regulator	1
Glycoside hydrolase; chitinase	1
Primosomal replication protein	1
Putative repressor protein	1
Replication protein P	1
Tail fibre protein	1
Tail length tape measure protein	1

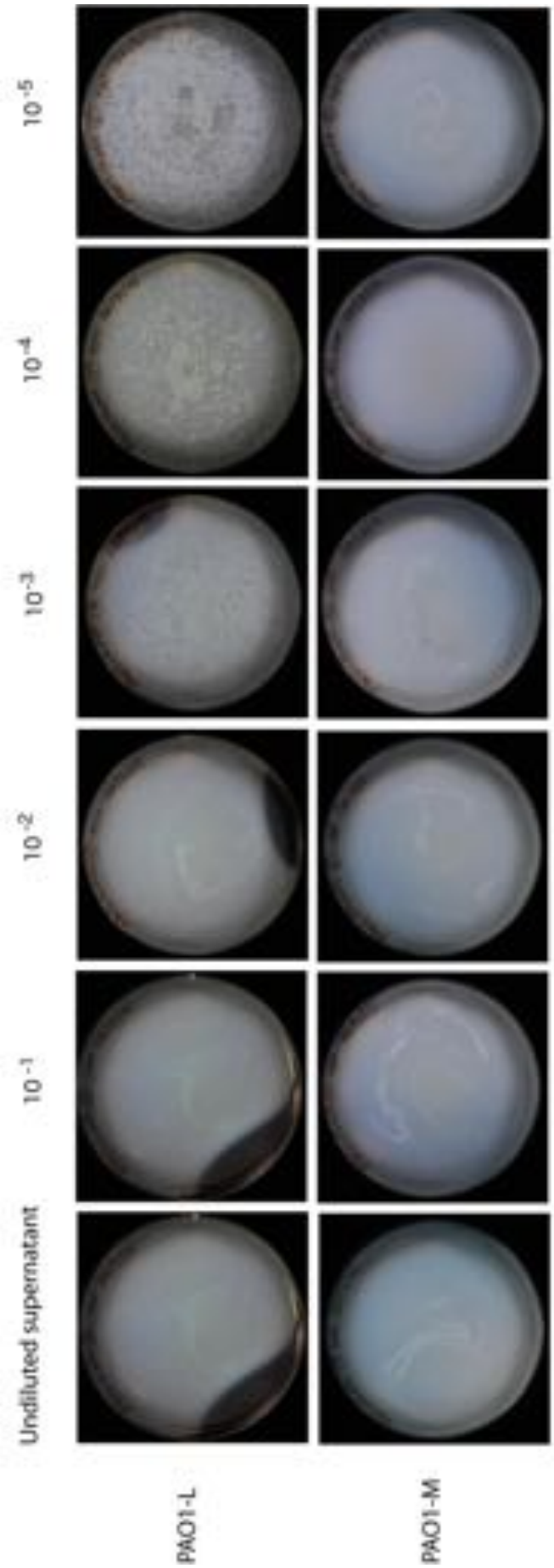

Figure 2.29: Bacteriophage plaque assay of PAO1-M prophage. *Bacterial cells of sublines PAO1-M and PAO1-L were grown with bacterial supernatant from PAO1-M, at varying serial dilutions. These were grown on soft top agar and incubated for 20 hours at 37 °C.*

Results of colony PCR of twelve of the resulting plaques at the sixth serial dilution of supernatant, using primers specific to the PAO1-M insertion (PAO1MIN_FW and PAO1MIN_RV, shown in Table 2.2), are shown in Figure 2.30B. The primers used did not yield a PCR product with PAO1-L template DNA, but did with PAO1-M template DNA (Figure 2.30A). Three out of the twelve colony PCR products were subjected to Sanger sequencing. This confirmed that the amplified product was located in the PAO1-M insertion sequence.

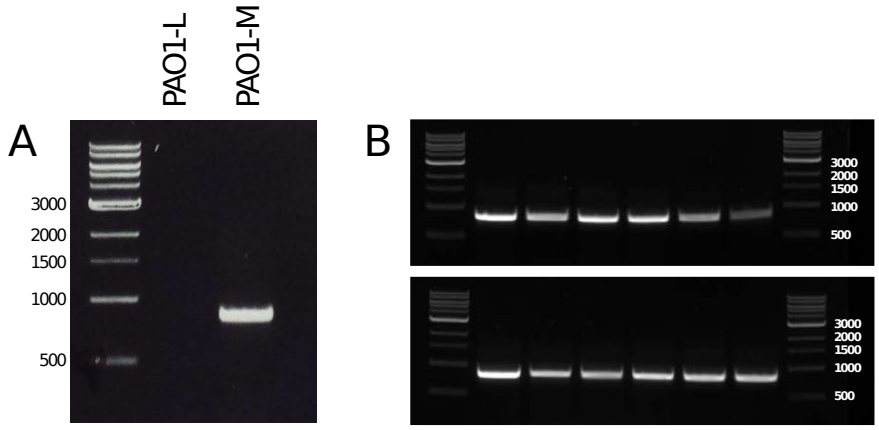

Figure 2.30: Colony PCR of plaques reveals presence of PAO1-M insertion sequence. ***A****: PCR of PAO1-L and PAO1-M genomic DNA with primers specific to the PAO1-M insertion sequence.* ***B****: Colony PCR of twelve plaques, following the growth of PAO1-L bacterial culture with supernatant from PAO1-M bacterial culture, with primers specific to the PAO1-M insertion sequence.*

2.3.12 Chloramphenicol resistant phenotype of PAO1 sublines

The original PAO1 strain was a spontaneous Cm resistant mutant of *P. aeruginosa* PAO, which was selected by its ability to grow on Cm at a concentration of 250 μg/ml (Holloway, 1955). In order to test if currently used PAO1 laboratory sublines are Cm resistant, the eight PAO1 sublines and strains PAO2 and PAO3 were grown on LB agar supplemented with 250 μg/ml and

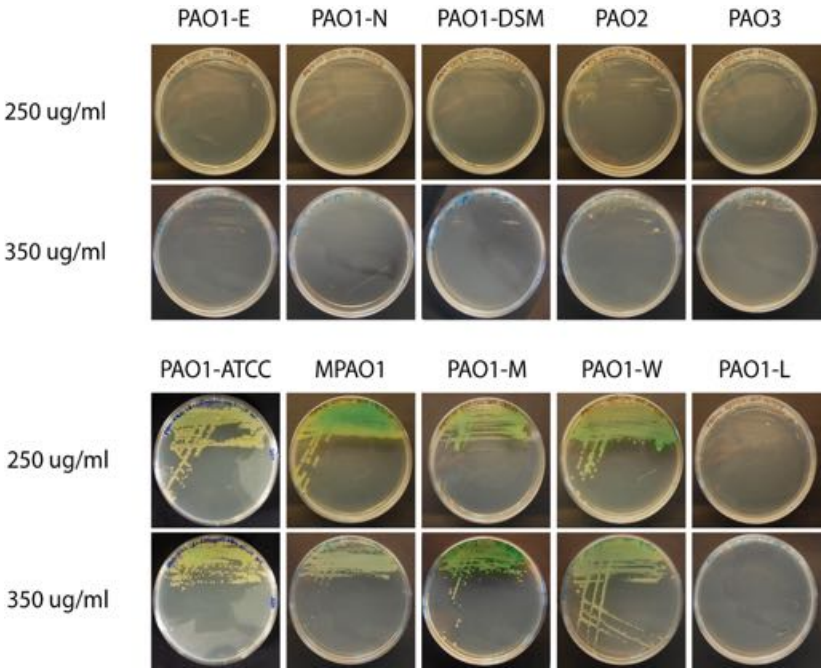

Figure 2.31: Growth of PAO1 sublines and strains PAO2 and PAO3 in the presence of chloramphenicol. *Each strain/subline was grown on LB plates with 250 μg/ml and 350 μg/ml Cm. Plates were incubated for 20 hours at 37 °C. Plates which did not display growth were incubated for a further 24 hours, after which growth was still not seen.*

350 μg/ml Cm for 20 hours at 37 °C (Figure 2.31). Plates which did not display bacterial growth were incubated for a further 24 hours, after which growth was still not seen.

Strains PAO2 and PAO3 were not able to grow on Cm, which is expected given that it is only PAO1 which is Cm resistant. However, surprisingly, four of the eight PAO1 sublines were unable to grow at 250 μg/ml Cm. Based on these findings, it appears that some laboratory sublines of PAO1 have reverted and lost the Cm resistance phenotype over time.

2.4 Discussion

The first whole genome sequence of the most commonly used strain in the study of *Pseudomonas aeruginosa*, strain PAO1, from the University of Washington (PAO1-UW) is currently widely used as a reference sequence in the field. PAO1 is a spontaneous Cm resistant mutant of strain

P. aeruginosa PAO, which was originally isolated from an infected wound in Australia (Holloway, 1955, 1969). It has been previously shown that the reference sequence differs from the PAO1 subline used to generate the physical map (PAO1-DSM) and the subline used to construct a widely used transposon library (MPAO1) by a number of SNPs and INDELs, as well as a 2.2 Mb chromosomal inversion and a 12 kb prophage insertion (Klockgether *et al.*, 2010). In this study this was further explored by examining the genetic diversity of a range of addtional commonly used PAO1 sublines, using whole genome sequencing and optical mapping. We also employed these approaches on strains PAO2 and PAO3, which are auxotroph mutants for serine and tryptophan, respectively, produced from the original *P. aeruginosa* PAO isolate by manganese mutagenesis (Holloway, 1955). Using this data it was possible to determine the likely genomic features of the ancestral *P. aeruginosa* PAO. This information should be taken into account when working with *P. aeruginosa* PAO1.

Analysis of the kmer content of whole genome sequencing reads revealed the over-representation of poly-A kmers at the 3' end of sequencing reads for all strains and sublines (Figure 2.6). This was particularly the case for the reverse reads (Figure 2.6B). It has been previously mentioned in the research community that this could be due to the sequencing instrument reading through the fragment to be sequenced, into the adapter (which explains the adapter contamination observed), and then into unknown space (Loman, 2013). This is due to the combination of short fragment sizes and long read lengths. Such artefacts at the 3' end of reads could potentially be detrimental to de novo genome assemblies, however for the purpose of read alignment and variant calling, the read aligner would trim such artefacts from the end of reads, as they would not map to the reference sequence. They would therefore be unlikely to affect variant calling.

The presence of the previously described RGP42 island was confirmed

in all strains and sublines used in this study, using data from whole genome sequencing and optical mapping (Figure 2.7). The RGP42 sequence was found to contain a large number of SNPs in all sequenced strains and sublines with respect to the previously published sequence (Klockgether *et al.*, 2010), which were likely to be due to single base pair errors in the existing RGP42 sequence (Table 2.5). These were due to the method of genome assembly used, namely the extension of RGP42 contigs generated from reads which did not map to the PAO1-UW reference sequence using reads which map to the RGP5 region of PAO1-UW (Figure 2.12). Alignment of the RGP42 region with the RGP5 region revealed that the single base pair errors in RGP42 were identical to the corresponding base in the RGP5 region (Table 2.5). This supports the idea that these single base errors were introduced by the assembly method used. It is possible that these errors were detected due to the improvement of read lengths and base quality of next generation sequencing technologies (Figures 2.2 and 2.3), which led to a higher coverage of the RGP42 island in this dataset compared to the dataset from (Klockgether *et al.*, 2010).

A total of 33 SNP loci were found to occur outside of the RGP42 region (Table 2.8). The results for these SNP loci were identical when using three different base quality thresholds at the SNP calling stage, and so these results are considered to be relatively reliable compared to results which were not identical at all quality thresholds (shown in Table 2.9). Almost half of these loci (12/33) corresponded to SNPs which were common to all ten sequenced strains and sublines, including PAO2 and PAO3. This is either due to single base pair misassemblies in the PAO1-UW reference sequence, or to SNPs in the PAO1-UW subline. Since a sample of this subline could not be obtained, it has not been possible to confirm which of the two possibilities is the case. However, given that these SNPs are conserved in all PAO1 sublines sequenced, as well as in strains PAO2 and PAO3, it is clear

that these SNPs would be conserved in the ancestral *P. aeruginosa* PAO. 14 of the 33 SNP loci consist of SNPs which are unique to individual sublines. In the case of PAO1 sublines, it is likely that such SNPs occurred after the distribution of the strains to their respective laboratories. In the case of PAO2 and PAO3, these SNPs are likely to have occurred as a result of manganese mutagenesis; the method used to obtain the auxotroph mutants (Holloway, 1955).

A further 20 SNP loci corresponded to SNP results which were not identical when different base quality thresholds were applied (Table 2.9). Six of these SNP loci were present when the lower threshold was applied (Q10) but not when the higher base quality thresholds (Q13, Q15) were applied. Visual inspection of mapped reads at these regions indicate the presence of a SNP. This shows that for this particular dataset, picking a high base quality threshold would have eliminated some potential true SNPs. In contrast, three of these 20 SNP loci were called when the higher base quality cut off was used, but not when the lower two quality cut offs were used. Visual inspection of the reads revealed that these corresponded to three low frequency SNPs, one of which was the *lasR* mutation in PAO1-DSM which was previously reported (Klockgether *et al.*, 2010). For this reason, it was decided that rather than picking a single quality cut off at which to call SNPs, that results from a variety of cut offs would be considered in this analysis.

INDEL calling revealed a total of 54 regions corresponding to short INDELs in the ten sequenced strains and sublines. INDELs were conserved between all sequenced sublines at 36 of these regions (Table 2.10). Given that these INDELs were conserved between all PAO1 sublines, as well as strains PAO2 and PAO3, it is likely that these regions correspond either to INDELs in PAO1-UW, or to misassemblies in PAO1-UW. Therefore the INDELs called should correspond to the sequence of the ancestral *P. aeruginosa*

PAO. More than 50% (10/18) of these INDEL regions corresponded to INDELs which were unique to a particular subline, of which all but three lead to out of frame mutations within protein coding regions. A summary of SNPs and INDELs unique to individual strains and sublines is shown in Table 2.15.

It is evident from Table 2.15 that PAO1-DSM has the highest number of SNPs and INDELs, all of which occur in protein coding regions. Six other PAO1 sublines contain at least one SNP or INDEL, of which four have at least one SNP or INDEL within a protein coding region. PAO1-E does not contain any unique SNPs or INDELs, however it does contain a number of SNPs and INDELs which are conserved in PAO1-N. From the data, it appears that all PAO1 sublines studied contain at least one mutation which was most likely gained after the distribution and propagation of the original PAO1 strain, of which all but one contain at least one mutation in a protein coding region. This includes PAO1-ATCC, which is a strain purchased from the ATCC Strain Collection (http://http://www.atcc.org), where it is labelled as PAO1 ATCC 15692; first mentioned by Bruce Holloway (1969).

PAO1-N and PAO1-E were found to contain an identical 58.6 kb deletion between genes PA0707 and PA0668.1, which includes the Cm acetyltransferase (*cat*) gene (Section 2.3.10). This has been confirmed by whole genome sequencing (Figure 2.22A), optical mapping (Figure 2.22B) and PCR (Figure 2.24). A 6 bp repeat sequence was identified at the start and at the end of the deletion (Figure 2.24C), and although this short sequence occurs multiple times in the genome sequence, as well as seven additional times within the deletion, the presence of the sequence at the start of the deletion and after the end of the deletion, so that one copy of the repeat remain after the deletion, could be indicative of a homologous recombination event.

A nucleotide BLAST of the deleted sequence in PAO1-N and PAO1-E revealed that it is conserved in *P. aeruginosa* and is not found in any

Mutation type	Genomic position in PAO-consensus	Reference base in PAO-consensus	Mutation in PAO-consensus	Effect on protein	Gene ID	Gene Name	Subline(s) affected
SNP	875031	C	G	L195V	PA4147	*acoR*	PAO1-DSM
SNP	3659765	T	C	I68V	PA1710	*exsC*	PAO1-DSM
SNP	3816223	C	T	R115W	PA1561	*aer*	PAO1-DSM
SNP	3957011	C	A	R216L	PA1430	*lasR*	PAO1-DSM
SNP	4108613	G	A	A70V	PA1296	-	PAO1-DSM
SNP	5036302	A	C	T20P	PA4499	-	PAO1-DSM
SNP	5041200	G	C	E266Q	PA4502	-	PAO1-DSM
SNP	5815280	G	A	A132T	PA5155	-	PAO1-DSM
SNP	2708163	C	G	R66P	PA2492	*mexT*	PAO1-ATCC
3 bp Ins.	2807991	G	GCTC	In frame ins.	PA2492	*mexT*	PAO1-ATCC
SNP	5688553	C	A	E458*	PA5040	*pilQ*	PAO1-ATCC
6 bp Ins.	417866	G	GCCGAGC	In frame ins.	PA0373	*ftsY*	PAO1-N
SNP	1505695	C	T	S65S	PA3578	-	PAO1-N
1 bp Ins.	2342111	T	TT	Out of frame ins.	-	-	MPAO1
SNP	2707503	G	A	A286V	PA2492	*mexT*	MPAO1
SNP	5069340	G	T	S64R	PA4525	*pilA*	PAO1-M
SNP	587916	A	G	N/A	-	-	PAO1-W
8 bp ins	2807716	T	-	Out of frame ins	PA2492	*mexT*	PAO1-L

Table 2.15: Mutations unique to each sequenced PAO1 subline. *Ins = insertion, Del = deletion*

other species (Figure 2.23). Phast analysis did not indicate the presence of a phage sequence in the deleted region. The regions before and after the deletion, i.e. the sequence of PAO1-E and PAO1-N at the deletion region, were used for a BLAST search to see if the deletion was conserved in any other Pseudomonas isolates. This search yielded no results, and so therefore it can be concluded that the deletion in PAO1-N and PAO1-E is not conserved in any other sequenced *Pseudomonas* isolate. Since the deleted sequence does not appear to be a phage or mobile genetic element, and is a sequence which is conserved in *P. aeruginosa* it could be that the deletion in PAO1-N and PAO1-E is an example of reductive evolution, where selection pressure on the deleted region is relieved.

The deletion in PAO1-N and PAO1-E results in the loss of one of the four ribosomal operons (*rrnB*). It has been shown previously that deletion of ribosomal operons in *Escherichia coli* (Asai *et al.*, 1999a,b) and *Bacillus subtilis* (Yano *et al.*, 2013), which contain seven and ten ribosomal RNA (*rrn*) operons respectively, leads to a reduction in growth rate (both), pro-

tein synthesis (*E. coli*) and motility (*B. subtilis*). A previous study comparing the number of *rrn* operons in *Pseudomonas* showed that all eight *P. aeruginosa* genomes studied contained four ribosomal operons (Bodilis *et al.*, 2012). Therefore it could be interesting to investigate the effect of the loss of ribosomal operons in *P. aeruginosa*, although no significant reduction in growth rate or motility has been observed in PAO1-N compared to PAO1-L (data not shown).

Given that the deletion is identical in PAO1-N and PAO1-E, and that the two sublines share a number of SNPs and INDELs, it is highly likely that these genetic changes occurred in a common ancestor of the two sublines. Inspection of whole genome sequencing data of strain PAN067, the first mutant generated from subline PAO1-N more than 20 years ago (Latifi *et al.*, 1995), reveals the presence of the deletion in this mutant. This suggests that the deletion occurred in the common ancestor of PAO1-N and PAO1-E before the two strains were distributed to the respective laboratories. However, it is unclear where the common ancestor originated from.

Optical mapping of the chromosomes of all strains and sublines of this study revealed that the PAO1-UW inversion between *rrnA* and *rrnB* is not conserved in any of the strains and sublines studied (Figure 2.20). It has previously been speculated that the 2.2 Mb inversion in PAO1-UW, with respect to the physical map, is not unique to this subline since PAO1 stocks from other laboratories contain the same inversion (Stover *et al.*, 2000). The results obtained in this study do not support this, although perhaps a larger sampling of PAO1 sublines, particularly from North America, would shed more light on this. Given that all PAO1 sublines studied, as well as strains PAO2 and PAO3, do not contain the 2.2 Mb inversion, it is unlikely that the ancestral *P. aeruginosa* PAO contained this feature. If there were other PAO1 sublines with this same inversion, these would be likely to have originated from a common ancestor which gained the inversion after the

distribution of PAO1 strains.

Optical mapping revealed a 40.1 kb insertion in PAO1-M, in the *mutS* gene, which was not found in any other strain or subline in the study (Figure 2.25). The insertion was found to occur after the 40th nucleotide of the *mutS* gene, which resulted in a truncated version of the gene at the 5' flanking end of the insertion (PA3620.1), and an almost complete copy of the *mutS* gene at the 3' flanking end of the insertion (PA3620.2) (Figure 2.26A). There are differences in the first and fifth amino acids of the protein sequence of PA3620.2, as well as an additional four residues at the start of the protein sequence compared to the MutS protein in PAO1-UW (Figure 2.26D). It is not known whether PAO1-M has a functional MutS protein.

Analysis of the insertion sequence revealed it was most similar to the sequence of Liverpool Epidemic Strain (LES) prophage 3 (Figure 2.28). Sequence annotation of this putative phage sequence revealed a number of phage proteins (Table 2.14). A bacteriophage plaque assay of the supernatant from PAO1-M with PAO1-L bacterial cells revealed that PAO1-M contains a phage which lyses PAO1-L cells (Figure 2.29). Colony PCR of twelve of the resulting plaques, using primers specific to the PAO1-M insertion, yielded PCR products (Figure 2.30). Sanger sequencing of these PCR products showed that these PCR products corresponded to part of the PAO1-M insertion sequence. Therefore it can be concluded that the PAO1-M insertion corresponds to a bacteriophage. The origin of the phage in this laboratory subline is not known, although BLAST results indicate a level of similarity with LES phages. The prophages of LES are thought to be specific to this strain, and the discovery of a LES-like prophage sequence in a non LES strain has implications for LES specific PCRs of regions of the accessory genome. This appears to be a novel prophage sequence unique to the PAO1-M laboratory subline.

The serine and tryptophan auxotroph phenotypes of PAO2 and PAO3,

respectively, were confirmed by growth on minimal medium and minimal medium supplemented with the appropriate amino acid (Figure 2.13). Although the auxotroph phenotypes of PAO2 and PAO3 were confirmed, it was found that neither PAO2 nor the prototroph PAO1-L were able to grow with 9 mM tryptophan (Figure 2.13B). Therefore it appears that high levels of tryptophan inhibit a metabolic process in *P. aeruginosa*. It has been shown that tryptophan inhibits prephenate dehydratase for the biosynthesis of phenylalanine in *B. subtilis* (Rebello and Jensen, 1970). A reduction in phenylalanine production would lead to a lack of growth on minimal media, as is seen in Figure 2.13B. However, further work would need to be carried out to confirm this.

Whole genome sequencing indicated a deletion in the *serA* gene in PAO2 and a SNP in the *trpC* gene in PAO3, both of which were confirmed by PCR. Wild type versions of *serA* and *trpC* were cloned into PAO2 and PAO3, respectively, which led to the restoration of the prototroph phenotype in both strains (Figure 2.18). Therefore the in frame deletion of 23 residues from the NAD binding Rossman fold of *serA* in PAO2 leads to the serine auxotroph phenotype, and the E61K mutation in *trpC* of PAO3 leads to the tryptophan auxotroph phenotype.

The original PAO1 strain was a Cm resistant mutant of *P. aeruginosa* PAO (Holloway, 1955). In order to verify if currently used PAO1 laboratory sublines have conserved this phenotype, all sublines were grown on nutrient medium containing Cm (Figure 2.31), revealing some PAO1 sublines to be Cm sensitive. From Tables 2.4, 2.8, 2.9 and 2.11, it is clear that there are a number of SNPs and INDELs in the sequenced strains and sublines which occur in genes related to antibiotic resistance, some of which could lead to loss of the Cm resistance phenotype (Table 2.16).

Table 2.16: Mutations relating to antibiotic resistance in the sequenced strains and sublines. *Ins = insertion, Del = deletion*

Mutation type	Genomic position in PAO-consensus	Reference base in PAO-consensus	Mutation in PAO-consensus	Effect on protein	Gene ID	Gene Name	Strain/subline(s) affected	Chloramphenicol resistant?
SNP	2704842	G	C	Y326*	PA2494	*mexF*	PAO1-N, PAO1-E	No
SNP	2707503	G	A	A286V	PA2492	*mexT*	MPAO1	Yes
SNP	2708163	C	G	R66P	PA2492	*mexT*	PAO1-ATCC	Yes
SNP	2709203	G	A	D249N	PA2491	*mexS*	PAO1-DSM, PAO2, PAO3	No
8 bp ins	2807716	T	-	Out of frame ins	PA2492	*mexT*	PAO1-L	No
3 bp ins	2807991	G	GCTC	In frame ins	PA2492	*mexT*	PAO1-ATCC	Yes
58.6 kb del	4734853-4793425	-	-	Loss of *cat*	PA0706	*cat*	PAO1-N, PAO1-E	No

PAO1-DSM, PAO2 and PAO3 share a mutation in the *mexS* gene, which is not found in any of the other sublines sequenced. Given that all three strains don't seem to be Cm sensitive, that PAO2 and PAO3 are derivatives of *P. aeruginosa* PAO, and all three do not contain any other mutations in genes associated with antibiotic resistance, it is possible that the *mexS* gene corresponds to the unmapped *chl-2* locus which is mutated in PAO1 and leads to the Cm resistance phenotype. Mutations in the *mexS* gene (PA2491), which is located immediately upstream of *mexT*, have previously been shown to promote expression of the *mexEF-oprN* multidrug efflux pump, leading to an increase in resistance to Cm, as well as quinolones and imipenem (Sobel *et al.*, 2005). Given this hypothesis about the *chl-2* locus, it could be postulated that PAO1-DSM is in fact a derivative of PAO rather than PAO1. This is supported to some extent by the presence of two SNPs which are shared between PAO1-DSM and PAO2, both of which are found in coding regions and one of which leads to a change in the protein sequence in a predicted serine/threonine transporter.

PAO1-N and PAO1-E not only share the loss of the *cat* gene as a result of the 58.6 kb deletion, but also share a SNP which causes a premature stop codon in the *mexF* gene (previously reported by (Luong *et al.*, 2014)). This would lead to a truncated form of the 1062 amino acid protein. The protein MexF forms the energy dependent efflux pump, or RND pump, of the MexEF-OprN efflux pump, which confers resistance to a range of antibiotics including Cm (Table 1.4). The loss of this RND pump would lead to a build up of Cm in the cell. The *cat* gene encodes the enzyme Cm acetyltransferase, which prevents the Cm from binding to bacterial ribosomes by modifying the antibiotic. The loss of the *cat* gene, as well as the loss of function of MexF, would both lead to a Cm sensitive phenotype, as is seen with PAO1-N and PAO1-E.

The global regulator *mexT* has been shown to enhance expression of

the MexEF-OprN efflux pump, which is inactive in wild type cells, and decrease expression of the carbapenem-permeable porin OprD (Kohler *et al.*, 1999). The overexpression of this efflux pump and the repression of the porin leads to what is described as an *nfxC* mutant phenotype, whereby cells display resistance to imipenem and chloramphencol, and hyper susceptibility to β-lactams and aminoglycosides (Fukuda *et al.*, 1990). PAO1-L contains an 8 bp insertion in *mexT*, which has been previously described in other sublines including PAO1-UW (Maseda *et al.*, 2000). The removal of this insertion (an 8 bp deletion) has been shown to lead to an *nfxC* mutant phenotype, due to the activation of MexT, leading to overexpression of the multidrug efflux pump MexEF-OprN (Maseda *et al.*, 2000). Therefore it would be expected that the insertion inactivates MexT, and therefore the efflux pump is not expressed, which correlates with the Cm sensitive phenotype displayed by PAO1-L. It can therefore also be assumed that PAO1-UW is also Cm sensitive, given that it contains the same 8 bp insertion in the *mexT* gene as PAO1-L.

There is a SNP in the *mexT* gene of MPAO1, and both a SNP and an in frame insertion in the *mexT* gene of PAO1-ATCC. However, it appears that these genetic changes do not affect efflux pump expression in both of these sublines, as both are seen to be Cm resistant. SNPs and INDEls were not detected in known antibiotic resistance genes of PAO1-M and PAO1-W, and both of these sublines display the Cm resistant phenotype.

Random amplified polymorphic DNA (RAPD) analysis was carried out to see if the large scale genomic changes in PAO1 sublines, as well as small SNPs and INDELs, affect the results of an often used typing method (Figure 2.19). RAPD profiles of all sublines were identical with both sets of primers used, and differed from the clinical isolate *P. aeruginosa* Nottingham K13. This shows that RAPD is not a discriminatory method for typing *P. aeruginosa* strains with large insertions and deletions. This emphasises

the importance of developing methods such as whole genome sequencing for use in the clinic, to carry out deeper analyses of infectious agents and to detect events such as the acquisition of phage sequences or the gain and loss of mobile genetic elements.

From this analysis it has been determined that the laboratory strain *P. aeruginosa* PAO1 exists as highly variable sublines which contain genetic differences, not only compared the reference sequence of PAO1-UW, but also compared to one another. This includes isolates in laboratory stocks, as well as isolates which are publicly available from strain collections. Therefore the whole genome sequencing of laboratory isolates and isolates in strain collections should become a more routine exercise. Efforts such as the joint project between Public Health England and the Wellcome Trust Sanger Institute to sequence 3000 bacterial and 500 viral genomes from the NCTC and NCPV collections, respectively, will provide a wealth of data for researchers working with strains in these collections, which includes three *P. aeruginosa* strains (NCTC10332, NCTC9433 and NCTC10662) (WTSI, 2015). However, for researchers working with laboratory stocks of PAO1, the need to carry out whole genome sequencing of their laboratory isolates is becoming more apparent.

A recent study by Lee *et al.* (2015) described the transposon sequencing (Tn-seq) of subline MPAO1 grown in three different media, to identify genes which were essential for the survival of the organism. The list of identified essential genes was compared with all genes identified in this study as containing genetic differences in any of the strains or sublines. Of all the essential genes identified by Lee *et al.* (2015), only gene PA4685 was identified in this study as containing a genetic change. This was a large deletion which results in loss of almost the entire gene in a number of PAO1 sublines, including MPAO1 (Table 2.4).

This highlights the issue of using a reference sequence which does not

correspond to the subline being studied, as the loss of a large proportion of the gene PA4685 in this case would mean that the transposon would not insert here. However, since it was assumed that this gene exists in the subline being studied, the gene was seen to be essential due to a lack of signal. In this study this finding is not detrimental, as this only accounts for one gene out of the 352 identified as being essential in all three of the media studied. However, it highlights the importance of sequencing strains used in laboratory studies and comparing these to the reference sequence before drawing conclusions in analyses. Furthermore, it appears that many researchers are not aware of the variability between PAO1 sublines, and refer to PAO1 as a uniform and genetically stable model strain in publications. This study shows that the variability between laboratory sublines is very high, and emphasises the need to take this variability into account.

Chapter 3

Whole genome sequencing of fourty nine *Pseudomonas aeruginosa* wound isolates

3.1 Introduction

P. aeruginosa is a ubiquitous organism, with strains found in a variety of environmental niches, in addition to animal and human hosts. It is in humans and animals where *P. aeruginosa* is known to act as an opportunistic pathogen, affecting immunocompromised individuals, neonates, cancer patients, patients with wounds such as burns or ulcers, and sufferers of cystic fibrosis (CF). *P. aeruginosa* is currently most commonly associated with hospital-acquired infections, where the pathogen is often acquired from an environmental source such as a water outlet (Berrouane *et al.*, 2000; Quick *et al.*, 2014; Walker and Moore, 2015).

P. aeruginosa is one of the most commonly isolated bacteria from chronic wounds, and is often resistant to antimicrobials due to the tendency of the organism to aggregate with other *P. aeruginosa* cells as well as other bacteria in the wound, forming biofilms. The bacterial diversity of wounds has been described in various studies (Dowd *et al.*, 2008; Frank *et al.*, 2009; Wolcott *et al.*, 2009; Smith *et al.*, 2010; Tuttle *et al.*, 2011), and not only are there multiple species present in a single wound, but also multiple strains of the same species. This means it is often hard to decide which antimicrobials would be suitable for treatment, as it is often difficult to pinpoint the major causative agent and its antibiotic susceptibilities. Culture based methods only identify a subset of the bacteria present, and although culture-independent methods such as metagenomics generate a complete picture of bacterial diversity, it is currently difficult to make clinical decisions based on this wealth of data. In addition to this the method of wound sampling also affects the result obtained, and varies for different types of wound (Bowler *et al.*, 2001).

P. aeruginosa is able to exploit a variety of nutrient sources, and is able to adapt to changes in environmental conditions over time. This is reflected in the large genome size of *P. aeruginosa* compared to other bac-

teria, between 5.2 and 7 Mb (Wiehlmann *et al.*, 2007), and its large number of regulatory genes, with more than 500 identified in *P. aeruginosa* strain PAO1 (Stover *et al.*, 2000). The genome sizes of isolates of *P. aeruginosa* vary highly due to differences in the accessory genome (Mathee *et al.*, 2008).

The transition from the external environment to host environment means there are nutritional and physiochemical changes, which *P. aeruginosa* must adapt to in order to survive. In addition to this, the organism needs to survive the host immune system, the presence of antibiotics and the presence of other competing microorganisms. A number of studies have looked into the genetic changes associated with adaptation to the CF lung (Smith *et al.*, 2006; Bragonzi *et al.*, 2009; Huse *et al.*, 2010; Cramer *et al.*, 2011; Fothergill *et al.*, 2014). Smith *et al.* (2006) compared two clonal *P. aeruginosa* isolates from two different timepoints (6 months and 96 months) from a single CF patient and found a number of mutations in virulence factors and their regulators. This included genes involved in O-antigen biosynthesis, multidrug efflux and quorum sensing, among others, and the associated phenotypes were also observed (i.e. loss of serotype-specific antigenicity, loss of pyoverdine production and loss of secreted protease activities). Mutations in quorum sensing genes as a result of adaptation to the CF lung have also been reported by Fothergill *et al.* (2014), with the most commonly observed adaptation being a loss of function mutation in the *lasR* gene (D'Argenio *et al.*, 2007). A study of 44 morphologically identical *P. aeruginosa* isolates from a single CF patient sputum sample revealed that intra-population recombination plays a significant role in the short term generation of genetic and phenotypic diversity within individual patients, in particular quorum sensing phenotypes (Darch *et al.*, 2015). Large deletions (Cramer *et al.*, 2011) and insertions/loss of MGEs such as prophage (Chung *et al.*, 2012) have also been reported over the course of *P. aeruginosa* infections in CF patients.

Infections are often polymicrobial, and consist of more than one *P. aeruginosa* strain or isolate, with a variety of phenotypes among these cells. Therefore certain strains may benefit from the activity of other strains, acting as "cheats" in the population (Jiricny *et al.*, 2010). For example, strains which do not produce quorum sensing signal molecules (QSSMs) have been shown to incur a fitness advantage by avoiding the metabolic cost of QSSM production and QS behaviour such as the secretion of exoenzymes, however are still able to benefit from QSSMs produced by other cooperator strains in the population. The presence of "cheats" has been shown to lead to an increased susceptibility to antibiotics, therefore it is evident that these strains require the presence of cooperators in order to propagate (Popat *et al.*, 2012). Antimicrobial treatment may alter the population dynamics of cheats and cooperating bacteria, such that killing off the less virulent, antibiotic susceptible cheats would allow more virulent strains to dominate in the population, having a negative outcome on an infection in the long term. Therefore understanding the types of strains present at sites of infection, as well as their virulence characteristics, would be useful to research effective treatments for the clearance of infection.

Quorum sensing (QS) is thought to influence the regulation and timing of the production of several virulence factors in *P. aeruginosa*, and could potentially have a role in antibiotic resistance, swarming and biofilm formation (Camara *et al.*, 2002). Numerous studies have found QS diversity in isolates both within and between CF patients (Wilder *et al.*, 2009; Chung *et al.*, 2012; Workentine *et al.*, 2013; Darch *et al.*, 2015), including isolates which were found to be QS deficient (Tingpej *et al.*, 2007; Hansen *et al.*, 2012). A recent study of QSSM production in clinical CF isolates of *P. aeruginosa* from a single sputum sample showed a positive correlation between production of signal molecules 2-heptyl-4-quinoline (HHQ), *N*-butanoyl-L-homoserine lactone (C_4-HSL), 2-heptyl-4-

hydroxyquinoline *N*-oxide (HQNO), *Pseudomonas* quinolone signal (PQS) and *N*-3-oxo-dodecanoyl-L-homoserine lactone (3-oxo-C_{12}-HSL). Identical recombining loci were found to be associated with altered production of the QSSMs PQS and HQNO, and it was therefore hypothesised that these two phenotypes could be intrinsically linked (Darch *et al.*, 2015).

McKeown (2013) studied the QSSM and virulence factor production in a cohort of forty nine *P. aeruginosa* wound colonising isolates from different patients and different wound types. It can be seen in Figure 3.1 that a diverse array of QS profiles was observed in this set of isolates, including QS deficient isolates, as has been shown in previous studies on isolates from CF infections. Of note was the novel finding that two of the strains were able to produce AHL signal molecules, but not AQs, where previous studies had only found AHL-deficient and AQ proficient strains (Wilder *et al.*, 2009; Chung *et al.*, 2012; Hansen *et al.*, 2012).

A lot is known about the phenotypic diversity of *P. aeruginosa*, however with the recent drop in price of whole genome sequencing (WGS), as well as the increasing portability of WGS instruments, such as the Illumina MiSeq and Ion Torrent (now Life Technologies) Personal Genome Machine (PGM) desktop sequencers and the USB stick sized MinION from Oxford Nanopore, the study of genomic diversity of *P. aeruginosa* is now much more accessible to researchers. There have been a multitude of genomics studies carrying out WGS of a group of related *P. aeruginosa* isolates (Jeukens *et al.*, 2014; Darch *et al.*, 2015; Williams *et al.*, 2015), including isolates involved in hospital outbreaks (Snyder *et al.*, 2013; Feliziani *et al.*, 2014; Davis *et al.*, 2015).

Stewart *et al.* (2014) carried out WGS of a diverse set of 12 *P. aeruginosa* isolates, which included CF lung isolates which had adapted to the host, as well as four environmental isolates. These isolates were compared with 43 publicly available complete genome sequences. This comparative genomics

Table 3.1: Heat map of the relative QS molecule and protein levels found in supernatants of the 49 wound isolates. *Figure created using data from (McKeown, 2013). Isolates were categorised as producing high (red), medium (white) or low (blue) levels of each QSSM after 8 and 24 hours of incubation, as described in (McKeown, 2013).*

Isolate	C4-HSL		3-oxo-C12-HSL		HQNO		HHQ		PQS		Pyocyanin		LasB elastase		Protease		Isolate
	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr	
1																	1
2																	2
3																	3
4																	4
5																	5
6																	6
7																	7
8																	8
9																	9
10																	10
11																	11
12																	12
13																	13
14																	14
15																	15
16																	16
17																	17
18																	18
19																	19
20																	20
21																	21
22																	22
23																	23
24																	24
25																	25
26																	26
27																	27
28																	28
29																	29
30																	30
31																	31
32																	32
33																	33
34																	34
35																	35
36																	36
37																	37
38																	38
39																	39
40																	40
41																	41
42																	42
43																	43
44																	44
45																	45
46																	46
47																	47
48																	48
49																	49
PAO1																	PAO1

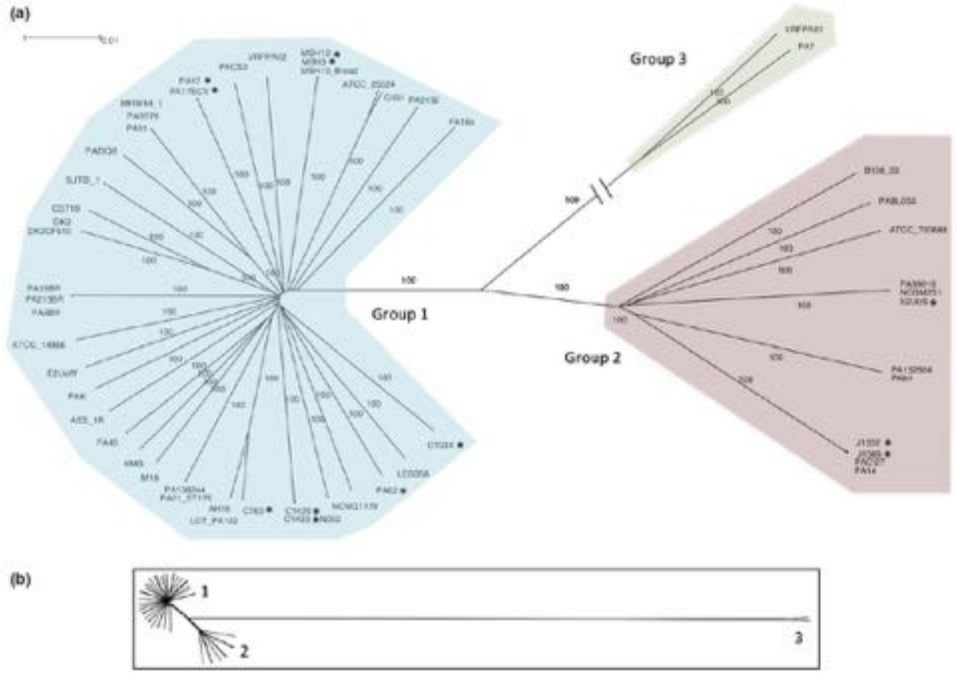

Figure 3.1: ***A:*** *Unrooted core genome SNP tree of 55 P. aeruginosa isolates shows the division of these isolates into three major subgroups; shown in blue (group 1), pink (group 2) and green (group 3). Strains labelled with an asterisk were sequenced in the study. Group 3 is truncated for clarity. The true distances are shown in* ***B****. (Figure 1 in Stewart* et al.*, 2014.)*

study revealed that this set of isolates formed three major subgroups (Figure 3.1). Group 1, the largest of the three groups, contained the well known strains PAO1, DK2, LESB58 and PAK, and the smaller group 2 contained strain PA14. The third group consisted only of strains PA7, a multidrug resistant isolate from Argentina and a known taxonomic outlier, and VRF-PAO1, a blood isolate from India. The 55 isolates were found to produce 35 distinct branches in the resulting core genome phylogeny, indicating that further survey sequencing would populate these branches.

3.2 Materials and Methods

3.2.1 Whole genome sequencing

Genomic DNA was isolated from an overnight culture using the DNeasy Blood and Tissue Kit (QIAGEN, Hilden, Germany). Genomic DNA (500

ng) was mechanically fragmented for 40 sec using a Covaris M220 (Covaris, Woburn MA, USA) with default settings. Fragmented DNA was transferred to a PCR tube and library synthesis was performed with the Kapa Hyperprep kit (Kapa biosystems, Wilmington MA, USA) according to manufacturer's instructions. TruSeq HT adapters (Illumina, SanDiego CA, USA) were used to barcode the library, which was sequenced with 47 other *Pseudomonas* isolates on a single Illumina MiSeq 300 bp paired-end run[1].

3.2.2 *De novo* assembly

De novo assembly for each isolate was carried out using the A5-miseq pipeline version 20140521 (Coil *et al.*, 2015) [2]. This pipeline includes quality control of reads.

3.2.3 Phylogenetic analysis

Core genome alignments and phylogenetic trees were generated using Parsnp, with the -c flag, and with the reference genome (-r flag) set to "!", to pick a random sequence from the input as the reference (PA-W10 was randomly picked in this analysis). Contig sequences for each isolate were used as the input. Parsnp is part of the Harvest (version 1.1) package (Treangen *et al.*, 2014).

3.2.4 Read mapping, realignment and variant calling

Whole genome sequencing reads were aligned to the PAO1-UW sequence (Genbank accession: NC_002516) using Smalt version 0.7.0.1 (Ponstingl, 2012) with a kmer of 12 and a step size of 1. SNPs were determined using

[1] Library preparation and whole genome sequencing was carried out by Irena Kukavica, Université Laval.

[2] *De novo* assembly was carried out using bash scripts from Jerome Laroche, Université Laval.

Samtools mpileup, using a base quality threshold of 13 (parameter -Q). Regions with zero read coverage were detected using the genomeCoverageBed functionality of bedtools with the -bga option (version 2.17.0, Quinlan and Hall, 2010).

3.2.5 Comparison with existing genome sequences

Accession numbers of all reference sequences can be found in Table 5.1. Assembled contigs were mapped to the PAO1-UW reference sequence using the Blast Ring Image Generator (BRIG, Alikhan *et al.*, 2011), which utilises BLAST (BLAST+ version 2.2.29).

3.2.6 Gene prediction and annotation

Contig sequences for each isolate were submitted to the RAST webserver (Aziz *et al.*, 2008) for gene annotation.

3.2.7 Characterisation of insertion sequences

Insertion sequences were characterised by carrying out a BLAST search against the nucleotide collection (nt) database at NCBI (Johnson *et al.*, 2008) and against the IS Finder database (**?**) using default parameters.

3.2.8 Characterisation of quorum sensing genes

For each quorum sensing gene, the nucleotide sequence from PAO1-UW (taken from http://www.pseudomonas.com) was used as the query in a nucleotide BLAST search (BLAST+ version 2.2.29) against the contig sequences of all sequenced isolates. The Perl script blast2fasta.pl (Section 5.7 in Appendix) was used to obtain the nucleotide sequence of the corresponding BLAST hits from each isolate for each gene. The nucleotide sequences were then translated to the protein sequence using the translate-fasta script

of the AMOS (version 3.0.0) package. Protein sequences were aligned using MUSCLE (version 3.8.31), and aligned sequences viewed with Jalview (version 2.8.2). Functional domains and residues in protein sequences were annotated using the NCBI Conserved Domain Search (Marchler-Bauer *et al.*, 2015).

3.3 Results

3.3.1 Identification and collection of *P. aeruginosa* wound isolates

Fourty nine *Pseudomonas aeruginosa* wound isolates were collected by Dr. Mathew Diggle and colleagues at the Department of Clinical Microbiology, Queen's Medical Centre, Nottingham. The quorum sensing (QS) and virulence phenotypes of these isolates have been previously characterised (McKeown, 2013). The isolates were identified as *P. aeruginosa* and selected for collection based on the following criteria: isolates were Gram negative bacilli, oxidase positive, had the correct clonal appearance on cystein lactose electrolyte deficient (CLED) media (a matted colony surface with rough periphery), and produced green pigmentation. Isolates were able to grow on Pseudomonas Isolation Agar (PIA), and a subset of the isolates (PA-W26, 27, 28, 31, 34, 36, 45 and 47) were subjected to 16S ribosomal RNA (rRNA) PCR and sequencing, which confirmed that these were *P. aeruginosa* isolates[3]. The isolates originated from wounds that were co-colonised with other bacteria at the time of sample collection; therefore the *P. aeruginosa* isolates may have only been present in the wounds in a transitory manner before being outcompeted and eradicated from the bacterial infection. A limited amount of basic wound location and patient informa-

[3]This initial characterisation was carried out by Dr. Mathew Diggle and colleagues, Dept of Clinical Microbiology, Queen's Medical Centre, Nottingham

tion was available for these strains, shown in Table 3.2. The 49 strains were isolated from 49 different patients (data not shown).

Table 3.2: *Basic clinical strain information for each isolate. Strain information was not available for isolate PA-W49. Each isolate was sampled from a different patient.*

Isolate	Site of isolation	Age	Gender	Inpatient (IP)/ outpatient (OP)	Ward	Date isolated
PA-W1	Chronic right leg ulcer (Green exudate)	87	M	OP	-	20/06/2012
PA-W2	Spine pressure sore	74	M	IP	D8	20/06/2012
PA-W3	Left leg ulcer	91	F	OP	-	20/06/2012
PA-W4	Left leg ulcer	95	M	OP	-	20/06/2012
PA-W5	Left leg	91	F	OP	-	20/06/2012
PA-W6	Post-op wound	71	M	OP	-	20/06/2012
PA-W7	Skin	4	F	OP	-	20/06/2012
PA-W8	Left shin ulcer	46	M	OP	-	20/06/2012
PA-W9	Left leg ulcer	75	M	OP	-	20/06/2012
PA-W10	Left foot - burns	56	M	IP	B3 (Burns)	21/06/2012
PA-W11	Left leg ulcer	91	M	OP	-	21/06/2012
PA-W12	Toenail	83	F	OP	-	21/06/2012
PA-W13	Swab from wound	74	F	OP	-	21/06/2012
PA-W14	Right foot ulcer	56	F	OP	-	21/06/2012
PA-W15	Right leg skin	91	M	IP	D57	22/06/2012
PA-W16	Back	51	M	IP	Southwell	22/06/2012
PA-W17	Toe	58	M	IP	Papplewick	22/06/2012
PA-W18	Left heel ulcer (pre-op)	54	M	IP	Elizabeth Suite	22/06/2012
PA-W19	Skin, post-op	65	M	OP	-	22/06/2012
PA-W20	Leg ulcer	84	F	OP	-	22/06/2012
PA-W21	Abdomen wound swab	62	F	IP	A&E	23/06/2012
PA-W22	Right leg skin	85	F	IP	D57	25/06/2012
PA-W23	Wound post op	64	F	IP	E12	25/06/2012
PA-W24	Chest	69	M	OP	-	12/06/2012
PA-W25	Abdominal skin	49	F	IP	Morton	13/06/2012
PA-W26	Right leg skin	86	F	OP	-	13/06/2012
PA-W27	CV line	3	F	IP	E38	13/06/2012
PA-W28	Leg	81	F	OP	-	13/06/2012
PA-W29	Right foot	74	M	IP	Patience1	13/06/2012
PA-W30	Skin ulcer	64	M	OP	-	13/06/2012
PA-W31	Toe	61	M	OP	-	13/06/2012
PA-W32	Ankle	81	F	OP	-	13/06/2012
PA-W33	Groin	89	F	OP	-	13/06/2012
PA-W34	Right shin ulcer	81	M	OP	-	13/06/2012
PA-W35	Wound swab (foot)	77	M	OP	-	13/06/2012
PA-W36	Leg ulcer	71	M	OP	-	13/06/2012
PA-W37	Ulceration between toes	74	M	OP	-	13/06/2012
PA-W38	Right leg	82	F	OP	-	14/06/2012
PA-W39	Wound site not given	74	M	OP	-	14/06/2012
PA-W40	Toe	79	F	OP	-	14/06/2012
PA-W41	Left leg skin	65	F	PO	-	14/06/2012
PA-W42	Toe	84	M	IP	B49	14/06/2012
PA-W43	Left shin	31	M	OP	-	15/06/2012
PA-W44	Leg ulcer	70	F	OP	-	15/06/2012
PA-W45	Heel	91	M	OP	-	15/06/2012
PA-W46	Heel	86	M	OP	-	15/06/2012
PA-W47	Left calf, burns, failed graft	93	M	IP	B3 (Burns)	16/06/2012
PA-W48	Leg ulcer skin	37	M	IP	Southwell	16/06/2012

3.3.2 *De novo* assembly and gene annotation

The raw Illumina data for each isolate was assembled into contigs using the A5-miseq assembly pipeline. Gene prediction and annotation was carried out using the RAST web server. The assembly and gene annotation statistics are shown in Table 3.3. The assembly sizes range from 6.2 to 7.57 Mb, and the number of predicted gene sequences range from 5836 to 7040. Isolate PA-W40 was found to have the largest genome assembly size and the largest number of annotated genes, however this was also the most fragmented genome assembly.

3.3.3 Phylogenetic analysis

Harvest was used to analyse the core genomes of the 49 wound isolates, as well as a few isolates from the previously determined major subgroups of *P. aeruginosa* (Stewart *et al.*, 2014). The resulting core genome phylogeny is shown in Figure 3.2. As was previously seen by Stewart *et al.* (2014), the 56 *P. aeruginosa* sequences form three major groups, with PA7 and VRFPAO1 forming the distantly related group 3. Isolates PAO1, LESB58, DK2 and PA14 fit into the same groups as the Stewart *et al.* (2014) analysis. PANOT106, an isolate from a six year long hospital outbreak in City Hospital, Nottingham (Snyder *et al.*, 2013), fits into group 1 of the core genome phylogeny, and appears to be most closely related to PA-W19 in terms of core genome SNPs. 48 of the 49 wound isolates fit into groups 1 and 2 of the phylogeny, whereas isolate PA-W1 appears to form its own distantly related group based on this phylogeny. Phylogenetic analysis of a set of 390 *P. aeruginosa* isolates, including these 49 isolates, shows that PA-W1 fits into group 3 alongside PA7, VRFPAO1 and a number of additional *P. aeruginosa* strains (Freschi *et al.*, 2015). This group does, however, remain the least populated of the three groups, containing 12/390 strains. 16/49 isolates were found to be contained within four clades of the phyloge-

Isolate	Assembly size (Mb)	No. of contigs	Avg. contig size (kb)	Largest contig (kb)	N50 (kb)	N50(n)	Gaps (N)	No. of predicted genes
1	6.31	36	175.32	905.05	489.65	5	0	5868
2	6.51	44	147.90	937.78	396.33	6	0	6059
3	6.76	61	110.89	775.16	212.63	10	0	6348
4	6.79	87	78.05	579.50	230.71	10	0	6347
5	7.07	47	150.48	775.41	461.44	6	0	6675
6	6.30	31	203.07	886.80	425.37	5	0	5880
7	7.24	71	102.00	662.79	255.86	9	0	6723
8	6.39	24	266.17	1754.69	670.68	3	0	5948
9	6.54	75	87.25	715.18	475.31	6	0	6105
10	6.71	47	142.70	688.58	360.94	7	0	6287
11	6.63	25	265.09	1411.33	778.21	4	0	6217
12	6.34	34	186.40	1098.17	663.09	4	0	5892
13	6.23	31	200.98	1265.79	928.78	3	0	5836
14	6.79	71	95.57	543.37	239.28	9	0	6330
15	6.45	33	195.52	909.04	426.92	5	0	6044
16	6.51	27	241.06	958.19	483.65	5	0	6092
17	6.74	55	122.57	704.36	340.38	7	0	6263
18	6.92	56	123.58	1164.70	530.59	5	0	6498
19	6.45	34	189.68	921.11	659.18	4	0	5958
20	6.70	46	145.72	743.17	351.33	6	0	6228
21	6.79	41	165.71	1574.46	409.25	5	0	6368
22	6.30	28	225.10	910.63	508.39	5	16	5868
23	6.61	51	129.69	773.87	328.32	7	0	6176
24	6.97	68	102.43	517.57	216.34	11	0	6558
25	6.79	47	144.49	865.36	421.48	6	0	6381
26	6.74	36	187.14	1111.61	467.02	5	0	6263
27	6.74	38	177.28	859.24	467.13	6	0	6259
28	6.78	43	157.72	898.49	424.71	6	0	6225
29	6.46	43	150.27	729.15	317.26	6	0	6005
30	6.81	32	212.88	994.45	455.19	5	0	6391
31	7.07	51	138.55	867.57	567.95	6	0	6662
32	6.88	73	94.26	877.82	256.58	9	0	6374
33	6.98	46	151.85	1100.26	388.15	6	0	6470
34	6.58	51	129.11	557.42	303.69	9	0	6194
35	6.44	28	229.84	1615.67	396.53	5	0	6029
36	7.04	83	84.80	755.97	271.68	9	0	6616
37	6.32	49	129.03	734.55	347.30	6	0	5896
38	6.80	52	130.75	602.83	423.06	7	0	6377
39	7.06	51	138.35	649.94	335.30	8	0	6655
40	7.57	139	54.46	514.34	170.99	14	0	7040
41	6.70	58	115.53	1467.55	298.13	5	0	6276
42	7.08	49	144.43	867.67	432.82	6	0	6677
43	6.86	51	134.48	713.18	306.34	7	0	6332
44	6.39	36	177.41	1296.60	425.20	5	0	5924
45	6.51	45	144.56	786.92	471.54	6	0	6101
46	6.40	51	125.56	910.73	382.28	6	0	5957
47	6.82	72	94.72	424.19	224.87	11	0	6392
48	7.00	83	84.31	692.25	281.29	9	0	6502
49	6.33	48	131.79	773.84	348.64	7	0	5891

Table 3.3: *Assembly and gene prediction statistics for each isolate. N50 is the length (N) where 50% of bases in the genome assembly are found within contigs of length N or less.*

netic tree, indicating the presence of four clusters of highly related isolates within the sample of 49 isolates (red brackets in Figure 3.2). No obvious patterns were observed between these subgroups and the wound type, site of location or patient information when compared with information in Table 3.2.

It has previously been found that the 49 isolates display a great level of diversity in terms of their QSSM production, as can been seen in Figure 3.1 (McKeown, 2013). The QSSM production data from this study was mapped onto the phylogenetic tree (Figure 3.3), however no correlation was seen between phylogenetic group and QS phenotype. Although only data for QSSM production after 24 hours is shown, for simplicity, there were also no patterns observed between QSSM production after 8 hours and phylogenetic grouping.

Figure 3.2: Core genome phylogeny of the 49 P. aeruginosa wound isolates. *Unrooted maximum likelihood tree based on core genome SNPs (as determined by Parsnp) of the 49 P. aeruginosa wound isolates and 7 publicly available P. aeruginosa genomes. The tree was calculated using FastTree2 with 100 bootstraps.* ***A****: Core genome phylogeny where group three has been truncated for the purposes of the figure. Publicly available genome sequences are shown in dark blue bold font. Bootstrap values for nodes were 100 unless otherwise shown. Some bootstrap values in group 1 are shown in colour for clarity.* ***B****: Zoomed out view of the tree in* ***A****. Clades of related isolates, in terms of core genome phylogeny, are marked in red brackets.*

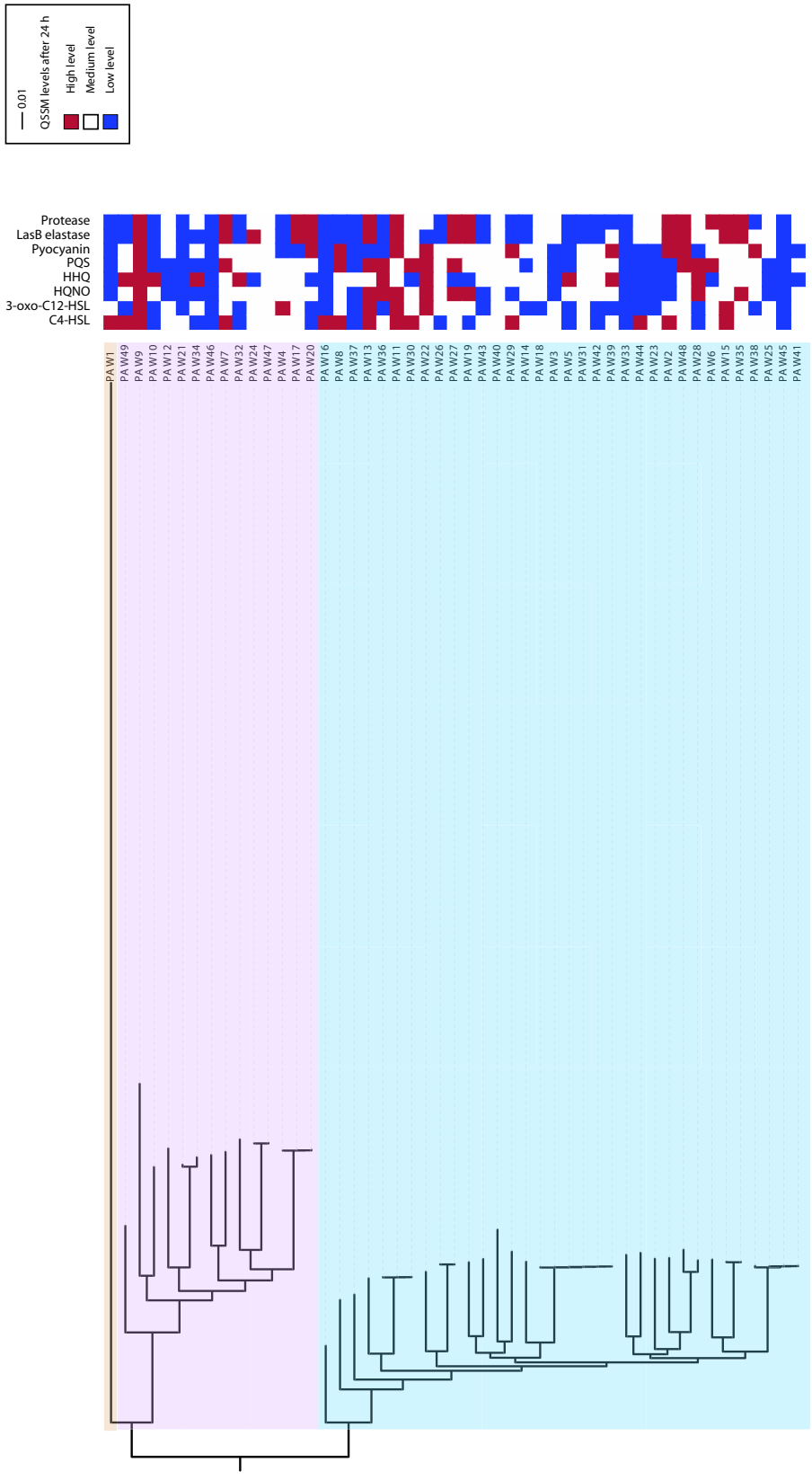

Figure 3.3: The mapping of QSSM levels of the 49 isolates onto the phylogenetic tree. *The unrooted core genome phylogeny of 49 isolates sequenced in this study has been rooted arbitrarily for clarity. The three major groups are coloured as Figure 3.2 (Group 1 in blue, Group 2 in pink, Group 3 in peach), with PA-W1 now coloured as Group 3. QSSM levels after 24 hours are shown, with high levels shown in red, medium levels in white, and low levels in blue. Figure created using iTOL (Letunic and Bork, 2011).*

3.3.4 Comparison with complete reference sequences

Mapping reads from each isolate to the PAO1-UW reference sequence revealed that between 94.88 and 98.7% of the genome sequence was covered by at least 5 reads from each isolate. A BLAST comparison of genome assembly contigs of all isolates to the PAO1-UW genome sequence can be seen in Figure 3.4. It can be immediately seen that there is one region which is variable compared to PAO1-UW in all isolates apart from PA-W14, PA-W28 and PA-W48; namely *wbp* genes (genes PA3146-PA3160), which encode the B-band O antigen (O-Ag). Two other regions which were variable in most isolates were genes PA2221-PA2231, which contains the gene *pslA*, which is involved the synthesis of Psl (*Pseudomonas* synthetic locus) exopolysaccharide, which is involved in surface attachment (Byrd *et al.*, 2010), and a region including genes PA1381-PA1393, which contains three predicted glycosyl transferases (PA1389 - PA1391). These three glycosyl transferases were variable in all isolates except PA-W13, PA-W28 and PA-W48.

A large number of differences were seen between these 49 isolates and the PAO1-UW reference sequence. The remainder of this thesis chapter will focus on differences in key quorum sensing genes in this set of isolates.

3.3.5 Analysis of *las* quorum sensing genes

Analysis of mutations which lead to a change in protein sequence, as well as SNPs in intergenic regions of the *las* quorum sensing genes in the 49 isolates revealed a high number of mutations within the *lasR* gene and the upstream intergenic region, with a low number of mutations in the *lasI* gene (Figure 3.5). Interestingly, none of the 49 isolates were found to contain a non-synonymous mutation in the *rsaL* gene. The key mutations in the *las* quorum sensing genes and the 3-oxo-C_{12}-HSL production for each isolate can be seen in Table 3.4. There are a number of mutations which correlate with a reduced production of 3-oxo-C_{12}-HSL by the isolates compared to

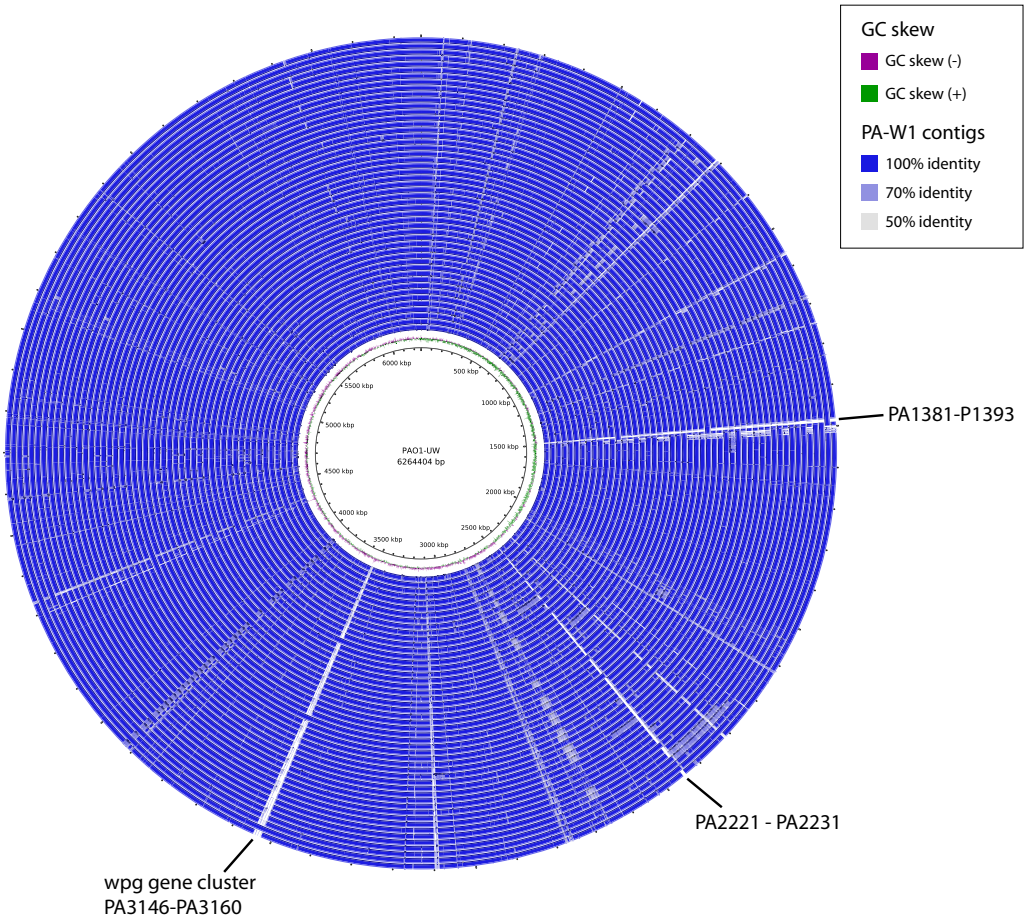

Figure 3.4: Whole genome comparison of the 49 sequenced isolates with PAO1-UW. *Contigs from the 49 P. aeruginosa isolates are shown in blue, with the innermost ring corresponding to PA-W1, and the outermost ring corresponding to PA-W49. GC skew is shown on the innermost ring in purple and green. Figure generated using BRIG (Alikhan* et al.*, 2011).*

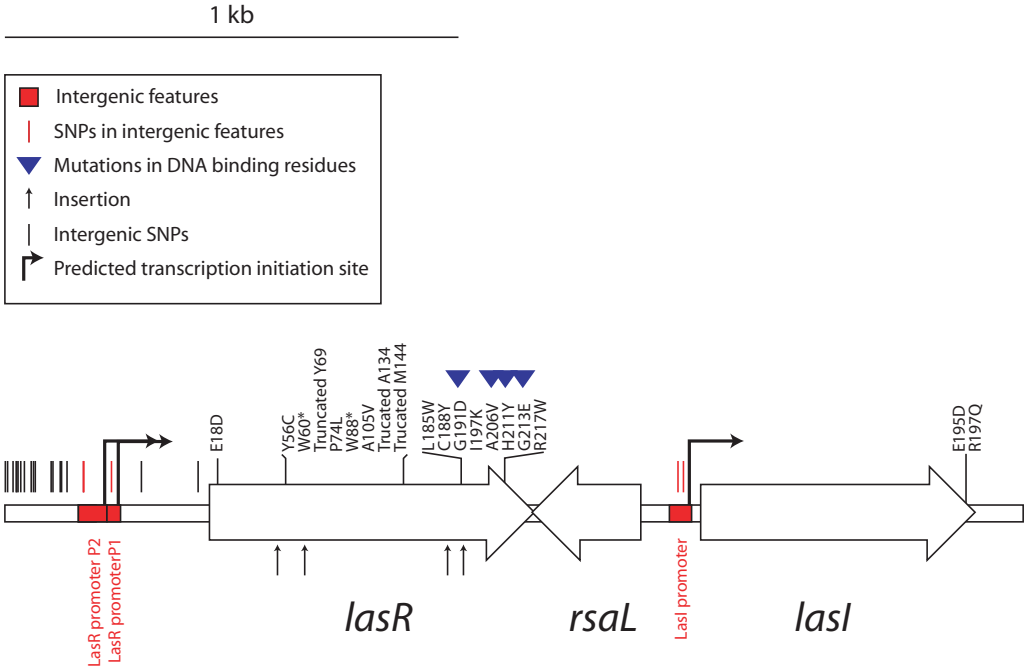

Figure 3.5: SNPs in the las quorum sensing genes of P. aeruginosa in the 49 sequenced isolates. *Arrows correspond to genes, with the direction of the arrow indicating the relative orientation of the gene in the chromosome. Intergenic regions are indicated by the rectangle running across the arrows. Promoters are coloured in red, with SNPs in these intergenic features shown as red vertical lines. Mutations in annotated DNA binding residues are indicated by blue triangles. The upward arrows correspond to sites where large insertions occur within in the lasR gene in six of the 49 isolates.*

PAO1. These are caused by *lasR* mutations such as insertions, out of frame deletions, and large deletions encompassing all quorum sensing genes. Mutations in both *lasR* and *lasI* promoters did not appear to correlate with a reduction in 3-oxo-C_{12}-HSL production in this set of isolates.

3.3.6 Out of frame deletions in *lasR*

Isolates PA-W8, 33, 34, 46 and 47 were found to contain out of frame deletions in the *lasR* gene, and in all five isolates a reduction in production of 3-oxo-C_{12}-HSL was observed compared to PAO1 (Table 3.4). Of note, isolates PA-W46 and 47, and isolates PA-W33 and 34 were found to contain identical out of frame deletions, however were not genetically identical in terms of core genome SNPs. PA-W46 and 47 both fall into group 2 in the

Table 3.4: 3-oxo-C_{12}-HSL production and las quorum sensing system mutations in the 49 isolates. *Isolates were categorised as producing high (red), medium (white) or low (blue) levels of each QSSM after 8 and 24 hours of incubation, as described in (McKeown, 2013).*

Isolate	3-oxo-C_{12}-HSL		Mutations			
	8 hr	24 hr	LasR	RsaL	LasI	Intergenic
1			E18D		E195D, R197Q	lasI promoter
2						
3						
4						
5						
6						
7						*lasR* promoter P1
8			Y69 out of frame deletion			*lasR* promoter P2
9						*lasR* promoter P1
10			P74L			*lasR* promoter P1
11						
12			A206V			lasR promoter P1
13			R217W			
14			A59*			
15						*lasI* promoter
16			C188Y			
17						
18			Insertion sequence			
19						
20						
21			Insertion sequence			*lasR* promoter P1
22			A105V			
23			G213E			
24						*lasR* promoter P1
25						
26			Y56C			
27			Y56C			*lasR* promoter P1
28			L185W			
29						
30						
31			Insertion sequence			
32			Deletion of entire las system	Deletion of entire las system	Deletion of entire las system	
33			A134 out of frame deletion			
34			A134 out of frame deletion			*lasR* promoter P1
35						lasI promoter
36			Insertion sequence			
37			G191D			
38			A59*			
39			Insertion sequence			
40						
41						
42			Insertion sequence			
43			H211Y			
44						
45			F87*			
46			M144 out of frame deletion			*lasR* promoter P1
47			M144 out of frame deletion			*lasR* promoter P1
48						*lasR* promoter P1
49			I197K			*lasR* promoter P1
PAO1						

core genome phylogeny (Figure 3.2), but have greater similarity to other isolates than each other in terms of core genome SNPs. PA-W33 and 34 are found in into two distinct phylogenetic groups and are therefore very different from one another genetically, in terms of core genome SNPs. These were isolates from separate patients and different wound types (Table 3.2).

3.3.7 Mutations in DNA binding residues of LasR

Isolates PA-W12, 23, 37 and 43 were found to contain non-synonymous mutations which would lead to changes in DNA binding residues in the LasR protein. These isolates displayed a reduction in 3-oxo-C_{12}-HSL production compared to PAO1, indicating that these mutations interfere with the function of LasR (Table 3.4). In PA-W23, 37 and 43, the production of all other measured QSSMs was also reduced compared to PAO1 (Table 3.5).

Table 3.5: Heat map of the relative QSSM levels found in supernatants of isolates with mutations in DNA binding residues of LasR. *Figure created using data from (McKeown, 2013). Isolates were categorised as producing high (red), medium (white) or low (blue) levels of each QSSM after 8 and 24 hours of incubation, as described in (McKeown, 2013).*

Isolate	C4-HSL		3-oxo-C12-HSL		HQNO		HHQ		PQS		Pyocyanin		LasB elastase		Protease		Isolate	Genotype LasR Protein
	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr		
12																	12	A206V
23																	23	G123E
37																	37	G191D
43																	43	H211Y
PAO1																	PAO1	

Table 3.6: Heat map of the relative QSSM levels found in supernatant of PA-W32, an isolate with a 68 kb deletion encompassing the las quorum sensing system *Figure created using data from (McKeown, 2013). Isolates were categorised as producing high (red), medium (white) or low (blue) levels of each QSSM after 8 and 24 hours of incubation, as described in (McKeown, 2013).*

Isolate	C4-HSL		3-oxo-C12-HSL		HQNO		HHQ		PQS		Pyocyanin		LasB elastase		Protease		Isolate	Genotype LasR Protein
	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr		
32																	32	
PAO1																	PAO1	

Table 3.7: Heat map of the relative QSSM levels found in supernatants of isolates with large insertions in lasR. *Figure created using data from (McKeown, 2013). Isolates were categorised as producing high (red), medium (white) or low (blue) levels of each QSSM after 8 and 24 hours of incubation, as described in (McKeown, 2013).*

Isolate	C4-HSL		3-oxo-C12-HSL		HQNO		HHQ		PQS		Pyocyanin		LasB elastase		Protease		Isolate	Genotype LasR Protein
	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr	8 hr	24 hr		
18																	18	Insertion sequence
21																	21	Insertion sequence
31																	31	Insertion sequence
36																	36	Insertion sequence
39																	39	Insertion sequence
42																	42	Insertion sequence
PAO1																	PAO1	

3.3.8 Deletion of *las* quorum sensing genes

One of the isolates, PA-W32, was found to be missing the *lasR*, *lasI* and *rsaL* genes, due to a 68 kb deletion which encompassed these genes (Figure 3.7). This was confirmed by mapping the contigs from the genome assembly for isolate PA-W32 to the PAO1-UW reference sequence using Abacas (**?**). Contigs 8 and 22 were found to map either side of the putative deletion, up to and including genes PA1382 and PA1435, respectively, however no contigs were found to map to the deleted region. This isolate did not produce 3-oxo-C_{12}-HSL, and produced low levels of other QSSMs compared to PAO1, with the exception of HHQ (Table 3.6).

3.3.9 *lasR* insertions

In six of the isolates (PA-W18, 21, 31, 36, 39 and 42) the *lasR* nucleotide sequence was found to be split across two contigs. In previous work, PCR amplification of the *lasR* gene did not yield a product for these particular isolates and PA-W32, whereas PCR products were obtained for all other isolates tested (Figure 3.6). Ends of contig pairs across which the *lasR* gene was split in these six isolates were used as a query in an BLAST search against the nucleotide collection (nt) database at NCBI. These results revealed the presence of transposase sequences before and after the split *lasR* genes fragments in these isolates. This is indicative of a transposon insertion in the *lasR* gene in isolates PA-W18, 21, 31, 36, 39 and 42. These results were confirmed by carrying out a BLAST search against the IS Finder database. These six isolates displayed reduced 3-oxo-C_{12}-HSL production compared to PAO1 (Table 3.7). All six of these isolates, apart from PA-W18 also showed reduction in C_4-HSL, LasB elastase and protease production compared to PAO1 (Figure 3.1). No differences were observed in the antibiotic resistance profiles of these six isolates compared to the remaining 43 (Figure **??**, Appendix).

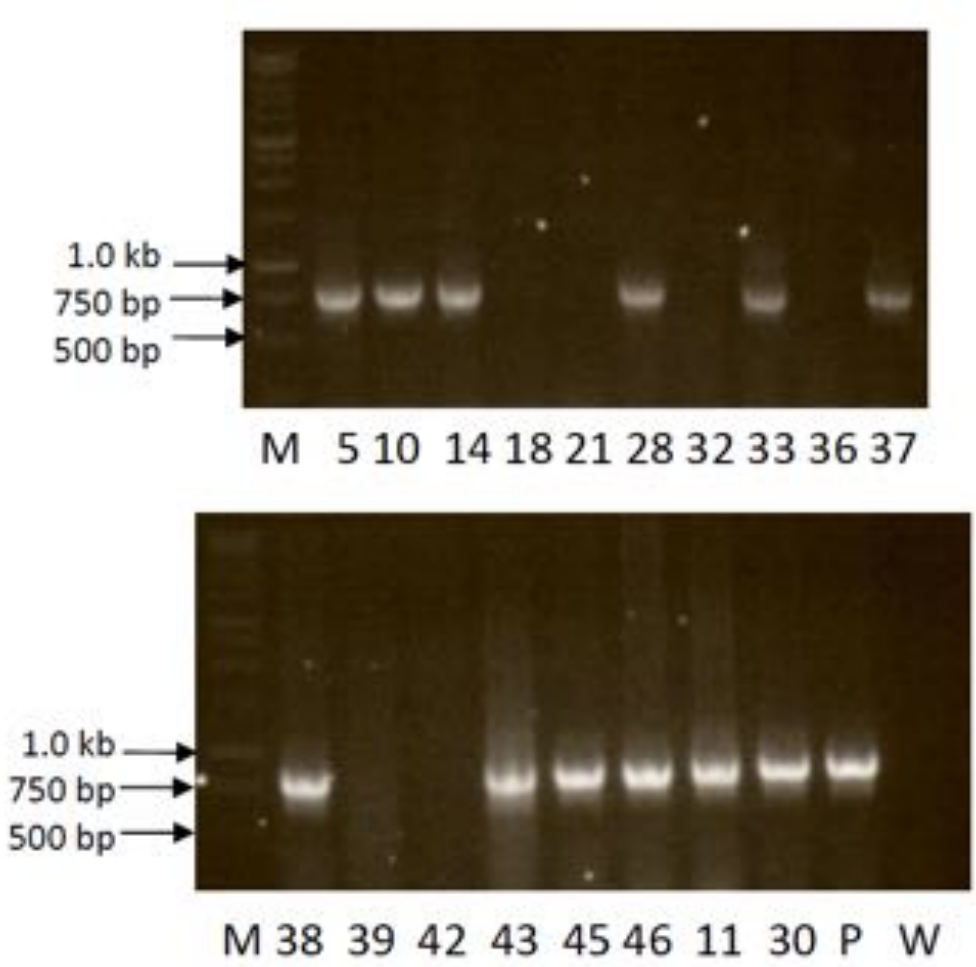

Figure 3.6: PCR of the lasR gene in a subset of the 49 isolates. *P = PAO1 positive control, W = water template-negative control, M = 1.0 kb DNA marker. Figure 5.7B in (McKeown, 2013).*

Figure 3.7: A 68 kb deletion in PA-W32 encompasses the las quorum sensing genes of P. aeruginosa. *Coverage of whole genome sequencing reads from PA-W32 are shown in the top track. The coding sequences in PAO1-UW at the deletion region are shown in the middle track. The coding sequences shown in green, and marked by the green arrow, correspond to LasR, RsaL and LasI. Genetic coordinates are with respect to PAO1-UW (NC_002516).*

The ends of contig pairs across which the *lasR* gene was split in each isolate were used as a query in a BLAST search against the contig sequences of the respective isolate, in order to find a putative insertion sequence within the genome assembly for each isolate (Figure 3.8). Contigs were considered for further characterisation if *lasR* insertion fragments at the ends of contig pairs (shown in blue in Figure 3.8) had a BLAST hit with greater than 99% sequence identity to two ends (or two different regions) of a contig in the genome assembly.

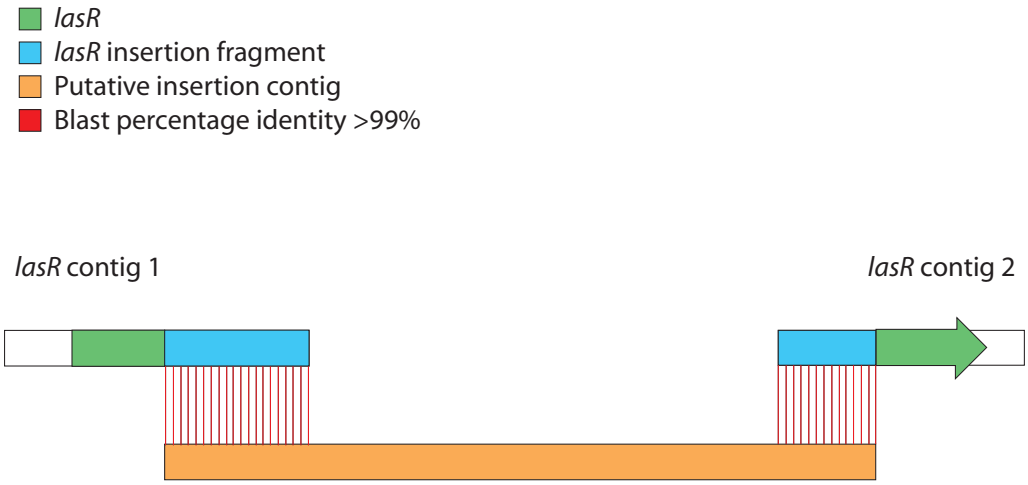

Figure 3.8: Diagram showing first step of characterisation of putative lasR insertion sequences in six of the P. aeruginosa isolates (not to scale). *The two contigs across which the lasR gene is split in PA-W18, 21, 31, 36, 39 and 42 are shown on the top row and labelled "lasR contig 1" and "lasR contig 2". The lasR gene fragments on these contigs are shown in green and the fragments of the insertion sequence on these contigs are shown in blue. A BLAST search was carried out with these lasR contigs against the remaining contigs in the genome assembly. Contigs with BLAST hits of >99% sequence identity to the lasR contigs, as shown in the diagram, were deemed to correspond to the putative lasR insertion sequence.*

These putative insertion contigs, as well as the *lasR* insertion fragments at the ends of contig pairs, were then used as the query sequences in BLAST searches of both the nucleotide collection (nt) database at NCBI, and the IS Finder database. Only those contigs where both NCBI BLAST and IS Finder results matched those of the *lasR* insertion fragments were considered to correspond to the putative insertion sequence in each isolate. The results of the characterisation of the putative *lasR* insertion sequences in these six

isolates can be seen in Table 3.9.

In all isolates except PA-W31 and PA-W36, this resulted in a single contig being identified as a putative insertion sequence. For isolate PA-W31, there were no contigs which fit the BLAST criteria shown in Figure 3.8. However, BLAST results for each *lasR* insertion fragment contained the other *lasR* insertion fragment, i.e. *lasR* insertion fragment 1 had a BLAST hit to *lasR* insertion fragment 2, and vice versa. The overlapping sequence was found to contain a transposase gene, and had sequence similarity to a known *P. aeruginosa* insertion sequence (Table 3.9).

In the case of isolate PA-W36, there are multiple contigs which could correspond to the insertion sequence, according to the aforementioned characterisation method. All of these contigs had the same best NCBI BLAST hit, namely to the insertion sequence RP10 in *P. stutzeri*, and all had the same result with IS Finder (Table 3.9). The size of the insertion sequence is unclear, as the possible insertion sequence contigs, where there is a BLAST hit with 100% sequence identity between the start and end of the contig and both *lasR* insertion fragments in all cases, range in size from 685 bp to 138.3 kb (Table 3.8).

Table 3.8: Sizes of the putative lasR insertion sequences in PA-W36.

Scaffold	Scaffold size (bp)	Region with BLAST hit to *lasR* insertion fragment 1	Region with BLAST hit to *lasR* insertion fragment 2
Scaffold 17	138,310	1-297	138,105-138,310
Scaffold 18	127,885	1-296	127,609-127,885
Scaffold 24	124,043	1-297	123,766-124,043
Scaffold 37	24,007	1-278	23,712-24,007
Scaffold 75	1,113	1-297	817-1095
Scaffold 80	685	1-294	409-685

Table 3.9: Characteristics of the putative lasR insertion sequences.

Isolate	Size of putative insertion (bp)	Number of predicted genes	Types of genes	NCBI BLAST result	Accession	Percentage Identity	Query coverage	IS finder result
PA-W18	36772	29	Multiple copper/heavy metal resistance genes	*P. putida* H8234, complete genome	CP005976	98	99	*P. aeruginosa* ISPa42
PA-W21	1325	2	Two mobile element proteins	*Bordetella petrii* strain DSM 12804, complete genome	AM902716.1	99	100	*Ralstonia metallidurans* ISRme15
PA-W31	1505	1	One transposase	*P. aeruginosa* strain Ps133, ISPa45	JX440361.1	100	100	*P. aeruginosa* ISPa45
PA-W36	-	-	-	*P. stutzeri* A1501 complete genome	CP000304.1	-	-	*Pseudomonas* sp. ISPsp8
PA-W39	1505	1	One transposase	*P. aeruginosa* strain Ps133, ISPa45	JX440361.1	100	100	*P. aeruginosa* ISPa45
PA-W42	1505	1	One transposase	*P. aeruginosa* strain Ps133, ISPa45	JX440361.1	100	100	*P. aeruginosa* ISPa45

The locations of the putative insertion sequences in *lasR* in these six isolates is shown in Figure 3.9. All insertions were found to occur either within or bordering predicted functional domains in the LasR protein sequence, and a loss of function of the LasR protein in these isolates is indicated by the reduction in the production of acylhomoserine lactones (AHLs), and the loss of elastase and protease activity (Table 3.7). It was found that isolates PA-W31, 39 and 42 contain an insertion at the same position in *lasR*.

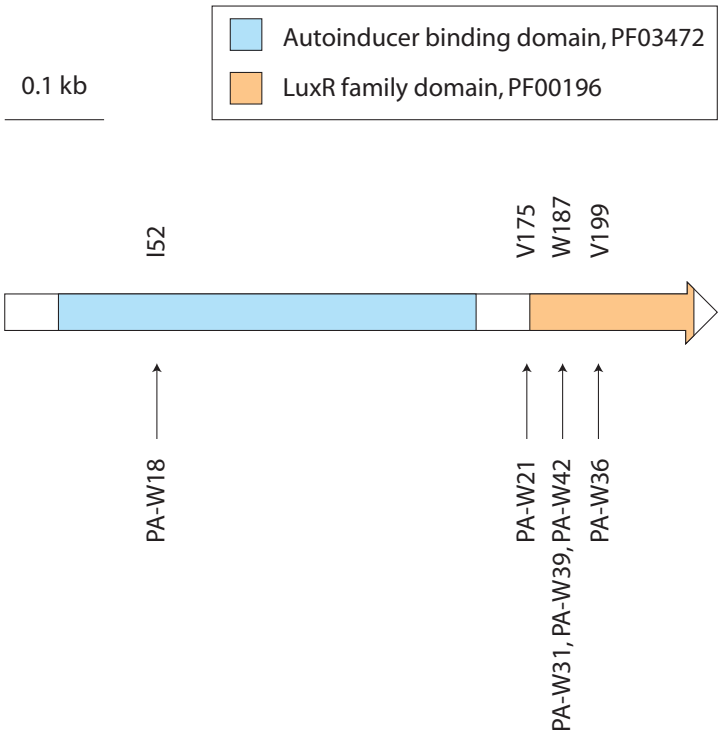

Figure 3.9: Locations of putative insertions in lasR in six of the isolates. *The sequence of lasR is indicated by the arrow, and the filled regions correspond to the predicted autoinducer binding domain (light blue) and the predicted LuxR family domain, which is the DNA binding domain of the protein (light orange). The arrows indicate the location of the putative insertions within the sequence, with the isolates in which the insertion is found shown below the arrow, and the amino acid after/within which the insertion occurs shown above the arrow.*

3.3.10 Analysis of *rhl* quorum sensing genes

Analysis of non-synonymous mutations, as well as SNPs in intergenic regions of the *rhl* quorum sensing system in the 49 isolates revealed mutations in both *rhlI* and *rhlR* genes, as well as in the *rhlR* promoter (Figure 3.10).

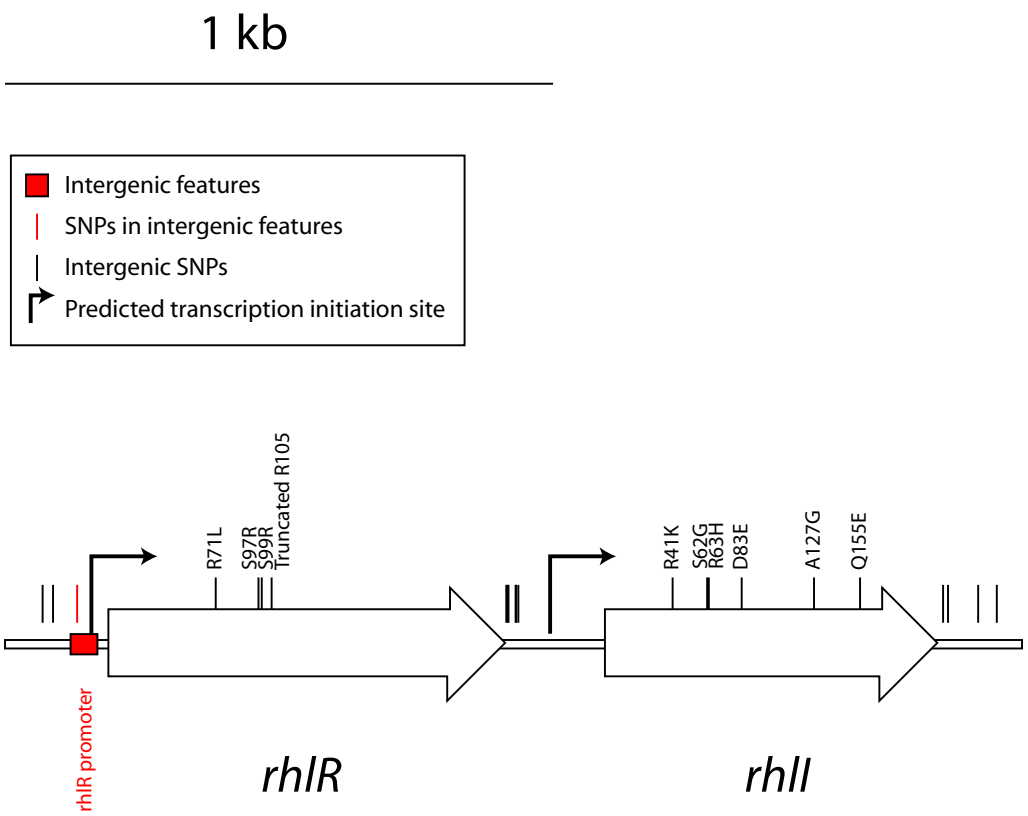

Figure 3.10: SNPs in the rhl quorum sensing genes of P. aeruginosa in the 49 sequenced isolates. *Arrows correspond to genes, with the direction of the arrow indicating the relative orientation of the gene in the chromosome. Intergenic regions are indicated by the rectangle running across the arrows. Promoters are coloured in red, with SNPs in these intergenic features shown as red vertical lines.*

A comparison of these mutations with the associated C_4-HSL phenotype is shown in Table 3.10.

The DNA binding domain of RhlR occurs after the 180th residue in the protein, and so the only mutation observed in these isolates which would likely affect function is the truncation of RhlR in PA-W45, which also correlates with a loss of C_4-HSL production in this isolate. Isolate PA-W10 contains a R71L substitution in RhlR and displays low levels of C_4-HSL production.

There are no annotations of functional residues in RhlI and so there are no clear loss of function mutations in this protein. There are, however mutations which appear to correlate with a reduction in C_4-HSL production, and a number of mutations which do not. The S62G mutation in RhlI is

Table 3.10: C_4-HSL production and rhl quorum sensing system mutations in the 49 isolates. *Isolates were categorised as producing high (red), medium (white) or low (blue) levels of each QSSM after 8 and 24 hours of incubation, as described in (McKeown, 2013).*

Isolate	C_4-HSL		Mutations		
	8 hr	24 hr	RhlI	RhlR	Intergenic
1			R63H, Q155E		*rhlR* promoter P1
2					
3					
4			S62G	S97R	
5					
6					*rhlAB* minor promoter
7			A127G, S62G		
8			D83E		
9			D83E		
10			R41K	R71L	
11					
12					
13					
14				S99R	
15					
16					
17			S62G	S97R	
18					
19					
20			S62G	S97R	
21					
22					
23					
24			R41K		
25					
26					
27					
28					
29					
30					
31					
32			S62G		
33					
34					
35					
36					
37					
38					
39					
40					
41					
42					
43					
44					
45				R105 truncated	
46					
47			R41K		
48					
49					
PAO1					

conserved between four isolates (PA-W4, 7, 17 and 20) and these isolates display a range of C_4-HSL production levels. The same is true of the R42K mutation, conserved between PA-W10, 24 and 47, however these do all display mid-low levels of C_4-HSL production. PA-W24 and 47 contain only the R42K mutation, and both display medium levels of C_4-HSL production, whereas PA-W10 also contains the R71L RhlR mutation and displays low levels of C_4-HSL production. The mutation D83E in RhlI is conserved between isolates PA-W8 and PA-W9. Both of these isolates display an identical low level of C_4-HSL production at 8 hours, followed by full C_4-HSL production at 24 hours; namely a delay in C_4-HSL production.

3.3.11 Analysis of *pqs* genes

Analysis of mutations which lead to a change in protein sequence, as well as SNPs in intergenic regions of the *pqs* genes in the 49 isolates revealed a large number of mutations in all genes of the *pqsABCDE* operon, however there were strikingly few mutations seen in the *pqsE* gene (Figures 3.11 and 3.12). A comparison of these mutations with the associated HHQ and PQS phenotypes are shown in Table 3.11.

Of all the mutations in the *pqs* genes in these isolates, two were found to occur at known dimer interface residues (Figure 3.11). These were the D109A mutation in *pqsB* in PA-W1 and the C93* premature stop codon in *pqsD* PA-W3. Both of these isolates display a loss of production of HHQ, and therefore also PQS (Table 3.11). Apart from PA-W1 and PA-W3, a number of isolates displayed a low level production of HHQ and PQS compared to PAO1, namely isolates PA-W16, 21, 23, 33, 44, 45 and 46. These isolates contained a number of mutations in *pqs* genes which did not occur at known functional residues. Isolate PA-W48 was the only isolate to show an increased production of HHQ and PQS compared to PAO1, and the only mutation found in this isolate is a E267A substitution in PqsA.

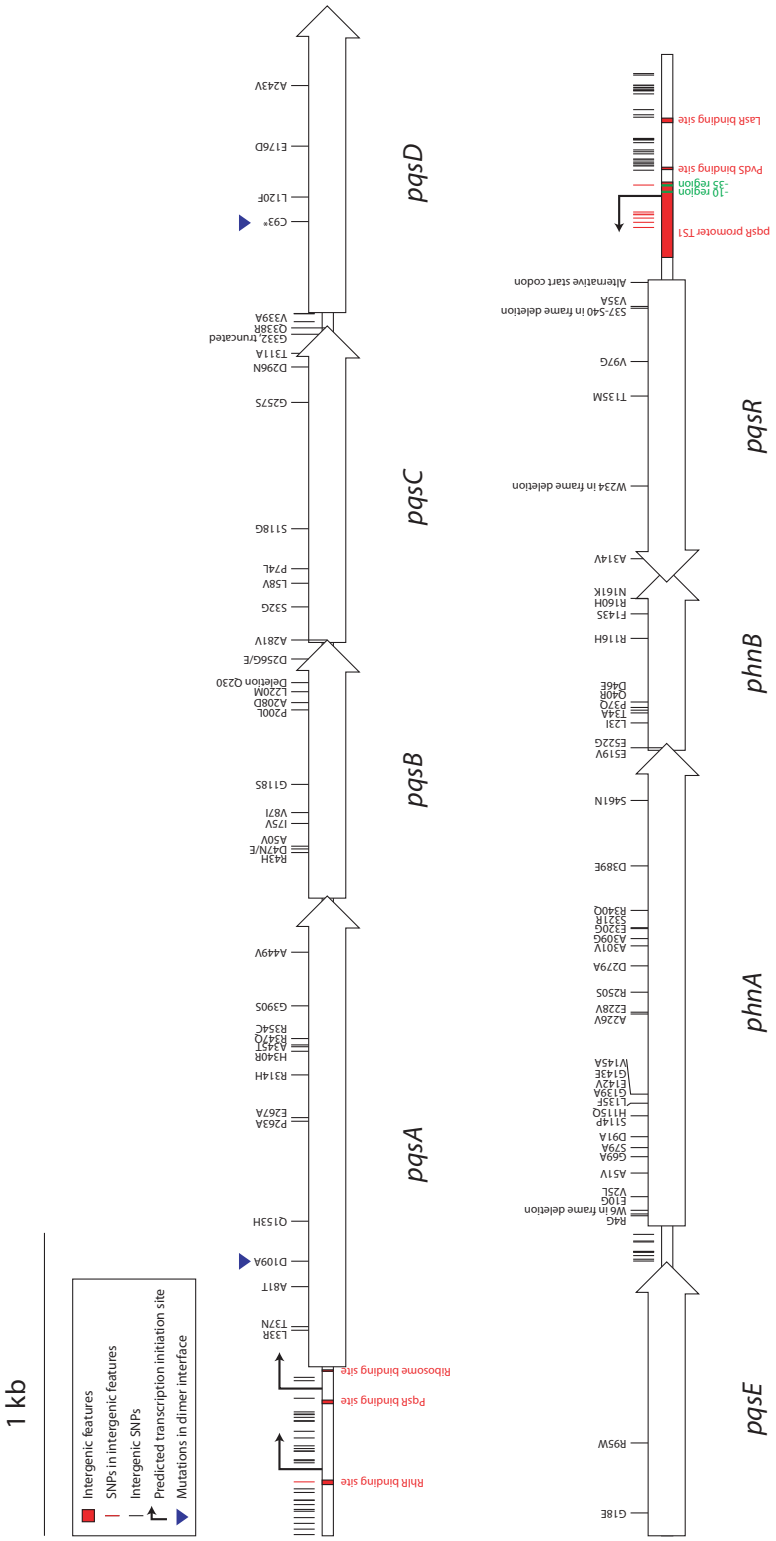

Figure 3.11: SNPs in the pqs genes of P. aeruginosa in the 49 sequenced isolates. *Arrows correspond to genes, with the direction of the arrow indicating the relative orientation of the gene in the chromosome. Intergenic regions are indicated by the rectangle running across the arrows. Promoters are coloured in red, with SNPs in these intergenic features shown as red vertical lines. Mutations in annotated dimer interface residues are indicated by blue triangles.*

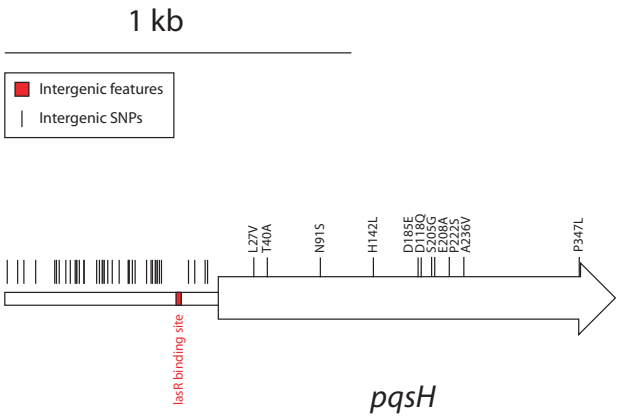

Figure 3.12: SNPs in PqsH of *P. aeruginosa* in the 49 isolates. The arrow correponds to the pqsH gene, with the direction of the arrow indicating the relative orientation of the gene in the chromosome. The intergenic region is indicated by the rectangle running through the arrow. The *lasR* binding site is coloured in red. Mutations in the intergenic regions are indicated by the black vertical lines above the intergenic region.

Table 3.11: HHQ and PQS production and pqs quorum sensing system mutations in the 49 isolates. *Isolates were categorised as producing high (red), medium (white) or low (blue) levels of each QSSM after 8 and 24 hours of incubation, as described in (McKeown, 2013). Mutations which were found in 4 or more isolates which did not share a common phenotype and were not associated with a putative functional residue were omitted from this table for clarity.*

Isolate	HHQ		PQS		Isolate	Mutations							
	8 hr	24 hr	8 hr	24 hr		PqsA	PqsB	PqsC	PqsD	PqsE	PqsH	PqsR	Intergenic
1					1	L33R, D109A, Q153H, H340R	D47E, I75V, L220M	L58V, P74L, S118G			L27V, H142L, D185E, L188Q, S205G, P222S		*pqsR* promoter
2					2	G390S				R95W			
3					3				C93*				
4					4								
5					5								
6					6			4bp deletion after G332		G18E	D185G		
7					7	A81T, E267A							
8					8								*pqsR* promoter
9					9		A208D				L188Q		
10					10						L188Q		
11					11		D47N						
12					12		R43H, A208D, in frame deletion Q230	T311A, Q338R				V35A	
13					13		G118S				P347L		
14					14								
15					15	R354C					T40A		
16					16			D296N				W234 missing	
17					17								
18					18								
19					19								
20					20								
21					21	A449V						Alternative start codon	*pqsR* promoter -35 region
22					22			S32G					
23					23			D296N, V339A				S37-S40 missing	
24					24								*pqsR* promoter
25					25								
26					26								
27					27								
28					28			G257S					
29					29	R354C							
30					30		D47N						
31					31								
32					32		A208D, A281V						
33					33							V97G	
34					34	A449V							*pqsR* promoter -35 region
35					35	R354C					T40A		
36					36		D47N						
37					37								
38					38								
39					39								
40					40	R314H					E208A		

Continued on next page

Table 3.11 – *Continued from previous page*

Isolate	HHQ		PQS		Isolate	Mutations							
	8 hr	24 hr	8 hr	24 hr		PqsA	PqsB	PqsC	PqsD	PqsE	PqsH	PqsR	Intergenic
41					41								
42					42								
43					43								
44					44								
45					45								
46					46	E267A					A236V	In frame repeat insertion, T135M	*pqsR* promoter
47					47								
48					48			G257S					*pqsR* promoter
49					49								
PAO1					PAO1	T37N							

3.3.12 Isolates PA-W1 and PA-W3 display an atypical quorum sensing phenotype

The alkyl quinolones (AQs) HHQ and HQNO are produced and regulated by *pqs* genes. These in turn are regulated by the *las* and *rhl* quorum sensing systems and their AHLs 3-oxo-C_{12}-HSL and C_4-HSL. Previously, AHL-deficient and AQ-proficient strains have been isolated from clinical samples, and occur due to mutations in the *las* quorum sensing genes. However, until now no AHL-proficient and AQ-deficient strains have been reported in the literature. Isolates PA-W1 and PA-W3 were found to produce medium to high levels of the AHLs 3-oxo-C_{12}-HSL and C_4-HSL, and did not produce the AQs HHQ and HQNO, as well as PQS and pyocyanin (McKeown, 2013 and Figure 3.1). It was therefore postulated that either there are mutations in the *pqs* genes in these isolates, or there is a previously undetermined regulator causing this phenotype. This section will look at the mutations in the quorum sensing genes of these isolates.

Analysis of *pqs* genes

Genes *pqsABCDE* and *pqsR* of PA-W1 and PA-W3 were compared with the orthologous sequences in PAO1-UW. PA-W3 was found to contain two mutations in these genes, a P200L mutation in PqsB, which was found to be conserved in multiple other isolates, and a premature stop codon (E93*) in PqsD, which was found to be unique to this isolate. This premature stop codon would result in a truncated form of the PqsD protein (Figure 3.13), with the truncated form of the protein lacking the active site. Without a functional PqsD protein, PA-W3 would be unable to produce HHQ, and therefore also PQS.

A number of unique non-synonymous substitutions were found in genes *pqsA*, *pqsB* and *pqsC* of PA-W1 compared to PAO1-UW (Table 3.12), however no unique non-synonymous substitutions were found in genes *pqsD*,

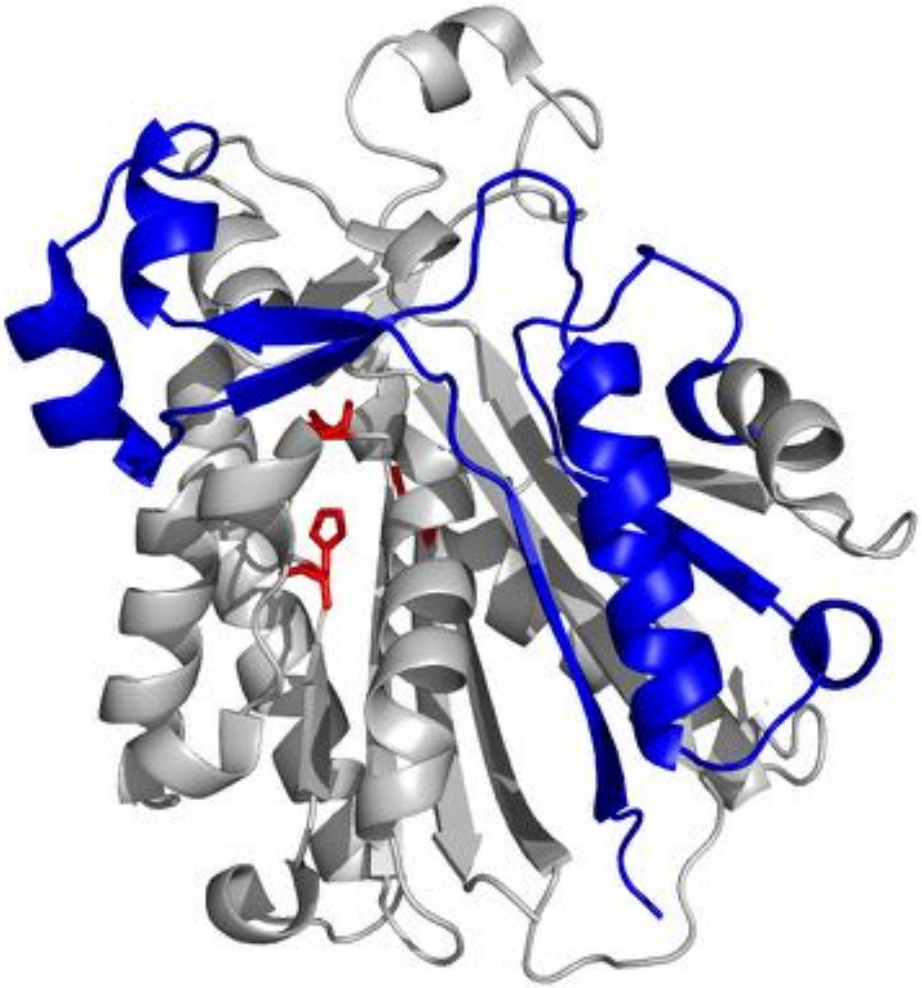

Figure 3.13: The premature stop codon in PqsD in PA-W3 would lead to a truncated form of the protein lacking the active site. *The structure shown is that of chain A of the PqsD dimer of P. aeruginosa (PDB: 3H76). The part of the protein prior to the premature stop codon in PA-W3 is shown in blue, and the remainder of the protein is shown in grey. The residues of the catalytic triad of the enzyme are shown as red sticks.*

Table 3.12: Amino acid changes in PqsA, B and C of PA-W1 compared to PAO1-UW, which are not conserved in any other isolate

Protein	Mutations only in PA-W1
PqsA	L33R, D109A, Q153H, H340R
PqsB	I75V, L220M
PqsC	L58V, P74L, S118G

pqsE or *pqsR*. The D109A mutation in PqsA would change a predicted dimer interface residue from the negatively charged aspartic acid to a hydrophobic alanine residue.

PA-W1 is the taxonomic outlier of the 49 *P. aeruginosa* wound isolates sequenced in this study, and has been found to fit into group 3 of the core genome phylogeny of *P. aeruginosa*, in the same group as taxanomic outliers PA7 and PAO1-VRF. PA7 has previously been described to have a frameshift mutation in the transcriptional activator of the *pqs* quorum sensing system, gene *pqsR* (Roy *et al.*, 2010). *pqsR* mutants have previously been shown to lead to a reduction in AQ and PQS levels (Cao *et al.*, 2001; Deziel *et al.*, 2005; Kesarwani *et al.*, 2011). However, PA-W1 was found not to contain the frameshift mutation in *pqsR* of PA7 (Figure 3.14). The nucleotide substitutions in the *pqsR* gene of PA-W1 compared to PAO1 are all silent mutations.

Figure 3.14: Multiple sequence alignment of the pqsR gene sequence of PA-W1, PAO1 and PA7 reveal the PA7 frameshift mutation is not conserved in PA-W1. *The alignment is coloured according to percentage identity, and the position of the frameshift mutation in PA7 is marked with a red asterisk.*

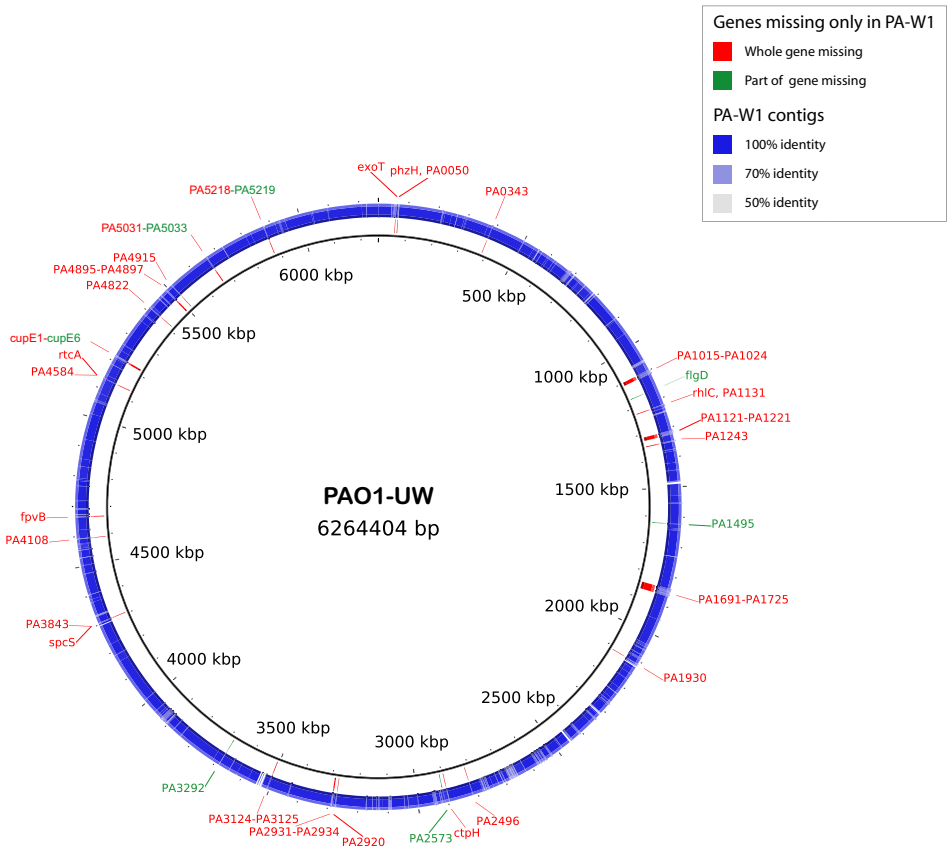

Figure 3.15: *Genomic regions of PAO1-UW which are not present in PA-W1. Contig alignments of PA-W1 to the PAO1-UW chromosome were carried out using BRIG (Alikhan* et al.*, 2011) and are shown in blue, with colour intensity varying according to percentage identity with the reference. Whole genes which are missing in PA-W1 are labelled in red and genes which are partially missing in PA-W1 are labelled in green.*

Regions missing in PA-W1

There are a number of SNPs and INDELs which are unique to PA-W1 (data not shown), as well as a number of regions which are missing in PA-W1 compared to the PAO1-UW reference sequence, and present in the remaining 48 isolates (Figure 3.15). These regions could be subject to further investigation to potentially locate a previously undetermined QS regulator responsible for the lack of AQ production in this isolate.

3.4 Discussion

Bacterial diversity has been shown to vary over the course of an infection (Dowd *et al.*, 2008; Frank *et al.*, 2009; Wolcott *et al.*, 2009; Smith *et al.*, 2010; Tuttle *et al.*, 2011). QS, virulence factor production and antibiotic resistance have previously been shown to vary not only within patients but also between different patients and sites of infection (Wilder *et al.*, 2009; Chung *et al.*, 2012; Workentine *et al.*, 2013; McKeown, 2013; Darch *et al.*, 2015). Until recently, WGS studies of *P. aeruginosa* have generally focused on groups of related *P. aeruginosa* isolates (Jeukens *et al.*, 2014; Darch *et al.*, 2015; Williams *et al.*, 2015), with a general bias in the field towards clinical isolates, particularly those from CF patients. However, *P. aeruginosa* is a ubiquitous organism with an ability to exploit a variety of nutrient sources and adapt to a range of environmental niches. A study by Stewart *et al.* (2014) on a set of CF lung isolates, as well as four environmental isolates, explored the genomic diversity of *P. aeruginosa* and was one of the first to correct the bias towards clinical isolates in genomics studies. This study aimed to further the understanding of genomic diversity of *P. aeruginosa* by analysing whole genomes of a set of 49 *P. aeruginosa* isolates which although were present at sites of clinical infection in 49 different patients at the same hospital, were not necessarily the cause of infection in each case. These isolates have previously been shown to display diversity in terms of QS and virulence phenotype, and two of the isolates were shown to produce high levels of AHLs and low levels of AQs, which is the opposite of what is usually found in clinical isolates. Therefore genes of the three main QS systems were looked at in greater detail in order to establish putative links between genotype and phenotype, providing an initial set of results upon which further laboratory studies can be carried out.

De novo assembly of the genomes of the 49 *P. aeruginosa* isolates revealed that the genome assembly sizes of these isolates generally fit previ-

ously reported genome sizes of *P. aeruginosa* (5.2-7 Mb, Wiehlmann *et al.*, 2007), apart from isolates PA-W7 and PA-W40, which have genome assembly sizes of 7.24 and 7.57 Mb, respectively (Table 3.3). However, the genome assemblies of these two isolates were also more fragmented compared to those of other isolates, i.e. were split into a larger number of contigs. Therefore it is unclear whether the large genome sizes of these two isolates reflect larger chromosome sizes, or if they are due to the limitations of *de novo* assembly. It could be the case that these two isolates contain frequent occurrences of large repetitive regions which are larger than the insert size of the WGS library, or that there are regions of the chromosome which were present at a lower coverage than others. These could be resolved in the future using WGS technologies such as PacBio Single Molecule Real Time (SMRT) sequencing (Pacific Biosciences, California, USA) or MinION nanopore sequencing (Oxford Nanopore Technologies, Oxford, UK), if required.

Construction of a core genome phylogeny of the 49 isolates revealed that these isolates form groups similar to those determined by Stewart *et al.* (2014) (Figure 3.2). 34 of the isolates were determined to be part of group 1 in the core genome phylogeny, 14 were found in group 2 and one isolate, PA-W1, was determined to be in group 3. This set of 49 *P. aeruginosa* isolates collected within a short time frame at a single hospital was found to display as much genomic diversity as the set of isolates used by Stewart *et al.* (2014), which were from various locations in the world. This is probably due to the already known high level of bacterial diversity in wounds, as well as the non-uniform and non biased method of sampling employed. In this study, isolates were not selected due to a particular disease phenotype, but were randomly sampled from a range of wound types from different patients. It is well known that *P. aeruginosa* is a ubiquitous organism, and the isolates in this study may have been present in a range of different environments

prior to being transferred to the wound environment, which would explain the high level of genomic diversity observed.

Upon initial construction of the core genome phylogeny, isolate PA-W1 appeared to be part of a distinct phylogenetic group than those which were previously reported. The 49 isolates in this study have been included as part of a larger consortium project to sequence 1000 *Pseudomonas* genomes. Thus far, 390 genomes have been sequenced, including the 49 Nottingham isolates described here. The core genome phylogeny of the 390 *Pseudomonas* genomes revealed that isolate PA-W1 is part of phylogenetic group 3 (Freschi *et al.*, 2015). 16/49 isolates were found to be contained within four clades of the phylogenetic tree (Figure 3.2). No obvious patterns were observed between these subgroups and the wound type, site of location or patient information when compared with information in Table 3.2. This finding is an example of how WGS enables closely related isolates to be distinguished from one another, which would not be possible by traditional typing methods. These findings highlights the importance of the efforts of the 1000 *Pseudomonas* genomes project towards fully understanding the genomic diversity of this ubiquitous organism; the greater the number isolates that are sequenced, the greater the understanding of this highly diverse genus will become. Given that *P. aeruginosa* is an important clinical pathogen, this may lead to increased understanding of the evolution of virulence in *P. aeruginosa*, which may be useful for identifying novel therapeutics.

QSSM levels for each isolate were mapped onto the core genome phylogeny, which did not reveal any correlations between QS phenotype and phylogenetic grouping (Figure 3.3). Previous studies have found that mutations in QS genes, such as *lasR*, are associated with adaptation of *P. aeruginosa* to the CF lung (Smith *et al.*, 2006; D'Argenio *et al.*, 2007; Fothergill *et al.*, 2014). These are thought to be examples of *P. aeruginosa* adapting to its environmental niche. In a study of *P. aeruginosa* isolates

from a single CF patient sputum sample, the isolates displayed a high level of diversity of QS phenotype, and patterns of acquisition or accrual of QS phenotypes along the evolutionary trajectory of the phylogeny were not observed. Mutations in quorum sensing genes were, however, found to be associated with recombination events (Darch *et al.*, 2015). Therefore the bacterial population in which each of the 49 wound isolates are found would be expected to have an effect on the types of mutations observed. Given that the isolates in this study were found in different patients and wound types, each isolate would adapt to its own unique environment consisting of its own unique bacterial population, and therefore the lack of correlation between QSSM production and phylogenetic grouping is not surprising.

A limited amount of patient and basic clinical strain information was available for each isolate (Table 3.2), however there were no links between this information and phylogenetic groupings of the isolates (data not shown). This is expected, as these isolates were not necessarily causing infection at the time of sampling, and were not clonally related.

Comparison of the genomes of the 49 isolates to the PAO1-UW genome sequence revealed that a number of isolates were missing genes involved in LPS biosynthesis (Figure 3.4). This included genes involved in O antigen (O-Ag) biosynthesis. In such a genetically diverse set of *P. aeruginosa* isolates, it is expected that multiple serotypes would be observed, and therefore a variability in LPS, and particularly O-Ag, biosynthesis genes would be expected. A previous WGS study of three *P. aeruginosa* isolates by Spencer *et al.* (2003), one environmental and two from CF patients, also noted low WGS read coverage of the O-Ag biosynthesis genes. Mutations in the O-Ag biosynthesis genes in these three isolates rendered them untypable.

Analysis of non-synonymous mutations in the *las* quorum sensing genes and SNPs in the intergenic regions revealed a high number of mutations in *lasR* and its upstream region, relatively few mutations in *lasI* and no

non-synonymous mutations in *rsaL* (Figure 3.5 and Table 3.4). A previous study of *lasR* mutations in 66 clinical and environmental isolates found that 81% of isolates had a *lasR* gene sequence identical to strain PAO1 or only with a silent mutation. It was also found that 15% of isolates had non-synonymous changes in *lasR* and 5% contained ISs in *lasR* (Cabrol *et al.*, 2003). In this study, a larger proportion of isolates were found to contain *lasR* mutations, with 42% of isolates containing either a non-synonymous mutation, and 12% of isolates containing an IS in the *lasR* gene.

Mutations in *lasI*, *lasR* and *rhlI* have previously been shown to lead to a reduction in virulence *in vivo*, with the most significant reduction in *lasI*/*rhlI* double mutants. Only isolate PA-W1 was found to contain mutations in LasI, and this isolate does not contain any non-synonymous mutations in RhlI. From this data it appears that genes *lasI*, the 3-oxo-C_{12}-HSL synthase, and its negative regulator *rsaL* appear to be protected from non-synonymous mutations. A protein BLAST search of the *lasI* and *rsaL* genes against all genomes in the Pseudomonas Genome Database (http://pseudomonas.com, accessed 5th August 2015) reveal that these proteins are conserved with 100% identity in all *P. aeruginosa* isolates in which they are found, apart from in *P. aeruginosa* PA7, in which LasI shares 99.5% sequence identity. The LasI protein sequence in PA7 differs from that of PAO1 by a singe mutation, R197Q; which is one of the two mutations found in PA-W1. This mutation occurs at the C terminal end of the protein, as does the second LasI mutation in PA-W1. Isolate PA-W32 was found to be missing the *las* genes, which were part of a 68 kb deletion in this isolate (Figure 3.7). This isolate was unable to produce AHLs and LasB elastase.

lasR mutants have previously been reported in clinical isolates (Hamood *et al.*, 1996; Cabrol *et al.*, 2003; Schaber *et al.*, 2004; Salunkhe *et al.*, 2005; Hoffman *et al.*, 2009) and have been shown to incur a selective advantage in certain environmental conditions compared to wild type strains (Heurlier

et al., 2005; Sandoz *et al.*, 2007). A variety of loss of function mutations in LasR were observed in the 49 *P. aeruginosa* wound isolates. Isolate PA-W10 was found to contain a P74L mutation in LasR, which has previously been shown to lead to loss of multimerisation of LasR, and severely affect protein function (Kiratisin *et al.*, 2002). PA-W10 is unable to produce both 3-oxo-C_{12}-HSL and C_4-HSL, which would be expected if the LasR protein were unable to function. PA-W12, 23, 37 and 43 were found to contain mutations which would lead to changes in predicted binding residues of LasR, and each of these isolates displayed a reduction in 3-oxo-C_{12}-HSL production compared to PAO1 (Table 3.5). All but PA-W12 also displayed a reduction in C_4-HSL production as a result of a mutation in a predicted DNA-binding residue.

Isolates PA-W18, 21, 31, 36, 39 and 42 were found to contain putative transposon-mediated insertions in the *lasR* gene, and all six of these isolates showed a reduction in 3-oxo-C_{12}-HSL production, with all but PA-W18 also displaying a reduction in C_4-HSL production (Table 3.7). Isolate PA-W21 showed a reduced production of all QSSMs. In all six isolates the insertion sequences were found to occur within predicted functional domains in the LasR protein sequence (Figure 3.9). The LasR insertions in isolates PA-W31, 39 and 42 appear to originate from *P. aeruginosa*, whereas insertions in isolates appear to PA-W18, 21 and 36 originate from other species (Table 3.9). This could be due to horizontal gene transfer from other organism present in either the wound from which they were isolated, or the environment from which they were originated. The putative insertion sequence in PA-W18 is 36.8 kb, and contains multiple copper and heavy metal resistance genes. The remaining putative insertion sequences consist of one or two transposase genes, apart from PA-W36, where a single contig corresponding to a putative insertion sequence could not be determined (Table 3.8).

The putative insertion sequences in PA-W31, 39 and 42 were found to occur in the same location within the *lasR* gene, and were also identical in sequence content (Table 3.9). These three isolates appear to be genetically related according to the core genome phylogeny, however were isolated from three different patients. The genome assembly statistics of these three isolates are strikingly similar, in terms of genome assembly size, number of contigs and number of predicted genes. It is unclear whether the three patients from whom these strains were isolated picked up this strain from a common source, or whether a sample handling error at some point in time means these three isolates are actually three copies of a single isolate. These three isolates do display differences in their QSSM profiles, although generally produce medium-low levels of each QSSM, apart from PA-W39 which produces high levels of HHQ.

The finding of transposon insertions in *lasR* in clinical isolates of *P. aeruginosa* highlights the importance of looking at *de novo* genome assemblies for larger structural variants than can be detected by read mapping techniques. These transposon insertions are, however, putative, and remain to be confirmed by PCR.

Analysis of non-synonymous mutations in genes *rhlI* and *rhlR* revealed a number of mutations in these genes as well as a mutation in the rhlR promoter (Figure 3.10 and Table 3.10). Predicted DNA binding residues in RhlR occur after the 180th residue in the protein sequence, and therefore an out of frame deletion in PA-W45 which would lead to a truncated form of the RhlR protein after residue R105, could affect the function of the RhlR protein. PA-W45 also contains a truncated version of the LasR protein due to a premature stop codon, however does not contain any non-synonymous mutations in *pqs* genes. This isolate produces low levels of all QSSM molecules tested. *lasR*/*rhlR* double mutants have previously been shown to be unable to produce AHLs, proteases, LasB elastase and PQS (Dekimpe and Deziel,

2009), which fits what is observed with PA-W45. There are a number of isolates which contain mutations in RhlI and RhlR and display a reduction in C_4-HSL production. These mutations require further characterisation in order to determine if they could directly influence phenotype.

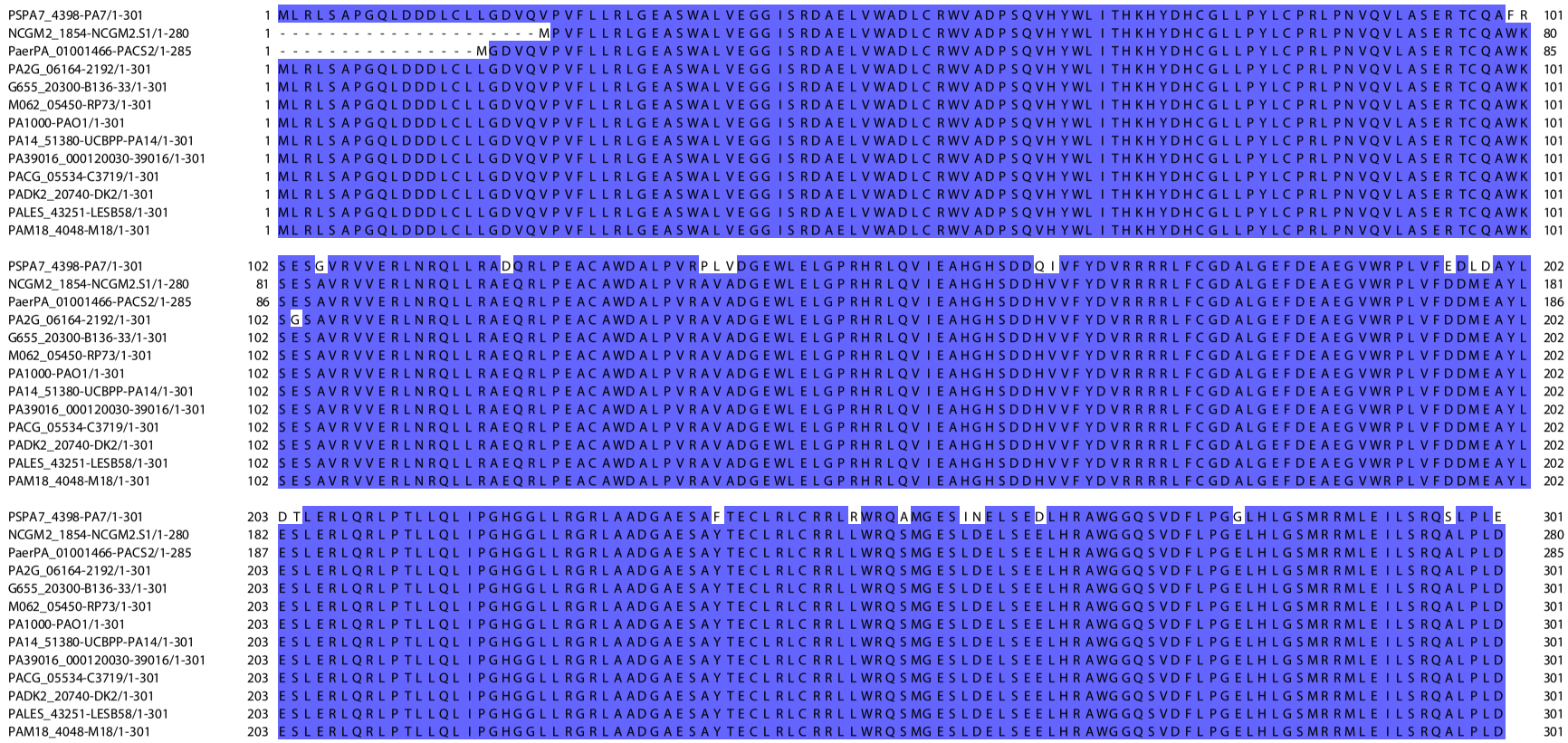

Figure 3.16: Multiple sequence alignment of PqsE protein sequences from Pseudmonas Genomes Database. *The multiple sequence alignment was carried out using CLUSTAL-W, and visualised using Jalview. Residues in the alignment are coloured by percentage identity.*

A number of non-synonymous mutations were observed in the *pqs* genes, however, interestingly there were no non-synonymous mutations observed in *pqsE* (Figures 3.11 and 3.12, and Table 3.11). This indicates that, like *lasI* and *rsaL*, *pqsE* may be protected from non-synonymous mutations. Mutations in the *lasI* gene were previously shown to lead to a reduction in virulence (Zhu *et al.*, 2004). RsaL is known to act as a global regulator of QS by acting as a repressor of LasI (Rampioni *et al.*, 2007). The function of PqsE in the biosynthesis of HHQ was recently determined (Drees and Fetzner, 2015). The lack of mutations in these genes in this set of highly diverse *P. aeruginosa* isolates indicates the importance of the function of these genes in *P. aeruginosa*.

BLAST comparison of the PqsE protein sequence with strains at the Pseudomonas Genome Database revealed that the PqsE protein sequence shared 100% sequence identity with all strains apart from PA7, *P. aeruginosa* 2192, *P. aeruginosa* NCGM1.S1 and *P. aeruginosa* PACS2 (Figure 3.16). In the case of NCGM1.S1 and PACS2, it may be that the gene model may be incorrect and needs to begin at the previous methionine in the sequence. This would leave only PA7 and strain 2192 where the PqsE protein sequence differs from the 49 *P. aeruginosa* wound isolates.

A number isolates which were found to contain mutations in *pqs* genes also displayed a reduction in HHQ and PQS production. Isolates PA-W1 and PA-W3 contained mutations at predicted dimer interface residues in *pqsA* and *pqsD* respectively. These two isolates were found previously to be AHL-proficient, AQ-deficient strains (McKeown, 2013). The mutation in *pqsD* of PA-W3 was found to lead a premature stop codon at residue E93 in the protein sequence (Figure 3.13). This premature stop codon would lead to a truncated form of the PqsD protein, and this could be the cause of the AQ-deficient phenotype, although this remains to be verified. The D109A substitution at the predicted dimer interface residue of PqsA

in isolate PA-W1 would change a negatively charged aspartic acid to a hydrophobic alanine residue. However, this dimer interface annotation at the NCBI Conserved Domains database (Marchler-Bauer *et al.*, 2015) is based on the structure of a PqsA homolog in *Burkholderia xenovorans*, where the protein forms a homodimer. Functional characterisation of PqsA in *P. aeruginosa* revealed that the protein is most likely to exist as a native monomer (Coleman *et al.*, 2008), and if this is the case then the D109A mutation in PA-W1 is unlikely to affect the function of the protein. The previously determined frameshift mutation in *pqsR* in PA7 was found not to be conserved in PA-W1 (Figure 3.14). In short, there were no obvious differences which could be the cause of the AHL-proficient, AQ-deficient phenotype in PA-W1. There were other non-synonymous mutations in *pqsA*, as well as in *pqsB* and *pqsC* which could be the cause of this phenotype, although further experiments are required to verify this. There were also a number of SNPs and INDELs unique to PA-W1 (data not shown), as well as deletions in PA-W1 compared to PAO1-UW which were not conserved in the 48 other *P. aeruginosa* wound isolates (Figure 3.15). There is scope for future work to determine the genotype which causes this distinct QS phenotype, as well as to determine if the premature stop codon in *pqsD* is the cause of the AHL-proficient, AQ-deficient phenotype in PA-W3.

Chapter 4

General Discussion

It has been proposed that *P. aeruginosa* PAO1 exists as genetically variable sublines in strain collections and laboratories, and the current reference sequence is thought to correspond to a distinct subline with its own genetic peculiarities. In this study the whole genomes of 8 PAO1 sublines, including the subline from the ATCC strain collection, ATCC 15692, thought to be the original PAO1 of Bruce Holloway, were sequenced, as well as strains PAO2 and PAO3, which correspond to auxotroph mutants derived from *P. aeruginosa* PAO in the 1950s.

Every PAO1 subline in this study has been found to contain mutations which are not conserved in any other subline, as well as, in the case of some sublines, prophage insertions and large chromosomal deletions. This confirms the previous hypothesis that PAO1 exists as highly variable sublines which vary between laboratories and strain collections. These are likely to correspond to mutations which have accrued in PAO1 sublines after they were distributed among research labs and strain collections. A large proportion of these mutations lead to changes in protein sequences, which have the potential to affect phenotype. This highlights the importance of using the same laboratory subline when attempting to reproduce research from other laboratories, as use of two distinct sublines for the same experiment may lead to discrepancies in results. A large 58.6 kb deletion was found to be

conserved between two PAO1 laboratory sublines, PAO1-N and PAO1-E, which were not known to have been related, or purchased from the same genetic stock centre. This highlights the importance of keeping a precise record of where laboratory strains are obtained from, and also, if possible, batch numbers of isolates.

It could be argued that researchers should abandon their current subline in favour of a single PAO1 strain from a single strain collection. However, results of this study also show unique mutations in a PAO1 isolate purchased from the ATCC stock centre, PAO1-ATCC 15692 first mentioned by Bruce Holloway (Holloway, 1969). It could be deduced that this strain is the "original" PAO1 strain given the literature, however in this study the ATCC strain has been found to contain non-synonymous mutations and premature stop codons in genes involved in antibiotic resistance and type IV pili formation, which are unique to this subline. The idea of abandoning PAO1 altogether, and moving to another *P. aeruginosa* strain for laboratory studies into this opportunistic pathogen could be proposed, however it could be the case that other *P. aeruginosa* strains also contain genetic changes unique to particular laboratories. In fact, this could be the case for any other bacterial strains, as well as cell lines, used routinely in laboratory experiments. Therefore it seems reasonable that laboratories should carry out WGS of their own laboratory strains, and take into account differences compared to the reference sequence for the strain.

As well as variability in laboratory sublines, the choice of reference sequence is also brought into question by this study. The comparison of the genome sequences and optical maps of PAO1 sublines with the reference sequence, PAO1-UW, revealed that the strains and sublines have more in common with one another than with the reference sequence; namely that PAO1-UW is an outlier. Since the original PAO1-UW isolate is not available, comparisons can only be made to the genomic sequence. Therefore

it is unclear if this observation due to errors in the genome assembly of PAO1-UW, carried out using whole genome shotgun methods prior to the availability of next generation technologies, or if the isolate selected for WGS in 2000 had accrued genetic changes over time. The exception to this is the 2.2 Mb inversion found in PAO1-UW compared to the physical map, which has been previously reported (Stover *et al.*, 2000; Klockgether *et al.*, 2010). In this study the inversion in PAO1-UW was not found to be conserved in any of the isolates studied. This study has revealed that the reference sequence of PAO1-UW is not representative of the range of PAO1 sublines included in this study.

Therefore differences between PAO1-UW and the ancestral *P. aeruginosa* PAO have been deduced, using genetic information from PAO1 sublines, and PAO-derived strains PAO2 and PAO3. These should be taken into consideration by researchers when carrying out genetic analyses in PAO1. For the purposes of this study, only the most reliable set of mutations in PAO1-UW, which are not conserved in any other PAO1 subline, PAO2 or PAO3, were taken into account for the deduction of the ancestral genomic sequence of *P. aeruginosa* PAO. These consist of SNPs shared by all strains and sublines in Table 2.8 and all INDELs in Table 2.10. However, there are some caveats associated with this ancestral sequence. These genetic mutations remain to be confirmed by PCR in all ten strains and sublines, and therefore at this stage are predictions. There may be additional mutations in the PAO1-UW reference sequence which were not conserved in all ten strains and sublines in this study, for example changes which were only conserved in eight or nine strains and sublines. A conservative approach was taken with determining the ancestral sequence of PAO, and so these mutations were not included and remain to be verified. There may be mutations in PAO1-UW which are not present in the deduced ancestral *P. aeruginosa* PAO sequence which could not be determined by this study, for example

in highly repetitive regions of the genomic sequence. New long read technologies such as the PacBio SMRT or the Oxford Nanopore MinION could be used to resolve the sequence of these regions in laboratory strains and sublines, in order to determine if the reference sequence contains any unique mutations in these regions compared to other sublines of PAO1.

Although *P. aeruginosa* PAO1 is the most commonly studied strain in research laboratories, it corresponds to a single isolate within a diverse species which is able to adapt to and survive in a range of ecological niches, including human hosts. Previous WGS studies of *P. aeruginosa* have focused on the host adaptation of *P. aeruginosa* in chronic CF infections, however given that *P. aeruginosa* is able to adapt to a variety of environmental niches, the characteristics of isolates from different types of infection, and therefore different environments, may not be identical. Therefore the need to study a variety of different strains of *P. aeruginosa* from a range of environments is evident.

In this study the bias towards the analysis of genetic diversity within *P. aeruginosa* from the CF lung was considered, and the genomic diversity of a set of 49 wound isolates, which were not necessarily causing infection at the time of sampling, was determined. It was found that these isolates were genetically diverse, with isolates spanning all three previously determined phylogenetic groups of *P. aeruginosa*. These isolates were generally found to fall into the previously reported range of *P. aeruginosa* genome sizes. Non-synonymous mutations in genes in the three major QS systems, *las*, *rhl* and *pqs*, were characterised and a number were found to correlate with observed QS phenotypes.

In this study *de novo* assemblies were generated and analysis of the coding sequences corresponding to QS genes revealed that six of the 49 isolates contained insertions within the *lasR* gene, which are thought to correspond to transposable elements. Although mutations in the *lasR* gene have been

reported in multiple studies, and are thought to be a method of adaptation of *P. aeruginosa* to the host environment, this one of few reports of transposable elements disrupting the *lasR* gene in clinical isolates of *P. aeruginosa*. This could possibly be due to most studies using a resequencing approach focused on identifying SNPs and INDELs, compared to de novo assembly of bacterial genomes. This is likely due to the cost of WGS, where until recently it was more economical to sequence more bacterial genomes on a single run of a sequencing instrument, thereby getting a low genome coverage per isolate, which was sufficient for a resequencing approach but not for a de novo assembly approach. However, with the ever-reducing costs as well as increases in yield and read lengths of WGS technologies, it is now more affordable to yield a high enough genome coverage for a *de novo* assembly approach.

Carrying out WGS of a set of 49 clinical isolates from a variety of wound types, whose diverse QS phenotypes have been characterised has led to the identification of a number of mutations within QS genes which correlate with QS phenotype. Two of the 49 isolates were previously determined to be AHL-proficient and AQ-deficient, which was a novel QS phenotype at the time for *P. aeruginosa*. These two isolates were found to contain mutations in *pqs* genes, involved in the biosynthesis of the AQ HHQ. A mutation in the *pqsD* gene of PA-W3 was found to lead to a premature stop codon, which could lead to a loss of function of PqsD, and therefore explain the lack of production of HHQ. However, there was no simple explanation for the lack of HHQ production in isolate PA-W1. This isolate could be the subject of further study, with the potential to identify a novel regulator of AQ biosynthesis, and therefore QS.

Although WGS technologies have been around for almost ten years, *P. aeruginosa* genomic research has focused on clinical isolates and bacteriak populations from CF infections, in order to track the evolution of

these isolates within infections. In this study, it has been shown that WGS can, and arguably should, be used as a routine method to characterise genetic variability in laboratory isolates. This is especially the case with the low cost of WGS and the availability of web servers such as the Galaxy server (https://usegalaxy.org/) and the PATRIC workspace (http://www.patricbrc.org/), where users can upload their WGS data and carry out analyses using a graphical user interface, making a simple analysis such as the characterisation of SNPs and short INDELs or *de novo* genome assembly accessible to wet lab scientists who are not familiar with the use of open source software in a command line environment.

This study has also worked to address the bias of WGS analyses of *P. aeruginosa* towards isolates from CF infections by characterising the genomic diversity of a set of clinical wound isolates. The genomes of the isolates in this part of the study form part of the 1000 *Pseudomonas* genomes project, which aims to sequence the genomes of 1000 human, animal and environmental isolates of *P. aeruginosa* (Freschi *et al.*, 2015). This is the highest resolution study of *P. aeruginosa* genomic diversity to date, and will allow a large scale characterisation of the core and accessory genomes, as well as the virulence factors and antibiotic resistance genes of this species. The potential impact of this project on our understanding of *P. aeruginosa* as a species is huge, and could lead to novel insights into this opportunistic pathogen, including novel strategies to counter *P. aeruginosa* infections.

Chapter 5

Appendix

5.1 Genome Accession Numbers

Table 5.1: *Accession numbers of reference sequences used in this study*

Genome	Accession Number	Reference
DK2	NC_018080	(Marvig *et al.*, 2013)
LESB58	NC_011770	(Winstanley *et al.*, 2009)
PA7	NC_009656	(Roy *et al.*, 2010)
PA14	NC_008463	(Lee *et al.*, 2006)
PANOT106	Obtained from authors	(Snyder *et al.*, 2013)
PAO1-UW	NC_002516	(Stover *et al.*, 2000)
RGP42	GQ141978	(Klockgether *et al.*, 2010)
VRFPAO1	AOBK00000000	Malathi and Madhavan, unpublished data

5.2 insertion.pl

```
#!/usr/bin/perl
# Make an insertion of one fasta sequence into another
# <insertion.pl>
# Author: Hardeep Naghra

my($genome, $insertion, $inspos, $outfile) = @ARGV;

if(!defined $outfile)
{
  print "\nusage: ./insertion.pl <genome_fasta>
     <insertion_fasta> <insertion_point> <outfile>\n\n";
  exit;
}
# fasta file of insertion sequence
open(INS, $insertion) or die "$!";
# fasta file of reference sequence
open(FILE, $genome) or die "$!";

my $insseq='';

# get whole insertion sequence in a string variable
while(<INS>)
{
  if($_!~/^>/)
  {
    chomp $_;
    $insseq.=$_;
  }
}

my $sequence='';
my $header='';
# get whole reference sequence in a string variable
while(<FILE>)
{
  if ($_=~/^>/)
  {
    chomp $_;
    $header=$_;
  }
  else
  {
    chomp $_;
    $sequence.=$_;
  }
}

my $genomelength=length($sequence);
my $inslength=length($insseq);
```

```
# divide reference into two parts where part 1 is the
   sequence before the insertion, and part 2 is the
   sequence after the insertion
my $firstsub=substr($sequence, 0, $inspos);
my $secondsub=substr($sequence, $inspos, $genomelength);

# make insertion
my $cat=$firstsub.$insseq.$secondsub;
my $catlength=length($cat);

# create outfile
open(OUT, ">$outfile") or die "$!";

# Print insertion statistics to Standard Error
print STDERR"Genome length: $genomelength\tInsertion
   length: $inslength\tTotal length: $catlength\n";
# Print insertion sequence to outfile
print OUT "$header\n$cat\n";
```

5.3 inversion.pl

```
#!/usr/bin/perl
# Script to make inversion in fasta sequence
# <inversion.pl>
# Author: Hardeep Naghra

my($genome, $start, $end, $outfile) = @ARGV;

if(!defined $outfile)
{
  print "\nusage: <genome_fasta> <inversion_start>
     <inversion_end> <outfile>\n\n";
  exit;
}
# open reference genome
open(FILE, $genome) or die "$!";

my $sequence='';
my $header='';
# store genome sequence and header as string variables
while(<FILE>)
{
  if ($_=~/^>/)
  {
    chomp $_;
    $header=$_;
  }
  else
  {
    chomp $_;
    $sequence.=$_;
  }
}

my $value1=$start-1;
my $tmpval=$end-$start;
my $value2=$tmpval+1;

my $length=length($sequence);

# sequence to be inverted
my $substring=substr($sequence, $value1, $value2);
my $sublength=length($substring);

# reverse complement the substring which is to be inverted
my $first=substr($sequence, 0, $start-1);
my $second=substr($sequence, $value1+$sublength ,
   $length);
my $rev=scalar reverse $substring;
$rev=~tr/ACGTacgt/TGCAtgca/;
```

```
# create inverted sequence as string variable
my $inverted=$first.$rev.$second;

# create and open outfile
open (OUT, ">$outfile") or die "$!";

# print sequence with inversion to outfile
print OUT "$header\n$inverted\n";
```

5.4 fastq_trim_5.pl

```
#!/usr/bin/perl
# clip a given number of bases from the 5' end of fastq
   reads
# <fastq_trim_5.pl>
# Author: Hardeep Naghra

my($fastq, $trim, $min)=@ARGV;

if(!defined $trim)
{
  print "\nusage: ./fastq_trim_5.pl <fastq file> <length
     to trim> <min read length>\n\n";
  exit;
}

open(FQ, $fastq) or die "$!";

while(<FQ>)
{
  my $seqhead=$_;
  my $seq=<FQ>;
  my $qualhead=<FQ>;
  my $qual=<FQ>;
  chomp $seqhead;
  chomp $seq;
  chomp $qualhead;
  chomp $qual;

  # clip sequence and quality score lines
  my $length=length($seq);
  my $clippedread=substr($seq, $trim, $length);
  my $clippedqual=substr($qual, $trim, $length);
  my $readlength=length($clippedread);

  # print to STDOUT
  if(defined $min)
  {
    if($readlength>=$min)
    {
      print
         "$seqhead\n$clippedread\n$qualhead\n$clippedqual\n";
    }
  }
  else
  {
    print
       "$seqhead\n$clippedread\n$qualhead\n$clippedqual\n";
  }
```

```
}
```

5.5 filter_vcf.pl

```
#!/usr/bin/perl -w
use strict;

my($vcf, $cutoff)=@ARGV;

if(!$vcf)
{
  print "Usage: <vcf> <cut-off>\n";
  exit;
}
if(!$cutoff)
{
  $cutoff=5;
  print STDERR "Cut-off not specified, so will use the
     default cut-off of $cutoff\n";
  print STDERR "If you require a different coverage
     cut-off... usage: <vcf> <cutoff>\n";
}

# infile
open (FH, $vcf) or die "$!";

# outfile of filtered SNPs
open (OUT, ">$vcf.depth_$cutoff.pass") or die "$!";

# outfile of sub-quality SNPs
open (BAD, ">$vcf.depth_$cutoff.fail") or die "$!";

while(<FH>)
{
  # skip comment lines
  next if ($_=~/^#/);

  # skip INDELs
  next if ($_=~/INDEL/);

  # put columns into array
  my @line= split /\t/;

  # only interested in lines in VCF files with SNP info
  next if ($line[4] eq ".");

  # coverage cutoff at SNP locus (this is 5 unless
     specified otherwise)
  next unless $_=~/DP=(\d+)/;
  my $dp=$1;
  if ($dp<$cutoff)
  {
```

```
    print BAD $_;
  }
  elsif($dp>=$cutoff)
  {
    # print filtered VCF lines to outfile
    print OUT $_;
  }
}

# inform user of outfile
print STDERR "Filtered vcf in file
   \"$vcf.depth_$cutoff.pass\"\n";
print STDERR "Discarded SNPs in file
   \"$vcf.depth_$cutoff.fail\"\n";
```

5.6 filter_vcf_dp4.pl

```
#!/usr/bin/perl -w
use strict;

my($vcf, $cutoff)=@ARGV;

if(!$cutoff)
{
  print "Usage: <vcf> <dp4_cutoff>\n";
  exit;
}

# infile
open (FH, $vcf) or die "$!";

# outfile of filtered SNPs
open (OUT, ">$vcf.dp4_$cutoff.pass") or die "$!";

# outfile of sub-quality SNPs
open (BAD, ">$vcf.dp4_$cutoff.fail") or die "$!";

while(<FH>)
{
  # DP4 cut-off: 1) fwd ref 2) rev ref 3) fwd SNP 4) rev
     SNP
  if($_=~/DP4=(\d+),(\d+),(\d+),(\d+)/)
  {
    my $fwdref=$1;
    my $revref=$2;
    my $fwdsnp=$3;
    my $revsnp=$4;
    my $line=$_;

    if ($fwdsnp >= $cutoff && $revsnp >= $cutoff)
    {
      print OUT $line;
    }
    else
    {
      print BAD $line;
    }
  }
  else
  {
    print BAD "DEBUG: $_";
  }
}

# inform user of outfile
```

```
print STDERR "Filtered vcf in file
   \"$vcf.dp4_$cutoff.pass\"\n";
print STDERR "Discarded SNPs in file
   \"$vcf.dp4_$cutoff.fail\"\n";
```

5.7 blast2fasta.pl

```
#!/usr/bin/perl
# Take blast output file and get the sequence of the
   corresponding subject sequence hit

# Date 25/02/15

use strict;
use warnings;
use Bio::SeqIO;

my($blast, $fa)=@ARGV;

if(!defined $blast)
{
   print "\nusage: <blast> <fasta_sequence>\n\n";
   exit;
}

# parse blast output and get the appropriate sequence
open(BLAST, $blast) or die "$!";

my $count=1;

while(<BLAST>)
{
  chomp $_;
  my ($queryId, $subjectId, $percIdentity, $alnLength,
     $mismatchCount, $gapOpenCount, $queryStart,
     $queryEnd, $subjectStart, $subjectEnd, $eVal,
     $bitScore) = split(/\t/);

  # fasta sequence of scaffold (subject hit)
  my @scaffoldfa=`get_single_seq.pl '$subjectId' $fa`;

  # remove header of scaffold fa and make a sequence
     string
  shift @scaffoldfa;
  my $sequence=join('', @scaffoldfa);
  #remove spaces/newlines
  $sequence=~s/\n//g;
  $sequence=~s/ //g;

  # subsequence of scaffold which corresponds to the
     blast hit

  my $length=0;
  my $start=0;
  my $substring='';
```

```
  if($subjectEnd > $subjectStart)
  {
    $length = ($subjectEnd-$subjectStart)+1;
    $start=$subjectStart-1;
    $substring=substr($sequence, $start, $length);

    print ">$blast\_$count region
       $subjectStart..$subjectEnd\n$substring\n";
  }
  elsif($subjectStart > $subjectEnd)
  {
    $length = ($subjectStart-$subjectEnd)+1;
    $start = $subjectEnd-1;
    $substring=substr($sequence, $start, $length);
    $substring=~tr/AGCTN/TCGAN/;
    my $rev=reverse($substring);

    print ">$blast\_$count region
       $subjectEnd..$subjectStart\n$rev\n";
  }

$count++;

}
```

5.8 Antibiotic resistance profiles of 49 wound isolates

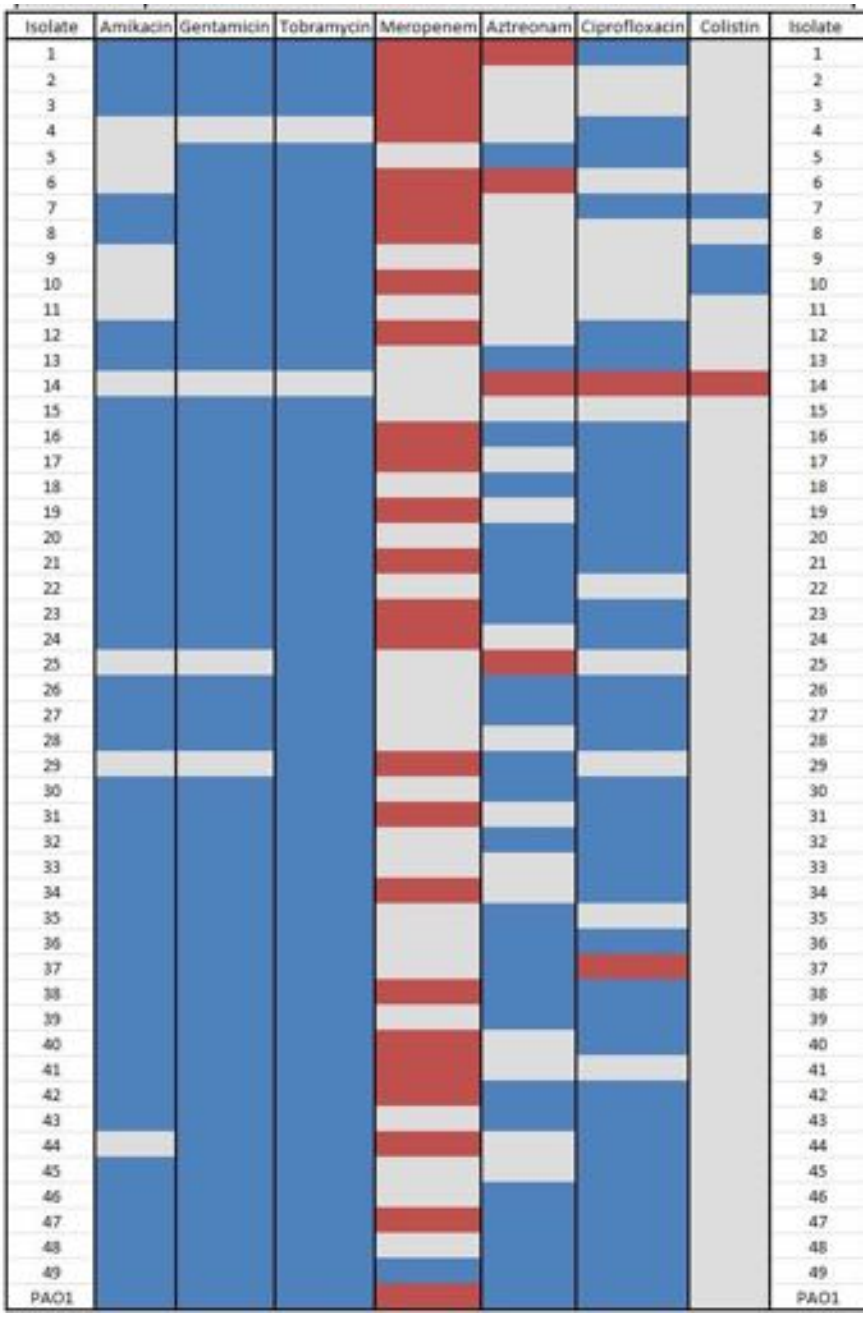

Figure 5.1: Heat map of the minimum inhibitory concentrations (MICs) of antibiotics added to wound isolates. *The heatmap is coloured whether the isolates were defined as being reisistant (red), intermediate (white) or sensitive (blue) to each antibiotic. PAO1 is included as a comparison. Figure 5.4 in (McKeown, 2013).*

Bibliography

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K. L., Tyson, G. W., and Nielsen, P. H. (2013). 'Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes.' *Nat. Biotechnol.*, **31**(6):533–538.

Albus, A. M., Pesci, E. C., Runyen-Janecky, L. J., West, S. E., and Iglewski, B. H. (1997). 'Vfr controls quorum sensing in *Pseudomonas aeruginosa*.' *J. Bacteriol.*, **179**(12):3928–3935.

Alikhan, N. F., Petty, N. K., Ben Zakour, N. L., and Beatson, S. A. (2011). 'BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons.' *BMC Genomics*, **12**:402.

Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Lahti, L., Loman, N. J., Andersson, A. F., and Quince, C. (2014). 'Binning metagenomic contigs by coverage and composition.' *Nat. Methods*, **11**(11):1144–1146.

Alontaga, A. Y., Rodriguez, J. C., Schonbrunn, E., Becker, A., Funke, T., Yukl, E. T., Hayashi, T., Stobaugh, J., Moenne-Loccoz, P., and Rivera, M. (2009). 'Structural characterization of the hemophore HasAp from *Pseudomonas aeruginosa*: NMR spectroscopy reveals protein-protein interactions between Holo-HasAp and hemoglobin.' *Biochemistry*, **48**(1):96–109.

Andersson, S. G. and Kurland, C. G. (1998). 'Reductive evolution of resident genomes.' *Trends Microbiol.*, **6**(7):263–268.

Angus, B. L., Carey, A. M., Caron, D. A., Kropinski, A. M., and Hancock, R. E. (1982). 'Outer membrane permeability in *Pseudomonas aeruginosa*: comparison of a wild-type with an antibiotic-supersusceptible mutant.' *Antimicrob. Agents Chemother.*, **21**(2):299–309.

Asai, T., Condon, C., Voulgaris, J., Zaporojets, D., Shen, B., Al-Omar, M., Squires, C., and Squires, C. L. (1999a). 'Construction and initial characterization of *Escherichia coli* strains with few or no intact chromosomal rRNA operons.' *J. Bacteriol.*, **181**(12):3803–3809.

Asai, T., Zaporojets, D., Squires, C., and Squires, C. L. (1999b). 'An *Escherichia coli* strain with all chromosomal rRNA operons inactivated: complete exchange of rRNA genes between bacteria.' *Proc. Natl. Acad. Sci. U.S.A.*, **96**(5):1971–1976.

Augustin, D. K., Song, Y., Baek, M. S., Sawa, Y., Singh, G., Taylor, B., Rubio-Mills, A., Flanagan, J. L., Wiener-Kronish, J. P., and Lynch, S. V. (2007). 'Presence or absence of lipopolysaccharide O antigens affects type III secretion by *Pseudomonas aeruginosa*.' *J. Bacteriol.*, **189**(6):2203–2209.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., *et al.* (2008). 'The RAST Server: rapid annotations using subsystems technology.' *BMC Genomics*, **9**:75.

Baba, T. and Schneewind, O. (1998). 'Instruments of microbial warfare: bacteriocin synthesis, toxicity and immunity.' *Trends Microbiol.*, **6**(2):66–71.

Bagge, N., Schuster, M., Hentzer, M., Ciofu, O., Givskov, M., Greenberg, E. P., and Høiby, N. (2004). '*Pseudomonas aeruginosa* biofilms exposed to imipenem exhibit changes in global gene expression and beta-lactamase and alginate production.' *Antimicrob. Agents Chemother.*, **48**(4):1175–1187.

Barben, J., Hafen, G., and Schmid, J. (2005). '*Pseudomonas aeruginosa* in public swimming pools and bathroom water of patients with cystic fibrosis.' *J. Cyst. Fibros.*, **4**(4):227–231.

Bashir, A., Klammer, A. A., Robins, W. P., Chin, C. S., Webster, D., Paxinos, E., Hsu, D., Ashby, M., Wang, S., Peluso, P., *et al.* (2012). 'A hybrid approach for the automated finishing of bacterial genomes.' *Nat. Biotechnol.*, **30**(7):701–707.

Beatson, S. A., Whitchurch, C. B., Sargent, J. L., Levesque, R. C., and Mattick, J. S. (2002). 'Differential regulation of twitching motility and elastase production by Vfr in *Pseudomonas aeruginosa*.' *J. Bacteriol.*, **184**(13):3605–3613.

Becerra, C., Albesa, I., and Eraso, A. J. (2001). 'Leukotoxicity of pyoverdin, production of reactive oxygen species, and effect of UV radiation.' *Biochem. Biophys. Res. Commun.*, **285**(2):414–418.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., *et al.* (2008). 'Accurate whole human genome sequencing using reversible terminator chemistry.' *Nature*, **456**(7218):53–59.

Berka, R. M. and Vasil, M. L. (1982). 'Phospholipase C (heat-labile hemolysin) of *Pseudomonas aeruginosa*: purification and preliminary characterization.' *J. Bacteriol.*, **152**(1):239–245.

Berrouane, Y. F., McNutt, L. A., Buschelman, B. J., Rhomberg, P. R., Sanford, M. D., Hollis, R. J., Pfaller, M. A., and Herwaldt, L. A. (2000).

'Outbreak of severe *Pseudomonas aeruginosa* infections caused by a contaminated drain in a whirlpool bathtub.' *Clin. Infect. Dis.*, **31**(6):1331–1337.

Biosciences, P. (2014). 'New chemistry boosts average read length to 10 kb - 15 kb for PacBio RS II.'

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., *et al.* (1997). 'The complete genome sequence of *Escherichia coli* K-12.' *Science*, **277**(5331):1453–1462.

Bodilis, J., Nsigue-Meilo, S., Besaury, L., and Quillet, L. (2012). 'Variable copy number, intra-genomic heterogeneities and lateral transfers of the 16S rRNA gene in Pseudomonas.' *PLoS ONE*, **7**(4):e35647.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). 'Trimmomatic: a flexible trimmer for Illumina sequence data.' *Bioinformatics*, **30**(15):2114–2120.

Bottomley, M. J., Muraglia, E., Bazzo, R., and Carfi, A. (2007). 'Molecular insights into quorum sensing in the human pathogen *Pseudomonas aeruginosa* from the structure of the virulence regulator LasR bound to its autoinducer.' *J. Biol. Chem.*, **282**(18):13592–13600.

Bowler, P. G., Duerden, B. I., and Armstrong, D. G. (2001). 'Wound microbiology and associated approaches to wound management.' *Clin. Microbiol. Rev.*, **14**(2):244–269.

Brackman, G., Cos, P., Maes, L., Nelis, H. J., and Coenye, T. (2011). 'Quorum sensing inhibitors increase the susceptibility of bacterial biofilms to antibiotics *in vitro* and *in vivo*.' *Antimicrob. Agents Chemother.*, **55**(6):2655–2661.

Bragonzi, A., Paroni, M., Nonis, A., Cramer, N., Montanari, S., Rejman, J., Di Serio, C., Doring, G., and Tummler, B. (2009). '*Pseudomonas aeruginosa* microevolution during cystic fibrosis lung infection establishes clones with adapted virulence.' *Am. J. Respir. Crit. Care Med.*, **180**(2):138–145.

Bredenbruch, F., Geffers, R., Nimtz, M., Buer, J., and Haussler, S. (2006). 'The *Pseudomonas aeruginosa* quinolone signal (PQS) has an iron-chelating activity.' *Environ. Microbiol.*, **8**(8):1318–1329.

Brouwer, S., Pustelny, C., Ritter, C., Klinkert, B., Narberhaus, F., and Haussler, S. (2014). 'The PqsR and RhlR transcriptional regulators determine the level of *Pseudomonas* quinolone signal synthesis in *Pseudomonas aeruginosa* by producing two different *pqsABCDE* mRNA isoforms.' *J. Bacteriol.*, **196**(23):4163–4171.

Burrus, V., Pavlovic, G., Decaris, B., and Guedon, G. (2002). 'Conjugative transposons: the tip of the iceberg.' *Mol. Microbiol.*, **46**(3):601–610.

Byrd, M. S., Pang, B., Mishra, M., Swords, W. E., and Wozniak, D. J. (2010). 'The exopolysaccharide Psl facilitates surface adherence and NF-kappaB activation in A549 cells.' *MBio*, **1**(3):e00140–10.

Cabrol, S., Olliver, A., Pier, G. B., Andremont, A., and Ruimy, R. (2003). 'Transcription of quorum-sensing system genes in clinical and environmental isolates of *Pseudomonas aeruginosa*.' *J. Bacteriol.*, **185**(24):7222–7230.

Camara, M., Williams, P., and Hardman, A. (2002). 'Controlling infection by tuning in and turning down the volume of bacterial small-talk.' *Lancet Infect Dis*, **2**(11):667–676.

Cao, H., Krishnan, G., Goumnerov, B., Tsongalis, J., Tompkins, R., and Rahme, L. G. (2001). 'A quorum sensing-associated virulence gene of *Pseudomonas aeruginosa* encodes a LysR-like transcription regulator with a unique self-regulatory mechanism.' *Proc. Natl. Acad. Sci. U.S.A.*, **98**(25):14613–14618.

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G., and Parkhill, J. (2005). 'ACT: the Artemis Comparison Tool.' *Bioinformatics*, **21**(16):3422–3423.

Cheng, K., Smyth, R. L., Govan, J. R., Doherty, C., Winstanley, C., Denning, N., Heaf, D. P., van Saene, H., and Hart, C. A. (1996). 'Spread of beta-lactam-resistant *Pseudomonas aeruginosa* in a cystic fibrosis clinic.' *Lancet*, **348**(9028):639–642.

Christiaen, S. E., Matthijs, N., Zhang, X. H., Nelis, H. J., Bossier, P., and Coenye, T. (2014). 'Bacteria that inhibit quorum sensing decrease biofilm formation and virulence in *Pseudomonas aeruginosa* PAO1.' *Pathog Dis*, **70**(3):271–279.

Chugani, S. A., Whiteley, M., Lee, K. M., D'Argenio, D., Manoil, C., and Greenberg, E. P. (2001). 'QscR, a modulator of quorum-sensing signal synthesis and virulence in *Pseudomonas aeruginosa*.' *Proc. Natl. Acad. Sci. U.S.A.*, **98**(5):2752–2757.

Chung, J. C., Becq, J., Fraser, L., Schulz-Trieglaff, O., Bond, N. J., Foweraker, J., Bruce, K. D., Smith, G. P., and Welch, M. (2012). 'Genomic variation among contemporary *Pseudomonas aeruginosa* isolates from chronically infected cystic fibrosis patients.' *J. Bacteriol.*, **194**(18):4857–4866.

Clarke, J., Wu, H. C., Jayasinghe, L., Patel, A., Reid, S., and Bayley, H. (2009). 'Continuous base identification for single-molecule nanopore DNA sequencing.' *Nat Nanotechnol*, **4**(4):265–270.

Clarke, P. and Richmond, M. (1975). *Genetics and Biochemistry of Pseudomonas*. John Wiley & Sons Ltd. ISBN 0471158968.

Cobben, N. A., Drent, M., Jonkers, M., Wouters, E. F., Vaneechoutte, M., and Stobberingh, E. E. (1996). 'Outbreak of severe *Pseudomonas aeruginosa* respiratory infections due to contaminated nebulizers.' *J. Hosp. Infect.*, **33**(1):63–70.

Coil, D., Jospin, G., and Darling, A. E. (2015). 'A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data.' *Bioinformatics*, **31**(4):587–589.

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., *et al.* (1998). 'Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence.' *Nature*, **393**(6685):537–544.

Coleman, J. P., Hudson, L. L., McKnight, S. L., Farrow, J. M., Calfee, M. W., Lindsey, C. A., and Pesci, E. C. (2008). '*Pseudomonas aeruginosa* PqsA is an anthranilate-coenzyme A ligase.' *J. Bacteriol.*, **190**(4):1247–1255.

Conibear, T. C., Willcox, M. D., Flanagan, J. L., and Zhu, H. (2012). 'Characterization of protease IV expression in *Pseudomonas aeruginosa* clinical isolates.' *J. Med. Microbiol.*, **61**(Pt 2):180–190.

Conway, S. P., Brownlee, K. G., Denton, M., and Peckham, D. G. (2003). 'Antibiotic treatment of multidrug-resistant organisms in cystic fibrosis.' *Am J Respir Med*, **2**(4):321–332.

Costerton, J. W., Stewart, P. S., and Greenberg, E. P. (1999). 'Bacterial biofilms: a common cause of persistent infections.' *Science*, **284**(5418):1318–1322.

Cramer, N., Klockgether, J., Wrasman, K., Schmidt, M., Davenport, C. F., and Tummler, B. (2011). 'Microevolution of the major common *Pseudomonas aeruginosa* clones C and PA14 in cystic fibrosis lungs.' *Environ. Microbiol.*, **13**(7):1690–1704.

Crnich, C. J., Gordon, B., and Andes, D. (2003). 'Hot tub-associated necrotizing pneumonia due to *Pseudomonas aeruginosa*.' *Clin. Infect. Dis.*, **36**(3):e55–57.

Crouch Brewer, S., Wunderink, R. G., Jones, C. B., and Leeper, K. V. (1996). 'Ventilator-associated pneumonia due to *Pseudomonas aeruginosa*.' *Chest*, **109**(4):1019–1029.

Cryz, S. J., Pitt, T. L., Furer, E., and Germanier, R. (1984). 'Role of lipopolysaccharide in virulence of *Pseudomonas aeruginosa*.' *Infect. Immun.*, **44**(2):508–513.

Cuzick, A., Stirling, F. R., Lindsay, S. L., and Evans, T. J. (2006). 'The type III pseudomonal exotoxin U activates the c-Jun NH2-terminal kinase pathway and increases human epithelial interleukin-8 production.' *Infect. Immun.*, **74**(7):4104–4113.

Darch, S. E., McNally, A., Harrison, F., Corander, J., Barr, H. L., Paszkiewicz, K., Holden, S., Fogarty, A., Crusz, S. A., and Diggle, S. P. (2015). 'Recombination is a key driver of genomic and phenotypic diversity in a *Pseudomonas aeruginosa* population during cystic fibrosis infection.' *Sci Rep*, **5**:7649.

D'Argenio, D. A., Calfee, M. W., Rainey, P. B., and Pesci, E. C. (2002). 'Autolysis and autoaggregation in *Pseudomonas aeruginosa* colony morphology mutants.' *J. Bacteriol.*, **184**(23):6481–6489.

D'Argenio, D. A., Wu, M., Hoffman, L. R., Kulasekara, H. D., Deziel, E., Smith, E. E., Nguyen, H., Ernst, R. K., Larson Freeman, T. J., Spencer, D. H., *et al.* (2007). 'Growth phenotypes of *Pseudomonas aeruginosa lasR* mutants adapted to the airways of cystic fibrosis patients.' *Mol. Microbiol.*, **64**(2):512–533.

Davey, M. E., Caiazza, N. C., and O'Toole, G. A. (2003). 'Rhamnolipid surfactant production affects biofilm architecture in *Pseudomonas aeruginosa* PAO1.' *J. Bacteriol.*, **185**(3):1027–1036.

Davies, J. C. (2002). '*Pseudomonas aeruginosa* in cystic fibrosis: pathogenesis and persistence.' *Paediatr Respir Rev*, **3**(2):128–134.

Davis, R. J., Jensen, S. O., Van Hal, S., Espedido, B., Gordon, A., Farhat, R., and Chan, R. (2015). 'Whole Genome Sequencing in Real-Time Investigation and Management of a *Pseudomonas aeruginosa* Outbreak on a Neonatal Intensive Care Unit.' *Infect Control Hosp Epidemiol*:1–7.

Dekimpe, V. and Deziel, E. (2009). 'Revisiting the quorum-sensing hierarchy in *Pseudomonas aeruginosa*: the transcriptional regulator RhlR regulates LasR-specific factors.' *Microbiology (Reading, Engl.)*, **155**(Pt 3):712–723.

van Delden, C., Comte, R., and Bally, A. M. (2001). 'Stringent response activates quorum sensing and modulates cell density-dependent gene expression in *Pseudomonas aeruginosa*.' *J. Bacteriol.*, **183**(18):5376–5384.

Desper, R. and Gascuel, O. (2004). 'Theoretical foundation of the balanced minimum evolution method of phylogenetic inference and its relationship to weighted least-squares tree fitting.' *Mol. Biol. Evol.*, **21**(3):587–598.

Deziel, E., Gopalan, S., Tampakaki, A. P., Lepine, F., Padfield, K. E., Saucier, M., Xiao, G., and Rahme, L. G. (2005). 'The contribution of MvfR to *Pseudomonas aeruginosa* pathogenesis and quorum sensing circuitry regulation: multiple quorum sensing-regulated genes are modulated without affecting *lasRI*, *rhlRI* or the production of *N*-acyl-L-homoserine lactones.' *Mol. Microbiol.*, **55**(4):998–1014.

Diggle, S. P., Matthijs, S., Wright, V. J., Fletcher, M. P., Chhabra, S. R., Lamont, I. L., Kong, X., Hider, R. C., Cornelis, P., Camara, M., *et al.* (2007). 'The *Pseudomonas aeruginosa* 4-quinolone signal molecules HHQ

and PQS play multifunctional roles in quorum sensing and iron entrapment.' *Chem. Biol.*, **14**(1):87–96.

Diggle, S. P., Winzer, K., Chhabra, S. R., Worrall, K. E., Camara, M., and Williams, P. (2003). 'The *Pseudomonas aeruginosa* quinolone signal molecule overcomes the cell density-dependency of the quorum sensing hierarchy, regulates *rhl*-dependent genes at the onset of stationary phase and can be produced in the absence of LasR.' *Mol. Microbiol.*, **50**(1):29–43.

van Dijk, E. L., Auger, H., Jaszczyszyn, Y., and Thermes, C. (2014). 'Ten years of next-generation sequencing technology.' *Trends Genet.*, **30**(9):418–426.

Domingues, S., Harms, K., Fricke, W. F., Johnsen, P. J., da Silva, G. J., and Nielsen, K. M. (2012). 'Natural transformation facilitates transfer of transposons, integrons and gene cassettes between bacterial species.' *PLoS Pathog.*, **8**(8):e1002837.

Doran, A. G. and Creevey, C. J. (2013). 'Snpdat: easy and rapid annotation of results from de novo snp discovery projects for model and non-model organisms.' *BMC Bioinformatics*, **14**:45.

Dowd, S. E., Sun, Y., Secor, P. R., Rhoads, D. D., Wolcott, B. M., James, G. A., and Wolcott, R. D. (2008). 'Survey of bacterial diversity in chronic wounds using pyrosequencing, DGGE, and full ribosome shotgun sequencing.' *BMC Microbiol.*, **8**:43.

Drees, S. L. and Fetzner, S. (2015). 'PqsE of *Pseudomonas aeruginosa* Acts as Pathway-Specific Thioesterase in the Biosynthesis of Alkylquinolone Signaling Molecules.' *Chem. Biol.*, **22**(5):611–618.

Earl, D., Bradnam, K., St John, J., Darling, A., Lin, D., Fass, J., Yu, H. O., Buffalo, V., Zerbino, D. R., Diekhans, M., *et al.* (2011). 'Assemblathon 1: a competitive assessment of de novo short read assembly methods.' *Genome Res.*, **21**(12):2224–2241.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., *et al.* (2009). 'Real-time DNA sequencing from single polymerase molecules.' *Science*, **323**(5910):133–138.

Engel, L. S., Hill, J. M., Caballero, A. R., Green, L. C., and O'Callaghan, R. J. (1998a). 'Protease IV, a unique extracellular protease and virulence factor from *Pseudomonas aeruginosa*.' *J. Biol. Chem.*, **273**(27):16792–16797.

Engel, L. S., Hill, J. M., Moreau, J. M., Green, L. C., Hobden, J. A., and O'Callaghan, R. J. (1998b). '*Pseudomonas aeruginosa* protease IV produces corneal damage and contributes to bacterial virulence.' *Invest. Ophthalmol. Vis. Sci.*, **39**(3):662–665.

Erickson, D. L., Lines, J. L., Pesci, E. C., Venturi, V., and Storey, D. G. (2004). '*Pseudomonas aeruginosa relA* contributes to virulence in *Drosophila melanogaster*.' *Infect. Immun.*, **72**(10):5638–5645.

Ewing, B. and Green, P. (1998). 'Base-calling of automated sequencer traces using phred. II. Error probabilities.' *Genome Res.*, **8**(3):186–194.

Fedoroff, N. V. (2012). 'Presidential address. Transposable elements, epigenetics, and genome evolution.' *Science*, **338**(6108):758–767.

Feliziani, S., Marvig, R. L., Lujan, A. M., Moyano, A. J., Di Rienzo, J. A., Krogh Johansen, H., Molin, S., and Smania, A. M. (2014). 'Coexistence and within-host evolution of diversified lineages of hypermutable *Pseudomonas aeruginosa* in long-term cystic fibrosis infections.' *PLoS Genet.*, **10**(10):e1004651.

Fetar, H., Gilmour, C., Klinoski, R., Daigle, D. M., Dean, C. R., and Poole, K. (2011). '*mexEF-oprN* multidrug efflux operon of *Pseudomonas aeruginosa*: regulation by the MexT activator in response to nitrosative stress and chloramphenicol.' *Antimicrob. Agents Chemother.*, **55**(2):508–514.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., and Merrick, J. M. (1995). 'Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.' *Science*, **269**(5223):496–512.

Fleiszig, S. M., Efron, N., and Pier, G. B. (1992). 'Extended contact lens wear enhances *Pseudomonas aeruginosa* adherence to human corneal epithelium.' *Invest. Ophthalmol. Vis. Sci.*, **33**(10):2908–2916.

Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A., Korlach, J., and Turner, S. W. (2010). 'Direct detection of DNA methylation during single-molecule, real-time sequencing.' *Nat. Methods*, **7**(6):461–465.

Fothergill, J. L., Mowat, E., Ledson, M. J., Walshaw, M. J., and Winstanley, C. (2010). 'Fluctuations in phenotypes and genotypes within populations of *Pseudomonas aeruginosa* in the cystic fibrosis lung during pulmonary exacerbations.' *J. Med. Microbiol.*, **59**(Pt 4):472–481.

Fothergill, J. L., Neill, D. R., Loman, N., Winstanley, C., and Kadioglu, A. (2014). '*Pseudomonas aeruginosa* adaptation in the nasopharyngeal reservoir leads to migration and persistence in the lungs.' *Nat Commun*, **5**:4780.

Frank, D. N., Wysocki, A., Specht-Glick, D. D., Rooney, A., Feldman, R. A., St Amand, A. L., Pace, N. R., and Trent, J. D. (2009). 'Microbial diversity in chronic open wounds.' *Wound Repair Regen*, **17**(2):163–172.

Freschi, L., Jeukens, J., Kukavica-Ibrulj, I., Boyle, B., Dupont, M., Laroche, J., Larose, S., Maaroufi, H., Fothergill, J. L., Moore, M., *et al.* (2015). 'Clinical utilisation of genomics data produced by the international *Pseudomonas aeruginosa* consortium.' *Frontiers in Microbiology*.

Frithz-Lindsten, E., Du, Y., Rosqvist, R., and Forsberg, A. (1997). 'Intracellular targeting of exoenzyme S of *Pseudomonas aeruginosa* via type III-dependent translocation induces phagocytosis resistance, cytotoxicity and disruption of actin microfilaments.' *Mol. Microbiol.*, **25**(6):1125–1139.

Fukuda, H., Hosaka, M., Hirai, K., and Iyobe, S. (1990). 'New norfloxacin resistance gene in *Pseudomonas aeruginosa* PAO.' *Antimicrob. Agents Chemother.*, **34**(9):1757–1761.

Fuller, C. W., Middendorf, L. R., Benner, S. A., Church, G. M., Harris, T., Huang, X., Jovanovich, S. B., Nelson, J. R., Schloss, J. A., Schwartz, D. C., *et al.* (2009). 'The challenges of sequencing by synthesis.' *Nat. Biotechnol.*, **27**(11):1013–1023.

Fuqua, W. C., Winans, S. C., and Greenberg, E. P. (1994). 'Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators.' *J. Bacteriol.*, **176**(2):269–275.

Gallagher, L. A. and Manoil, C. (2001). '*Pseudomonas aeruginosa* PAO1 kills *Caenorhabditis elegans* by cyanide poisoning.' *J. Bacteriol.*, **183**(21):6207–6214.

Gallagher, L. A., McKnight, S. L., Kuznetsova, M. S., Pesci, E. C., and Manoil, C. (2002). 'Functions required for extracellular quinolone signaling by *Pseudomonas aeruginosa*.' *J. Bacteriol.*, **184**(23):6472–6480.

Gambello, M. J. and Iglewski, B. H. (1991). 'Cloning and characterization of the *Pseudomonas aeruginosa lasR* gene, a transcriptional activator of elastase expression.' *J. Bacteriol.*, **173**(9):3000–3009.

Gambello, M. J., Kaye, S., and Iglewski, B. H. (1993). 'LasR of *Pseudomonas aeruginosa* is a transcriptional activator of the alkaline protease gene (*apr*) and an enhancer of exotoxin A expression.' *Infect. Immun.*, **61**(4):1180–1184.

Gardy, J. L., Johnston, J. C., Ho Sui, S. J., Cook, V. J., Shah, L., Brodkin, E., Rempel, S., Moore, R., Zhao, Y., Holt, R., *et al.* (2011). 'Whole-genome sequencing and social-network analysis of a tuberculosis outbreak.' *N. Engl. J. Med.*, **364**(8):730–739.

Gerth, M. L., Nigon, L. V., and Patrick, W. M. (2012). 'Characterization of the indole-3-glycerol phosphate synthase from *Pseudomonas aeruginosa* PAO1.' *Protein J.*, **31**(5):359–365.

Gessard, G. (1984). 'Classics in infectious diseases. On the blue and green coloration that appears on bandages. By Carle Gessard (1850-1925).' *Rev. Infect. Dis.*, **6 Suppl 3**:S775–776.

Giwercman, B., Jensen, E. T., Høiby, N., Kharazmi, A., and Costerton, J. W. (1991). 'Induction of beta-lactamase production in *Pseudomonas aeruginosa* biofilm.' *Antimicrob. Agents Chemother.*, **35**(5):1008–1010.

Gould, T. A., Schweizer, H. P., and Churchill, M. E. (2004). 'Structure of the *Pseudomonas aeruginosa* acyl-homoserinelactone synthase LasI.' *Mol. Microbiol.*, **53**(4):1135–1146.

Grada, A. and Weinbrecht, K. (2013). 'Next-generation sequencing: methodology and application.' *J. Invest. Dermatol.*, **133**(8):e11.

Green, M. R. M. R., Sambrook, J., and Laboratory, C. S. H. (2012). *Molecular cloning : a laboratory manual*. Cold Spring Harbor, N.Y. : Cold Spring Harbor Laboratory Press, 4th ed. / michael r. green, joseph sambrook edition. ISBN 9781936113415 (3 v. set : cased). Rev. ed. of: Molecular cloning : a laboratory manual / Joseph Sambrook, David W. Russell. 2001.

Green, S. K., Schroth, M. N., Cho, J. J., Kominos, S. K., and Vitanza-jack, V. B. (1974). 'Agricultural plants and soil as a reservoir for *Pseudomonas aeruginosa*.' *Appl Microbiol*, **28**(6):987–991.

Grigoriev, A. (1998). 'Analyzing genomes with cumulative skew diagrams.' *Nucleic Acids Res.*, **26**(10):2286–2290.

Haas, D. and Holloway, B. W. (1976). 'R factor variants with enhanced sex factor activity in *Pseudomonas aeruginosa*.' *Mol. Gen. Genet.*, **144**(3):243–251.

Haenni, M., Hocquet, D., Ponsin, C., Cholley, P., Guyeux, C., Madec, J. Y., and Bertrand, X. (2015). 'Population structure and antimicrobial susceptibility of *Pseudomonas aeruginosa* from animal infections in France.' *BMC Vet. Res.*, **11**(1):9.

Hamood, A. N., Griswold, J., and Colmer, J. (1996). 'Characterization of elastase-deficient clinical isolates of *Pseudomonas aeruginosa*.' *Infect. Immun.*, **64**(8):3154–3160.

Hansen, S. K., Rau, M. H., Johansen, H. K., Ciofu, O., Jelsbak, L., Yang, L., Folkesson, A., Jarmer, H. Ø., Aanaes, K., von Buchwald, C., *et al.* (2012). 'Evolution and diversification of *Pseudomonas aeruginosa* in the paranasal sinuses of cystic fibrosis children have implications for chronic lung infection.' *ISME J*, **6**(1):31–45.

Harris, S. R., Feil, E. J., Holden, M. T., Quail, M. A., Nickerson, E. K., Chantratita, N., Gardete, S., Tavares, A., Day, N., Lindsay, J. A., *et al.* (2010). 'Evolution of MRSA during hospital transmission and intercontinental spread.' *Science*, **327**(5964):469–474.

Haussler, S. and Becker, T. (2008). 'The *Pseudomonas* quinolone signal (PQS) balances life and death in *Pseudomonas aeruginosa* populations.' *PLoS Pathog.*, **4**(9):e1000166.

Hayden, E. C. (2014). 'Data from pocket-sized genome sequencer unveiled.'

He, J., Baldini, R. L., Deziel, E., Saucier, M., Zhang, Q., Liberati, N. T., Lee, D., Urbach, J., Goodman, H. M., and Rahme, L. G. (2004). 'The broad host range pathogen *Pseudomonas aeruginosa* strain PA14 carries two pathogenicity islands harboring plant and animal virulence genes.' *Proc. Natl. Acad. Sci. U.S.A.*, **101**(8):2530–2535.

Heeb, S., Fletcher, M. P., Chhabra, S. R., Diggle, S. P., Williams, P., and Camara, M. (2011). 'Quinolones: from antibiotics to autoinducers.' *FEMS Microbiol. Rev.*, **35**(2):247–274.

Heeb, S., Itoh, Y., Nishijyo, T., Schnider, U., Keel, C., Wade, J., Walsh, U., O'Gara, F., and Haas, D. (2000). 'Small, stable shuttle vectors based on the minimal pVS1 replicon for use in gram-negative, plant-associated bacteria.' *Mol. Plant Microbe Interact.*, **13**(2):232–237.

Hentzer, M., Riedel, K., Rasmussen, T. B., Heydorn, A., Andersen, J. B., Parsek, M. R., Rice, S. A., Eberl, L., Molin, S., Høiby, N., *et al.* (2002). 'Inhibition of quorum sensing in *Pseudomonas aeruginosa* biofilm bacteria by a halogenated furanone compound.' *Microbiology*, **148**(Pt 1):87–102.

Hentzer, M., Wu, H., Andersen, J. B., Riedel, K., Rasmussen, T. B., Bagge, N., Kumar, N., Schembri, M. A., Song, Z., Kristoffersen, P., *et al.* (2003). 'Attenuation of *Pseudomonas aeruginosa* virulence by quorum sensing inhibitors.' *EMBO J.*, **22**(15):3803–3815.

Heurlier, K., Denervaud, V., Haenni, M., Guy, L., Krishnapillai, V., and Haas, D. (2005). 'Quorum-sensing-negative (*lasR*) mutants of *Pseudomonas aeruginosa* avoid cell lysis and death.' *J. Bacteriol.*, **187**(14):4875–4883.

Hobman, J. L., Penn, C. W., and Pallen, M. J. (2007). 'Laboratory strains of *Escherichia coli*: model citizens or deceitful delinquents growing old disgracefully?' *Mol. Microbiol.*, **64**(4):881–885.

Hoffman, L. R., Kulasekara, H. D., Emerson, J., Houston, L. S., Burns, J. L., Ramsey, B. W., and Miller, S. I. (2009). '*Pseudomonas aeruginosa lasR* mutants are associated with cystic fibrosis lung disease progression.' *J. Cyst. Fibros.*, **8**(1):66–70.

Holloway, B. W. (1955). 'Genetic recombination in *Pseudomonas aeruginosa*.' *J. Gen. Microbiol.*, **13**(3):572–581.

Holloway, B. W. (1969). 'Genetics of *Pseudomonas*.' *Bacteriol Rev*, **33**(3):419–443.

Holloway, B. W. and Jennings, P. A. (1958). 'An infectious fertility factor for *Pseudomonas aeruginosa*.' *Nature*, **181**(4612):855–856.

Holloway, B. W., Krishnapillai, V., and Morgan, A. F. (1979). 'Chromosomal genetics of Pseudomonas.' *Microbiol. Rev.*, **43**(1):73–102.

Holloway, B. W. and Rolfe, B. (1964). 'Host Genome Control In Host-Induced Modification Of *Pseudomonas aeruginosa* Phages.' *Virology*, **23**:595–602.

Hooton, T. M., Bradley, S. F., Cardenas, D. D., Colgan, R., Geerlings, S. E., Rice, J. C., Saint, S., Schaeffer, A. J., Tambayh, P. A., Tenke, P., *et al.* (2010). 'Diagnosis, prevention, and treatment of catheter-associated urinary tract infection in adults: 2009 International Clinical Practice Guidelines from the Infectious Diseases Society of America.' *Clin. Infect. Dis.*, **50**(5):625–663.

Howe, K. and Wood, J. M. (2015). 'Using optical mapping data for the improvement of vertebrate genome assemblies.' *Gigascience*, **4**:10.

Hupkova, M., Blahova, J., and Krcmery, V. (1994). 'Transfer, by conjugation and transduction, of resistance to ceftazidime and cefotaxime from the same clinical isolate of *Pseudomonas aeruginosa*.' *Zentralbl. Bakteriol.*, **281**(2):196–200.

Huse, H. K., Kwon, T., Zlosnik, J. E., Speert, D. P., Marcotte, E. M., and Whiteley, M. (2010). 'Parallel evolution in *Pseudomonas aeruginosa* over 39,000 generations *in vivo*.' *MBio*, **1**(4).

International Human Genome Sequencing Consortium (2004). 'Finishing the euchromatic sequence of the human genome.' *Nature*, **431**(7011):931–945.

Isaac, J. H. and Holloway, B. W. (1968). 'Control of pyrimidine biosynthesis in *Pseudomonas aeruginosa*.' *J. Bacteriol.*, **96**(5):1732–1741.

Iyobe, S., Sagai, H., and Mitsuhashi, S. (1981). 'Tn2001, a transposon encoding chloramphenicol resistance in *Pseudomonas aeruginosa*.' *J. Bacteriol.*, **146**(1):141–148.

Jacobs, M. A., Alwood, A., Thaipisuttikul, I., Spencer, D., Haugen, E., Ernst, S., Will, O., Kaul, R., Raymond, C., Levy, R., *et al.* (2003). 'Comprehensive transposon mutant library of *Pseudomonas aeruginosa*.' *Proc. Natl. Acad. Sci. U.S.A.*, **100**(24):14339–14344.

James, C. E., Davies, E. V., Fothergill, J. L., Walshaw, M. J., Beale, C. M., Brockhurst, M. A., and Winstanley, C. (2015). 'Lytic activity by temperate phages of *Pseudomonas aeruginosa* in long-term cystic fibrosis chronic lung infections.' *ISME J*, **9**(6):1391–1398.

Jefferies, J. M., Cooper, T., Yam, T., and Clarke, S. C. (2012). '*Pseudomonas aeruginosa* outbreaks in the neonatal intensive care unit–a systematic review of risk factors and environmental sources.' *J. Med. Microbiol.*, **61**(Pt 8):1052–1061.

Jeukens, J., Boyle, B., Kukavica-Ibrulj, I., Ouellet, M. M., Aaron, S. D., Charette, S. J., Fothergill, J. L., Tucker, N. P., Winstanley, C., and Levesque, R. C. (2014). 'Comparative genomics of isolates of a

Pseudomonas aeruginosa epidemic strain associated with chronic lung infections of cystic fibrosis patients.' *PLoS ONE*, **9**(2):e87611.

Jiricny, N., Diggle, S. P., West, S. A., Evans, B. A., Ballantyne, G., Ross-Gillespie, A., and Griffin, A. S. (2010). 'Fitness correlates with the extent of cheating in a bacterium.' *J. Evol. Biol.*, **23**(4):738–747.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T. L. (2008). 'NCBI BLAST: a better web interface.' *Nucleic Acids Res.*, **36**(Web Server issue):5–9.

Judge, K., Harris, S. R., Reuter, S., Parkhill, J., and Peacock, S. J. (2015). 'Early insights into the potential of the Oxford Nanopore MinION for the detection of antimicrobial resistance genes.' *J. Antimicrob. Chemother.*

Juhas, M., Wiehlmann, L., Huber, B., Jordan, D., Lauber, J., Salunkhe, P., Limpert, A. S., von Gotz, F., Steinmetz, I., Eberl, L., *et al.* (2004). 'Global regulation of quorum sensing and virulence by VqsR in *Pseudomonas aeruginosa*.' *Microbiology (Reading, Engl.)*, **150**(Pt 4):831–841.

Juhas, M., Wiehlmann, L., Salunkhe, P., Lauber, J., Buer, J., and Tummler, B. (2005). 'GeneChip expression analysis of the VqsR regulon of *Pseudomonas aeruginosa* TB.' *FEMS Microbiol. Lett.*, **242**(2):287–295.

Kasianowicz, J. J., Brandin, E., Branton, D., Deamer, D. W., and Deamer, D. W. (1996). 'Characterization of individual polynucleotide molecules using a membrane channel.' *Proc. Natl. Acad. Sci. U.S.A.*, **93**(24):13770–13773.

Kesarwani, M., Hazan, R., He, J., Que, Y. A., Que, Y., Apidianakis, Y., Lesic, B., Xiao, G., Dekimpe, V., Milot, S., *et al.* (2011). 'A quorum sensing regulated small volatile molecule reduces acute virulence and promotes chronic infection phenotypes.' *PLoS Pathog.*, **7**(8):e1002192.

Kessler, E., Safrin, M., Olson, J. C., and Ohman, D. E. (1993). 'Secreted LasA of *Pseudomonas aeruginosa* is a staphylolytic protease.' *J. Biol. Chem.*, **268**(10):7503–7508.

Kiratisin, P., Tucker, K. D., and Passador, L. (2002). 'LasR, a transcriptional activator of *Pseudomonas aeruginosa* virulence genes, functions as a multimer.' *J. Bacteriol.*, **184**(17):4912–4919.

Kiss, P., Lantos, J., Lanyi, B., and Fodre, Z. (1985). 'Characterization of *Pseudomonas aeruginosa* isolated from drinking water by serogrouping, phage sensitivity and pyocin pattern.' *Acta Microbiol. Hung.*, **32**(1):99–106.

Klockgether, J., Cramer, N., Wiehlmann, L., Davenport, C. F., and Tummler, B. (2011). '*Pseudomonas aeruginosa* Genomic Structure and Diversity.' *Front Microbiol*, **2**:150.

Klockgether, J., Munder, A., Neugebauer, J., Davenport, C. F., Stanke, F., Larbig, K. D., Heeb, S., Schock, U., Pohl, T. M., Wiehlmann, L., *et al.* (2010). 'Genome diversity of *Pseudomonas aeruginosa* PAO1 laboratory strains.' *J. Bacteriol.*, **192**(4):1113–1121.

Knirel, Y. A., Bystrova, O. V., Kocharova, N. A., Zahringer, U., and Pier, G. B. (2006). 'Conserved and variable structural features in the lipopolysaccharide of *Pseudomonas aeruginosa*.' *J. Endotoxin Res.*, **12**(6):324–336.

Kohler, T., Epp, S. F., Curty, L. K., and Pechere, J. C. (1999). 'Characterization of MexT, the regulator of the MexE-MexF-OprN multidrug efflux system of *Pseudomonas aeruginosa*.' *J. Bacteriol.*, **181**(20):6300–6305.

Kohler, T., Guanella, R., Carlet, J., and van Delden, C. (2010). 'Quorum sensing-dependent virulence during *Pseudomonas aeruginosa* colonisation and pneumonia in mechanically ventilated patients.' *Thorax*, **65**(8):703–710.

Kong, K. F., Jayawardena, S. R., Indulkar, S. D., Del Puerto, A., Koh, C. L., Høiby, N., and Mathee, K. (2005). '*Pseudomonas aeruginosa* AmpR is a global transcriptional factor that regulates expression of AmpC and PoxB beta-lactamases, proteases, quorum sensing, and other virulence factors.' *Antimicrob. Agents Chemother.*, **49**(11):4567–4575.

Kos, V. N., Desjardins, C. A., Griggs, A., Cerqueira, G., Van Tonder, A., Holden, M. T., Godfrey, P., Palmer, K. L., Bodi, K., Mongodin, E. F., *et al.* (2012). 'Comparative genomics of vancomycin-resistant *Staphylococcus aureus* strains and their positions within the clade most commonly associated with Methicillin-resistant *S. aureus* hospital-acquired infection in the United States.' *MBio*, **3**(3).

Koser, C. U., Holden, M. T., Ellington, M. J., Cartwright, E. J., Brown, N. M., Ogilvy-Stuart, A. L., Hsu, L. Y., Chewapreecha, C., Croucher, N. J., Harris, S. R., *et al.* (2012). 'Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak.' *N. Engl. J. Med.*, **366**(24):2267–2275.

Kresse, A. U., Dinesh, S. D., Larbig, K., and Romling, U. (2003). 'Impact of large chromosomal inversions on the adaptation and evolution of *Pseudomonas aeruginosa* chronically colonizing cystic fibrosis lungs.' *Mol. Microbiol.*, **47**(1):145–158.

Kung, V. L., Ozer, E. A., and Hauser, A. R. (2010). 'The accessory genome of Pseudomonas aeruginosa.' *Microbiol. Mol. Biol. Rev.*, **74**(4):621–641.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borchert, S., *et al.* (1997). 'The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*.' *Nature*, **390**(6657):249–256.

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S. L. (2004). 'Versatile and open software for comparing large genomes.' *Genome Biol.*, **5**(2):R12.

Laarman, A. J., Bardoel, B. W., Ruyken, M., Fernie, J., Milder, F. J., van Strijp, J. A., and Rooijakkers, S. H. (2012). '*Pseudomonas aeruginosa* alkaline protease blocks complement activation via the classical and lectin pathways.' *J. Immunol.*, **188**(1):386–393.

Lam, J. S., Taylor, V. L., Islam, S. T., Hao, Y., and Kocincova, D. (2011). 'Genetic and Functional Diversity of *Pseudomonas aeruginosa* Lipopolysaccharide.' *Front Microbiol*, **2**:118.

Lamb, J. R., Patel, H., Montminy, T., Wagner, V. E., and Iglewski, B. H. (2003). 'Functional domains of the RhlR transcriptional regulator of *Pseudomonas aeruginosa*.' *J. Bacteriol.*, **185**(24):7129–7139.

de Lamballerie, X., Zandotti, C., Vignoli, C., Bollet, C., and de Micco, P. (1992). 'A one-step microbial DNA extraction method using "Chelex 100" suitable for gene amplification.' *Res. Microbiol.*, **143**(8):785–790.

Lamont, I. L., Beare, P. A., Ochsner, U., Vasil, A. I., and Vasil, M. L. (2002). 'Siderophore-mediated signaling regulates virulence factor production in *Pseudomonas aeruginosa*.' *Proc. Natl. Acad. Sci. U.S.A.*, **99**(10):7072–7077.

Latifi, A., Foglino, M., Tanaka, K., Williams, P., and Lazdunski, A. (1996). '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.*, **21**(6):1137–1146.

Latifi, A., Winson, M. K., Foglino, M., Bycroft, B. W., Stewart, G. S., Lazdunski, A., and Williams, P. (1995). 'Multiple homologues of LuxR and LuxI control expression of virulence determinants and secondary metabolites through quorum sensing in *Pseudomonas aeruginosa* PAO1.' *Mol. Microbiol.*, **17**(2):333–343.

Latreille, P., Norton, S., Goldman, B. S., Henkhaus, J., Miller, N., Barbazuk, B., Bode, H. B., Darby, C., Du, Z., Forst, S., *et al.* (2007). 'Optical mapping as a routine tool for bacterial genome sequence finishing.' *BMC Genomics*, **8**:321.

Le, S., Yao, X., Lu, S., Tan, Y., Rao, X., Li, M., Jin, X., Wang, J., Zhao, Y., Wu, N. C., *et al.* (2014). 'Chromosomal DNA deletion confers phage resistance to *Pseudomonas aeruginosa*.' *Sci Rep*, **4**:4738.

Ledergerber, C. and Dessimoz, C. (2011). 'Base-calling for next-generation sequencing platforms.' *Brief. Bioinformatics*, **12**(5):489–497.

Lee, D. G., Urbach, J. M., Wu, G., Liberati, N. T., Feinbaum, R. L., Miyata, S., Diggins, L. T., He, J., Saucier, M., Deziel, E., *et al.* (2006). 'Genomic analysis reveals that Pseudomonas aeruginosa virulence is combinatorial.' *Genome Biol.*, **7**(10):R90.

Lee, S. A., Gallagher, L. A., Thongdee, M., Staudinger, B. J., Lippman, S., Singh, P. K., and Manoil, C. (2015). 'General and condition-specific essential functions of *Pseudomonas aeruginosa*.' *Proc. Natl. Acad. Sci. U.S.A.*, **112**(16):5189–5194.

Lequette, Y., Lee, J. H., Ledgham, F., Lazdunski, A., and Greenberg, E. P. (2006). 'A distinct QscR regulon in the *Pseudomonas aeruginosa* quorum-sensing circuit.' *J. Bacteriol.*, **188**(9):3365–3370.

Letunic, I. and Bork, P. (2011). 'Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy.' *Nucleic Acids Res.*, **39**(Web Server issue):W475–478.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). 'The Sequence Alignment/Map format and SAMtools.' *Bioinformatics*, **25**(16):2078–2079.

Li, H. and Homer, N. (2010). 'A survey of sequence alignment algorithms for next-generation sequencing.' *Brief. Bioinformatics*, **11**(5):473–483.

Liang, H., Li, L., Kong, W., Shen, L., and Duan, K. (2009). 'Identification of a novel regulator of the quorum-sensing systems in *Pseudomonas aeruginosa*.' *FEMS Microbiol. Lett.*, **293**(2):196–204.

Liao, Y. C., Lin, S. H., and Lin, H. H. (2015). 'Completing bacterial genome assemblies: strategy and performance comparisons.' *Sci Rep*, **5**:8747.

Lipovy, B., Rihova, H., Hanslianova, M., Gregorova, N., Suchanek, I., and Brychta, P. (2010). 'Prevalence and resistance of *Pseudomonas aeruginosa* in severely burned patients: a 10-year retrospective study.' *Acta Chir Plast*, **52**(2-4):39–43.

Llamas, M. A., van der Sar, A., Chu, B. C., Sparrius, M., Vogel, H. J., and Bitter, W. (2009). 'A Novel extracytoplasmic function (ECF) sigma factor regulates virulence in *Pseudomonas aeruginosa*.' *PLoS Pathog.*, **5**(9):e1000572.

Lodge, J. M., Minchin, S. D., Piddock, L. J., and Busby, S. J. (1990). 'Cloning, sequencing and analysis of the structural gene and regulatory region of the *Pseudomonas aeruginosa* chromosomal ampC beta-lactamase.' *Biochem. J.*, **272**(3):627–631.

Loman, N. (2013). 'Adaptor trim or die: Experiences with Nextera libraries.' `http://nickloman.github.io/high-throughput%20sequencing/genomics/bioinformatics/2013/04/17/adaptor-trim-or-die-experiences-with-nextera-libraries/`, Accessed 10 September 2013.

de Lorenzo, V., Cases, I., Herrero, M., and Timmis, K. N. (1993). 'Early and late responses of TOL promoters to pathway inducers: identification of postexponential promoters in Pseudomonas putida with lacZ-tet bicistronic reporters.' *J. Bacteriol.*, **175**(21):6902–6907.

Loutit, J. S. and Marinus, M. G. (1968). 'Investigation of the mating system of *Pseudomonas aeruginosa* strain 1. II. Mapping of a number of early markers.' *Genet. Res.*, **12**(1):37–44.

Luong, P. M., Shogan, B. D., Zaborin, A., Belogortseva, N., Shrout, J. D., Zaborina, O., and Alverdy, J. C. (2014). 'Emergence of the P2 phenotype in *Pseudomonas aeruginosa* PAO1 strains involves various mutations in *mexT* or *mexF*.' *J. Bacteriol.*, **196**(2):504–513.

Mackelprang, R., Waldrop, M. P., DeAngelis, K. M., David, M. M., Chavarria, K. L., Blazewicz, S. J., Rubin, E. M., and Jansson, J. K. (2011). 'Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw.' *Nature*, **480**(7377):368–371.

Madigan, M. and Martinko, J. M. (2006). *Brock Biology of Microorganisms*. Pearson Education International. ISBN 0131968939.

Mahajan-Miklos, S., Tan, M. W., Rahme, L. G., and Ausubel, F. M. (1999). 'Molecular mechanisms of bacterial virulence elucidated using a *Pseudomonas aeruginosa-Caenorhabditis elegans* pathogenesis model.' *Cell*, **96**(1):47–56.

Mahenthiralingam, E., Campbell, M. E., Foster, J., Lam, J. S., and Speert, D. P. (1996). 'Random amplified polymorphic DNA typing of *Pseudomonas aeruginosa* isolates recovered from patients with cystic fibrosis.' *J. Clin. Microbiol.*, **34**(5):1129–1135.

Mahillon, J. and Chandler, M. (1998). 'Insertion sequences.' *Microbiol. Mol. Biol. Rev.*, **62**(3):725–774.

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., *et al.* (2015). 'CDD: NCBI's conserved domain database.' *Nucleic Acids Res.*, **43**(Database issue):D222–226.

Marvig, R. L., Johansen, H. K., Molin, S., and Jelsbak, L. (2013). 'Genome analysis of a transmissible lineage of pseudomonas aeruginosa reveals pathoadaptive mutations and distinct evolutionary paths of hypermutators.' *PLoS Genet.*, **9**(9):e1003741.

Maseda, H., Saito, K., Nakajima, A., and Nakae, T. (2000). 'Variation of the *mexT* gene, a regulator of the MexEF-OprN efflux pump expression in wild-type strains of *Pseudomonas aeruginosa*.' *FEMS Microbiol. Lett.*, **192**(1):107–112.

Mashburn, L. M. and Whiteley, M. (2005). 'Membrane vesicles traffic signals and facilitate group activities in a prokaryote.' *Nature*, **437**(7057):422–425.

Mashburn-Warren, L., Howe, J., Garidel, P., Richter, W., Steiniger, F., Roessle, M., Brandenburg, K., and Whiteley, M. (2008). 'Interaction of quorum signals with outer membrane lipids: insights into prokaryotic membrane vesicle formation.' *Mol. Microbiol.*, **69**(2):491–502.

Mathee, K., Narasimhan, G., Valdes, C., Qiu, X., Matewish, J. M., Koehrsen, M., Rokas, A., Yandava, C. N., Engels, R., Zeng, E., *et al.* (2008). 'Dynamics of *Pseudomonas aeruginosa* genome evolution.' *Proc. Natl. Acad. Sci. U.S.A.*, **105**(8):3100–3105.

McClintock, B. (1950). 'The origin and behavior of mutable loci in maize.' *Proc. Natl. Acad. Sci. U.S.A.*, **36**(6):344–355.

McEvoy, C. R., Tsuji, B., Gao, W., Seemann, T., Porter, J. L., Doig, K., Ngo, D., Howden, B. P., and Stinear, T. P. (2013). 'Decreased vancomycin susceptibility in *Staphylococcus aureus* caused by IS256 tempering of WalKR expression.' *Antimicrob. Agents Chemother.*, **57**(7):3240–3249.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., *et al.* (2010). 'The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.' *Genome Res.*, **20**(9):1297–1303.

McKenna, M. (2013). 'Antibiotic resistance: the last resort.' *Nature*, **499**(7459):394–396.

McKeown, J. (2013). 'In search for new links and inhibitors between quorum sensing and virulence in *Pseudomonas aeruginosa* strains.' PhD Thesis, University of Nottingham, UK.

McKnight, S. L., Iglewski, B. H., and Pesci, E. C. (2000). 'The *Pseudomonas* quinolone signal regulates *rhl* quorum sensing in *Pseudomonas aeruginosa*.' *J. Bacteriol.*, **182**(10):2702–2708.

Mendelowitz, L. and Pop, M. (2014). 'Computational methods for optical mapping.' *Gigascience*, **3**(1):33.

Mutreja, A., Kim, D. W., Thomson, N. R., Connor, T. R., Lee, J. H., Kariuki, S., Croucher, N. J., Choi, S. Y., Harris, S. R., Lebens, M., *et al.* (2011). 'Evidence for several waves of global transmission in the seventh cholera pandemic.' *Nature*, **477**(7365):462–465.

Nielsen, H. B., Almeida, M., Juncker, A. S., Rasmussen, S., Li, J., Sunagawa, S., Plichta, D. R., Gautier, L., Pedersen, A. G., Le Chatelier, E., *et al.* (2014). 'Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes.' *Nat. Biotechnol.*, **32**(8):822–828.

Nikaido, H. (1998). 'Multiple antibiotic resistance and efflux.' *Curr. Opin. Microbiol.*, **1**(5):516–523.

Ochsner, U. A. and Reiser, J. (1995). 'Autoinducer-mediated regulation of rhamnolipid biosurfactant synthesis in *Pseudomonas aeruginosa*.' *Proc. Natl. Acad. Sci. U.S.A.*, **92**(14):6424–6428.

Ohmagari, N., Hanna, H., Graviss, L., Hackett, B., Perego, C., Gonzalez, V., Dvorak, T., Hogan, H., Hachem, R., Rolston, K., *et al.* (2005). 'Risk factors for infections with multidrug-resistant *Pseudomonas aeruginosa* in patients with cancer.' *Cancer*, **104**(1):205–212.

Ohman, D. E. and Chakrabarty, A. M. (1982). 'Utilization of human respiratory secretions by mucoid *Pseudomonas aeruginosa* of cystic fibrosis origin.' *Infect. Immun.*, **37**(2):662–669.

O'Loughlin, C. T., Miller, L. C., Siryaporn, A., Drescher, K., Semmelhack, M. F., and Bassler, B. L. (2013). 'A quorum-sensing inhibitor blocks *Pseudomonas aeruginosa* virulence and biofilm formation.' *Proc. Natl. Acad. Sci. U.S.A.*, **110**(44):17981–17986.

O'Toole, G. A. and Kolter, R. (1998). 'Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development.' *Mol. Microbiol.*, **30**(2):295–304.

Otto, T. D., Dillon, G. P., Degrave, W. S., and Berriman, M. (2011). 'RATT: Rapid Annotation Transfer Tool.' *Nucleic Acids Res.*, **39**(9):e57.

Parkhill, J., Wren, B. W., Thomson, N. R., Titball, R. W., Holden, M. T., Prentice, M. B., Sebaihia, M., James, K. D., Churcher, C., Mungall, K. L., *et al.* (2001). 'Genome sequence of *Yersinia pestis*, the causative agent of plague.' *Nature*, **413**(6855):523–527.

Passador, L., Cook, J. M., Gambello, M. J., Rust, L., and Iglewski, B. H. (1993). 'Expression of *Pseudomonas aeruginosa* virulence genes requires cell-to-cell communication.' *Science*, **260**(5111):1127–1130.

Pavlovskis, O. R., Iglewski, B. H., and Pollack, M. (1978). 'Mechanism of action of *Pseudomonas aeruginosa* exotoxin A in experimental mouse infections: adenosine diphosphate ribosylation of elongation factor 2.' *Infect. Immun.*, **19**(1):29–33.

Peacock, S. (2014). 'Health care: Bring microbial sequencing to hospitals.' *Nature*, **509**(7502):557–559.

Pearson, J. P., Passador, L., Iglewski, B. H., and Greenberg, E. P. (1995). 'A second *N*-acylhomoserine lactone signal produced by *Pseudomonas aeruginosa*.' *Proc. Natl. Acad. Sci. U.S.A.*, **92**(5):1490–1494.

Pedersen, S. S., Høiby, N., Espersen, F., and Koch, C. (1992). 'Role of alginate in infection with mucoid *Pseudomonas aeruginosa* in cystic fibrosis.' *Thorax*, **47**(1):6–13.

Persat, A., Inclan, Y. F., Engel, J. N., Stone, H. A., and Gitai, Z. (2015). 'Type IV pili mechanochemically regulate virulence factors in *Pseudomonas aeruginosa*.' *Proc. Natl. Acad. Sci. U.S.A.*, **112**(24):7563–7568.

Pesci, E. C., Milbank, J. B., Pearson, J. P., McKnight, S., Kende, A. S., Greenberg, E. P., and Iglewski, B. H. (1999). 'Quinolone signaling in the cell-to-cell communication system of *Pseudomonas aeruginosa*.' *Proc. Natl. Acad. Sci. U.S.A.*, **96**(20):11229–11234.

Pesci, E. C., Pearson, J. P., Seed, P. C., and Iglewski, B. H. (1997). 'Regulation of *las* and *rhl* quorum sensing in *Pseudomonas aeruginosa*.' *J. Bacteriol.*, **179**(10):3127–3132.

Pessi, G., Williams, F., Hindle, Z., Heurlier, K., Holden, M. T., Camara, M., Haas, D., and Williams, P. (2001). 'The global posttranscriptional regulator RsmA modulates production of virulence determinants and *N*-acylhomoserine lactones in *Pseudomonas aeruginosa*.' *J. Bacteriol.*, **183**(22):6676–6683.

PHE (2014). 'Polymicrobial bacteraemia and fungaemia in England, Wales and Northern Ireland, 2013.' *Health Protection Report*, **8**(48).

Pier, G. B. (2007). '*Pseudomonas aeruginosa* lipopolysaccharide: a major virulence factor, initiator of inflammation and target for effective immunity.' *Int. J. Med. Microbiol.*, **297**(5):277–295.

Pohl, S., Klockgether, J., Eckweiler, D., Khaledi, A., Schniederjans, M., Chouvarine, P., Tummler, B., and Haussler, S. (2014). 'The extensive set of accessory *Pseudomonas aeruginosa* genomic components.' *FEMS Microbiol. Lett.*, **356**(2):235–241.

Ponstingl, H. (2012). *SMALT - Sequence Mapping and Alignment Tool*. `http://www.sanger.ac.uk/resources/software/smalt/`.

Poole, K. (2001). 'Multidrug efflux pumps and antimicrobial resistance in *Pseudomonas aeruginosa* and related organisms.' *J. Mol. Microbiol. Biotechnol.*, **3**(2):255–264.

Popat, R., Crusz, S. A., Messina, M., Williams, P., West, S. A., and Diggle, S. P. (2012). 'Quorum-sensing and cheating in bacterial biofilms.' *Proc. Biol. Sci.*, **279**(1748):4765–4771.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., *et al.* (2010). 'A human gut microbial gene catalogue established by metagenomic sequencing.' *Nature*, **464**(7285):59–65.

Qin, N., Yang, F., Li, A., Prifti, E., Chen, Y., Shao, L., Guo, J., Le Chatelier, E., Yao, J., Wu, L., *et al.* (2014). 'Alterations of the human gut microbiome in liver cirrhosis.' *Nature*, **513**(7516):59–64.

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P., and Gu, Y. (2012). 'A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers.' *BMC Genomics*, **13**:341.

Quick, J., Cumley, N., Wearn, C. M., Niebel, M., Constantinidou, C., Thomas, C. M., Pallen, M. J., Moiemen, N. S., Bamford, A., Oppenheim, B., *et al.* (2014). 'Seeking the source of *Pseudomonas aeruginosa* infections in a recently opened hospital: an observational study using whole-genome sequencing.' *BMJ Open*, **4**(11):e006278.

Quinlan, A. R. and Hall, I. M. (2010). 'BEDTools: a flexible suite of utilities for comparing genomic features.' *Bioinformatics*, **26**(6):841–842.

Raeside, C., Gaffe, J., Deatherage, D. E., Tenaillon, O., Briska, A. M., Ptashkin, R. N., Cruveiller, S., Medigue, C., Lenski, R. E., Barrick, J. E., *et al.* (2014). 'Large chromosomal rearrangements during a long-term evolution experiment with Escherichia coli.' *MBio*, **5**(5):e01377–01314.

Rampioni, G., Bertani, I., Zennaro, E., Polticelli, F., Venturi, V., and Leoni, L. (2006). 'The quorum-sensing negative regulator RsaL of *Pseudomonas aeruginosa* binds to the *lasI* promoter.' *J. Bacteriol.*, **188**(2):815–819.

Rampioni, G., Schuster, M., Greenberg, E. P., Bertani, I., Grasso, M., Venturi, V., Zennaro, E., and Leoni, L. (2007). 'RsaL provides quorum sensing homeostasis and functions as a global regulator of gene expression in *Pseudomonas aeruginosa*.' *Mol. Microbiol.*, **66**(6):1557–1565.

Ratnaningsih, E., Dharmsthiti, S., Krishnapillai, V., Morgan, A., Sinclair, M., and Holloway, B. W. (1990). 'A combined physical and genetic map of *Pseudomonas aeruginosa* PAO.' *J. Gen. Microbiol.*, **136**(12):2351–2357.

Rau, M. H., Marvig, R. L., Ehrlich, G. D., Molin, S., and Jelsbak, L. (2012). 'Deletion and acquisition of genomic content during early stage adaptation of *Pseudomonas aeruginosa* to a human host environment.' *Environ. Microbiol.*, **14**(8):2200–2211.

Raychaudhuri, A., Jerga, A., and Tipton, P. A. (2005). 'Chemical mechanism and substrate specificity of RhlI, an acylhomoserine lactone synthase from *Pseudomonas aeruginosa*.' *Biochemistry*, **44**(8):2974–2981.

Rebello, J. L. and Jensen, R. A. (1970). 'Metabolic interlock. The multimetabolite control of prephenate dehydratase activity in *Bacillus subtilis*.' *J. Biol. Chem.*, **245**(15):3738–3744.

Reid, A. J., Blake, D. P., Ansari, H. R., Billington, K., Browne, H. P., Bryant, J., Dunn, M., Hung, S. S., Kawahara, F., Miranda-Saavedra, D., *et al.* (2014). 'Genomic analysis of the causative agents of coccidiosis in domestic chickens.' *Genome Res.*, **24**(10):1676–1685.

Reimmann, C., Beyeler, M., Latifi, A., Winteler, H., Foglino, M., Lazdunski, A., and Haas, D. (1997). 'The global activator GacA of *Pseudomonas aeruginosa* PAO positively controls the production of the autoinducer *N*-butyryl-homoserine lactone and the formation of the virulence factors pyocyanin, cyanide, and lipase.' *Mol. Microbiol.*, **24**(2):309–319.

Riley, M. C., Kirkup, B. C., Johnson, J. D., Lesho, E. P., and Ockenhouse, C. F. (2011). 'Rapid whole genome optical mapping of *Plasmodium falciparum*.' *Malar. J.*, **10**:252.

Rocchetta, H. L., Burrows, L. L., and Lam, J. S. (1999). 'Genetics of O-antigen biosynthesis in *Pseudomonas aeruginosa*.' *Microbiol. Mol. Biol. Rev.*, **63**(3):523–553.

Rocha, E. P. and Danchin, A. (2003). 'Essentiality, not expressiveness, drives gene-strand bias in bacteria.' *Nat. Genet.*, **34**(4):377–378.

Romling, U., Grothues, D., Bautsch, W., and Tummler, B. (1989). 'A physical genome map of *Pseudomonas aeruginosa* PAO.' *EMBO J.*, **8**(13):4081–4089.

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M., and Nyren, P. (1996). 'Real-time DNA sequencing using detection of pyrophosphate release.' *Anal. Biochem.*, **242**(1):84–89.

Rose, P. W., Prlic, A., Bi, C., Bluhm, W. F., Christie, C. H., Dutta, S., Green, R. K., Goodsell, D. S., Westbrook, J. D., Woo, J., *et al.* (2015). 'The RCSB Protein Data Bank: views of structural biology for basic and applied research and education.' *Nucleic Acids Res.*, **43**(Database issue):D345–356.

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H., Johnson, K., Milgrew, M. J., Edwards, M., *et al.* (2011). 'An integrated semiconductor device enabling non-optical genome sequencing.' *Nature*, **475**(7356):348–352.

Roy, P. H., Tetu, S. G., Larouche, A., Elbourne, L., Tremblay, S., Ren, Q., Dodson, R., Harkins, D., Shay, R., Watkins, K., *et al.* (2010). 'Complete genome sequence of the multiresistant taxonomic outlier Pseudomonas aeruginosa PA7.' *PLoS ONE*, **5**(1):e8842.

Royle, P. L., Matsumoto, H., and Holloway, B. W. (1981). 'Genetic circularity of the *Pseudomonas aeruginosa* PAO chromosome.' *J. Bacteriol.*, **145**(1):145–155.

RQIA (2012). 'Independent review of incidents of *Pseudomonas aeruginosa* infection in neonatal units in Northern Ireland: Final Report.'

Rumbaugh, K. P., Griswold, J. A., Iglewski, B. H., and Hamood, A. N. (1999). 'Contribution of quorum sensing to the virulence of *Pseudomonas aeruginosa* in burn wound infections.' *Infect. Immun.*, **67**(11):5854–5862.

Rust, L., Pesci, E. C., and Iglewski, B. H. (1996). 'Analysis of the *Pseudomonas aeruginosa* elastase (*lasB*) regulatory region.' *J. Bacteriol.*, **178**(4):1134–1140.

Sabirova, J. S., Xavier, B. B., Ieven, M., Goossens, H., and Malhotra-Kumar, S. (2014). 'Whole genome mapping as a fast-track tool to assess

genomic stability of sequenced *Staphylococcus aureus* strains.' *BMC Res Notes*, **7**:704.

Sakuragi, Y. and Kolter, R. (2007). 'Quorum-sensing regulation of the biofilm matrix genes (*pel*) of *Pseudomonas aeruginosa*.' *J. Bacteriol.*, **189**(14):5383–5386.

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., Turner, P., Parkhill, J., Loman, N. J., and Walker, A. W. (2014). 'Reagent and laboratory contamination can critically impact sequence-based microbiome analyses.' *BMC Biol.*, **12**:87.

Salunkhe, P., Smart, C. H., Morgan, J. A., Panagea, S., Walshaw, M. J., Hart, C. A., Geffers, R., Tummler, B., and Winstanley, C. (2005). 'A cystic fibrosis epidemic strain of *Pseudomonas aeruginosa* displays enhanced virulence and antimicrobial resistance.' *J. Bacteriol.*, **187**(14):4908–4920.

Sambrook, J. and Russell, D. W. (2001). *Molecular cloning: a laboratory manual, 3rd ed.* Cold Srping Harbor Laboratory Press, Cold Spring Harbor, New York.

Sandoz, K. M., Mitzimberg, S. M., and Schuster, M. (2007). 'Social cheating in *Pseudomonas aeruginosa* quorum sensing.' *Proc. Natl. Acad. Sci. U.S.A.*, **104**(40):15876–15881.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). 'DNA sequencing with chain-terminating inhibitors.' *Proc. Natl. Acad. Sci. U.S.A.*, **74**(12):5463–5467.

Scales, B. S. and Huffnagle, G. B. (2013). 'The microbiome in wound repair and tissue fibrosis.' *J. Pathol.*, **229**(2):323–331.

Schaber, J. A., Carty, N. L., McDonald, N. A., Graham, E. D., Cheluvappa, R., Griswold, J. A., and Hamood, A. N. (2004). 'Analysis of quorum sensing-deficient clinical isolates of *Pseudomonas aeruginosa*.' *J. Med. Microbiol.*, **53**(Pt 9):841–853.

Schertzer, J. W., Brown, S. A., and Whiteley, M. (2010). 'Oxygen levels rapidly modulate *Pseudomonas aeruginosa* social behaviours via substrate limitation of PqsH.' *Mol. Microbiol.*, **77**(6):1527–1538.

Schuster, M. and Greenberg, E. P. (2007). 'Early activation of quorum sensing in *Pseudomonas aeruginosa* reveals the architecture of a complex regulon.' *BMC Genomics*, **8**:287.

Schuster, M., Lostroh, C. P., Ogi, T., and Greenberg, E. P. (2003). 'Identification, timing, and signal specificity of *Pseudomonas aeruginosa* quorum-controlled genes: a transcriptome analysis.' *J. Bacteriol.*, **185**(7):2066–2079.

Schwartz, D. C., Li, X., Hernandez, L. I., Ramnarain, S. P., Huff, E. J., and Wang, Y. K. (1993). 'Ordered restriction maps of *Saccharomyces cerevisiae* chromosomes constructed by optical mapping.' *Science*, **262**(5130):110–114.

Seed, P. C., Passador, L., and Iglewski, B. H. (1995). 'Activation of the *Pseudomonas aeruginosa lasI* gene by LasR and the *Pseudomonas* autoinducer PAI: an autoinduction regulatory hierarchy.' *J. Bacteriol.*, **177**(3):654–659.

Shapiro, J. A. (1998). 'Thinking about bacterial populations as multicellular organisms.' *Annu. Rev. Microbiol.*, **52**:81–104.

Shaw, W. V. (1967). 'The enzymatic acetylation of chloramphenicol by extracts of R factor-resistant *Escherichia coli*.' *J. Biol. Chem.*, **242**(4):687–693.

Shepp, D. H., Tang, I. T., Ramundo, M. B., and Kaplan, M. K. (1994). 'Serious *Pseudomonas aeruginosa* infection in AIDS.' *J. Acquir. Immune Defic. Syndr.*, **7**(8):823–831.

Sinclair, M. I. and Holloway, B. W. (1982). 'A chromosomally located transposon in *Pseudomonas aeruginosa*.' *J. Bacteriol.*, **151**(2):569–579.

Singh, P. K., Schaefer, A. L., Parsek, M. R., Moninger, T. O., Welsh, M. J., and Greenberg, E. P. (2000). 'Quorum-sensing signals indicate that cystic fibrosis lungs are infected with bacterial biofilms.' *Nature*, **407**(6805):762–764.

Smith, D. M., Snow, D. E., Rees, E., Zischkau, A. M., Hanson, J. D., Wolcott, R. D., Sun, Y., White, J., Kumar, S., and Dowd, S. E. (2010). 'Evaluation of the bacterial diversity of pressure ulcers using bTEFAP pyrosequencing.' *BMC Med Genomics*, **3**:41.

Smith, E. E., Buckley, D. G., Wu, Z., Saenphimmachak, C., Hoffman, L. R., D'Argenio, D. A., Miller, S. I., Ramsey, B. W., Speert, D. P., Moskowitz, S. M., *et al.* (2006). 'Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients.' *Proc. Natl. Acad. Sci. U.S.A.*, **103**(22):8487–8492.

Snitkin, E. S., Zelazny, A. M., Thomas, P. J., Stock, F., Henderson, D. K., Palmore, T. N., and Segre, J. A. (2012). 'Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing.' *Sci Transl Med*, **4**(148):148ra116.

Snyder, L. A., Loman, N. J., Faraj, L. A., Levi, K., Weinstock, G., Boswell, T. C., Pallen, M. J., and Ala'Aldeen, D. A. (2013). 'Epidemiological investigation of *Pseudomonas aeruginosa* isolates from a six-year-long hospital outbreak using high-throughput whole genome sequencing.' *Euro Surveill.*, **18**(42).

Sobel, M. L., Neshat, S., and Poole, K. (2005). 'Mutations in PA2491 (*mexS*) promote MexT-dependent *mexEF-oprN* expression and multidrug resistance in a clinical strain of *Pseudomonas aeruginosa*.' *J. Bacteriol.*, **187**(4):1246–1253.

Spencer, D. H., Kas, A., Smith, E. E., Raymond, C. K., Sims, E. H., Hastings, M., Burns, J. L., Kaul, R., and Olson, M. V. (2003). 'Whole-genome sequence variation among multiple isolates of *Pseudomonas aeruginosa*.' *J. Bacteriol.*, **185**(4):1316–1325.

Steidle, A., Allesen-Holm, M., Riedel, K., Berg, G., Givskov, M., Molin, S., and Eberl, L. (2002). 'Identification and characterization of an *N*-acylhomoserine lactone-dependent quorum-sensing system in *Pseudomonas putida* strain IsoF.' *Appl. Environ. Microbiol.*, **68**(12):6371–6382.

Stevens, T. C., Ochoa, C. D., Morrow, K. A., Robson, M. J., Prasain, N., Zhou, C., Alvarez, D. F., Frank, D. W., Balczon, R., and Stevens, T. (2014). 'The *Pseudomonas aeruginosa* exoenzyme Y impairs endothelial cell proliferation and vascular repair following lung injury.' *Am. J. Physiol. Lung Cell Mol. Physiol.*, **306**(10):L915–924.

Stewart, L., Ford, A., Sangal, V., Jeukens, J., Boyle, B., Kukavica-Ibrulj, I., Caim, S., Crossman, L., Hoskisson, P. A., Levesque, R., *et al.* (2014). 'Draft genomes of 12 host-adapted and environmental isolates of *Pseudomonas aeruginosa* and their positions in the core genome phylogeny.' *Pathog Dis*, **71**(1):20–25.

Stewart, P. S. and Costerton, J. W. (2001). 'Antibiotic resistance of bacteria in biofilms.' *Lancet*, **358**(9276):135–138.

Stintzi, A., Evans, K., Meyer, J. M., and Poole, K. (1998). 'Quorum-sensing and siderophore biosynthesis in *Pseudomonas aeruginosa*: *lasR/lasI* mutants exhibit reduced pyoverdine biosynthesis.' *FEMS Microbiol. Lett.*, **166**(2):341–345.

Stokes, H. W., Elbourne, L. D., and Hall, R. M. (2007). 'Tn1403, a multiple-antibiotic resistance transposon made up of three distinct transposons.' *Antimicrob. Agents Chemother.*, **51**(5):1827–1829.

Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., Brinkman, F. S., Hufnagle, W. O., Kowalik, D. J., Lagrou, M., *et al.* (2000). 'Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen.' *Nature*, **406**(6799):959–964.

Sundin, C., Hallberg, B., and Forsberg, A. (2004). 'ADP-ribosylation by exoenzyme T of *Pseudomonas aeruginosa* induces an irreversible effect on the host cell cytoskeleton *in vivo*.' *FEMS Microbiol. Lett.*, **234**(1):87–91.

Sutherland, I. (2001). 'Biofilm exopolysaccharides: a strong and sticky framework.' *Microbiology (Reading, Engl.)*, **147**(Pt 1):3–9.

Tang, H. B., DiMango, E., Bryan, R., Gambello, M., Iglewski, B. H., Goldberg, J. B., and Prince, A. (1996). 'Contribution of specific *Pseudomonas aeruginosa* virulence factors to pathogenesis of pneumonia in a neonatal mouse model of infection.' *Infect. Immun.*, **64**(1):37–43.

Tingpej, P., Smith, L., Rose, B., Zhu, H., Conibear, T., Al Nassafi, K., Manos, J., Elkins, M., Bye, P., Willcox, M., *et al.* (2007). 'Phenotypic characterization of clonal and nonclonal *Pseudomonas aeruginosa* strains isolated from lungs of adults with cystic fibrosis.' *J. Clin. Microbiol.*, **45**(6):1697–1704.

Tipton, K. A., Coleman, J. P., and Pesci, E. C. (2015). 'Post-transcriptional regulation of gene PA5507 controls *Pseudomonas* quinolone signal concentration in P. aeruginosa.' *Mol. Microbiol.*, **96**(3):670–683.

Tobey, K. L. and Grant, G. A. (1986). 'The nucleotide sequence of the *serA* gene of *Escherichia coli* and the amino acid sequence of the encoded protein, D-3-phosphoglycerate dehydrogenase.' *J. Biol. Chem.*, **261**(26):12179–12183.

Toder, D. S., Gambello, M. J., and Iglewski, B. H. (1991). '*Pseudomonas aeruginosa* LasA: a second elastase under the transcriptional control of *lasR*.' *Mol. Microbiol.*, **5**(8):2003–2010.

Treangen, T. J., Ondov, B. D., Koren, S., and Phillippy, A. M. (2014). 'The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes.' *Genome Biol.*, **15**(11):524.

Tringe, S. G. and Rubin, E. M. (2005). 'Metagenomics: DNA sequencing of environmental samples.' *Nat. Rev. Genet.*, **6**(11):805–814.

Tsai, I. J., Zarowiecki, M., Holroyd, N., Garciarrubio, A., Sanchez-Flores, A., Brooks, K. L., Tracey, A., Bobes, R. J., Fragoso, G., Sciutto, E., *et al.* (2013). 'The genomes of four tapeworm species reveal adaptations to parasitism.' *Nature*, **496**(7443):57–63.

Tuttle, M. S., Mostow, E., Mukherjee, P., Hu, F. Z., Melton-Kreft, R., Ehrlich, G. D., Dowd, S. E., and Ghannoum, M. A. (2011). 'Characterization of bacterial communities in venous insufficiency wounds by use of conventional culture and molecular diagnostic methods.' *J. Clin. Microbiol.*, **49**(11):3812–3819.

Uwate, M., Ichise, Y. K., Shirai, A., Omasa, T., Nakae, T., and Maseda, H. (2013). 'Two routes of MexS-MexT-mediated regulation of MexEF-OprN and MexAB-OprM efflux pump expression in *Pseudomonas aeruginosa*.' *Microbiol. Immunol.*, **57**(4):263–272.

Valot, B., Guyeux, C., Rolland, J. Y., Mazouzi, K., Bertrand, X., and Hocquet, D. (2015). 'What It Takes to Be a *Pseudomonas aeruginosa*? The Core Genome of the Opportunistic Pathogen Updated.' *PLoS ONE*, **10**(5):e0126468.

Van Amersfoort, E. S., Van Berkel, T. J., and Kuiper, J. (2003). 'Receptors, mediators, and mechanisms involved in bacterial sepsis and septic shock.' *Clin. Microbiol. Rev.*, **16**(3):379–414.

Van Delden, C. and Iglewski, B. H. (1998). 'Cell-to-cell signaling and *Pseudomonas aeruginosa* infections.' *Emerging Infect. Dis.*, **4**(4):551–560.

Vogel, L., Jones, G., Triep, S., Koek, A., and Dijkshoorn, L. (1999). 'RAPD typing of *Klebsiella pneumoniae*, *Klebsiella oxytoca*, *Serratia marcescens* and *Pseudomonas aeruginosa* isolates using standardized reagents.' *Clin. Microbiol. Infect.*, **5**(5):270–276.

Wade, D. S., Calfee, M. W., Rocha, E. R., Ling, E. A., Engstrom, E., Coleman, J. P., and Pesci, E. C. (2005). 'Regulation of *Pseudomonas* quinolone signal synthesis in *Pseudomonas aeruginosa*.' *J. Bacteriol.*, **187**(13):4372–4380.

Wagner, V. E., Bushnell, D., Passador, L., Brooks, A. I., and Iglewski, B. H. (2003). 'Microarray analysis of *Pseudomonas aeruginosa* quorum-sensing regulons: effects of growth phase and environment.' *J. Bacteriol.*, **185**(7):2080–2095.

Waite, R. D. and Curtis, M. A. (2009). '*Pseudomonas aeruginosa* PAO1 pyocin production affects population dynamics within mixed-culture biofilms.' *J. Bacteriol.*, **191**(4):1349–1354.

Walker, J. and Moore, G. (2015). '*Pseudomonas aeruginosa* in hospital water systems: biofilms, guidelines, and practicalities.' *J. Hosp. Infect.*, **89**(4):324–327.

Wang, J. and Liu, J. H. (2004). 'Mutations in the chloramphenicol acetyltransferase (S61G, Y105C) increase accumulated amounts and resistance in *Pseudomonas aeruginosa*.' *FEMS Microbiol. Lett.*, **236**(2):197–204.

Wang, Y., Leung, H. C., Yiu, S. M., and Chin, F. Y. (2012). 'MetaCluster 5.0: a two-round binning approach for metagenomic data for low-abundance species in a noisy sample.' *Bioinformatics*, **28**(18):i356–i362.

Watters, C., DeLeon, K., Trivedi, U., Griswold, J. A., Lyte, M., Hampel, K. J., Wargo, M. J., and Rumbaugh, K. P. (2013). '*Pseudomonas aeruginosa* biofilms perturb wound resolution and antibiotic tolerance in diabetic mice.' *Med. Microbiol. Immunol.*, **202**(2):131–141.

Webb, J. S., Thompson, L. S., James, S., Charlton, T., Tolker-Nielsen, T., Koch, B., Givskov, M., and Kjelleberg, S. (2003). 'Cell death in *Pseudomonas aeruginosa* biofilm development.' *J. Bacteriol.*, **185**(15):4585–4592.

West, S. E., Sample, A. K., and Runyen-Janecky, L. J. (1994). 'The vfr gene product, required for *Pseudomonas aeruginosa* exotoxin A and protease production, belongs to the cyclic AMP receptor protein family.' *J. Bacteriol.*, **176**(24):7532–7542.

White, P. A., Stokes, H. W., Bunny, K. L., and Hall, R. M. (1999). 'Characterisation of a chloramphenicol acetyltransferase determinant found in the chromosome of *Pseudomonas aeruginosa*.' *FEMS Microbiol. Lett.*, **175**(1):27–35.

Whiteley, M., Lee, K. M., and Greenberg, E. P. (1999). 'Identification of genes controlled by quorum sensing in *Pseudomonas aeruginosa*.' *Proc. Natl. Acad. Sci. U.S.A.*, **96**(24):13904–13909.

Wiehlmann, L., Wagner, G., Cramer, N., Siebert, B., Gudowius, P., Morales, G., Kohler, T., van Delden, C., Weinel, C., Slickers, P., *et al.* (2007). 'Population structure of *Pseudomonas aeruginosa*.' *Proc. Natl. Acad. Sci. U.S.A.*, **104**(19):8101–8106.

Wilder, C. N., Allada, G., and Schuster, M. (2009). 'Instantaneous within-patient diversity of *Pseudomonas aeruginosa* quorum-sensing populations from cystic fibrosis lung infections.' *Infect. Immun.*, **77**(12):5631–5639.

Williams, D., Evans, B., Haldenby, S., Walshaw, M. J., Brockhurst, M. A., Winstanley, C., and Paterson, S. (2015). 'Divergent, coexisting *Pseudomonas aeruginosa* lineages in chronic cystic fibrosis lung infections.' *Am. J. Respir. Crit. Care Med.*, **191**(7):775–785.

Williams, P., Winzer, K., Chan, W. C., and Camara, M. (2007). 'Look who's talking: communication and quorum sensing in the bacterial world.' *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, **362**(1483):1119–1134.

Winson, M. K., Camara, M., Latifi, A., Foglino, M., Chhabra, S. R., Daykin, M., Bally, M., Chapon, V., Salmond, G. P., and Bycroft, B. W. (1995). 'Multiple *N*-acyl-L-homoserine lactone signal molecules regulate production of virulence determinants and secondary metabolites in *Pseudomonas aeruginosa*.' *Proc. Natl. Acad. Sci. U.S.A.*, **92**(20):9427–9431.

Winsor, G. L., Lam, D. K., Fleming, L., Lo, R., Whiteside, M. D., Yu, N. Y., Hancock, R. E., and Brinkman, F. S. (2011). 'Pseudomonas Genome Database: improved comparative analysis and population genomics capability for *Pseudomonas* genomes.' *Nucleic Acids Res.*, **39**(Database issue):596–600.

Winstanley, C., Langille, M. G., Fothergill, J. L., Kukavica-Ibrulj, I., Paradis-Bleau, C., Sanschagrin, F., Thomson, N. R., Winsor, G. L., Quail, M. A., Lennard, N., *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**(1):12–23.

Wise, J. (2012). 'Three babies die in **Pseudomonas** outbreak at Belfast neonatal unit.' *BMJ*, **344**:e592.

Wisseman, C. L., Smadel, J. E., Hahn, F. E., and Hopps, H. E. (1954). 'Mode of action of chloramphenicol. I. Action of chloramphenicol on assimilation of ammonia and on synthesis of proteins and nucleic acids in *Escherichia coli*.' *J. Bacteriol.*, **67**(6):662–673.

Wolcott, R. D., Gontcharova, V., Sun, Y., Zischakau, A., and Dowd, S. E. (2009). 'Bacterial diversity in surgical site infections: not just aerobic cocci any more.' *J Wound Care*, **18**(8):317–323.

Workentine, M. and Surette, M. G. (2011). 'Complex *Pseudomonas* population structure in cystic fibrosis airway infections.' *Am. J. Respir. Crit. Care Med.*, **183**(12):1581–1583.

Workentine, M. L., Sibley, C. D., Glezerson, B., Purighalla, S., Norgaard-Gron, J. C., Parkins, M. D., Rabin, H. R., and Surette, M. G. (2013). 'Phenotypic heterogeneity of *Pseudomonas aeruginosa* populations in a cystic fibrosis patient.' *PLoS ONE*, **8**(4):e60225.

Wozniak, R. A. and Waldor, M. K. (2010). 'Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow.' *Nat. Rev. Microbiol.*, **8**(8):552–563.

WTSI, P. (2015). 'Public Health England reference collections.' `https://www.sanger.ac.uk/resources/downloads/bacteria/nctc/#t_4`, Accessed 27 September 2015.

Xiao, G., He, J., and Rahme, L. G. (2006). 'Mutation analysis of the *Pseudomonas aeruginosa mvfR* and *pqsABCDE* gene promoters demonstrates complex quorum-sensing circuitry.' *Microbiology*, **152**(Pt 6):1679–1686.

Yan, Q., Wu, X. G., Wei, H. L., Wang, H. M., and Zhang, L. Q. (2009). 'Differential control of the PcoI/PcoR quorum-sensing system in *Pseudomonas fluorescens* 2P24 by sigma factor RpoS and the GacS/-GacA two-component regulatory system.' *Microbiol. Res.*, **164**(1):18–26.

Yano, K., Wada, T., Suzuki, S., Tagami, K., Matsumoto, T., Shiwa, Y., Ishige, T., Kawaguchi, Y., Masuda, K., Akanuma, G., *et al.* (2013). 'Multiple rRNA operons are essential for efficient cell growth and sporulation as well as outgrowth in *Bacillus subtilis*.' *Microbiology*, **159**(Pt 11):2225–2236.

Zerbino, D. R. and Birney, E. (2008). 'Velvet: algorithms for de novo short read assembly using de Bruijn graphs.' *Genome Res.*, **18**(5):821–829.

Zhang, Z. and Pierson, L. S. (2001). 'A second quorum-sensing system regulates cell surface properties but not phenazine antibiotic production in *Pseudomonas aureofaciens*.' *Appl. Environ. Microbiol.*, **67**(9):4305–4315.

Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). 'PHAST: a fast phage search tool.' *Nucleic Acids Res.*, **39**(Web Server issue):W347–352.

Zhu, H., Bandara, R., Conibear, T. C., Thuruthyil, S. J., Rice, S. A., Kjelleberg, S., Givskov, M., and Willcox, M. D. (2004). '*Pseudomonas aeruginosa* with *lasI* quorum-sensing deficiency during corneal infection.' *Invest. Ophthalmol. Vis. Sci.*, **45**(6):1897–1903.

Zulianello, L., Canard, C., Kohler, T., Caille, D., Lacroix, J. S., and Meda, P. (2006). 'Rhamnolipids are virulence factors that promote early infiltration of primary human airway epithelia by *Pseudomonas aeruginosa*.' *Infect. Immun.*, **74**(6):3134–3147.