Regulation of the PatAB efflux pump in *Streptococcus pneumoniae*

by

Alison Baylay

A thesis submitted to the University of Birmingham for the degree of DOCTOR OF PHILOSOPHY

Antimicrobials Research Group
School of Immunity and Infection
College of Medical and Dental Sciences
University of Birmingham

September 2013

Abstract

Streptococcus pneumoniae is the primary cause of community acquired pneumonia and represents a substantial disease burden worldwide. The PatAB ABC transporter is a multidrug efflux pump in this organism, encoded by the *patA* and *patB* genes, and over-expression of PatAB confers resistance to fluoroquinolone antibiotics.

This thesis investigates the clinical relevance and the regulation of PatAB expression. A strong association was found between cross-resistance to fluoroquinolones and dyes and over-expression of *patAB* in pneumococcal clinical isolates, supporting previous assertions that PatAB is the main fluoroquinolone efflux pump in S. pneumoniae. Two mechanisms that caused increased expression of *patA* and *patB* were identified by whole genome sequencing of multidrug resistant mutants of *S. pneumoniae* R6. Firstly, a novel duplication of a nine kilobase genomic region, including *patA* and *patB*, was identified in one mutant. Further investigation suggested that duplication caused high expression of the second copy of *patAB* by bringing these genes under the control of a tRNA gene promoter. Secondly, mutations were identified in the terminator stem-loop of a predicted transcriptional attenuator upstream of *patA* in three other mutants. Transcriptional fusion of mutated attenuator sequences with a GFP reporter gene suggested that these mutations increased transcription from the *patA* promoter.

To Andy

Acknowledgements

I am extremely grateful to my supervisor Professor Laura Piddock for giving me this opportunity, for always believing in me, and pushing me to be more confident in myself and my abilities. I would also like to thank all members of the ARG, old and new, for making the group such a lovely place to work. Particular thanks go to Dr Mark Webber, Dr Jessica Blair and Dr Vito Ricci for their help and advice and for answering all my silly questions so patiently! Thanks also go to Helen, Amelia, Grace, Steffi and Jen for their help and support during various crises and for generally making the lab a fun place to be. Also, thanks (again) to Jess and Helen for their excellent proof reading skills.

I would also like to thank Professor Tim Mitchell for useful feedback and advice on working with pneumococci, and Dr Andrea Mitchell for giving me access to equipment. A big thank you to Al Ivens and Nick Loman for giving me advice on handling genome sequencing data.

I am also very grateful to my wonderful friends Lia, Jess and Maria for generally being amazing, and for always being there no matter how busy we all are and how long it's been.

I would like to say a massive thank you to my parents, and my brother Chris, who have always encouraged me to do my best and have supported me in everything I've done. I am also extremely grateful for the last minute proof-reading!

And finally a special thank you to Andy for his continual love, support and seemingly endless patience, and for keeping me going in the last few months with inspiring pep talks and wonderful dinners. I couldn't have done this without you...

Clarification of Contribution to Collaborative Work

The sections detailed below were the result of collaborative work. The contribution of both parties to each section of work is clarified here.

Chapter 3:

Experiments and data collection, apart from measurement of grwoth rates of isolates, were carried out by Mark Garvey and Kim Long Ryan Wong (under the supervision of Mark Garvey). Alison Baylay carried out all data analysis, measured growth rates of isolates and generated all figures.

Genome sequencing: sections 4.3 and 4.7.1

Genomic DNA for sequencing was extracted by Alison Baylay. The GenePool (Edinburgh) prepared Illumina sequencing libraries, carried out Illumina sequencing, performed image analysis to generate sequence reads, and performed preliminary filtering and quality control. All subsequent analysis of Illumina sequencing data (alignment against a reference, de novo assembly, etc) was carried out by Alison Baylay.

Figure 5.5

Under the supervision of Alison Baylay, undergraduate student Adna Farah (University of Birmingham) transformed R6 cells with PCR amplimers of varying sizes from *patAB* over-expressing mutants to determine the genomic region containing mutations causing *patAB* upregulation. All figures in the present thesis were generated by Alison Baylay.

Contents

List of Figures

List of Tables

List of Abbreviations

ABC	ATP-binding cassette
BSA	Bovine Serum Albumin
cfu	colony forming units
CSF	Cerebrospinal fluid
GFP	Green Fluorescent Protein
kb	kilobase
MATE	Multidrug and toxic compound extrusion
MFS	Major Facilitator Superfamily
MIC	Minimum inhibitory concentration
NBD	Nucleotide Binding Domain
OD	Optical Density
PBS	Phosphate buffered saline
PCV	Pneumococcal Conjugate Vaccine
PPP	Prokaryotic Promoter Prediction
QRDR	quinolone resistance determining region
qRT-PCR	quantitative real-time PCR
RND	Resistance Nodulation Division
SDW	Sterile Distilled Water
SMR	Small Multidrug Resistance
SNP	Single nucleotide polymorphism
spc	spectinomycin
TBE	Tris borate EDTA buffer
TE	Tris EDTA buffer

1 Introduction

1.1 *Streptococcus pneumoniae*

Streptococcus pneumoniae, also known as pneumococcus, is a Gram positive, coccoid bacterium that colonises the human nasopharynx. Primarily, it exists as a harmless commensal organism, carried asymptomatically by varying percentages of the population (Bogaert *et al.*, 2004), but becomes an important human pathogen when it evades host defensive mechanisms and spreads to normally sterile areas of the body (Musher, 2004). It is the most common bacterial cause of community-acquired pneumonia (Butler, 2004) and bacterial meningitis (Mook-Kanamori *et al.*, 2011).

1.1.1 Biology, phylogeny, culture and identification of the pneumococcus

The streptococci belong to the *Lactobacillales* order of the phylum Firmicutes, which are Gram positive bacteria with a low-GC content (Slonczewski and Foster, 2009). Members of the *Lactobacillales* are facultative anaerobes, and derive energy for growth by fermentation of glucose to lactic acid. The order contains *Enterococcus*, *Lactobacillus*, *Lactococcus* and *Streptococcus* (Slonczewski and Foster, 2009). Streptococci are coccoid, and divide along a single axis, so appear as characteristic Gram positive diplococcoid pairs or chains when Gram stained. All streptococci are non-motile and non-

spore-forming. They are do not produce catalase, allowing them to be distinguished from catalase-positive bacteria such as staphylococci by a lack of reaction when exposed to hydrogen peroxide. This also means that streptococci must be cultured on a medium containing a source of catalase, such as blood, to neutralise hydrogen peroxide produced as a by-product of aerobic metabolism.

Streptococci can be broadly separated into groups based on haemolytic activity when grown on blood agar. Pyogenic streptococci, including the important human pathogens *S. agalactiae* and *S. pyogenes*, are β-haemolytic (Facklam, 2002). When β-haemolytic species are grown on blood agar a clear halo forms around the colony due to complete lysis of red blood cells (Slonczewski and Foster, 2009). The majority of the remaining streptococci are α-haemolytic, which is the incomplete breakdown of haemoglobin to methaemoglobin, resulting in formation of a green halo around a colony growing on blood agar (Slonczewski and Foster, 2009). The α-haemolytic streptococci have historically been difficult to classify due to minimal phenotypic differences between species, and many were classified into a large polyphyletic group known as the viridans streptococci (Kawamura *et al.*, 1995). Later analysis of 16S rRNA sequences showed that the viridans group could be split into five clusters: mitis, bovis, anginosus, sanguis and parasanguis (Kawamura *et al.*, 1995). *S. pneumoniae* is α-haemolytic, but was classified separately from the viridans streptococci due to its unique phenotypic characteristics of sensitivity to the antibiotic optochin and solubility in bile (Facklam, 2002). However, 16S rRNA sequencing showed that *S. pneumoniae* clusters within the mitis group of the viridans streptococci, and is very closely related to *S. mitis* and *S. oralis* (Facklam, 2002; Kawamura *et al.*, 1995).

Most pneumococcal strains are aerotolerant, although there are some strains that require totally anaerobic conditions for growth (Slonczewski and Foster, 2009). Pneumococci produce H_2O_2 as a byproduct of metabolism, which is not neutralised as streptococci are catalase-negative. Production of H_2O_2 is oxygen-dependent, meaning that a CO_2-

enriched atmosphere is required for optimal growth even for aerotolerant pneumococci, as well as a source of catalase such as blood.

Pneumococci are Gram positive, which means their cell wall consists of a single plasma membrane, surrounded by a thick layer of peptidoglycan. *S. pneumoniae* incorporates choline into its cell wall (Brundish and Baddiley, 1968; Tomasz, 1967), which is a trait shared by several other bacteria found in the nasopharynx. Uniquely, pneumococci have an auxotrophic requirement for choline in the growth medium (Rane and Subbarow, 1940). Incorporation of choline has been suggested to be important for maintaining cell shape (Horne and Tomasz, 1993). More importantly, pneumococci express proteins that specifically bind to phosphorylchloline moieties as a method of attachment to the cell surface, some of which are involved in pathogenesis (Swiatlo *et al.*, 2004). Pneumococci undergo autolysis in stationary phase (Lewis, 2000). Autolysis is a programmed cell death response mediated by the major autolysin LytA. LytA is a murein hydrolase, which breaks down peptidoglycan causing cell lysis (Mosser and Tomasz, 1970).

S. pneumoniae produces a polysaccharide capsule, which is attached to the cell wall. Over 90 different capsule types have been observed to date, which differ in the sugar moieties that make up the polysaccharide subunits (Bentley *et al.*, 2006). Some consist of only one or two sugars, while others have complex subunit structures (Bentley *et al.*, 2006). The presence of a capsule is essential for invasive pneumococcal disease, although unencapsulated isolates have been observed in nasopharyngeal carriage studies (Marsh *et al.*, 2010; Salter *et al.*, 2012). The type of capsule expressed by a strain determines its ability to cause disease (Briles *et al.*, 1992; Brueggemann *et al.*, 2004; Sjöström *et al.*, 2006). Expression of the pneumococcal capsule is tightly regulated, and is increased or decreased during different stages of infection. Low capsular expression promotes adherence to epithelial tissues (Talbot *et al.*, 1996), while increased production of capsule protects the cell from opsonisation and therefore complement-mediated clearance by the innate immune system (Abeyta *et al.*, 2003).

1.2 Pneumococcal transformation

Streptococcus pneumoniae is naturally transformable, meaning that it is able to take up DNA from its external environment by entering a competent state. Naturally competent cells can take up approximately 10^6 times more DNA than can be introduced by artificial methods such as electroporation or ionic shock (Lacks, 2004). *S. pneumoniae* is able to take up the equivalent of 10% of its cellular DNA content by natural transformation (Hotchkiss, 1954).

1.2.1 Competence

In cultured cells, transformation takes place only at certain cell densities or stage of growth, showing that competence is under regulatory control (Hotchkiss, 1954). In *S. pneumoniae*, and other closely related streptococcal species, competence is regulated by a two-stage system.

The first is a quorum sensing system that recognises when a streptococcal population has reached a sufficient cell density for release and uptake of DNA to be advantageous (Figure 1.1; Novak *et al.*, 1999). This is controlled by two operons, *comAB* and *comCDE*. ComC encodes a peptide precursor that is processed to a 17 amino acid competence-stimulating peptide (CSP) (Håvarstein *et al.*, 1995). CSP is exported from the cell by the ComAB ABC transporter (Håvarstein *et al.*, 1995; Hui and Morrison, 1991). Extracellular CSP is sensed by a two-component system formed from ComD and ComE (Håvarstein *et al.*, 1996; Pestova *et al.*, 1996). Phosphorylated ComE up-regulates expression of both the *comAB* and *comCDE* operons, thereby promoting increased production, excretion and detection of CSP.

ComE also up-regulates the expression of an alternative sigma factor, ComX, which

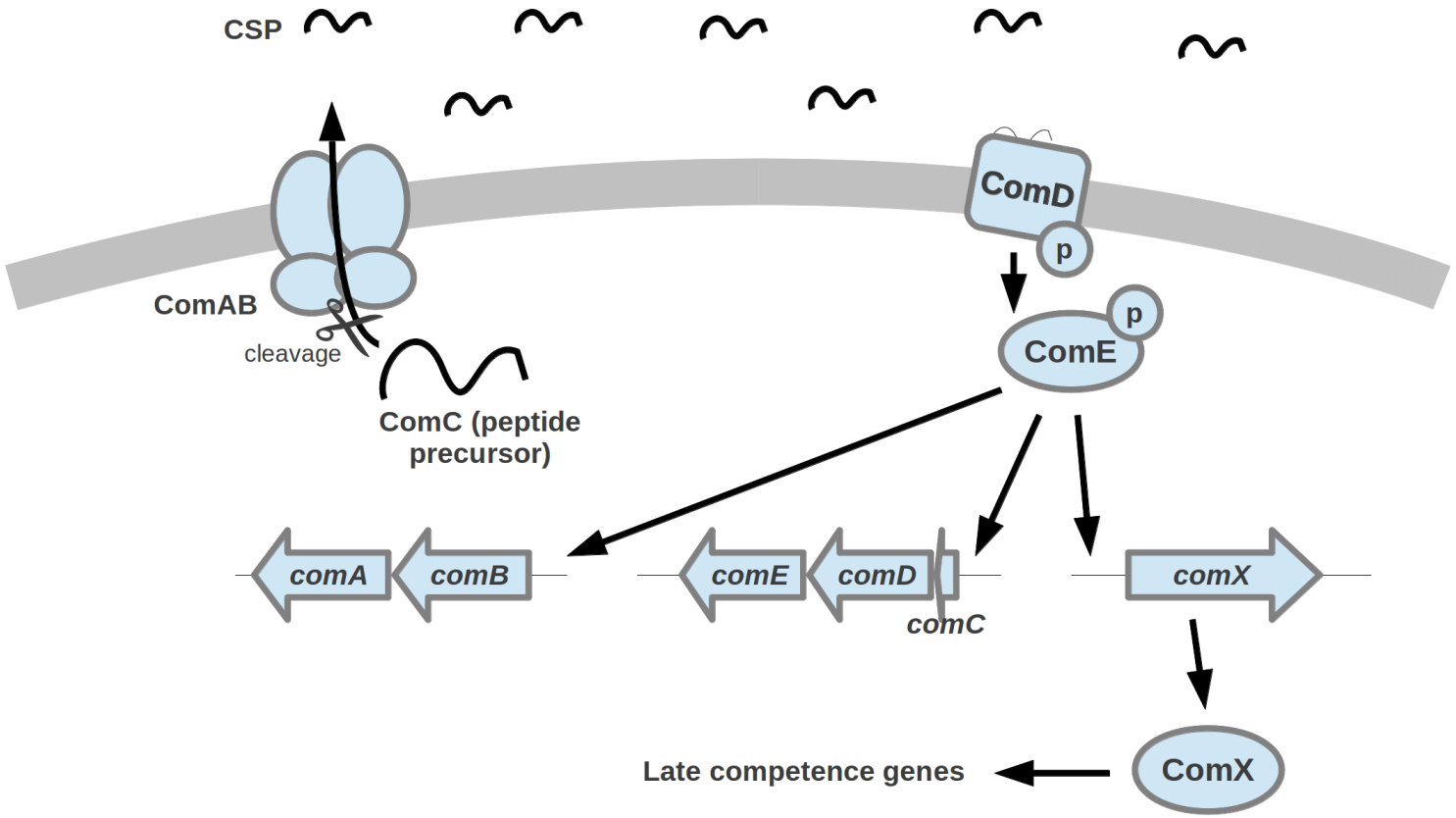


Figure 1.1. The quorum sensing system controlling development of the competent state in *S. pneumoniae* (adapted from Johnsborg and Håvarstein, 2009)

controls the second stage of competence activation (Lee and Morrison, 1999). ComX promotes transcription from promoters containing a combox sequence, TACGAATA, which is found upstream of about 60 genes (Campbell *et al.*, 1998). Over 180 genes have been identified that have altered expression during competence development, but only 23 of these are directly involved in DNA uptake and recombination (Johnsborg and Håvarstein, 2009). Other competence regulated genes include stress response genes such as chaperonins, heat shock proteins, membrane transporters, the major autolysin *lytA* and several genes of unknown function (Bartilson *et al.*, 2001; Peterson *et al.*, 2004; Rimini *et al.*, 2000). The diversity of function of competence-induced genes has led to the suggestion that competence is a pneumococcal stress response analogous to the SOS response found in many other bacteria (Claverys *et al.*, 2006). Some authors have suggested that the competent state should be referred to as X-state to reflect the fact that it is involved in more than just DNA uptake (Claverys *et al.*, 2006).

Release of DNA into the extracellular medium occurs during competence (Steinmoen *et al.*, 2002). It was originally thought that a subpopulation of cells underwent spontaneous autolysis during competence as an evolutionary strategy to promote genetic diversity within the species. However, it has since been shown that pneumococci in the competent state actively lyse non-competent pneumococci colonising the same niche (Guiral *et al.*, 2005). This lysis is a form of predation that has been termed fratricide, sobrinicide or allolysis (Johnsborg and Håvarstein, 2009). Fratricide is mediated by a two-peptide bacteriocin, CibAB, produced during competence (Guiral *et al.*, 2005).

1.2.2 DNA uptake and recombination

Exogenous DNA is internalised by pneumococci using a dedicated uptake system, components of which are upregulated during competence. Double-stranded DNA binds to a pseudopilus on the pneumococcal surface, then one strand of the bound DNA is cleaved

into fragments and taken up into the cell, while the other is degraded (Claverys *et al.*, 2009). In the cytoplasm, single-stranded DNA binding proteins associate with the DNA to form a nucleoprotein filament (Morrison *et al.*, 2007). The internalised DNA then recombines with the pneumococcal chromosome at regions of similar sequence in a process mediated by the RecA recombinase (Bergé *et al.*, 2003; Mortier-Barrière *et al.*, 1998).

During recombination, the Hex mismatch repair system functions to resolve polymorphisms between the new DNA and the pneumococcal chromosome (Claverys *et al.*, 1983). The Hex system, encoded by *hexA* and *hexB*, recognises and corrects purine-purine or pyrimidine-pyrimidine transition mutations (AT to GC) with greater efficiency than transversion mutations (AT to CG, AT to TA, or CG to GC), which means that transversion mutations are transformed with a higher efficiency than transitions (Claverys *et al.*, 1983). However, the Hex system becomes saturated as the number of mismatches between the donor and recipient DNA increases, meaning that mismatch repair plays less of a role in acquisition of DNA from more divergent sources (Humbert *et al.*, 1995).

The sizes of internalised and recombined DNA fragments have been estimated by several methods. In the 1960s, recombination sizes were estimated to be about two kb by measuring genetic linkage between selectable markers (Lacks, 1966). Other studies suggested sizes of three to six kb using donor DNA labelled with deuterium and nitrogen-15 (Fox and Allen, 1964; Gurney Jr. and Fox, 1968). Recent studies identified putative recombination events by comparing the genomes of closely related pneumococcal isolates. Examination of recombinations in the phylogeny of one particular pneumococcal lineage estimated a mean recombination size of 6.3 kb (Croucher *et al.*, 2011a), while comparison of co-colonising strains from a paediatric infection identified recombination events of between six and 28 kb (Hiller *et al.*, 2011).

Croucher *et al* (2012) investigated the recombination process at high resolution using whole genome sequencing of 124 pneumococcal isolates generated through *in vitro*

transformation experiments. In this study, the capsule biosynthesis operon in a serotype 23F PMEN1 strain was replaced with a kanamycin cassette by transformation with genomic DNA from a TIGR4 derivative. Analysis of recombinations at this selected locus by defining regions of donor and recipient alleles showed that multiple non-contiguous recombinations occurred from the same donor DNA molecule. The average combined length of these non-contiguous recombinations was 5.9 kb, which is close to the average length of DNA fragments imported during competence, estimated at 6.6 kb by Morrison and Guild (1972). Croucher and co-workers also identified numerous secondary, unselected recombinations distributed randomly in the genome sequences of the 124 transformants. The sizes of these secondary recombination events were typically less than 5kb and followed an exponentially decaying distribution with an average length of 2.3 kb. The frequency of secondary recombinations increased with DNA concentration, implying that recombinations can occur simultaneously in the pneumococcal cell from different DNA molecules.

1.2.3 Evolutionary implications of transformation

Natural transformation is advantageous for *S. pneumoniae* for three main reasons.

Firstly, natural transformation allows maintenance of genetic diversity among pneumococcal species. Analysis of homologous gene clusters from 17 pneumococcal genomes by Hiller et al (2007) showed that the total set of genes existing among pneumococcal species, all of which can be accessed by individual *S. pneumoniae* strains by natural transformation, is much larger than the genome of any individual bacterium. This is known as the distributed genome hypothesis. Of 3170 homologous gene clusters found in the 17 pneumococcal genomes studied by Hiller and co-workers, only 46% were found in all 17 genomes, thereby forming the core pneumococcal genome. The rest of the gene clusters were spread among one or more genomes, with each individual genome

consisting of 21-32% non-core genes.

Secondly, natural transformation allows the maintenance of diversity in certain key traits among pneumococcal populations by allelic substitution. The most clinically relevant example of this is the switching of capsular types. The genes encoding capsule biosynthetic enzymes are encoded by the same genetic locus in all pneumococcal strains (Bentley *et al.*, 2006), meaning that a recipient cell can change its capsule type by uptake of donor DNA from a strain expressing a different serotype. This allows *S. pneumoniae* strains to evade the host adaptive immune response directed against a particular capsular polysaccharide. This has become particularly relevant since the introduction of the PCV7 vaccine. Evidence of capsular switching has been observed in some pneumococcal lineages associated with disease or antibiotic resistance allowing them to evade the vaccine (Croucher *et al.*, 2011a; Jefferies and Smith, 2004).

Finally, the human nasopharynx is host to a heterogenous population of bacterial species, which can include more than one strain of *S. pneumoniae*, closely related species of oral *Streptococcus*, and more diverse organisms such as *Haemophilus influenzae* (Bogaert *et al.*, 2004). Natural transformation allows the spread of advantageous gene variants throughout the population of the nasopharynx. For example, mosaic genes formed by repeated gene transfers between closely related *Streptococcus* species have been shown to be involved in pneumococcal resistance to penicillin (Dowson *et al.*, 1993, discussed further in section 1.7.1).

1.3 Pneumococcal carriage, disease and treatment

1.3.1 Colonisation

Colonisation is the key step in pneumococcal infection, as development pneumococcal disease requires the establishment of the carrier state (Bogaert *et al.*, 2004). A colonisation event consists of three stages: acquisition, carriage and termination of carriage (Crook *et al.*, 2004). During acquisition, pneumococci enter the nasopharynx via respiratory droplets and become established by adhesion to mucosal epithelia (Musher, 2003). Carriage then persists for a period of weeks or months, before the organism is finally cleared by the immune system (Ekdahl *et al.*, 1997; Högberg *et al.*, 2007). Duration of carriage varies according to the serotype of the colonising isolate (Högberg *et al.*, 2007).

1.3.1.1 Epidemiology

Rates of pneumococcal carriage vary widely according to age, geographical area and socioeconomic and genetic background, but are highest in children (Bogaert *et al.*, 2004). Estimation of true carriage rates is difficult due to variation in methods of nasopharyngeal swabbing between studies (Crook *et al.*, 2004). For example, recorded rates of carriage measured in studies of healthy children by transnasal swabbing reviewed by Bogaert *et al* (2004) ranged from 8% to 70%. However, it is clear that pneumococcal carriage in children is highly dependent on age. Several long-term studies that have measured colonisation of children over time have shown that pneumococcal colonisation increases markedly during the first year of life, then declines steadily from about four years of age due to maturation of the immune system (Aniansson *et al.*, 1992; Bogaert *et al.*, 2003; Daw *et al.*, 1997; Leino *et al.*, 2001; Mühlemann *et al.*, 2003).

Risk factors for pneumococcal colonisation in healthy children include number of older

siblings, socioeconomic status and ethnicity (Bogaert *et al.*, 2004). Time spent in institutions with other children has also been identified as an important risk factor. Several studies have associated time spent in day care centres with increased colonisation (Bogaert *et al.*, 2001; Ciftçi *et al.*, 2001; Dunais *et al.*, 2011, 2003; Rodrigues *et al.*, 2009; Sá-Leão *et al.*, 2008). One particular study found a colonisation rate of 82% among infants in a French orphanage (Raymond *et al.*, 2000). Colonisation with antibiotic resistant strains of *S. pneumoniae* has been linked to recent antibiotic use (Dagan *et al.*, 1998).

The rate of carriage among adults is typically less than 10%, and duration of carriage of a particular serotype is reduced (Bogaert *et al.*, 2004; Högberg *et al.*, 2007). However, adult carriage rates increase in crowded conditions, such as hospitals, long-term care and prisons, and in adults taking care of small children (Crook *et al.*, 2004; de Galan *et al.*, 1999; Hendley *et al.*, 1975; Hoge *et al.*, 1994; Millar *et al.*, 1994).

1.3.1.2 Mechanisms of colonisation

To successfully colonise the nasopharynx pneumococci need to avoid clearance by the host immune systems or entrapment in mucus, and then adhere to respiratory epithelia. In addition, pneumococci must be able to withstand interspecies competition from other bacteria that colonise the same niche. They possess numerous characteristic features that allow them to do this.

The capsule is essential for colonisation and has been described as the most important virulence factor in *S. pneumoniae* (Bogaert *et al.*, 2004; Magee and Yother, 2001). Expression of capsule in the early stages of colonisation prevents recognition of underlying cell wall components by antibodies or complement components, which prevents phagocytosis (Abeyta *et al.*, 2003). The overall negative charge of capsular polysaccharides also reduces entrapment of the pneumococcus in nasal mucus (Nelson *et al.*, 2007).

However, capsule is downregulated in later stages of colonisation as it inhibits adhesion to epithelial cells (Magee and Yother, 2001; Weiser *et al.*, 1994).

Adhesion of pneumococci to mucosal epithelia is mediated by several cell surface components. Firstly, the pneumococcal cell wall contains phosphorylcholine, which binds to the platelet activating factor receptor (Cundell and Tuomanen, 1994; Fischer *et al.*, 1993). Two pneumococcal cell surface proteins, CbpA and PsaA, are also involved in pneumococcal adhesion. CbpA is part of a family of choline binding proteins which attach to the pneumococcal cell wall by non-covalent interactions with phosphorylcholine (Swiatlo *et al.*, 2004). CbpA promotes adhesion by directly binding to the polymeric immunoglobulin receptor found on mucosal surfaces (Hammerschmidt *et al.*, 1997; Zhang *et al.*, 2000). PsaA is a pneumococcal surface lipoprotein that binds to N-acetyl-glucosamine, which is a cell-surface carbohydrate found on non-inflamed epithelium (Berry and Paton, 1996; Bogaert *et al.*, 2004; Marra *et al.*, 2002). Finally, the LPXTG-anchored cell surface protein neuraminidase A (NanA) cleaves host cell surface oligosaccharides, exposing further N-acetyl-glucosamine for binding (Cámara *et al.*, 1994; Tong *et al.*, 2000).

NanA is also involved in interspecies competition. *Haemophilus influenzae* protects itself from the host immune system by decorating its glycoproteins with terminal sialic acid residues so that they resemble host polysaccharides (Harvey *et al.*, 2001). Removal of these residues by pneumococcal NanA reverses this molecular mimicry, causing the host immune system to attack *H. influenzae* (Shakhnovich *et al.*, 2002).

1.3.2 Pneumococcal disease

In healthy individuals, pneumococcal colonisation is contained within the nasopharynx by the active clearance of cells from the lower respiratory tract and the action of the immune system (Butler, 2004). However, under certain circumstances, these clearance mechanisms fail, for example if the respiratory tract is damaged by trauma or viral in-

fection, or if the individual is immunocompromised (Butler, 2004). Pneumococci are then able to spread passively to normally sterile areas of the respiratory tract and cause mucosal infections. *S. pneumoniae* is the most common bacterial cause of community acquired pneumonia, which occurs when bacteria spread to the alveoli. Pneumococci proliferate in the alveolar spaces, causing an inflammatory response. Fluid and white blood cells then accumulate in the alveoli, causing coughing, fatigue and shortness of breath (Musher, 2004). This inflammatory response is mediated by the pneumococcal toxin, pneumolysin, which lyses eukaryotic cells by forming pores in the membrane (Morgan *et al.*, 1995). *S. pneumoniae* may also spread to the middle ear, causing acute otitis media, which is a common disease in children. *S. pneumoniae* is the most common bacterial agent isolated in otitis media cases (Dowell *et al.*, 1999).

Pneumococci can also cross epithelia and cause invasive diseases. Invasive diseases are rare but are associated with high mortality. Bacteraemia is caused by entry of pneumococci to the circulatory system. Invasion can occur from the lungs as a complication of pneumonia (secondary bacteraemia) or directly from the nasopharynx resulting in a primary bacteraemia with no obvious focus of infection (known as occult bacteraemia) (Musher, 2004). Pneumococci can also cross the blood-brain barrier and cause meningitis. *S. pneumoniae* is the most common cause of bacterial meningitis in adults, since the introduction of vaccines against *H. influenzae* (Quagliarello and Scheld, 1997).

Pneumococcal infections are frequently associated with the extremes of age. Children under the age of five are at risk of infection due to their high levels of nasopharyngeal colonisation, while the elderly are susceptible due to low immune function (Dockrell *et al.*, 2012; Mook-Kanamori *et al.*, 2011).

Incidence of pneumococcal disease follows a seasonal pattern that peaks in the winter months and correlates with the frequency of influenza virus infections (Kim *et al.*, 1996). Synergy has been demonstrated between influenza infection and pneumococcal

pneumonia, and many deaths during influenza epidemics have been retrospectively ascribed to secondary infection with *S. pneumoniae* (Chien *et al.*, 2009; Geddes, 2009). The influenza virus produces a neuraminidase protein similar in structure to the protein produced by pneumococci (Peltola and McCullers, 2004). Cleavage of sialic acid residues from host cell surface polysaccharides by the viral enzyme exposes receptors to which pneumococci can adhere, which promotes development of pneumonia and accounts for the synergy between the two infections (McCullers and Bartmess, 2003).

1.3.3 Prevention and treatment of pneumococcal infections

1.3.3.1 Vaccines

Pneumococcal disease can be reduced by immunisation of children under five years old with polysaccharide conjugate vaccines (PCV). These vaccines consist of pneumococcal capsular polysaccharides conjugated to a non-toxic version of the diptheria toxin (Weil-Olivier *et al.*, 2012). PCVs cannot cover all 90 pneumococcal serotypes, so are targeted to the serotypes most commonly associated with invasive pneumococcal disease and antibiotic resistance (Hausdorff *et al.*, 2000; Tyrrell *et al.*, 2009). PCV7 was introduced in the US in 2000, and was adopted into national immunisation programs in Europe between 2006 and 2008 (Pelton *et al.*, 2004; Weil-Olivier *et al.*, 2012). This vaccine covers serotypes 4, 6B, 9V, 14, 18C, 19F and 23F (Hausdorff *et al.*, 2000). In general, since the introduction of PCV7, incidence of community acquired pneumonia and invasive pneumococcal disease has decreased, and the proportions of isolates with PCV7 serotypes has declined among invasive pneumococcal disease isolates (Gladstone *et al.*, 2011; Weil-Olivier *et al.*, 2012). However, there has been an increase in relative prevalence of isolates with serotypes not included in PCV7, particularly serotypes 1, 7F and 19A (SáâĂŘLeão and Nunes, 2009; Tyrrell *et al.*, 2009). To counteract this problem, PCV13, an extended version of the vaccine which contains six additional serotypes (1, 3, 5, 6A, 7F and 19A),

has been developed and replaced PCV7 in the US and UK in 2010 (Jefferies *et al.*, 2011).

1.3.3.2 Antibiotic treatment

A variety of bacterial agents can cause pneumonia and invasive infections, and diagnostic tests to rapidly identify the causative organism are not yet available. As a result, these diseases are usually treated empirically with broad spectrum antibiotics (Jones *et al.*, 2010).

The recommended antibiotic treatment for community acquired pneumonia depends on the severity of the disease, the previous health of the patient and whether they are treated in hospital or as an outpatient. The joint guidelines from the Infectious Diseases Society of America and the American Thoracic Society (Mandell *et al.*, 2007) state that outpatients who were previously healthy and have no risk factors for carrying antibiotic resistant *S. pneumoniae* should be treated with a macrolide antibiotic, such as azithromycin, clindamycin or erythromycin. Outpatients with co-morbidities, such as influenza infection, chronic obstructive pulmonary disease, diabetes or renal failure, should be treated with a respiratory fluoroquinolone such as moxifloxacin, gemifloxacin or levofloxacin, or with a combination of high dose amoxicillin (a β-lactam antibiotic) and a macrolide. This is also recommended for patients who have received antibiotic therapy within the previous three months, which is a risk factor for carriage of antibiotic resistant *S. pneumoniae*. Hospitalised patients are treated with either a respiratory fluoroquinolone or a combination of a β-lactam (cefotaxime, ceftriaxone or ampicillin-sulbactam) and a macrolide.

1.4 Antibiotic Resistance

1.4.1 Mechanisms of antibiotic resistance.

Antibiotics inhibit growth of bacteria by specifically binding to bacterial cellular targets and inhibiting their function. Antibiotics block essential cellular pathways such as protein synthesis or cell wall biosynthesis, which may kill the bacterium (bacteriocidal antibiotic) or prevent further growth (bacteriostatic antibiotic). Bacteria have evolved a multitude of mechanisms to circumvent this molecular inhibition, and these can be roughly categorised into four general mechanisms of resistance.

Antibiotic inactivation. Bacteria may produce enzymes that either degrade the antibiotic molecule or modify it by addition of chemical groups such that it can no longer bind to its cellular target. The primary mechanism of resistance to β-lactam antibiotics in many bacteria is production of a β-lactamase enzyme that hydrolyses the vital β-lactam ring structure (Ambler, 1980; Bush *et al.*, 1995). These enzymes are usually secreted extracellularly so can destroy antibiotic molecules before they reach the cell. A large family of enzymes that modify aminoglycoside antibiotics by addition of acetyl, phosphate or nucleotidyl groups has been characterised (Ramirez and Tolmasky, 2010). Other families of group transfer enzymes that confer antibiotic resistance include the macrolide kinases and the chloramphenicol acetyltransferases (Matsuoka *et al.*, 1998; Nakamura *et al.*, 2000; Noguchi *et al.*, 1995; Schwarz *et al.*, 2004).

Alteration or protection of antibiotic target. The molecular target of an antibiotic may be modified, either by mutation or post-translationally, such that the antibiotic can no longer bind but the function of the target is maintained. For example, resistance to fluoroquinolone antibiotics is typically conferred by mutations in the topoisomerase

genes that prevent antibiotic binding (Ruiz, 2003). Resistance to ribosome-targeting antibiotics such as aminoglycosides and macrolides is often conferred by acquisition of genes encoding methylase enzymes, which methylate key ribosomal nucleotides and block antibiotic binding (Weisblum, 1995a). There are also examples of dedicated proteins that protect intracellular targets from antibiotic binding. Tet(M) and Tet(O) proteins are analogues of the ribosomal elongation factors EF-G and EF-Tu, which cause resistance to tetracyline by dislodging tetracyline molecules from the ribosome (Thaker *et al.*, 2010). Similarly, Qnr proteins have been discovered that protect topoisomerase enzymes from inhibition by quinolones (Martínez-Martínez *et al.*, 1998; Tran and Jacoby, 2002).

Metabolic bypass. Bacteria may also circumvent the antibiotic inhibition of a particular metabolic pathway by expressing a modified enzyme that allows the pathway to function by an alternative route. The most clinically relevant example of this is the acquisition of a methicillin-insensitive penicillin binding protein, PBP2x, by methicillin-resistant *Staphylococcus aureus* which allows cell wall synthesis to continue when the native PBP enzyme is inhibited (Fuda *et al.*, 2004; Pinho *et al.*, 2001). Another example is resistance to sulfonamides and trimethoprim, which inhibit the dihydropteroate synthase and dihydrofolate reductase enzymes of the folate synthesis pathway, respectively. Acquisition of genes encoding antibiotic insensitive versions of these enzymes allows continuation of folate synthesis in the presence of antibiotic (Huovinen, 2001).

Reduced intracellular accumulation of antibiotic. Finally, decreased susceptibility to a variety of antibiotics may be conferred by reducing the amount of antibiotic that accumulates within the cell. This means that a higher external concentration of antibiotic is required to achieve the intracellular concentration required for inhibition, which may not be feasible clinically. Reduced intracellular accumulation of antibiotic can be achieved in two ways. Antibiotic molecules may be actively exported from the cell by

efflux pumps, and this mechanism is discussed in detail in section 1.5. Alternatively, the permeability of the cell may be reduced, preventing the entry of antibiotic molecules. Intracellular accumulation of antibiotic is typically lower in Gram negative bacteria than in Gram positive bacteria due to the presence of the outer membrane which acts as a permeability barrier (Piddock, 2006a). The permeability of the outer membrane can be futher reduced by altered expression of porin proteins (Livermore, 2001; Nikaido, 2003).

1.4.2 Evolution of antibiotic resistance

1.4.2.1 Intrinsic resistance

Intrinsic or innate resistance is defined as a relatively high tolerance to an antibiotic due to a feature that is shared by all members of a particular bacterial species. For example, *Pseudomonas aeruginosa* is naturally resistant to several antibiotics due to its low permeability. This is mediated by production of highly selective porin proteins and expression of several multidrug efflux pumps (Livermore, 2001; Ochs *et al.*, 1999; Sugawara *et al.*, 2006). Intrinsic resistance to some β-lactams is found in several species of *Enterobacter* and *Citrobacter* due to their expression of a chromosomally-encoded cephalosporinase enzyme, AmpC (Kaye *et al.*, 2001; Lindberg *et al.*, 1987; Woodford *et al.*, 2007). Similarly, some species of *Enterococcus* possess an operon that encodes vancomycin resistance genes (Leclercq *et al.*, 1992; Navarro and Courvalin, 1994). Intrinsic antibiotic resistance is evolutionarily ancient and is common in soil-dwelling bacteria, which regularly come into contact with antibiotic-producing organisms such as fungi (Dantas *et al.*, 2008; D'Costa *et al.*, 2006).

1.4.2.2 Acquired resistance

Acquired resistance refers to antibiotic resistance emerging in bacterial species that do not possess innate resistance to the antibacterial agent being used. Resistance may be acquired by spontaneous mutation, for example in the antibiotic binding site of the target protein, or by horizontal transfer of a resistance determinant from a resistant strain.

Mutational resistance. Development of antibiotic resistance by mutation has been observed for every class of antibiotics in current clinical use (Woodford and Ellington, 2007), although it is not always the predominant mechanism of resistance. Mutational resistance is particularly important in bacterial species that do not commonly participate in horizontal gene transfer, such as *Mycobacterium tuberculosis* (Gillespie, 2002), and for resistance to synthetic antibiotics, such as fluroquinolones, for which there is not a reservoir of ancient resistance genes in the environment (Ruiz, 2003).

Horizontal gene transfer Bacteria can acquire resistance genes that are disseminated on mobile genetic elements such as plasmids, transposons and integrons (Carattoli, 2001; Salyers *et al.*, 1995), or from DNA acquired from other members of the same species by natural transformation (Johnsborg *et al.*, 2007). Transfer of resistance genes by bacteriophage transduction also has a clinically relevant role in antibiotic resistance development in bacteria (Colomer-Lluch *et al.*, 2011; Modi *et al.*, 2013; Parsley *et al.*, 2010). A wide variety of resistance determinants are associated with mobile genetic elements. The most clinically relevant example is the widespread dissemination of extended spectrum β-lactamases and carbapenemases among the *Enterobacteriaceae* which are mobilised on conjugative plasmids (Hossain *et al.*, 2004; Nordmann *et al.*, 2011; Yong *et al.*, 2009).

1.4.2.3 Role of gene amplification in bacterial adaptation, including antibiotic resistance

Gene duplication and amplification is the transient increase in copy number of a chromosomal region. The role of gene duplication events in eukaryotic evolution has been extensively studied, but the role of gene amplification in bacterial adaptation has been somewhat overlooked. It has been suggested that gene duplication and amplification in bacteria has a role in adaptation to stressful conditions, including antibiotic exposure (Andersson and Hughes, 2009).

In bacteria, duplications of chromosomal regions can occur spontaneously at a high frequency compared to the emergence of point mutations. Analysis of transient duplications of 38 metabolic genes carried out by Anderson and Roth (1981) in unselected cultures of *Salmonella* Typhimurium suggested that duplications occured spontaneously at frequencies of between 10^{-2} and 10^{-4} per cell, depending on the chromosomal site investigated. The authors estimated that, in an unselected bacterial culture, approximately 10% of cells will possess a duplication somewhere in their genome. Duplications therefore arise with higher frequency than point mutations, suggesting that gene duplication, not point mutation, provides the majority of the initial genetic diversity required for bacterial populations to survive external stresses (Andersson and Hughes, 2009).

Gene duplications can be formed by both RecA-dependent and RecA-independent mechanisms (Andersson and Hughes, 2009; Whoriskey *et al.*, 1987). RecA-dependent mechanisms involve recombination between repeat regions of at least 40 nucleotides (Shen and Huang, 1986). In particular, recombination between copies of the ribosomal RNA genes has been observed as a common mechanism of duplication formation (Anderson and Roth, 1981; Lehner and Hill, 1980). Other repeat sequences that can mediate duplication formation by RecA-dependent recombination are IS elements and REP sequences (Haack and Roth, 1995; Jessop and Clugston, 1985; Lin *et al.*, 1984). RecA-independent

duplications do not require long repeat regions at the duplication boundaries, and examples have been observed where there was no sequence homology at the duplication join point at all (Reams and Neidle, 2004; Seoane *et al.*, 2003). Suggested mechanisms for RecA-independent duplication formation are strand slippage during rolling-circle replication, and illegitimate end joining by DNA gyrase (Ikeda *et al.*, 2004; Lovett *et al.*, 1993; Seoane *et al.*, 2003; Trinh and Sinden, 1993).

Once formed, duplications can be amplified to form tandem arrays of direct repeats (Edlund *et al.*, 1979; Jessop and Clugston, 1985; Reams and Neidle, 2004; Tlsty *et al.*, 1984; Whoriskey *et al.*, 1987). Two mechanisms have been proposed to mediate this. The first is RecA-dependent homologous recombination between copies of the repeats, and the second is rolling circle replication (Andersson and Hughes, 2009). The number of copies of a chromosomal region that can be generated by amplification varies widely and is limited by the effects of increasing gene dosage and the energetic requirements of replicating the extra DNA. Gene amplications of up to 40 copies have been reported (Sandegren and Andersson, 2009).

A few examples of antibiotic resistance mediated by chromosomal gene amplification have been reported in the past three decades. The first observation was amplification of a β-lactamase, *ampC*, in *E. coli* following *in vitro* selection for increased ampicillin resistance (Normark *et al.*, 1977). A 9-18 kb region of the genome, including the *ampC* gene, was amplified between 2 and 40 times, mediated by short, 12-13bp direct repeat sequences. Increased β-lactam resistance was observed in a strain of *Yersinia enterocolitica* following growth in ampicillin, due to five-fold amplification of a 28 kb region containing the *blaA* gene (Seoane *et al.*, 2003). Growth of a methicillin-resistant *S. aureus* in a high concentration of methicillin led to further increases in resistance, due to two- to three-fold amplification of the *SCCmec* element (Matthews and Stewart, 1988). Transient resistance to multiple antibiotic agents was observed in *E. coli* strains with decreased genomic stability caused by mutations in the *lon* protease, due to very

large (149-300 kb), unstable chromosomal duplications including the genes encoding the AcrAB-TolC efflux pump (Nicoloff *et al.*, 2006, 2007). The only report to date of a large (>1 kb) gene amplification associated with resistance in a clinical isolate was that of a clinical isolate of *Streptococcus agalactiae* in which a four-fold amplification of the *fol* operon was observed (Brochet *et al.*, 2008). This caused an increase in resistance to sulfonamide antibiotics and trimethoprim, due to overproduction of enzymes involved in the folate synthesis pathway.

Gene duplications and amplifications are usually transient and are readily lost in the absence of selective pressure due to recombination between copies of the repeats (Andersson and Hughes, 2009). This instability may explain why the involvement of gene amplification in antibiotic resistance in clinical isolates is not well characterised.

1.5 Efflux as a mechanism of multidrug resistance

Efflux pumps are membrane-spanning transporter proteins that actively export compounds from the cell. Efflux pumps are evolutionarily ancient and in many cases their physiological role in the cell is unclear. One function is thought to be to protect the bacterium from toxic substances found in the natural environment (Piddock, 2006b). For example, some enteric pathogens express efflux pumps that confer resistance to bile salts, which improves survival of the bacterium in the gut (Lin *et al.*, 2003). However, efflux pumps can also transport antibiotic compounds and resistance occurs when pumps are over-expressed (Poole, 2005).

Some efflux pumps are specific for particular antibiotics, while others have much wider substrate specificities and are known as multidrug efflux pumps (Lewis, 1994). Multiple specificity means that over-expression of an efflux pump selected with one agent could indirectly lead to resistance to multiple clinically relevant antibiotics.

Multidrug efflux pumps can be classified as primary or secondary transporters based on the mechanism used to power substrate efflux. The ATP-Binding Cassette transporters are primary transporters as they couple transport of substrates with hydrolysis of ATP (discussed further in section 1.5.2). Secondary transporters couple transport of substrates to the proton motive force or sodium ion gradients. The balance between the use of primary and secondary transporters varies depending on the metabolism of the bacterial species. Fermentative bacteria tend to rely to a greater extent on primary transporters whereas aerobic bacteria contain more secondary transporters (Paulsen *et al.*, 1998, 2000).

1.5.1 Secondary transporters

Secondary transporters can be divided into four superfamilies based on structure (Piddock, 2006a). These are the major facilitator superfamily (MFS), the resistance-nodulation-division (RND) family, the multidrug and toxic compound extrusion (MATE) family and the small multidrug resistance (SMR) family.

MFS efflux pumps are the largest family of secondary transporters involved in multidrug efflux (Pao *et al.*, 1998). MFS transporters are monomeric transporters with 12 or 14 membrane-spanning domains, which power substrate efflux using the proton motive force (Paulsen *et al.*, 1996). Bmr from *Bacillus subtilis* (Neyfakh *et al.*, 1991), and NorA from *Staphylococcus aureus* (Neyfakh *et al.*, 1993) are examples of 12 transmembrane domain transporters. The QacA and LmrS pumps from *S. aureus*, which export cationic lipophilic drugs and linezolid, respectively, are examples of 14 transmembrane domain MFS transporters (Floyd *et al.*, 2010; Littlejohn *et al.*, 1992).

RND efflux pumps are exclusively found in Gram negative bacteria (Poole, 2005). They are tripartite efflux pumps consisting of an inner membrane transporter protein, and a porin located in the outer membrane, which are joined by a periplasmic adapter protein

(Piddock, 2006b). RND pumps export substrates by coupling transport to the proton motive force by functional rotation of the three subunits of the inner membrane pump (Piddock, 2006a). The best characterised system is the AcrAB-TolC efflux pump of *E. coli* and *Salmonella* species, which confers resistance to various structurally diverse antibiotics, such as β-lactams, fluoroquinolones and aminoglycosides (Piddock, 2006a).

MATE efflux pumps are the most recently identified group of secondary transporters involved in multidrug resistance. The first example to be characterised was NorM from *Vibrio parahaemolyticus*, which was identified in 1998 (Morita *et al.*, 1998). There are 20 characterised members of the family, but genomic analyses suggest that they are widespread and found in bacteria, archaea and eukaryotes (Kuroda and Tsuchiya, 2009). MATE transporters consist of a single transporter protein with 12 transmembrane domains (Brown *et al.*, 1999). Unlike MFS transporters, many MATE transporters use sodium ion gradients to power substrate transport instead of the proton motive force (Kuroda and Tsuchiya, 2009). Examples include MepA in *S. aureus*, which confers decreased susceptibility to tigecycline (McAleese *et al.*, 2005), and NorM in *Neisseria* species, which increases resistance to cationic antimicrobial compounds (Rouquette-Loughlin *et al.*, 2003).

The SMR family consists of small, 100-140 amino acid proteins with four membrane spanning domains, which form functional transporters by dimerisation in the membrane. The first protein of this family to be identified was the Smr transporter of *S. aureus*, which is carried on plasmids and confers resistance to various compounds, including organic cations and dyes (Grinius *et al.*, 1992; Littlejohn *et al.*, 1992). There are now more than 250 members of the family, most of which confer resistance to cationic compounts such as quarternary ammonium compounds and ethidium bromide (Bay *et al.*, 2008). Smr from *S. aureus* and EmrE from *E. coli* transport clinically relevant antibiotics such as ampicillin, erythromycin and tetracycline (Bay *et al.*, 2008; Yerushalmi *et al.*, 1995).

1.5.2 ABC transporters

The ABC transporter superfamily is the largest known paralogous protein family and examples are found in all three kingdoms of life (Bouige *et al.*, 2002; Higgins, 2001). The superfamily can be divided into three groups based on function: uptake transporters, efflux transporters, and ATPases not involved in membrane transport (Dassa and Bouige, 2001).

1.5.2.1 Structure and function

ABC transporters consist of two transmembrane domains, each formed from six transmembrane helices, and two nucleotide binding domains (NBDs) (Higgins *et al.*, 1986). The way these domains are coded in the genome differs between subfamilies. For bacterial uptake transporters, the transmembrane and nucleotide binding domains are encoded by separate genes, and the transporter is a tetramer of four separate proteins (Higgins, 2001). In eukaryotes the four domains of the ABC transporter are usually fused and are encoded by a single gene (Lubelski *et al.*, 2007). ABC transporters involved in substrate efflux in bacteria are often formed from homo- or heterodimerisation of two half-transporters, each consisting of one nucleotide binding domain fused to one transmembrane domain (Lubelski *et al.*, 2007).

The nucleotide binding domains of ABC transporters are highly conserved, while the transmembrane domains are more variable allowing transport of a variety of substrates (Cuthbertson *et al.*, 2010). There are several important conserved motifs in the nucleotide binding domain. The C-loop is a highly conserved motif (LSGGQ) which is required to classification of a protein as a member of the ABC superfamily (Seeger and van Veen, 2009). The Walker A and B motifs are found in all P-loop ATPase enzymes, and form the ATP-binding site. To be fuctionally active, a dimer of NBDs is required

where the C-loop of one monomer complements the ATP binding site of the other (Dawson and Locher, 2006; Hopfner *et al.*, 2000). Other conserved sequences involved in ATP binding and hydrolysis have also been identified, such as a conserved histidine residue (H-loop) and a conserved glutamine residue (Q-loop) (Lubelski *et al.*, 2007; Seeger and van Veen, 2009).

Crystal structures of the MsbA lipid A transporter from *E. coli* and the Sav1866 multidrug transporter from *S. aureus* have been used to elucidate the mechanism of substrate transport by ABC efflux pumps (Dawson and Locher, 2006; Dawson *et al.*, 2007). Nucleotide binding and hydrolysis confers conformational changes in the transmembrane domains, switching them from a cytoplasmic-facing conformation to an outward-facing conformation, which transports bound substrates from the cytoplasm to the outside of the cell (Chang, 2003).

1.5.2.2 Role in multidrug resistance

The first multidrug ABC transporter to be described was the LmrA transporter found in *Lactococcus lactis* in 1996 (van Veen *et al.*, 1996). *Lactococcus lactis* is an important bacterium used in the dairy industry, but has also proved to be a useful model organism for studying ABC transporter-mediated multidrug resistance. Heterologous expression of LmrA in *E. coli* conferred decreased susceptibility to a variety of structurally diverse compounds, including aminoglycosides, macrolides, quinolones, tetracyclines, chloramphenicol, ethidium bromide and rhodamine 6G (van Veen *et al.*, 1996). This transporter may also be involved in protection against salt stress and exposure to cadmium, sodium laureate and ethanol (Achard-Joris *et al.*, 2005; Bourdineaud *et al.*, 2004).

LmrA homologues have been found in several other industrially important bacteria. *Oenococcus oeni* is involved in wine production and expresses the LmrA homologue OmrA, which confers resistance to ethanol, sodium laureate and other wine compounds

(Bourdineaud *et al.*, 2004). *Lactobacillus brevis* causes beer spoilage and produces an LmrA homologue HorA, which confers resistance to hop compounds (Sakamoto *et al.*, 2001). Both of these transporters confer a classic multidrug resistance phenotype when heterologously expressed (Bourdineaud *et al.*, 2004; Sakamoto *et al.*, 2001).

Sav1866 is a homodimeric ABC transporter from *S. aureus* that is homologous to LmrA. The role of this transporter in multidrug resistance in *S. aureus* has not been confirmed, but crystal structures of this transporter have been instrumental in determining the molecular mechanism of substrate transport in bacterial ABC multidrug efflux transporters (Dawson and Locher, 2006; Dawson *et al.*, 2007).

The characterisation of the role of LmrA in multidrug resistance was mostly carried out by heterologous expression in *E. coli*. Subsequent studies have suggested that multidrug efflux in *Lactococcus lactis* itself is mainly mediated by another ABC transporter LmrCD, with LmrA playing a relatively minor role (Lubelski *et al.*, 2006). LmrCD is a heterodimeric transporter formed from two half-transporters encoded by *lmrC* and *lmrD*, which are transcriptionally linked in a small operon regulated by LmrR (Agustiandari *et al.*, 2008). Over-expression of LmrCD confers decreased susceptibility to several toxic compounds including ethidium, daunomycin, rhodamine 6G, Hoescht 33342 and cholate (Lubelski *et al.*, 2004), while inactivation of *lmrC* or *lmrD* confers hypersusceptibility to these agents (Lubelski *et al.*, 2006).

Homologues of LmrCD have been identified in some pathogenic Gram positive bacteria. *Enterococcus faecalis* has a high intrinsic resistance to many antibiotics and is an opportunistic, nosocomial human pathogen, causing diseases such as endocarditis and urinary tract infections (Lee *et al.*, 2003). The LmrCD homologue EfrAB has been linked to ciprofloxacin and doxycycline resistance in *E. faecalis* (Lee *et al.*, 2003). The PatAB transporter in *S. pneumoniae*, which is the focus of this study and is discussed further in section 1.7.2, is also a LmrCD homologue, as is the related SatAB transporter in *S.*

suis (section 1.7.2.5).

A few ABC transporters have been implicated in antibiotic resistance in Gram negative bacteria. The MacAB transporter in *E. coli* confers decreased susceptibility to some classes of macrolide antibiotics. Over-expression of *macAB* in an *E. coli* mutant lacking the major RND transporter AcrAB resulted in increased resistance to macrolides with 14 and 15-membered ring structures, such as erythromycin and azithromycin, but not to macrolides with 16-membered rings such as leucomycin (Kobayashi *et al.*, 2001). Unlike LmrCD, MacAB is not a heterodimeric ABC transporter formed from two different half transporters. Instead, MacB forms a homodimeric transporter, while MacA is a periplasmic adapter protein which stimulates ATPase activity of MacB and provides a link between MacB and the outer membrane porin TolC (Lin *et al.*, 2009). A similar system has been characterised in *Neisseria gonorrhoeae* (Rouquette-Loughlin *et al.*, 2005). In addition, a classic homodimeric ABC transporter VcaM from *Vibrio cholerae* has been shown to transport tetracycline, ciprofloxacin and norfloxacin, which are drugs used to treat cholera, as well as doxorubicin and daunomycin (Huda *et al.*, 2003).

1.5.3 Role of multidrug efflux pumps beyond resistance

Many multidrug efflux pumps have been shown to have physiological roles beyond their contribution to antibiotic resistance, including pathogenicity, stress responses and biofilm formation (Nishino *et al.*, 2009). In some cases, physiological roles of efflux pumps can be directly related to their multidrug transport function. For example, the AcrAB-TolC RND efflux pump, and its homologue CmeAB in *Campylobacter jejuni*, export bile salts (Lin *et al.*, 2003; Thanassi *et al.*). This increases the resistance of the bacteria to toxic compounds in bile, allowing them to better colonise the gut. Strains of *Salmonella* lacking AcrB or TolC are attenuated in their ability to colonise chickens (Buckley *et al.*, 2006) Requirement of TolC for bile resistance has also been shown in *V. cholerae* (Bina

and Mekalanos, 2001).

A role in virulence of *Salmonella* has also been observed for MacAB, which was shown to be the multidrug efflux pump with the greatest contribution towards ability to cause disease in a mouse model of infection (Nishino *et al.*, 2006). MacAB is regulated by PhoPQ, a two-component regulatory system that controls expression of several genes linked to invasion of epithelial cells and which has previously been linked to bile resistance (Behlau and Miller, 1993; Nishino *et al.*, 2006; van Velkinburgh and Gunn, 1999).

In other cases, the link between observed physiological roles of efflux pumps and substrate transport is less clear. Lack of a functional AcrAB-TolC efflux pump in *Salmonella* Typhimurium substantially decreases the ability of the bacterium to invade human macrophages and intestinal cells *in vitro* (Buckley *et al.*, 2006). Inactivation of components of the AcrAB-TolC efflux pump has also been shown to alter expression of genes encoded in the *Salmonella* pathogenicity island 1, and of operons involved in chemotaxis and motility (Webber *et al.*, 2009). Increased expression of efflux pumps, including AcrAB-TolC, has also been observed in bacteria residing inside macrophages, suggesting that efflux pumps are important for intracellular survival (Nishino *et al.*, 2009). *P. aeruginosa* lacking components of the MexAB-OprM RND efflux pump also show reduced invasion. Attentuation of invasion in mutants lacking MexAB-OprM was alleviated by exogenous addition of sterilised media previously used to grow wild type cultures, which suggests that MexAB-OprM exports an unknown factor required for virulence (Hirakata *et al.*, 2002).

Multidrug efflux pumps have also been linked to bacterial responses to other cellular stresses. In *P. aeruginosa*, the MexXY-OprM RND efflux pump is upregulated in response to oxidative stress, and the MexCD-OprJ system is upregulated by membrane-damaging agents (Fraud *et al.*, 2008). AcrAB-TolC is over-expressed in *E. coli* in response to ethanol exposure and high salt concentration (Ma *et al.*, 1995).

The regulatory pathways controlling the expression of AcrAB-TolC and MexAB-OprM have been well studied (Fraud and Poole, 2011; Morita *et al.*, 2006; Nishino *et al.*, 2009). These regulatory mechanisms are complex and involve both local and global regulators, which is consistent with the fundamental roles of these pumps in the response of the organism to various environmental conditions. Expression of *acrAB* is controlled by the local repressor AcrR, but is also regulated by the global regulators MarR, MarA, SoxS and Rob in *E. coli*, and additionally by RamR and RamA in *Salmonella*. These global regulators also control expression of many other genes involved in response to changing environmental conditions, bacterial virulence and biofilm formation. Via this global regulatory network, expression of efflux pumps can be induced by exposure to a diverse array of compounds (Lawler *et al.*, 2013; Nikaido *et al.*, 2008; Pomposiello *et al.*, 2001; Semchyshyn *et al.*, 2005).

1.5.4 Efflux pump inhibitors

Due to the role of efflux pumps in bacterial resistance to various antimicrobial agents, and the lack of development of new antibiotics in recent decades, there has been much interest in developing efflux pump inhibitors. Specific inhibition of efflux pumps would act synergistically with existing antibiotics and help circumvent resistance. Efflux pump inhibitors are defined as small molecules that specifically inhibit efflux transporters, either competitively or non-competitively, and this distinguishes them from compounds that inhibit efflux non-specifically by dissipating the proton motive force, such as CCCP (Lomovskaya and Watkins, 2001).

The plant alkaloid reserpine, originally extracted from the roots of *Rauwolfia vomitoria* (Poisson *et al.*, 1954), was used as an antihypertensive drug due to its ability to inhibit the vesicular monoamine transporter that controls the transport of biogenic amines into storage organelles in mammalian cells (Ahmed *et al.*, 1993). Reserpine also inhibits

bacterial efflux pumps such as Bmr in *B. subtilis*, which are similar in structure to the monoamine transporter (Ahmed *et al.*, 1993). Reserpine cannot be used therapeutically as an efflux inhibitor as the concentrations required for this are neurotoxic, but it remains a useful research tool for investigating efflux-mediated resistance (Ahmed *et al.*, 1993).

Many compounds that inhibit bacterial efflux activity have been identified from a variety of different sources. Some are modified versions of drugs known to have activity against transporters in human cells, for example, P-glycoprotein inhibitors used as anticancer agents (Leitner *et al.*, 2011) and antipsychotic drugs such as phenothiazines (Kaatz *et al.*, 2003). Other efflux pump inhibitors have been identified in plant extracts and from other natural sources (Stavri *et al.*, 2007). However, none of these potential efflux inhibitors is yet licensed for clinical use.

1.6 Antibiotic resistance in *S. pneumoniae*

Pneumococcal infections are usually treated with antibiotics from the β-lactam, macrolide and fluoroquinolone classes, either alone or in combination (section 1.3.3.2). Resistance to these agents has increased in recent decades, although there is considerable variation in reports from different geographical areas. Incidence of pneumococcal disease caused by antibiotic resistance has recently stabilised, and in some regions decreased, with the introduction of the PCV7 vaccine, as the serotypes most commonly associated with antibiotic resistance were included in the vaccine. However, the prevalence of non-vaccine serotypes has since increased, and antibiotic resistance is increasing within this group, particularly in serotype 19A isolates (Croucher *et al.*, 2011b; Kyaw *et al.*, 2006). This trend may be reversed by introduction of the PCV13 vaccine, which includes 19A among other serotypes, but surveillance data detailing the effects of this new vaccine on antibiotic resistance is not yet available.

1.6.1 β-lactams and macrolides

The main mechanism of resistance to β-lactams in *S. pneumoniae* is through the acquisition of penicillin-binding proteins (PBPs) which have a reduced affinity for β-lactam antibiotics (Dowson *et al.*, 1993; Laible and Hakenbeck, 1987; Laible *et al.*, 1991). This is thought to have arisen through recombination of *pbp* genes with those from other species of *Streptococcus*, in particular *S. mitis* (Dowson *et al.*, 1993), resulting in the formation of mosaic genes. The first case of penicillin resistance in *S. pneumoniae* was reported in 1965 (Kislak *et al.*, 1965), about 20 years after the introduction of β-lactam antibiotics, but the prevalence of resistant strains remained low throughout the 1960s with less than 1% of isolates showing resistance (Swartz, 2002).

Rates of penicillin resistance are difficult to monitor as clinical breakpoints differ between disease presentation and method of antibiotic administration. However, data from global surveillance studies suggests that the proportion of isolates globally that were not susceptible to penicillin rose from 10.4% in 1996 (Felmingham and Grüneberg, 2000) to 22.3% in 2002 (Felmingham *et al.*, 2002). Following this, global data suggest that penicillin resistance remained stable at 21.8% between 2002 and 2005 (Schito and Felmingham, 2005).

Global increases in penicillin resistance have been caused primarily due to the worldwide spread of highly successful international clones which carry penicillin resistance determinants (McGee *et al.*, 2001). This has occurred due to acquisition of recombinant *pbp* genes by strains that are also highly transmissible. The first example to be discovered was the Spain^{23F}-1 clone (designated PMEN1 by the Pneumococcal Molecular epidemiology Network), which in 1998 accounted for nearly 40% of all penicillin resistant isolates in the USA (Corso *et al.*, 1998; McGee *et al.*, 2001). Incidence of disease caused by this clone has diminished due to inclusion of serotype 23F in PCV7, but a shift in the population towards serotype 19A capsular switch variants of the lineage has been

observed in the USA since about 2005 (Croucher *et al.*, 2011b). A more recent example is that of the Spain^{9V}-3 clonal cluster (ST156, PMEN3) which represented about 50% of penicillin non-susceptible isolates in Sweden in 2003 (Sjöström *et al.*, 2007) and 48% in Poland in 2005 (Sadowy *et al.*, 2010). This clonal type possesses the *rlrA* genetic islet which encodes a pilus (Hava and Camilli, 2002; Sjöström *et al.*, 2007). Piliated isolates were shown to outcompete non-piliated isolates, suggesting that piliation promotes nasopharyngeal colonisation (Sjöström *et al.*, 2007). It was shown that, in Sweden, penicillin non-susceptibility was over-represented in piliated isolates compared to the general population, and the combination of the two features was suggested to result in penicillin-nonsusceptible strains particularly capable of global dissemination (Sjöström *et al.*, 2007).

Macrolide antibiotics bind to the 50S subunit of the bacterial ribosome, blocking the elongation step of protein synthesis (Weisblum, 1995b). In *S. pneumoniae*, resistance can be conferred by two mechanisms. The first is methylation of 23S rRNA by a methylase encoded by the *erm* gene, which prevents macrolide binding (Weisblum, 1995a). The second mechanism of pneumococcal macrolide resistance is over-expression of the Mef efflux pumps, which are 44 kDa MFS transporter proteins (Clancy *et al.*, 1996). Two variants have been observed in *S. pneumoniae*: MefA and MefE (Gay *et al.*, 2000; Santagati *et al.*, 2000).

Both the *erm* and *mef* genes are located on mobile genetic elements. The *ermB* gene is found on the conjugative transposon Tn*1545*, which may also carry other resistance determinants such as *tetM* (Trieu-Cuot *et al.*, 1990). The *mefA* gene has been found on the defective transposon Tn*1207.1* (Santagati *et al.*, 2000), while *mefE* is carried by the macrolide efflux genetic assembly (mega) element (Gay and Stephens, 2001),

Macrolide resistance spreads by a combination of clonal expansion and horizontal gene transfer. For example, carriage of macrolide resistance elements was associated with

the internationally successful PMEN1 lineage, but acquisition and switching of these elements has occured multiple times throughout the phylogeny (Croucher *et al.*, 2011b).

Global prevalance of macrolide resistance has increased from 16.5% in 1996 (Felmingham and Grüneberg, 2000) to 31.8% in 2002 (Felmingham *et al.*, 2002). From 2002 to 2005 macrolide resistance continued to rise to 36.3% despite the level of penicillin resistance remaining stable (Schito and Felmingham, 2005).

Worryingly, high levels of multidrug resistance in *S. pneumoniae* isolates have been observed worldwide. Multidrug resistance, defined as resistance to two or more of the antimicrobials penicillin, second-generation cephalosporins, macrolides, tetracyclines, and co-trimoxazole, was observed in 38.6% of isolates in the PROTEKT study (Schito and Felmingham, 2005). A particularly high level of co-resistance was observed between penicillin and macrolides, with 73.8% of penicillin resistant isolates also showing resistance to azithromycin (Schito and Felmingham, 2005).

1.6.2 Fluoroquinolone resistance

The quinolones are a class of synthetic antibiotics with a broad spectrum of activity. Naladixic acid was the first quinolone to be used clinically (Lesher *et al.*, 1962). The second generation of quinolones, the fluoroquinolones, were developed by fluoridation of a particular carbon atom in the quinolone structure, which improved the antimicrobial and pharmacodynamic properties of the drug (Ball, 2000). Norfloxacin was introduced into clinical practice in 1986, followed by ciprofloxacin in 1987 (Paton and Reeves, 1988). Third and fourth generation quinolones have been developed by further modification of fluoroquinolone structures to improve potency and spectrum of activity. Third generation quinolones include levofloxacin, gatifloxacin and moxifloxacin (Barrett, 2000; Perry *et al.*, 1999), and gemifloxacin is a clinically important fourth generation quinolone (Blondeau and Missaghi, 2004).

Quinolone antibiotics inhibit the action of the type II topoisomerases, topoisomerase IV and DNA gyrase. DNA gyrase is a tetrameric enzyme formed from two GyrA and two GyrB subunits, which catalyses formation of negative supercoils in DNA, aiding DNA replication (Gellert *et al.*, 1976). Topoisomerase IV is also tetrameric and is formed from two ParC and two ParE subunits, and its main cellular role is to decatenate replicated chromosomes (Adams *et al.*, 1992; Kato *et al.*, 1990). Both enzymes create and religate double-stranded breaks in the DNA (Drlica and Zhao, 1997). Quinolone antibiotics bind to the enzyme DNA complex and prevent religation of double stranded DNA breaks, which leads to cell death (Drlica *et al.*, 2008).

In *S. pneumoniae*, as in other bacteria, fluoroquinolone resistance is mediated by mutations in the quinolone resistance determining regions (QRDRs) of the topoisomerase IV subunit ParC and the DNA gyrase subunit GyrA, which prevent binding of fluoroquinolone antibiotics (Janoir *et al.*, 1996; Pan *et al.*, 1996).

Initial studies on development of fluoroquinolone resistance in *S. pneumoniae* showed that high level ciprofloxacin resistance was caused by mutations in the QRDR regions of both ParC and GyrA (Schmitz *et al.*, 2002). The mutations most commonly associated with this phenotype were changes of serine 79 of ParC and serine 81 of GyrA to tyrosine or phenylalanine respectively (de la Campa *et al.*, 2004; Janoir *et al.*, 1996). Mutation of ParC alone caused a modest increase in ciprofloxacin resistance (Janoir *et al.*, 1996). In *in vitro* selection experiments, mutations in *parC* emerged before mutations in *gyrA*, which supported the proposal that ParC is the primary target of ciprofloxacin in Gram positive bacteria (Janoir *et al.*, 1996; Muñoz and De La Campa, 1996; Pan *et al.*, 1996). This situation contrasts with that in Gram negative bacteria where GyrA is the primary fluoroquinolone target (Yoshida *et al.*, 1990).

This picture has since become more complicated with the introduction of new fluoroquinolone antibiotics and the characterisation of more fluoroquinolone resistant clinical

isolates. Alternative mutations in ParC (A63T) and GyrA (E85K) have been found, as well as mutations in GyrB and ParE that have been shown to contribute to resistance (Weigel *et al.*, 2001). There are also reports of fluoroquinolone resistance conferred by recombinant *parC* alleles formed by recombination with other mitis group streptococci (Balsalobre *et al.*, 2003; Stanhope *et al.*, 2005). These recombinant alleles are rare (less than 11% of ciprofloxacin resistant isolates tested), and the reason for this is unclear. Previous studies suggest that recombination events are more frequent than mutational events, other viridans streptococcal species are present in the nasopharynx and possession of recombinant *parC* does not seem to impose a fitness cost (Balsalobre and De La Campa, 2008).

Although topoisomerase mutations cause cross-resistance to several fluoroquinolone antibiotics, the impact varies by agent and the number of mutations present. For example, in a set of fluoroquinolone resistant isolates characterised by Weigel and co-workers, a previously uncharacterised mutation in GyrB increased resistance to gatifloxacin but had no effect on moxifloxacin resistance. Conversely, in the same study a GyrA mutation had a particularly strong association with moxifloxacin resistance but less so with gatifloxacin. This latter observation was linked to other studies that suggested that moxifloxacin preferentially targets GyrA instead of ParC in *S. pneumoniae*. The impact of particular topoisomerase mutations on fluoroquinolone resistance was further complicated by observation of a putative efflux mechanism for fluoroquinolones in *S. pneumoniae*, discussed further in section 1.7 below. However, mutations in ParC are most frequently isolated, consistently reduce susceptibility to all agents, and predispose isolates to further selection of high level antibiotic resistance (Muñoz and De La Campa, 1996).

Global prevalence of fluoroquinolone resistance remains low (<1%) (Canton *et al.*, 2003), although there have been reports of high incidences in some locations such as Spain and Hong Kong (García-Rey *et al.*, 2000; Ho *et al.*, 2001). A long-term surveillance

study carried out in Canada showed a steady increase in the prevalence of ciprofloxacin resistant isolates from 1% in 1997 to 4.2% in 2005 (Adam *et al.*, 2007). Similar increases were seen in levels of resistance to levofloxacin, gatifloxacin, gemifloxacin and moxifloxacin. However, subsequent study extended the analysis to 2009 and showed that fluoroquinolone resistance stabilised due to changes in prescribing practices from ciprofloxacin to the newer fluoroquinolones (Patel *et al.*, 2011). The PneumoWorld surveillance study measured antimicrobial susceptibility of *S. pneumoniae* in eight European countries and showed that, although fluoroquinolone resistance rates were low (<2%), they varied between countries with the highest rates found in Spain and Italy (1.2-1.3%) (Reinert *et al.*, 2005). Resistance rates correlated with the level of fluoroquinolone use in each country. Several surveillance studies have found that the risk of being infected with a fluoroquinolone resistant isolate is highest in hospitals and nursing homes, and if a patient has been treated with fluoroquinolones in the previous three months (Ho *et al.*, 2001; Vanderkooi *et al.*, 2005).

Association of fluoroquinolone resistance with particular clonal lineages is controversial. A study of levofloxacin resistant clinical isolates from the Active Bacterial Core Surveillance program (CDC, USA) between 1998 and 2002 found that over half of the isolates were related to five international clones, with the Spain^{23F}-1 clone particularly highly represented (Pletz *et al.*, 2004). However, several other analyses, including the global PROTEKT surveillance program, have shown that the clonal types of sets of fluoroquinolone resistant clinical isolates are heterogenous (Adam *et al.*, 2007; Canton *et al.*, 2003). In a phylogenetic analysis of the PMEN1 lineage Croucher *et al* (2011b) observed that multiple topoisomerase mutations had emerged independently. Mutations causing fluoroquinolone resistance can emerge spontaneously during therapy and have been associated with treatment failures when levofloxacin or ciprofloxacin was used (Fuller and Low, 2005).

Several studies have shown that some topoisomerase mutations confer a fitness defect

compared to fluoroquinolone susceptible strains (Balsalobre and De La Campa, 2008; Gillespie *et al.*, 2003; Johnson *et al.*, 2005; Rozen *et al.*, 2007). The extent of the growth defect varies depending on the mutation. Mutation of serine 79 of ParC to phenylalanine conferred a reduction in fitness of approximately 8% compared to R6, while mutation of the same residue to tyrosine did not affect growth (Rozen *et al.*, 2007). The opposite effect was observed when serine 81 of GyrA was mutated to tyrosine or phenylalanine, and a growth defect was also observed for GyrA E85K.

However, fitness costs conferred by one topoisomerase mutation can be partially or fully alleviated by acquisition of a second mutation (Balsalobre *et al.*, 2011; Rozen *et al.*, 2007). Balsalobre and co-workers (2011) linked this to DNA supercoiling. In this study, the fitness defect conferred by a high-cost GyrA mutation (E85K) was compensated by acquisition of a recombinant *parC* allele. The authors measured global DNA supercoiling in the single and double mutants and showed that it was disrupted in the *gyrA* mutant but restored when the recombinant *parC* gene was added. As topoisomerase IV and DNA gyrase have opposing effects on DNA supercoiling, it seems that expression of one non-optimal enzyme disrupts the overall supercoiling state, but when the activity of the opposing enzyme is also reduced the supercoiling state is restored.

1.7 Efflux mediated fluoroquinolone resistance in *S. pneumoniae* and background to this project

The existence of an efflux pump in *S. pneumoniae* that confers resistance to fluoroquinolones was originally inferred from several reports of pneumococcal isolates that showed resistance to ciprofloxacin and norfloxacin that was either reversed by addition of efflux inhibitors or could be attributed to lower accumulation of drug.

A study by Piddock and co-workers (1997) showed that three ciprofloxacin resistant

mutants of *S. pneumoniae* accumulated lower levels of ciprofloxacin than their respective parental strains. No mutations were found in the topoisomerase genes, suggesting that there was another mechanism for ciprofloxacin resistance.

Baranova and Neyfakh (1997) observed that exposure of a pneumococcal strain to a subinhibitory concentration of reserpine resulted in a 2.25 to 3.5-fold decrease in susceptibility to ciprofloxacin, norfloxacin, and ethidium bromide. A mutant selected for increased resistance to ethidium bromide showed cross-resistance to ciprofloxacin and norfloxacin that could be reversed by addition of reserpine. Ethidium bromide resistance in the mutant was due to increased efflux, leading the authors to hypothesise that *S. pneumoniae* possesses a multidrug transporter that can transport both ethidium bromide and fluoroquinolones.

Brenwald and co-workers also selected *S. pneumoniae* mutants with decreased susceptibility to ciprofloxacin and norfloxacin (Brenwald *et al.*, 1997). Addition of reserpine reversed fluoroquinolone resistance in three mutants selected on norfloxacin, and these mutants were found to be cross-resistant to ciprofloxacin, ethidium bromide and acriflavine. A further five mutants with reserpine-sensitive norfloxacin resistance were selected in a subsequent study, and a representative mutant, 1N27, was again shown to have cross-resistance to ethidium bromide (Brenwald *et al.*, 1998). Using this observed phenotype as a marker for strains with increased efflux, the authors screened 1,037 clinical isolates of *S. pneumoniae* for susceptibility to norfloxacin, ciprofloxacin, ethidium bromide and acriflavine, using an agar dilution method. Of 273 isolates that showed reduced susceptibility to the fluoroquinolone antibiotics (MIC of norfloxacin $>$ 8 µg/ml and MIC of ciprofloxacin $>$ 1 µg/ml), 98 isolates showed a four-fold or greater reduction in norfloxacin resistance in the presence of reserpine, combined with elevated resistance to ethidium bromide and acriflavine.

The results from these three studies suggest that a multidrug efflux transporter that

can transport fluoroquinolones exists in *S. pneumoniae*, and is a potential mechanism of resistance in clinical isolates.

1.7.1 PmrA

Subsequent work on R6N, an R6 transformant of the 1N27 efflux mutant selected by Brenwald *et al* (1998), identified over-expression of *pmrA*, a gene encoding an efflux pump of the major facilitator superfamily (Gill *et al.*, 1999). The PmrA protein showed 24% amino acid identity to the known NorA and Bmr multidrug efflux pumps of *Staphylococcus aureus* and *Bacillus subtilis*. Gill and co-workers inactivated *pmrA* in R6N by insertion of a chloramphenicol resistance cassette. This reduced MICs of norfloxacin, ciprofloxacin, ethidium bromide and acriflavine to the levels observed in R6, leading the authors to suggest PmrA as a candidate for the pneumococcal multidrug efflux pump that confers fluoroquinolone resistance (Gill *et al.*, 1999). However, later studies provided evidence suggesting that PmrA is not the main fluoroquinolone efflux pump in *S. pneumoniae*.

Piddock and co-workers (2002) measured expression of *pmrA* in a set of clinical isolates of which 34 were ciprofloxacin resistant and 12 were ciprofloxacin susceptible. Thirty-eight of these isolates were norfloxacin resistant and, for eighteen of these, norfloxacin resistance was reduced four-fold or more in the presence of reserpine. Expression of *pmrA* in the isolates varied widely, and there was no correlation between high expression of *pmrA* and reserpine-sensitive norfloxacin resistance.

Brenwald *et al* (2003) investigated two mutants, LEV1 and GEM11, which had a phenotype consistent with increased efflux. The mutants had been selected in a previous study by exposure to levofloxacin and gemifloxacin (Davies *et al.*, 1999). To investigate resistance in these isolates due to efflux, without the confounding influence of topoisomerase mutations, R6 was transformed with genomic DNA from each isolate and transformants

were selected using ethidium bromide, a known substrate of efflux pumps. The *pmrA* gene was not expressed at a level significantly higher than wild-type R6 in the R6LEV1 and R6GEM11 transformants, as determined by Northern blotting.

Another piece of evidence for a second fluoroquinolone efflux transporter in *S. pneumoniae* came from a study by Pestova and co-workers (2002), which looked at two strains where *pmrA* had been inactivated by insertion of a chloramphenicol resistance gene. Inactivation of *pmrA* in a wild type strain did not affect ethidium bromide or fluoroquinolone resistance, while in a mutant with increased efflux *pmrA* inactivation restored susceptibility to ethidium bromide, ciprofloxacin and norfloxacin to wild type levels. However, the authors observed that an effect of reserpine on ciprofloxacin and ethidium bromide susceptibility was not completely eliminated in either of the strains lacking PmrA, suggesting that another reserpine-sensitive transporter is present.

Futhermore, both Pestova *et al* (2002) and Brenwald *et al* (2003) found that mutants with increased resistance to ethidium bromide could be selected from strains lacking functional PmrA. In both cases, the selected mutants showed cross-resistance to fluoroquinolones and the phenotype was reversed in the presence of reserpine, which is suggestive of increased efflux.

There was considerable variation observed in the contribution of PmrA to efflux-mediated fluoroquinolone resistance between different strains and studies, and several lines of evidence suggested that at least two fluoroquinolone efflux systems exist in *S. pneumoniae*. Similar evidence for the existence of more than one multidrug transporter with overlapping substrate specificities had been previously observed in *B. subtilis* (Ahmed *et al.*, 1995) and *S. aureus* (Kaatz *et al.*, 2000).

1.7.2 PatA and PatB

1.7.2.1 Discovery and role in efflux mediated fluoroquinolone resistance

Two independent studies showed that the ABC transporter genes *patA* and *patB* were involved in efflux of fluoroquinolones.

Robertson *et al* (2005) systematically inactivated thirteen putative efflux pump genes in *S. pneumoniae* strain TIGR4. Inactivation of the genes encoding PatA and PatB resulted in significant increases in susceptibiliy to ethidium bromide, acriflavine, ciprofloxacin and norfloxacin, as well as the non-quinolone GyrB inhibitor novobiocin, a quarternary ammonium compound berberine, and the macrolide antibiotic erythromycin (Robertson *et al*., 2005). Robertson *et al* also constructed a double mutant where both *patA* and *patB* were inactivated, but this did not result in any further decreases in susceptibility to any of the tested agents. These results were later confirmed by Tocci *et al* (2013).

A different approach to identifying transporters involved in fluoroquinolone resistance was taken by Marrer *et al* (2006a). The transcriptome of M22, a ciprofloxacin resistant mutant selected from strain M4 (NCTC 7465) that had a phenotype suggestive of efflux (Piddock and Johnson, 2002), was determined using a microarray. Gene expression in M22 was compared to its parental strain M4 at three different growth phases and after exposure to three different concentrations of ciprofloxacin. A set of 22 genes were expressed at a higher level in M22 than in M4 under all conditions tested, and this included *patA* and *patB*. Increased expression of these genes was also observed in M4 following exposure to ciprofloxacin, suggesting that ciprofloxacin induces expression of *patA* and *patB*. Modest inductions were also seen in M22 when this strain was exposed to ciprofloxacin. As well as *patA* and *patB*, several other genes showed altered expression in M22 compared to M4, including genes involved in DNA repair and replication, some other transporters, genes involved in isoleucine and valine biosynthesis and several genes

encoding hypothetical proteins. This pleiotropic phenotype of M22 prompted the authors to suggest that the over-expression of *patA* and *patB* was mediated by a mutation in a global regulator, analagous to mutations found in MarR in *E. coli* which leads to altered expression of more than 60 genes, some of which are involved in antibiotic efflux (Barbosa and Levy, 2000).

A follow-up study by the same authors focused more specifically on the role of *patA* and *patB* in ciprofloxacin resistance in M22 and M4 (Marrer *et al.*, 2006b). The *patA* and *patB* genes were disrupted in R6 by insertion of an erythromycin resistance gene. In M22, *patB* was disrupted by the same method, but mutants with an insertion in *patA* could not be obtained. Growth of the R6 *patA* disruptant was impaired, suggesting that *patA* plays an important role in the cell and may be essential in the M22 genetic background. Disruption of either gene in R6 caused a fourfold reduction in resistance to ciprofloxacin, and disruption of *patB* in M22 partially reduced resistance to ciprofloxacin and norfloxacin, although MICs did not revert to the levels observed in M4. The M22 *patB* disruptant strain was shown to accumulate double the amount of ciprofloxacin accumulated by M22.

Additionally, Marrer *et al* (2006b) found that expression of *pmrA* was undetectable in both M4 and M22 by Northern blotting or quantative competitive reverse transcriptase PCR, and subsequent sequencing of the gene revealed that it contains an internal deletion of 84 nucleotides (28 amino acids) in these strains compared to R6. Based on a structural model of the protein (Gill *et al.*, 1999), this deletion was predicted to remove most of the sixth transmembrane helix of the transporter, which would presumably render it non-functional.

The involvement of *patAB* over-expression in ciprofloxacin resistance was further emphasised by Avrain and co-workers (2007). In this study, mutants were selected using four different fluoroquinolones from two parental strains, ATCC 49619, which is a fluo-

roquinolone susceptible type strain, and SP32, a clinical isolate known to over-express *pmrA*. When each strain was exposed to half its MIC of ciprofloxacin for 13 days, the resulting mutants showed increased resistance to ciprofloxacin, and also modest decreases in susceptibility to levofloxacin and gemifloxacin. Resistance to ciprofloxacin and levofloxacin could be partially reversed by addition of reserpine. Mutants derived from both parental strains were shown by quantitative real-time PCR to over-express *patA* and *patB*. In mutants of SP32, which originally expressed *pmrA* at a level 11-fold higher than ATCC49619, *pmrA* expression had decreased while expression of *patA* and *patB* was significantly increased. In contrast to ciprofloxacin, exposure to moxifloxacin, levofloxacin and garenoxacin selected for mutations in QRDRs of ParC, ParE and GyrA instead of changes in expression of *patA* and *patB*. This resulted in larger increases in MIC of all fluoroquinolones tested, and these MICs were unchanged by addition of reserpine. The authors concluded that the mechanism of fluoroquinolone resistance depended on the selecting antibiotic, rather than the initial strain. Ciprofloxacin selected for increased efflux instead of topoisomerase mutations, and this was mediated by *patAB*, not *pmrA*.

Over-expression of *patA* and *patB* could also be selected by exposure to the efflux inhibitor reserpine (Garvey and Piddock, 2008). Garvey and Piddock characterised two mutants with reduced susceptibility to reserpine, M169 selected from R6 and M168 selected from M4. As well as reduced susceptibility to reserpine, these mutants also showed cross-resistance to the fluoroquinolones ciprofloxacin, norfloxacin and levofloxacin, the dyes ethidium bromide and acriflavine and the biocide cetrimide. Both mutants were shown to over-express *patA* and *patB*. Both mutants were selected by a single round of exposure to 100 µg/ml reserpine, which suggested that the phenotype was conferred by a single genetic change. The mutation frequencies, defined as the number of reserpine-resistanct mutants isolated divided by the total number of viable cells per millilitre, were 7.3×10^6 and 3.6×10^5 for R6 and M4, respectively. These mutation frequencies led the authors to conclude that *patA* and *patB* over-expression was mediated by a mutation

in a single gene, hypothesised to be in a regulatory gene. However, as will be described in this thesis, the mutation causing reserpine resistance in a selected mutant of R6 was in fact a duplication instead of a point mutation, which may explain the differences in mutation frequencies observed between the two strains. When genomic DNA from M168 and M169 was transformed into R6, reserpine resistant transformants with the same phenotype as M168 and M169 could be recovered after a single round of transformation, giving strains M186 and M184, respectively. This supports the hypothesis that *patAB* over-expression is conferred by a single genetic change in these strains.

Garvey and Piddock (2008) also investigated the effect of inactivation of *patA* or *patB* in R6, M169, M184 and M186. This was carried out by insertion of a *magellan2* minitransposon into each gene. Inactivation of either *patA* or *patB* reversed multidrug resistance in the strains over-expressing these genes. In the case of ciprofloxacin, ethidium bromide and acriflavine, the resulting mutants were more susceptible to these agents than wild type R6. A similar hypersusceptible phenotype was observed when *patA* or *patB* was inactivated in R6. In strains over-expressing *patA* and *patB*, only inactivation of *patA* was able to restore reserpine resistance to wild type levels.

A double mutant where both *patA* and *patB* were inactivated by replacement with chloramphenicol and kanamycin resistance genes, respectively, was later constructed and analysed by Boncoeur and co-workers (2012). Reduction in growth rate of the double mutant compared to wild type R6 in the presence of 0.025, 0.05 and 0.1 μg/ml ethidium bromide was not significantly different from either of the single mutants. In contrast, when mutants were exposed to 100 ng/ml ciprofloxacin or 400 ng/ml norfloxacin, the growth of the double mutant was inhibited to a greater degree than either of the single mutants, particularly in the presence of ciprofloxacin.

A growth defect was observed when *patA* (but not *patB*) was inactivated in R6 in Garvey's study, but this was not observed in the Boncoeur 2012 study. Deletion of *patA* in

TIGR4 also did not affect growth (Robertson *et al.*, 2005). On the other hand, deletion of *patA* could not be successfully accomplished in M22 leading to suggestions of essentiality in this genetic background (Marrer *et al.*, 2006b). These observed differences in phenotype could be explained by differences in growth conditions; for example, Boncoeur *et al* grew strains in Todd-Hewitt broth whereas Garvey and Piddock used Brain Heart Infusion. Alternatively, the differences in the gene inactivation methods used could have an effect on growth. Garvey and Piddock used *in vitro Mariner* mutagenesis, while Boncoeur *et al* replaced *patA* and *patB* with chloramphenicol and kanamycin resistance genes, respectively, by overlap extension PCR. Robertson *et al* also used overlap extension PCR to replace *patA* with a spectinomycin resistance gene and *patB* with a chloramphenicol resistance gene. Another possibility is that there could be strain-specific differences in the essentiality of *patA*. Although Boncouer *et al* and Garvey and Piddock both used R6 as the wild type strain, there may still be small differences in genomes of the R6 parental strains used by the two different laboratories.

1.7.2.2 patA and patB expression as a response to stress

Previous studies have shown that expression of *patA* and *patB* is induced in the presence of sub-inhibitory concentrations of ciprofloxacin and norfloxacin (Marrer *et al.*, 2006a). A recent study by El Garch *et al* (2010) extended these observations by showing that over-expression of *patA* and *patB* may be part of a general stress response in *S. pneumoniae*. Firstly, they demonstrated that, not only can expression of *patA* and *patB* be induced by ciprofloxacin and norfloxacin, which are good substrates for the PatAB efflux pump (Avrain *et al.*, 2007), but this expression can also be induced by three other fluoroquinolones: levofloxacin, moxifloxacin and gemifloxacin. These fluoroquinolones are more hydrophobic than ciprofloxacin and norfloxacin and are not considered good substrates for efflux (Avrain *et al.*, 2007). This suggests that transient induction of *patA* and *patB* is a reaction to cell stress caused by challenge with fluoroquinolone antibiotics,

rather than an example of an efflux pump being upregulated in the presence of its substrate. In addition, it was shown that mitomycin C, a DNA damaging agent that is a known stimulator of the competence pathway in *S. pneumoniae*, also induced expression of *patA* and *patB*. Increases in expression of *patA* and *patB* by mitomycin C correlated strongly with the up-regulation of two late competence genes, *ssbB* and *recA*. Induction of the competence pathway has been suggested to be a general response to cell stress in *S. pneumoniae*, analogous to the SOS response characterised in *E. coli* (Claverys *et al.*, 2006; Little and Mount, 1982). Taken together, these results suggest that induction of *patA* and *patB* may be part of a general stress response to DNA damage, and that this could be linked to the competence pathway.

In the study by El Garch *et al* induction of *patA* and *patB* over-expression was observed in two laboratory mutants and two clinical isolates that over-expressed *patA* and *patB*, showing that *patA* and *patB* expression can be induced by fluoroquinolones in a variety of genetic backgrounds, and in mutants where the basal level of expression is already high. This observation contradicts that of Marrer *et al* (2006a) who did not observe further induction of *patA* and *patB* expression in the ciprofloxacin-selected M4 mutant, M22, in the presence of ciprofloxacin or norfloxacin. However, M22 had a highly pleiotropic phenotype so this discrepancy could be explained by the mutation causing *patA* and *patB* upregulation in M22 occurring in a global regulator higher up in the pathway controlling the expression of these genes. The mutants studied by El Garch *et al* may be mutated in local regulators, allowing global stress response mechanisms to further affect the expression of the genes. Alternatively, the mutants studied by El Garch *et al* may be mutated such that expression of *patA* and *patB* is partially de-repressed, while expression in M22 may be fully constitutive.

Van Opijnen and Camilli (2012) recently used Tn-seq to link pneumococcal genes to phenotypes. The technique involves high-throughput sequencing of transposon insertion libraries before and after application of a test condition. Changes in the frequency of

transposon insertions in each gene in the output library compared to the input library provides information regarding the importance of that gene for survival of the bacterium in that test condition (van Opijnen *et al.*, 2009). Van Opijnen and Camilli generated transposon insertion libraries in pneumococcal strain TIGR4 and exposed them to a variety of conditions including growth on different carbon sources, antibiotic exposure, and other cell stresses such as altered pH and temperature. Survival of transposon mutants in the nasopharynx and lung was also tested using a murine model of infection. As would be expected based on the results of previous studies, mutants with a transposon insertion in *patA* or *patB* were found to be significantly depleted in libraries exposed to norfloxacin. However, the Tn-seq analysis also suggested a role for *patA* and *patB* in tolerance of alkaline pH. The role of *patAB* in alkaline tolerance seemed to be separate from the role in norfloxacin tolerance, and led the authors to suggest that this is the first example of a dual-function ABC transporter.

Regulons have been determined for several pneumococcal global transcriptional regulators such as CiaRH, ComX, StkP, PsaR and SczA (Halfmann *et al.*, 2007; Kloosterman *et al.*, 2007, 2008; Peterson *et al.*, 2000; Saskova *et al.*, 2007). The *patA* and *patB* genes have not been identified among these regulons. One possible global regulatory mechanism that could control *patAB* expression is DNA supercoiling. Ferrándiz *et al* (2010) investigated differences in gene expression across the whole S. pneumoniae R6 genome in the presence or absence of novobiocin, a drug that causes relaxation of supercoiled DNA by inhibiting DNA gyrase. This analysis revealed that the R6 genome contains 15 distinct gene clusters, representing 13% of the genome, which show a coordinated regulatory response to the level of DNA supercoiling, independent of operon organisation. The *patA* and *patB* genes are located within topology-reacting cluster 15, a cluster that is upregulated in response to DNA relaxation.

1.7.2.3 Role of *patA* and *patB* in linezolid resistance, and contribution to virulence

Over-expression of *patA* and *patB* was subsequently observed in three out of four independent *S. pneumoniae* mutants selected for resistance to the oxazolidinone antibiotic linezolid (Feng *et al.*, 2009). Two linezolid resistant mutants were selected from parental strains R6 and CCRI 1974 by exposure to progressively increasing concentrations of linezolid from one to 32 $\mu g/ml$. All four mutants were found to have acquired point mutations in all four copies of the 23S rRNA, which was a previously known mechanism of resistance (Prystowsky *et al.*, 2001). However, one of the two R6 mutants and both of the CCRI 1974 mutants also possessed a mutation upstream of *patA*. In the R6 mutant and one of the CCRI 1974 mutants this was 46 bp upstream of the *patA* start codon, while in the remaining CCRI 1974 mutant the mutation was located 32 bp upstream. Both of these mutations were associated with higher expression of *patA* and *patB*, as shown by qRT-PCR. The CCRI 1974 mutants were also shown to be cross-resistant to ciprofloxacin. However, when *patA* or *patB* were inactivated in one of the CCRI 1974 mutants, linezolid resistance was only partially decreased, leading the authors to conclude that PatA and PatB have a role in linezolid resistance but that it is only minor compared to the effect of mutations in 23S rRNA.

A subsequent study by the same group used whole genome transformation to analyse the role of *patAB* over-expression in linezolid resistance (Billal *et al.*, 2011). Genomic DNA from a new linezolid resistant mutant selected from CCRI 1974 was used to transform a linezolid sensitive CCRI 1974 and R6 recipient strains. Two to three rounds of transformation were required to recreate the previously observed increases in linezolid resistance in both genetic backgrounds, indicating that transfer of more than one mutation was required. Genome sequencing of the transformants and original mutant revealed that increases in linezolid resistance at each stage of transformation correlated with accu-

mulation of mutations in the four copies of the 16S rRNA, followed by acquisition of mutations in the 50S ribosomal proteins L3 and L16. The transformant with the CCRI 1974 genetic background also acquired a mutation in the ABC transporter gene spr1021 on the third round of transformation. The mutation 32 bp upstream of *patA*, which was present in the original mutant, was not transferred to either of the transformants tested, and transfer of this mutation to a linezolid susceptible strain by transformation with a specific PCR amplimer did not increase linezolid resistance. However, the mutations in the 23S rRNA genes conferred a growth defect, which was exacerbated by the acquisition of the spr1021 mutation. Transfer of the mutation upstream of *patA* into strains carrying these mutations was shown to significantly reduce this growth defect, although growth was not restored to the level of the linezolid susceptible parental strain. Over-expression of *patAB* therefore appears to compensate for fitness defects conferred by mutations causing linezolid resistance. The molecular basis for this compensatory effect has not been determined.

A similar mutation upstream of *patA* causing *patAB* over-expression was identified in a clinical isolate taken from the cerebrospinal fluid (CSF) of a meningitis patient (Croucher *et al.*, 2013). This mutation was not found in an isogenic isolate found in the bloodstream of the same patient, leading to the suggestion that upregulation of *patAB* contributes to pneumococcal survival in the CSF.

1.7.2.4 Structure of PatAB

Functional characterisation of PatA and PatB was carried out by Boncoeur and co-workers (2012). In this study, PatA and PatB were over-expressed in *E. coli*, purified and incorporated singly and in combination into inside-out vesicles. Sodium orthovanadate-sensitive transport of the fluorescent dye Hoechst 33342 was observed only when PatA and PatB were combined. This transport was abolished when the Walker A motif of

either PatA or PatB was mutated, showing that transport is ATPase-dependent, as expected. The oligomerisation state of purified PatA and PatB, both singly and in combination, was also assessed by Boncoeur *et al* using three methods: gel filtration, analytical ultracentrifugation and size exclusion chromatography. The results were similar and showed that PatA and PatB together produced stable and functional heterodimers, while PatA alone formed stable but non-functional homodimers. When PatB was purified alone it formed a mixture of monomers and non-functional homodimers, suggesting that PatB homodimers are unstable.

1.7.2.5 Homologues of *patA* and *patB* in *S. suis*

Over-expression of the *satA* and *satB* genes, which are homologues of *patA* and *patB*, has been reported in the zoonotic pathogen *S. suis* (Escudero *et al.*, 2011). The SatA and SatB proteins are 66% and 67% identical to PatA and PatB respectively. Escudero and co-workers examined a *S. suis* isolate with resistance to ciprofloxacin and norfloxacin that was reversed by addition of reserpine. They showed that *satA* and *satB* were upregulated in this isolate, but expression of *smrA*, which encodes a MFS transporter with 58% sequence identity to PmrA, was unchanged. Heterologous expression of *satAB* in *Bacillus subtilis* conferred increased resistance to norfloxacin and ciprofloxacin, but not to enrofloxacin, levofloxacin, moxifloxacin or naladixic acid. In contrast, heterologous expression of *smrA* did not alter the resistance of the *B. subtilis* host to any of the agents tested, and no substrates were identified using a Biolog phenotype microarray. The authors concluded that SatAB is a narrow spectrum ciprofloxacin and norfloxacin transporter. SmrA is not involved in fluoroquinolone transport in the strain tested, and its natural substrate remains unidentified. This situation is very similar to that observed in the pneumococcus. However, the genetic organisation of the *satAB* locus is different from the *patAB* locus in *S. pneumoniae*. Whereas *patA* and *patB* are separated by a degenerate transposase, in *S. suis satA* and *satB* are contiguous, and Escudero *et al*

(2011) showed that they are transcribed in the same mRNA molecule. Upstream of, but slightly overlapping with *satA* there is a gene encoding a MarR superfamily regulator, named *satR*. Further investigation of this gene showed that SatR is indeed a regulator of SatAB (Escudero *et al.*, 2013). However, no homologue of SatR is present in *S. pneumoniae*.

1.8 Hypotheses tested in this project

The hypotheses to be investigated in this project are:

- Over-expression of *patA* and *patB* is a clinically relevant mechanism of fluoroquinolone resistance.
- Specific regulatory pathways exist in *S. pneumoniae* that control the expression of *patAB* in response to stimuli such as DNA damage.
- Over-expression of *patA* and *patB* in *S. pneumoniae* be conferred by mutations in regulatory genes.

1.9 Aims and objectives

- To complete experiments and data analysis on a set of fluoroquinolone-resistant clinical isolates to determine the clinical relevance of over-expression of *patA* and *patB*.
- To identify the mutated protein(s) responsible for up-regulation of *patA* and *patB* expression in *S. pneumoniae* laboratory mutants and clinical isolates by whole genome sequencing.

- To further characterise identified regulators to elucidate parts of the regulatory network controlling *patAB* expression.

2 Materials and Methods

2.1 Bacterial strains and growth

All bacterial strains used in this study are shown in Table 2.1.

2.1.1 Growth conditions and media

S. pneumoniae strains and isolates were grown in Brain Heart Infusion (BHI) broth (Oxoid, Basingstoke, U.K.; cat. no. CM1135B), or on Columbia agar (Oxoid, Basingstoke, U.K.; cat. no. CM0331B) plates supplemented with 5% horse blood (cat. no. HB034; TCS Biosciences, U.K.). Broth cultures and plates were incubated statically at 37°C in an atmosphere of 7.5% CO_2. Growth was monitored by measurement of the optical density at 660 nm (OD_{660}) of a 1 ml sample of culture using a spectrophotometer (Geneflow, UK), compared to a blank measurement from 1 ml of sterile BHI broth.

E. coli strains used for construction of the pBAV1K-*gfp82* plasmid (section 2.9.1) were grown in Luria Bertani (LB) broth, or on LB agar plates, supplemented with appropriate antibiotics. Agar plates were incubated statically at 37°C, and broth cultures were incubated at 37°C with shaking at 180 rpm. Growth was monitored by measurement of the optical density at 600 nm (OD_{600}) of a 1 ml sample, compared to a blank measurement from 1 ml of sterile LB broth.

Table 2.1. *S. pneumoniae* strains and isolates used in this study

Laboratory code	Description/Phenotype	Source/Reference
R6	Unencapsulated wild type strain, ATCC BAA255	(Avery *et al.*, 1944; Hoskins *et al.*, 2001)
M3	Serotype 2 type strain NCTC7466	-
M4	Serotype 1 type strain NCTC7465	(Brown and Roberts, 1991)
M168	Reserpine resistant mutant derived from M4	(Garvey and Piddock, 2008)
M169	Reserpine resistant mutant derived from R6	(Garvey and Piddock, 2008)
M184	R6 recipient transformed with M169 DNA, selected on 100 µg/ml reserpine	(Garvey and Piddock, 2008)
M240	R6 *patB::magellan2*	(Garvey and Piddock, 2008)
M246	R6 *patA::magellan2*	(Garvey and Piddock, 2008)
M37-M42, M44-M54	Clinical isolates from Lung Investigation Unit	(Piddock *et al.*, 1998)
M65-M73	Clinical isolates from Queen Elizabeth Hospital, Birmingham	(Piddock *et al.*, 1998)
M74-M86	Clinical isolates from Centres for Disease Control and Prevention, Atlanta, GA	(Jorgensen *et al.*, 1999)
M87-M101	Clinical isolates from MRL Pharamceutical services	(Jones and Pfaller, 2000)
M296-M297	Ciprofloxacin-selected mutants of clinical isolate AMC-058, Albany Medical Centre	(Jumbe *et al.*, 2006)
M499	R6 *hrcA* G824T	This study
M500	R6 recipient transformed with M184 DNA, selected on 8 µg/ml norfloxacin	This study
M500-M503	R6 recipients transformed with M184 DNA, selected on 8 µg/ml ethidium bromide	This study
M504	R6 recipient transformed with M168 DNA, selected on 8µg/ml ethidium bromide	This study
M505	R6 recipient transformed with M101 DNA, selected on 8µg/ml ethidium bromide	This study
M506	R6 recipient transformed with M187 DNA, selected on 8µg/ml ethidium bromide	This study
M507	R6 recipient transformed with M174 DNA, selected on 8µg/ml ethidium bromide	This study
M508-M510	M184 *patA* copy 1::*magellan2*	This study

2.1.2 Storage of bacterial strains

S. pneumoniae strains were stored at -80°C on Protect beads (Technical Service Consultants Ltd., cat. no. TS/80-MX). To resuscitate strains for use, beads were placed in 3 ml brain heart infusion (BHI) broth supplemented with 5% sterile defibrinated horse blood and incubated statically overnight. Bacteria were then transferred to Columbia agar plates supplemented with 5% defibrinated horse blood under sterile conditions, using the streak plate method to obtain pure colonies, and incubated overnight. BHI broth (5 ml) was inoculated with up to ten pure colonies, and the culture was grown until an OD_{660} of 0.4-0.6 was reached (mid-logarithmic phase). Sterile glycerol was added to a final concentration of 10%, and 1 ml aliquots were frozen at -80°C for use as starter cultures in all subsequent experiments.

E. coli strains were also stored at -80°C on Protect beads. Beads were rubbed on the surface of an LB agar plate (Sigma-Aldrich Ltd., U.K.; cat. no. L2897) containing appropriate antibiotics, and the plates were incubated overnight at 37°C. Plates were sealed with parafilm and stored at 4°C for up to two weeks.

2.2 DNA extraction

2.2.1 Extraction of genomic DNA

DNA was extracted from all *S. pneumoniae* strains using the Promega Wizard Genomic DNA Purification kit according to the manufacturer's instructions, as follows. Frozen, logarithmic phase cells were diluted 1:10 in sterile prewarmed BHI broth (10 ml) and grown statically at 37°C in 5% CO2 to an OD_{660} of 0.6. Cells were harvested by centrifugation (10 mins, 2200 x g, room temperature), resuspended in 480 µl EDTA (50 mM) and transferred into a microcentrifuge tube. A 120 µl volume of lysozyme (10

mg/ml) was added and the mixture was incubated at 37°C for 30-60 minutes. Lysozyme-treated cells were harvested by centrifugation (12,100 x g, 3 min), resuspended in 600 µl of Promega Nuclei Lysis Solution and lysed by heating (80°C, 5 min). After cooling to room temperature, 3 µl Promega RNase was added and the lysates incubated at 37°C for 15-60 minutes. To remove proteins, 200 µl Promega Protein Precipitation Solution was added, the sample was incubated on ice for five minutes, then the precipitated protein was pelleted by centrifugation (12,100 x g, 3 min). The supernatant was transferred to a microcentrifuge tube containing 600 µl isopropanol, and mixed by inversion until precipitated DNA formed a visible mass. DNA was pelleted by centrifugation (1 min, 12,100 x g), washed with 70% ethanol and resuspended in 100 µl of sterile nuclease-free water (Gibco). Prepared DNA was stored at -20°C for up to four months. To assess yield and quality, 5 µl of prepared DNA was visualised on a 1% agarose gel.

2.2.2 Extraction of plasmid DNA

High copy number plasmid DNA was harvested from *E. coli* strains using the QIAgen® Miniprep kit (QIAgen, UK, Cat. no. 27104), according to the manufacturer's instructions. Purified plasmid DNA was stored at -20°C.

Plasmid DNA was extracted from *S. pneumoniae* strains using a QIAgen® Miniprep kit following the QIAgen® user-developed protocol for isolation of plasmid DNA from *Bacillus subtilis*. This protocol differs from the standard miniprep protocol by adding a lysozyme treatment to break down the thick Gram-positive cell wall. Cells from 10 ml of mid-logarithmic phase (OD_{660} 0.4-0.6) culture were harvested by centrifugation (10 minutes at 2,200 x *g*). Cells were resuspended in 250 µl of QIAgen® Buffer P1 containing a final concentration of 1 mg/ml lysozyme, and incubated for 10 minutes at 37°C. Plasmid extraction was then continued as described from step two of the standard QIAgen® miniprep protocol.

2.2.3 Agarose gel electrophoresis

Genomic DNA and plasmid preparations were visualised by electrophoresis on a 1% agarose gel. To prepare agarose gels, 1 g of electrophoresis grade agarose (Life Technologies, cat. no. 75000-500) was dissolved in 100 ml of 1% Tris-Borate-EDTA (TBE; Ambion, cat. no. AM9864) by heating in a 650 W microwave for 5 minutes. After allowing the solution to cool, either ethidium bromide (final concentration 0.1 µg/ml) or Midori Green Advance DNA Stain (5 µl; Geneflow, cat no. S6-0022) was added as a fluorescent dye to allow DNA visualisation. The agarose solution was poured into a gel tray, allowed to set and immersed in 1% TBE in a gel tank. Five µl of each sample was mixed with 2 µl of 5x Blue DNA loading buffer (Bioline, cat. no. BIO-37045) and loaded into individual wells of the gel along with HyperLadder^{TM} 1kb (Bioline, cat. no. BIO-33053) Samples were separated by electrophoresis for one hour at 100 V. DNA fragments were visualised by UV trans-illumination using a Gene Genius image analyser (Syngene, Cambridge, UK) and were quantified, where appropriate, using the GeneTools software package (Syngene).

2.3 Transformation of whole genomic DNA from *patAB* over-expressing strains into R6

2.3.1 Transformation

Donor DNA was extracted from the appropriate strain as described above. A 1 ml frozen stock of R6 was thawed and diluted in 9 ml BHI broth. The culture was incubated at 37°C until an OD_{660} of 0.4 was reached, corresponding to log-phase growth. The culture was then diluted 1:20 in 10 ml sterile Competence Medium, which consisted of 37 g/L

Todd-Hewitt broth (Oxoid, cat. no. CM0189), 5 g/L Yeast Extract (Oxoid, cat. no CM0019), 1.2 mM calcium chloride and 2 g/L bovine serum albumin (BSA; Sigma, cat. no. B6917). Competence stimulating peptide 1 (CSP1; Mimotopes) was then added to a final concentration of 100 ng/ml. For each transformation reaction a 500 µl aliquot of the diluted culture was transferred to a microcentrifuge tube, and up to 1 µg of the required donor DNA was added. In parallel, a control transformation was set up where the donor DNA was replaced with an equivalent volume of sterile water or R6 genomic DNA. The mixtures were incubated statically for three hours (37°C), and then 100 µl of each mixture was spread onto Columbia blood agar plates containing an appropriate selective agent and incubated overnight (37°C, 7.5% CO_2). Resulting transformants were sub-cultured onto fresh Columbia blood agar and again incubated overnight (37°C, 7.5% CO_2).

2.3.2 Confirmation of *patAB* over-expression phenotype of transformants

2.3.2.1 Antibiotics and determination of minimum inhibitory concentrations

The minimum inhibitory concentrations (MICs) of all antibiotics, dyes and efflux pump inhibitors used in this study were determined by the agar doubling dilution method, performed according to the BSAC guidelines (Andrews, 2001). Stock solutions (10 mg/ml) of the antibiotics ciprofloxacin, norfloxacin, chloramphenicol, tetracycline, spectinomycin and kanamycin (Sigma Aldrich, Dorset, U.K.) were made up according to the manufacturer's instructions. Stock solutions (10 mg/ml) of the dyes ethidium bromide and acriflavine (Sigma Aldrich, Dorset, U.K.) were made by reconstitution in sterile distilled water (SDW). To make a 1000 µg/ml stock solution of the efflux pump inhibitor reserpine (Sigma Aldrich, Dorset, U.K.), 0.01 g reserpine powder was dissolved in 8 ml

DMSO by vigorous vortexing (10 min), and then made up to a final volume of 10 ml by addition of SDW. Plates were prepared by adding appropriate volumes of 10000, 1000 and 100 µg/ml stock solutions of antibiotics, dyes or efflux pump inhibitors to 30 ml universal tubes to give a doubling dilution series in a final volume of 20 ml of agar. Twenty ml Iso-Sensitest agar (ISA) was added to each tube using an automatic dispenser (Jencons, U.K.), poured into a Petri dish and allowed to set. Solidified agar plates were then dried for 30 minutes in a 60°C oven. A 1 µl spot of each bacterial strain, grown to an $OD_{6}60$ of 0.4 in BHI broth, was inoculated onto the surface of each plate using a multipoint inoculator (Denley Tech Ltd., Sussex, U.K.). Inoculations were allowed to dry, then the plates were incubated overnight. The MIC of each agent for a particular strain was defined as the lowest concentration of antibiotic (µg/ml) at which no growth was observed; a faint haze of growth was ignored.

2.3.2.2 Measurement of intracellular accumulation of ethidium bromide

Ethidium bromide fluoresces when intercalated into DNA, meaning that its intracellular accumulation can be monitored by measuring the increase in fluorescence of a culture over time when exogenous ethidium bromide is added. Mid-logarithmic phase cultures (OD_{660} of 0.4) were pelleted by centrifugation at 2200 x g and resuspended in an equivalent volume of sterile PBS. Cell suspensions were adjusted to an OD_{660} of 0.1 by addition of more sterile PBS, then added in triplicate to a black microtitre tray as follows. For fluorescence readings without addition of ethidium bromide, 200 µl of cell suspension was added. For measurements with addition of ethidium bromide, 180 µl cell suspension was used. Sterile PBS was used to provide blank readings throughout. Fluorescence was measured every two minutes for a total of 30 minutes using a FLUOstar Optima plate reader (BMG Labtech, Aylesbury, UK; excitation wavelength: 530 nm, emission wavelengths: 600 nm). Twenty µl of a 250 µM stock of ethidium bromide was injected into each of the wells containing 180 µl of culture on the second cycle of measurement.

To analyse the data, average blank readings with and without ethidium bromide were calculated and subtracted as appropriate from the fluorescence readings for each well, and average blank corrected fluorescence with and without ethidium bromide was calculated from the three technical replicates for each tested strain. To examine ethidium bromide accumulation over time, the average and standard deviation of fluorescence from three biological replicates was calculated for each time point. To identify significant differences in steady-state levels of ethidium bromide accumulation between strains, a time-point was chosen such that fluorescence of all strains had reached a steady-state (10 minutes after addition of ethidium bromide unless stated otherwise), and strains were tested for a significant difference ($p<0.05$) from the wild-type strain using a two-tailed Student's t-test on three biological replicates. If more than five comparisons were performed simultaneously, a Bonferroni correction for multiple testing was applied.

2.3.2.3 Measurement of growth kinetics

For measurement of growth rates of strains under standard pneumococcal culture conditions (static incubation at 37°C, 7.5% CO_2), frozen starter cultures of the strains to be tested were thawed and diluted 1:100 in 50-100 ml BHI broth. Cultures were incubated statically at 37°C in 7.5% CO_2, and the OD_{660} of a 1 ml sample was measured every hour using a spectrophotometer. To calculate generation times, two time points were chosen during the exponential phase of growth (named t_0 and t_1), and generation times were calculated using the formula: $G_t = (t_1 - t_0) \times ln(2)/ln(OD_1/OD_0)$, where OD_0 and OD_1 are the measured OD_{660} values at t_0 and t_1 respectively.

Growth of some strains and isolates was measured in a high-throughput manner in a microtitre tray using the Fluostar Optima plate reader (BMG Labtech). Growing cultures were diluted to an OD_{660} of 0.1 in fresh sterile BHI broth, then diluted 1:10 in more sterile BHI. For each culture to be tested, three 200 µl aliquots of the diluted culture

were added to wells of a clear microtitre tray. The microtitre tray was sealed with a clear, breathable membrane (Sigma Aldrich, cat. no. Z380059), and incubated overnight in the Fluostar Optima at 37°C. Absorbance at 600 nm was measured every ten minutes, with five seconds of shaking before each reading to redistribute cells.

2.4 Measuring gene expression by quantitative real-time PCR

2.4.1 Extraction of total RNA

Three biological replicates of each strain to be tested were grown to an OD_{660} of 0.4 in BHI broth. Total RNA was extracted using a Promega SV Total RNA Isolation kit according to the manufacturer's instructions provided for RNA extraction from Gram positive bacteria, as follows. All procedures were carried out on ice, unless otherwise stated. One ml of each culture was transferred to an RNase-free microcentrifuge tube and the cells pelleted by centrifugation (14,000 x g, 2 minutes). The supernatant was discarded and the cells were resuspended in 100 µl freshly prepared TE buffer containing 3 mg/ml lysozyme. Resuspended cells were incubated for 10 minutes at room temperature, then 75 µl of RNA lysis buffer was added, followed by 350 µl RNA Dilution buffer, and the suspension was mixed by inversion. Two hundred µl of 95% ethanol was added to the lysate, which was then mixed by pipetting three to four times and transferred to a spin column held in a collection tube. Nucleic acids were bound to the columns by centrifugation (14,000 x g, 60 seconds). The flow-through was discarded and the columns were washed with 600 µl RNA Wash Solution (centifugation for 60 seconds, 14,000 x g). The flow-through was discarded and 50 µl of DNase incubation mix (40 µl Yellow Core Buffer, 5 µl 90 mM $MnCl_2$, 5 µl DNase I) was added to each column.

The columns were incubated for 15 minutes at room temperature. To stop the reaction, 200 µl DNase Stop Solution was added to each column. This solution was removed by centrifugation (14,000 x g, 60 seconds) and the columns were washed with 600 µl RNA Wash Solution as above. The flow-through was discarded and the column was washed with 250 µl RNA Wash Solution by centrifugation (14,000 x g, 2 min). The spin column was transferred to a fresh RNase-free microcentrifuge tube containing 1 µl RNaseOUT (Invitrogen). RNA was eluted in 100 µl nuclease-free water (Promega) by centrifugation (14,000 x g, 60 seconds). Prepared RNA was divided into aliquots and stored at -80°C until required.

Concentrations of RNA and contaminating DNA were measured using a Qubit fluorometer. Samples were prepared using the Qubit RNA Assay kit for RNA quantification, and the Qubit dsDNA HS Assay kit for DNA quantification, according to the manufacturer's instructions. Residual DNA contamination was removed by DNase using the TURBO DNA-*free*™kit (Ambion, cat. no. AM1907) according to the manufacturer's instructions.

2.4.2 cDNA synthesis

Prepared RNA was reverse-transcribed into cDNA using SuperScript® III First-Strand Synthesis System for RT-PCR (Invitrogen, cat. no. 18080-051) using random priming. RNA preparations were adjusted to the same concentration using nuclease-free water. For each RNA preparation to be reverse-transcribed, a reaction mixture was prepared containing up to 5 µg RNA, 1 µl random hexamers (10 mM) and 1 µl dNTP mix (10 mM) in a final volume of ten µl. The mixtures were incubated at 65° for five minutes, then on ice for five minutes. A cDNA Synthesis Mix was prepared that contained 2 µl 10X RT buffer, 4 µl $MgCl_2$ (25 mM), 2 µl DTT (0.1M), 1 µl RNaseOUT™ (40 U/µl) and 1 µl SuperScript® III reverse transcriptase (200 U/µl) for each reaction to be carried

out. Components were added to the cDNA Synthesis Mix in the indicated order. Ten µl of the cDNA Synthesis Mix was added to each RNA and primer mixture and mixed by brief centrifugation. The reaction mixtures were incubated for 10 minutes at 25°C, then for 50 minutes at 50°, and finally the reaction was terminated by incubating for five minutes at 85°C. The reactions were cooled on ice, and then stored at -20°C until required.

2.4.3 Primer design

Primers for qRT-PCR were designed using Primer3 software (Rozen and Skaletsky, 2000) with the following settings: product size between 60 and 120 nucleotides, GC content between 50 and 60%, optimum primer length of 20 bp, optimum primer melting temperature of 60°C, maximum 3' end self-complementarity score of one, and maximum allowed mononucleotide repeat length of 3. These settings were derived from the built-in qPCR settings of Primer3Plus, the web-based implementation of Primer3 (Untergasser *et al.*, 2007). For each gene to be tested, ten possible primer pairs were designed and, from these, a set of primer pairs were selected such that the predicted annealing temperatures of primers for all the tested genes were within 0.5°C of the annealing temperature of the primers for the reference gene (Table 2.2). Primer specificity was checked by comparing each primer sequence against the R6 genome using BLAST. Selected primers were synthesised by Invitrogen, and checked for specificity and function by carrying out a standard PCR reaction using R6 genomic DNA as a template.

2.4.4 Quantitative real-time PCR

All quantitative real-time PCR reactions contained 12.5 µl IQ™ SYBR® Green 2X supermix (Biorad), 1 µl each of forward and reverse primers (12.5 µM), 1 µl of appropriately diluted cDNA, and 9.5 µl nuclease-free water. To reduce pipetting errors, for each primer

Table 2.2. Primers used to measure gene expression by qRT-PCR

Target	Forward primer sequence	Reverse primer sequence	Amplicon size
Spr1888 (*hexA*)	TGTCTAGTGTGCCACGGATT	CGCTGCGCTAATCAAACTCT	113
Spr1887 (*patA*)	TCTTAGGCGCCCTCCTTACT	ATAGGCTGCGAGGACAACA	120
Spr1886	GGATTGGGAATCGTTTAGGG	AGAATCCAGTCCAGCGAAAG	101
Spr1885 (*patB*)	AGAAATGTGACGCTGGCTCT	TTCTGCTGGAGGTTGGTGT	107
Spr1884 (*guaA*)	GCGCTTCGTCAGAATAAACC	AGTCCTTGCCAGTGACCTTC	106
Spr1882 (*gpi*)	CAAGACGCTTTCACTCTTGG	CTGGTTGGTCAAATGGGTTG	100
Spr1881 (*gltX*)	CAACTGTTCGTTTGGCTGTC	ACCACCGATATTGCCACCT	95
spr1880	GGAGGGAACCTGTGGTGTTT	AGGTGGTTTGTCTCGAAAGG	118
Spr0979 (*rpoB*)	GTGAACAGCGGAATCTCCTT	CTGGTTCTTGGGCAATCTTC	120
16S rRNA	GGACAGAGGTGACAGGTGGT	GAGTGCCCAACTGAATGATG	116

combination used, IQ™ SYBR® Green 2X supermix, primers and water were combined in the proportions given above to make a master mix of sufficient volume for the number of cDNA samples to be tested. For each reaction, 24 µl of the master mix was transferred to a well of a white 96-well PCR plate (BioRad) and 1 µl of cDNA was added. On each PCR plate tested, three replicates of a reaction where the cDNA was replaced with nuclease-free water was included as a no-template control.

Real-time PCR reactions and fluorescence measurements were carried out using a CFX96 Real-Time PCR Detection System (BioRad), using the following conditions: 95°C for three minutes, followed by 40 cycles of 95°C for 10 seconds, 54.5°C for 30 seconds. Fluorescence of each well was measured at the end of each amplification cycle. A melt curve analysis was carried out after completion of the amplification cycles to check for the presence of non-specific amplification products. To do this, samples were heated from 55°C to 95°C in 0.5°C increments every five seconds, with a fluorescence measured following each temperature increase.

To measure the amplification efficiency of the primers and to determine the cDNA concentration to use in subsequent reactions, three biological replicates of cDNA from R6 were serially diluted 1:10 in nuclease-free water to give dilutions ranging from 1:1 to $1{:}10^6$. For each primer combination to be tested, two technical replicate reactions for each dilution of each biological replicate were prepared as described above. Real-time PCR was carried out as described above and the average Cq values for each cDNA dilution for each primer combination were plotted against the $\log_{10}$ of the starting quantity of cDNA using CFX Manager software (BioRad). Primer efficiency (E) was calculated as described by Pfaffl (2001) as the $E = 10^{-1/slope}$, where slope is the slope of the best fit line through the Cq values. The cDNA dilution to be used in subsequent experiments for each primer combination was chosen such that the Cq value obtained using R6 cDNA was close to 30 cycles. This value was chosen to ensure that Cq values from strains with markedly higher or lower levels of gene expression than R6 would be detectable within

the 40 cycles of the experiment.

To measure gene expression, two technical replicate real-time PCR reactions were prepared as described above for each primer combination, biological replicate and strain to be tested. The cDNA dilution used was the optimal dilution determined during the primer efficiency measurement described above. Real-time PCR was carried out as described above. Relative expression of genes of interest in each strain relative to the control strain (R6, unless stated otherwise) was calculated from the Cq values according to the Pfaffl method (Pfaffl, 2001), as follows. First, the mean Cq value of the two technical replicates was calculated. Then, for each primer combination used, the difference in Cq value between the control and the sample strain (ΔCq) was calculated. The relative expression ratio for each target gene was the calculated using the formula $ratio = (E_{target})^{\Delta Cq_{target}}/(E_{ref})^{\Delta Cq_{ref}}$, where E_{target} and E_{ref} are the primer efficiencies and ΔCq_{target} and ΔCq_{ref} are the ΔCq values for the target gene and the reference gene, respectively.

For visualisation of endpoint concentration of PCR amplimers by gel electrophoresis, PCR was carried out using the same cDNA and primer combinations as described above, but substituting Reddymix (Abgene, U.K.) for the IQ™ SYBR® Green 2X supermix. Each PCR reaction consisted of 22.5 µl Reddymix, 0.5 µl each of forward and reverse primers (25 µM) 0.5 µl of cDNA and 1 µl of nuclease-free water. Forty cycles of PCR were carried out in a standard thermal cycling PCR machine using the same conditions as above, but without fluorescence measurements. Five µl of each reaction was loaded into the wells of a 1% agarose gel, and PCR amplimers were separated by electrophoresis at 100 V for one hour. PCR amplimers were visualised using the Gene Genius image analyser.

2.5 Whole-genome sequencing and data analysis

2.5.1 Illumina sequencing and quality control

Extraction of genomic DNA for whole genome sequencing was carried out as described above. Whole genome sequencing was performed by GenePool (Edinburgh, U.K.) on two occasions. Genepool carried out Illumina sequencing library preparation using standard Illumina procedures. On the first occasion (2009), 50 cycles of paired-end sequencing were carried out on Illumina Solexa Genome Analyser IIx sequencing system, for two *S. pneumoniae* strains multiplexed in a single lane with eight other samples. On the second occasion (2012), 100 bp paired-end reads were generated using an Illumina HiSeq system for nine pneumococcal genomes multiplexed in a single lane. De-multiplexing of samples, conversion from Illumina image files to Fastq format, and preliminary quality control was carried out by Genepool.

The Fastq files obtained from Genepool containing the read sequences were checked for overall quality using FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/; Babraham Institute, Cambridge, UK), a tool that performs various quality control checks on raw high-throughput sequencing data.

2.5.2 Mapping reads against reference genome

Illumina sequencing reads were aligned against the R6 reference genome (Genbank accession: NC_003098) and single nucleotide polymorphisms and small indels detected by two methods.

Initial analysis of sequencing carried out in 2009 was carried out using the My xBase next generation sequencing pipeline (ng.xbase.ac.uk/my, Centre for Systems Biology,

University of Birmingham, U.K.). This pipeline uses the Maq alignment program to align reads against the reference sequence and to identify single nucleotide polymorphisms (SNPs) and small insertions and deletions (indels) (Li *et al.*, 2008). The pipeline was run in paired-end mode, using an estimated paired-end insert length of 250 bp. The output from this process is a table of mutations, which was then filtered by quality score (> 100) and depth of reads covering a putative mutation (> 50).

For later analyses, including reanalysis of the sequencing data from 2009 and analysis of new sequencing data generated in 2012, alignments and mutation detection were carried out locally as follows. If required, the base quality scores in the Fastq files were converted from phred64 (Illumina 1.5) encoding to phred33 (Sanger) encoding using a Python script written by Nick Loman. Alignment of reads was carried out using Bowtie2 version 2.0.0-beta5 (Langmead and Salzberg, 2012) using the `-very-sensitive-local` pre-set parameter set (equates to settings `-D 20 -R 3 -N 0 -L 20 -i S,1,0.50 -local`). The reads were aligned against an indexed version of the R6 (NC_003098) genome, generated using Bowtie2-build. Sam format alignments generated by Bowtie2 was piped to Samtools version 0.1.18 (Li and Durbin, 2009) and sorted using the Samtools sort command (command used: `bowtie2 -t -very-sensitive-local -x <R6 reference> -1 <reads1.fastq> -2 <reads2.fastq> | samtools view -bS - | samtools sort - <out.bam>`). Resulting Bam format alignments were then indexed using `samtools index <file>`. To detect mutations relative to the reference genome, a consensus pileup was generated from each sorted and indexed Bam file using Samtools mpileup against the published R6 genome indexed with Samtools faidx. The output of this command was piped to BCFtools version 0.1.17-dev to detect all putative SNPs and small indels (commmand used: `samtools mpileup -uf <faidx referenced R6 genome> <bam file> | bcftools view -bvcg - > <output file>`). The resulting set of variations was filtered based on desired quality score and/or read depth covering a variation using a custom Python script.

Read alignments and positions of detected variations were visualised by loading indexed Bam and Bcf files respectively into Artemis version 14.0.0 (Carver *et al.*, 2012).

To calculate the normalised depths of reads mapping to each gene in the genome, reads per kilobase per million reads mapped (RPKM) was obtained for each gene using Artemis. For each read alignment to be tested, the respective Bam file was loaded into Artemis against the R6 genome. All CDS features in the genome were selected, and RPKM values calculated using the "Analyse RPKM value of selected features" option in the Bamview window. RPKM values calculated from strains M184, M500, M501 and M503 were plotted against those from R6 using R.

For strains generated by transformation of R6 recipient cells with genomic DNA from a different genetic background, clusters of SNPs representing putative recombination events were defined using the following procedure. The number of SNPs in a 5 kb, centre-aligned sliding window was calculated for every position in the R6 reference genome using the rollsum function in the R package 'zoo' (**?**). Clusters were then defined as regions where this rolling sum of SNPs was consistently greater than two.

2.5.3 Local and global *de novo* assembly

The Velvet sequence assembler (version 1.2.03; Zerbino and Birney, 2008) was used to generate draft *de novo* assemblies of both whole genomes, or parts thereof.

To prepare sequencing reads for whole genome assembly, the paired end reads were combined into a single file where the first and second reads were interleaved using the shuffleSequences_fastq Perl script included with the Velvet distribution. Reads containing undefined bases ('N') and reads with low average quality (<20) were removed using the eliminate_n_paired.py and filter_reads.py Python scripts written by Nick Loman (github.com/nickloman/xbase). Where necessary, reads were trimmed to 75 bp using the trim_reads.py script written by Nick Loman (github.com/nickloman/xbase).

The prepared reads were then assembled with Velvet, using the VelvetOptimiser script written by Simon Gladman (included with the Velvet distribution) to determine optimal parameter settings. This script automates the assembly process as follows. First, velveth is used to generate hash tables from the prepared fastq file for all kmer values in a given range. Then draft assemblies are generated using velvetg for each velveth output using default settings. The assembly with the largest length-weighted median contig length (N50) is chosen and several iterations of velvetg with differing values for the expected-coverage and coverage-cutoff parameters are carried out to improve the assembly.

To generate local assemblies, reads meeting the desired criteria were extracted from the read alignment using Samtools and assembled as above.

To generate an assembly from M184 reads that did not align to the R6 reference genome, unaligned reads were extracted from the M184 read alignment using Samtools, and assembled as above.

2.6 Further characterisation of mutations identified by genome sequencing

2.6.1 Primer design and PCR

Sequences of primers used in this study are shown in tables 2.3 and 2.4. Unless specified otherwise, primers were designed using Primer3Plus, the web interface to the Primer3 program (Untergasser *et al.*, 2007), and synthesised by Invitrogen. Lyophilised primers were reconstituted in sterile nuclease-free water (Gibco) to a concentration of 100 µM and stored at -20°C. For use in PCR reactions, 25 µM working stocks were prepared and also stored at -20°C.

Unless otherwise specified, PCR reaction mixtures consisted of 22.5 µl of 1.1x ReddyMix

Table 2.3. PCR primers used in this study for confirmation of mutations in M184

Target	Forward primer		Reverse primer		Amplicon size
	Number	Sequence 5' to 3'	Number		
Intergenic: spr0262-261	1201	TCCAAGGACGTATGGTACTCG	1202	CGTGGCTCGCAATAATGA	249
Spr0267 (*sulB*)	179	TCAGAAGCGAGGCTAGAAGC	1180	TGGGCTGACAGACATTGGTA	202
Spr0304 (*pbpX*)	1652	CGTGCAGAACGTTTGAAAGA	1653	AGCTTTTGCCTGATGCTGTT	2476
Spr0453 (*hrcA* internal region)	902	GACGGAGATTCCGCAGATTA	903	TTTTGCATCTGATCCTCACG	222
Spr0453 (*hrcA* whole gene)	910	GGTCTTCAGATTGATTTTTAGCA	913	CCATGTTCGCCTCCTTTAAT	597
Intergenic: spr0561-562	1363	CGGTCCTGTTTTTGAGGAGA	1364	GAACGATAAACAAGGGCAAAA	195
Spr0627 (*lctO* mutation 1)	1181	AATGCCTGGCAGCAAAAA	1182	GCCAAATGCTGCTTTAGGAA	296
Spr0627 (*lctO* mutation 2)	1183	GCTGCAGGAGCTTCTGGTAT	1184	TCAAAGACTTGACGCACACC	257
Spr0807	1365	AAACAATATCGCTCATTATTTGGA	1366	GTTTTCTAAGACGGCGCAAT	240
Spr0989 (*murZ*)	1199	AGCCCGTGAACCTGAGATTA	1200	CCCTTCCAGGTGTTCGTAAA	229
Spr1141 (*topA*)	1185	TGTATCTGCTGCCCAAACAG	1186	TTTTTCCAATTTCCGTCTGC	204
Spr1165 (*pyrP*)	1367	GCTGCCAATGGTAGTCAGGT	1368	TCTCCCAACTCTGTCGGTTC	296
Spr1343 (*tufA*)	1187	ACGTCACTACGCTCACATCG	1188	CAAGCAATTCTTCGTCGTCA	221
spr1582	1189	GCTCGACAAGGTTCCTTCAA	1190	CTTCTTCTTTGCGGCGTTT	221
spr1591	1197	GGTCGGAAACCCAACTTTCT	1198	CATCTACACCAAAGCGCTGA	210
Spr1620	1369	GAAGGAAAAGCTTGGCAAAA	1370	TGCATAAGCTGCCGTATTTG	554
Spr1699 (*treP*)	1191	CCCTTTTTATACCAAACAAGTTGC	1192	TCCTAAAACAAACCGCATCC	203
spr1875	1195	TGTTGAGCAAATCCAAAACG	1196	CTGCTGTCGCAGTTGTCACT	266
Spr1881 (*gltX*)	1373	AGAAAGCAAAAAGGCAGACG	1374	CATCTTTTCCCAGATGTCCA	1535
Spr1945 (*pcpA*)	1437	GCTGGTCAAGATGGAAAACC	1438	TTCTCCAGCTGACTGCCTTT	239

Table 2.4. Remaining primers used in this study

Purpose	Number	Sequence	Number	Sequence
Check PCR for inserts in pBAV1K-gfp82	982	GCTAGTTGAACGCTTCCATC	1797	TAGTATCGACGGAGCCGATT
Amplification of $tRNA^{Glu}$ from M184 for cloning	1892	TAAGAATTCTTTAATGAGCG-ATAGAAGAGTCAG	1893	TAATCTAGAAGAAGTCTCCTTG-ATACAGCTAGATT
Duplication junction in M184	1772	GATAACGCGGTTGCAGAAGT	1796	CTTATTGGTGGGGAGAAGAA
Amplification of whole duplicated region from M184	1771	GATAGGGCAGAAGAGCATCC	1772	GATAACGCGGTTGCAGAAGT
Amplification of synthesised GeneArt sequences from pMA-T	1850	CGCGACGTAATACGACTCAC	1851	CCAGTCTTGTGCTCCAGGTA
patA-hexA intergenic sequence	1771	GATAGGGCAGAAGAGCATCC	1772	GATAACGCGGTTGCAGAAGT
patA whole gene	604	TCTTGCTCAGTCCATCATCGAA-TATA	605	CCGCTGTGGATTAGTTCATTTCC
patB upstream sequence	606	AGAATCCAGTCCAGCGAAAGCT	607	GAAAGAACGACCAGATGTTCCAAT
patA internal region	608	ATGTTGTCCTCGCAGCCTAT	609	ACGAACCGATGAACAAGAGG
patB internal region	610	TTGCTGGTTCGGCTGTACTT	612	AACTGCTGTCATCTGGCCTT
patA internal region	1357	CAGCATCGGTTCCTTGTC	1358	CAGATGAAGAGTTGGTTGGA
Primers binding upstream of rRNA loci 1 and 2	rRNA-1	CTAGAACGCGTCTTAAGCAA	rRNA-2	TCGTGATTTGGTACGCTTAG
Primers binding upstream of rRNA loci 3 and 4	rRNA-3	CTAGAACGCGTCTTAAGCAA	rRNA-4	CTTATTGGTGGGGAGAAGAA

PCR Master Mix (Thermo Scientific), 0.5 µl genomic DNA, 0.5 µl of each of the forward and reverse primers, and 1 µl of SDW to give a total reaction volume of 25 µl. In contamination controls, the genomic DNA was replaced with 1 µl SDW. PCR reactions were carried out in a thermal cycler. The standard reaction conditions consisted of the following stages: an initial denaturation (5 minutes at 95°C), 30 cycles of thermal cycling (95°C for 30 seconds, 51°C for 30 seconds, 72°C for two minutes), then a final extension (72°C for 10 minutes). Resulting PCR products were separated by electrophoresis on a 1% agarose gel, and visualised using the Genius image analyser (Syngene, Cambridge, U.K.).

2.6.2 PCR purification and sequencing

PCR purification was performed using a QIAquick® PCR purification kit (Qiagen) following the manufacturer's instructions. Purified PCR products were separated by electrophoresis on a 1% agarose gel, and visualised using the Genius image analyser (Syngene, Cambridge, U.K.). Sequencing of purified PCR amplimers was carried out using the appropriate PCR primers by the Functional Genomics department of the University of Birmingham, unless otherwise specified.

2.6.3 Protein structure prediction using SWISS-MODEL

The structure of the HrcA protein was predicted using the SWISS-MODEL automated protein structure homology-modelling server (Arnold *et al.*, 2006). Templates for modelling were found using a PSI-BLAST search of the spr0453 gene against the PDB database. A sequence-structure alignment between the target protein and template protein was produced using the FUGUE sequence-structure homology recognition program (Shi *et al.*, 2001) and the model was produced by the SWISS-MODEL server using the alignment mode.

2.6.4 Prediction of promoters and terminators

Putative promoter sequences were predicted using the prokarotic promoter prediction (PPP) webserver (http://bioinformatics.biol.rug.nl/websoftware/ppp). This program scans DNA sequences for Lactococcal-like Sigma A binding sites, using Hidden Markov models constructed from alignments of known Sigma A binding sites (Zomer *et al.*, 2007).

For analysis of the position of predicted Rho-independent terminators in the R6 genome sequence, a set of pre-computed Rho-independent terminator predictions were used that had been predicted by TransTermHP (Kingsford *et al.*, 2007). These were downloaded from http://transterm.cbcb.umd.edu/index.php on 31/12/2012.

2.6.5 PCR and sequencing of short interspersed repeats

Genomic positions of short interspersed repeats in the R6 genome were obtained from predictions made by Croucher *et al* (Croucher *et al.*, 2011b, supplementary data), and visualised in Artemis. Repeats to be checked for mutations were chosen by visual inspection of the M184 read alignment against R6, according to the following criteria: a) repeats were located between divergently transcribed genes or genes transcribed in the same direction, and b) the depth of reads aligned to a repeat sequence was at least twice that of the flanking sequence. Primers were designed to amplify each repeat region using Primer3 as described above (Appendix 2). PCR was carried out as described above using an annealing temperature of 51°C and an extension time of two minutes. PCR amplimers were visualised by agarose gel electrophoresis and purified as described above. Sanger sequencing of amplimers was carried out by Source Biosciences (Nottingham, UK).

2.7 Measuring segregation rate of *patA*-spr1880 duplication

To prepare starter cultures of M184 and the R6^{M184} transformants where 100% of the cells should contain the duplication, cultures were grown overnight on agar plates containing 8 µg/ml ethidium bromide. Cells were then scraped off the plates using sterile swabs, resuspended directly in 1 ml BHI broth containing 10% glycerol and frozen for future use.

To measure loss of the *patA*-spr1880 duplication over time, ethidium bromide-selected stocks of each culture were diluted into ten ml fresh BHI broth and grown for 12-14 hours. Growth was monitored by measurement of OD_{660} and cultures were diluted 1:1000 in fresh BHI broth when they approached an OD_{660} of 0.6, which corresponds to late logarithmic phase. At the start and end of the growth period 20 µl samples were taken and serially diluted in 180 µl PBS to a final dilution of $1{:}10^8$. Fifty µl samples of each dilution were spotted onto plain agar plates and incubated overnight to obtain single colonies. Approximately 60 colonies per strain were picked using a sterile inoculation loop and streaked in duplicate onto one plate containing 8µg/ml norfloxacin and one plain agar plate. Following overnight incubation, the percentage of colonies in the sample that retained norfloxacin resistance (P_{ret}) was calculated by comparing the number of colonies that grew on the selective plates compared to the control plates. For each value of P_{ret}, a 95% confidence interval was determined. As P_{ret} is a ratio of probabilities, and is therefore not normally distributed, the confidence interval (CI) was calculated for $log(P_{ret})$ using the formula $CI = \pm z_{\alpha/2}\sqrt{Var(log(P_{ret})}$. $Var(log(P_{ret})$ was estimated using the delta method, which is defined as $Var(log(\hat{p_1}/\hat{p_2})) = (1 - \hat{p_1})/\hat{p_1}n_1 + (1 - \hat{p_2})/\hat{p_2}n_2$. The final confidence interval for P_{ret} was then derived by taking the exponential of the upper and lower bounds of the CI for $Var(log(P_{ret}))$.

Average probabilities of loss of the *patA*-spr1880 duplication per cell per generation, and their associated confidence intervals, were calculated essentially as described by Brochet *et al.* (2008). If each cell in a population has a probability p of losing the duplication in each generation, then P_{ret} after n generations will be given by the formula $P_{ret} = (1 - p)^n$. The half-life of the duplication within the population was calculated by finding the number of generations (n) required for P_{ret} to reduce to 0.5, which was then converted to a time in days by multiplying by the generation time of the strain.

2.8 Inactivation of *patA* in M184

Inactivation of *patA* in M184 was carried out by transforming M184 recipient cells with a PCR amplimer corresponding to *patA* containing a *magellan2* mini-transposon inserted between bases 833 and 834. This was derived from the previously described strain M246 (R6 *patA::magellan2* Garvey and Piddock, 2008). To avoid including *patA* upstream sequence, an internal region of *patA*, which included the transposon insertion, was amplified from M246 by PCR using primers 608 and 1357 (Table 2.4).

M184 recipient cells were grown to an OD_{660} of 0.4 in BHI broth, then diluted 1:20 in competence medium (section 2.3.1) containing 100 ng/ml CSP. Purified M246 *patA* amplimer was serially diluted 1:10 six times in nuclease-free water. Twenty microlitres of each dilution was added to 500 µl M184 competent cells in a microcentrifuge tube, and these transformation reactions were incubated for three hours at 37°C. A transformation where nuclease-free water was added instead of DNA was prepared and incubated in parallel. Following incubation, 100 µl of each transformation reaction was spread onto Columbia blood agar plates containing 100 µg/ml spectinomycin, and incubated overnight. Colonies for further investigation were selected from the plates corresponding to the lowest concentrations of donor DNA that resulted in spectinomycin resistant colonies.

2.9 Measuring fluorescence from wild-type and mutated *patA* promoters by transcriptional fusion with promoterless *gfp*.

2.9.1 Construction of plasmid pBAV1K-*gfp82*

Plasmid pBAV1K-T5-*gfp* (Bryksin and Matsumura, 2010) was kindly provided by Andrea Mitchell (University of Birmingham). This plasmid contains a *gfp* gene controlled by a T5 promoter. This results in continuous expression of GFP, allowing this plasmid to be used as a fluorescent marker to study cellular localisation (Bryksin and Matsumura, 2010). To generate a plasmid that can be used to monitor activity of cloned promoters, this constitutive promoter needed to be removed. It was also desirable for the plasmid to encode an unstable variant of GFP, such that accumulation of GFP does not obscure subtle variations in expression from the promoter.

To fulfil these requirements, the T5 promoter and *gfp* gene from pBAV1K-T5-*gfp* were replaced by a *gfp* gene from plasmid pMW82, which has previously been used to measure inducible activity of promoters in *Salmonella* (Lawler *et al.*, 2013). The pBAV1K-T5-*gfp* plasmid was propagated by transformation into One Shot® *E. coli* TOP10 cells (Life Technologies, cat. no. C4040) by heat shock according to the manufacturer's instructions. The pMW82 and pBAV1K-T5-*gfp* plasmids were digested with restriction enzymes *EcoRI* and *PstI* (New England Biolabs) as follows. 30 µl of each purified plasmid was added to 5 µl of the supplied restriction enzyme buffer, 1 µl BSA (10 mg/ml), 4 µl nuclease-free water, 5 µl *EcoRI* and 5 µl *PstI*. This reaction was incubated at 37°C for one hour, then terminated by heating to 80°C for 10 minutes. Digested fragments were separated by electrophoresis on a 1% agarose gel, and bands corresponding to the 2800 bp plasmid backbone of pBAV1K-T5-*gfp* and the 824 bp *gfp* gene from pMW82 were

cut out of the gel and extracted using a QIAgen® gel extraction kit, according to the manufacturer's instructions. The plasmid backbone and the *gfp* gene were ligated using QuickStick® ligase (Bioline, cat. no. 27027), as follows. The purified fragments were combined in a 1:10 vector:insert ratio, and mixed with 5 µl 4x QuickStick ligase and 1 µl QuickStick ligase in a total reaction volume of 20 µl. The reaction was mixed by pipetting, and incubated at room temperature for 5 minutes. *E. coli* TOP10 cells were transformed with 2 µl of this ligation reaction, and cells containing the ligated plasmid were selected on LB agar containing 50 µg/ml kanamycin. Plasmids were extracted from transformants using the QIAprep® Miniprep kit, and screened for correct insertion of the pMW82 *gfp* gene by PCR using primers 982 and 1797, and sequencing of the resulting PCR amplimer.

2.9.2 Construction and cloning of promoter fragments

Wild-type and mutated DNA fragments corresponding to the region 12 to 144 bp upstream of *patA* were cloned into an *EcoRI-XbaI* cloning site upstream of the *gfp* gene in pBAV1K-*gfp82* as follows. Wild-type and mutated *patA* upstream regions, with added 5' *EcoRI* and 3' *XbaI* sites, were synthesised by GeneArt (Life Technologies). These fragments were supplied by GeneArt as inserts in pMA-T plasmids. The *patA* promoter fragments were amplified from pMA-T plasmids using primers 1850 and 1851. The resulting PCR amplimers were purified, and digested using *EcoRI* and *XbaI* as described above. *XbaI* digestion is blocked by DNA methylation by the Dam methylase, so to remove this methylation from the pBAV1K-*gfp82* vector, purified plasmid DNA was transformed into *E. coli* JJA cells (kindly provided by Yanina Sevastyanovich, University of Birmingham), which are *dam*$^{-}$. Plasmid pBAV1K-*gfp82* was extracted from JJA cells following overnight growth, and digested with *EcoRI* and *XbaI*. Digested plasmid and insert were combined in a 1:100 molar ratio, and ligated using QuickStick ligase as de-

scribed above. Two microlitres of ligation mixture was used to transform *E. coli* TOP10 cells, and transformants containing ligated plasmid were selected on LB agar containing 50 µg/ml kanamycin. Colonies were screened for plasmid containing the correct insert by PCR using primers 982 and 1797.

To add the duplicated tRNA^{Glu} gene from M184 upstream of the *patA* promoter in plasmid pBAV1K-*gfp82*-p^{patA}, to give plasmid pBAV1K-*gfp82*-tRNA-p^{patA}, the tRNA gene from M184 was amplified by PCR using the 3' primer 1892, and the 5' primer 1893 which incorporates an *EcoRI* site. A naturally occurring *SpeI* site exists upstream of the *patA* promoter sequence. The tRNA PCR amplimer and pBAV1K-*gfp82*-p^{patA} were therefore digested with *EcoRI* and *SpeI* (New England Biolabs) as described above. Digested plasmid and insert fragments were purified with a QIAquick PCR purification kit, mixed in a 100:1 insert:vector molar ratio, and ligated using QuickStick ligase as described above. *E. coli* TOP10 cells were transformed with 2 µl of ligation reaction, and transformants were selected on LB agar containing 50 µg/ml kanamycin. Colonies were screened for plasmid containing the inserted tRNA by PCR using primers 982 and 1797.

The pBAV1K-*gfp82* constructs were extracted from *E. coli* cells using a QIAprep® Miniprep kit, and transformed into *S. pneumoniae* R6 as follows. R6 recipient cells were grown to an OD_{660} of 0.4 in BHI broth, then diluted 1:20 in competence medium (section 2.3.1) supplemented with 100 ng/ml CSP. Twenty microlitres of each pBAV1K-*gfp82* contruct was added to 500 µl competent R6 cells, and the transformation mixtures were incubated for three hours at 37°C. R6 cells containing the desired plasmid were selected on Columbia blood agar plates containing 200 µg/ml kanamycin.

2.9.3 Fluorescence assays

To take single fluorescence readings from mid-logarithmic phase cultures that had been grown under normal laboratory conditions, the following method was used. Three biological replicates of each strain were grown statically to an OD_{660} of 0.5 in BHI broth (37°C, 7.5% CO_2). Cells from 1 ml aliquots of each culture were pelleted by centrifugation (14,000 x g, 3 min) and resuspended 200 µl PBS. Each resuspended culture was added to a well of a black microtitre tray, and fluorescence was measured using a Fluostar Optima using excitation and emission wavelengths of 492 and 520 nm respectively and a gain of 1452. Significant differences in fluorescence between strains ($p<0.5$) were determined using a two-tailed Student's t-test on at least three biological replicates. A Bonferroni correction for multiple testing was applied if more than five comparisons were performed simultaneously.

To study GFP expression over the whole growth cycle, fluorescence and growth measurements were taken simultaneously from cultures growing in BHI broth. Three biological replicates of each culture to be tested were grown to stationary phase (OD_{660} of 0.6-0.8) in BHI broth supplemented with 100µg/ml kanamycin. Cultures were then diluted 1:1000 in fresh BHI broth containing kanamycin and 200 µl of each culture was added to two wells of a black microtitre tray with a clear base. The plate was incubated in the Fluostar Optima at 37°C for twelve hours, and the script mode of the Optima software was used to take fluorescence and absorbance readings simultaneously every three minutes. Fluorescence was measured using the top optic with excitation and emission wavelengths of 492 and 520 nm respectively, and a gain of 1452. Absorbance was measured at 600 nm using the bottom optic. The plate was shaken for 5 seconds prior to each reading to evenly distribute the cells.

3 Over-expression of patA and patB is observed in clinical isolates

3.1 Background

Prior to this study it was well established that a subset of pneumococcal clinical isolates with fluoroquinolone resistance showed a phenotype indicative of increased efflux (Baranova and Neyfakh, 1997; Brenwald *et al.*, 1997, 1998; Piddock *et al.*, 1997, 2002). However, it had been shown that production of PmrA, the first fluoroquinolone efflux pump to be discovered in *S. pneumoniae*, was not strongly associated with efflux mediated fluoroquinolone resistance in clinical isolates (Piddock *et al.*, 2002).

At the start of this study in 2009 it had been shown that, in laboratory mutants of *S. pneumoniae*, increased expression of *patA* and *patB* resulted in decreased susceptibility to the fluoroquinolones ciprofloxacin and norfloxacin and the dyes ethidium bromide and acriflavine (Garvey and Piddock, 2008; Marrer *et al.*, 2006b). This resistance was reversed by addition of the efflux inhibitors reserpine and sodium orthovanadate, and the overall phenotype was similar to that observed in clinical isolates with increased efflux (Garvey and Piddock, 2008). This observation led to the suggestion that PatAB could be the predominant transporter conferring efflux-mediated fluoroquinolone resistance in pneumococcal clinical isolates. However, a clear link between increased expression of *patAB* and efflux-mediated fluoroquinolone resistance in clinical isolates had not been

definitively shown.

Also, the effect of *patAB* over-expression had not been investigated in *S. pneumoniae* isolates carrying mutations in the topisomerase genes. Topoisomerase mutations cause high-level resistance to fluoroquinolones, particularly when both *parC* and *gyrA* are mutated (Janoir *et al.*, 1996). In contrast, the increase in resistance caused by over-expression of *patAB* is much smaller (Garvey and Piddock, 2008; Marrer *et al.*, 2006b). Therefore, it was unclear whether PatAB-mediated fluoroquinolone efflux would be important in a clinical setting. Efflux pumps in several other organisms have been implicated in intrinsic resistance to antibiotics, such that their inactivation reduced resistance even in the presence of other resistance mechanisms (Baucheron *et al.*, 2002; Luo *et al.*, 2003; Markham and Neyfakh, 1996). However, it is also possible that possession of topoisomerase mutations could supercede any effect of *patAB* over-expression, rendering it irrelevant in the clinical environment.

To answer these questions, an analysis was carried out on unpublished data that had been previously collected by Mark Garvey and Ryan Wong for a set of 57 clinical isolates of *S. pneumoniae*. Of these, 46 had been included in a previous study that analysed the contribution of *pmrA* expression to fluoroquinolone resistance, during which fluoroquinolone MICs and sequences of the topoisomerase genes had been determined (Piddock *et al.*, 2002).

The dataset consisted of:

- MICs of ciprofloxacin, norfloxacin, ethidium bromide and acriflavine for all 57 isolates in the presence or absence of 20 μg/ml reserpine or 50 μM sodium orthovanadate.
- Expression levels of *patA* and *patB* for all isolates. These were determined by comparative reverse transcription-PCR, followed by quantification of amplimers by denaturing high-pressure liquid chromatography (Garvey and Piddock, 2008).

- Levels of intracellular accumulation of the fluorescent dye Hoechst 33342 for 28 representative isolates, determined by measurement of fluorescence as described previously (Garvey and Piddock, 2008).
- Sequences of the QRDRs of *parC*, *parE*, *gyrA* and *gyrB* for 55 isolates (Piddock *et al.*, 2002)
- MICs of ciprofloxacin, ethidium bromide and acriflavine, in the presence or absence of reserpine or sodium orthovanadate, for 21 isolates following inactivation of *patA* and *patB* by *in vitro Mariner* mutagenesis

The sequences of *patA* and *patB* had been determined for 21 of the clinical isolates, of which 13 over-expressed *patAB* and eight did not. Comparison of these sequences had not revealed any mutations in either the transporters themselves or the promoter regions that were associated with the isolates over-expressing *patA* and *patB* (Garvey and Wong, unpublished data).

3.2 Aims and Hypotheses

Based on the phenotypes of laboratory mutants over-expressing *patA* and *patB* (Garvey and Piddock, 2008; Marrer *et al.*, 2006b), it was hypothesised that clinical isolates with cross-resistance to both fluoroquinolones and dyes would show significantly higher expression of *patA* and *patB*. It was also predicted that this over-expression would result in decreased accumulation of the fluorescent dye Hoechst 33342.

Furthermore, it was hypothesised that PatAB contributes to intrinsic resistance of *S. pneumoniae* to fluoroquinolone antibiotics, meaning that inactivation of *patA* or *patB* should affect fluoroquinolone resistance in clinical isolates, despite the coexistence of topoisomerase mutations.

Therefore, the aim of this section of the study was to analyse the data collected from the clinical isolates to:

- Determine whether there is a significant association between expression of *patA* and *patB* and resistance to fluoroquinolones and dyes in these isolates
- Determine whether over-expression of *patAB* in these isolates is associated with decreased accumulation of Hoechst 33342
- Investigate the effect of chemical or genetic inactivation of the PatAB efflux pump on fluoroquinolone and dye resistance in this set of isolates

3.3 Grouping of isolates based on antibiotic resistance profile

Isolates were sorted into four groups based on their resistance to the fluoroquinolones ciprofloxacin and norfloxacin and the dyes ethidium bromide and acriflavine. Ciprofloxacin resistance was defined as a ciprofloxacin MIC of more than 2 µg/ml, and norfloxacin resistance as and MIC of more than 8 µg/ml. Ethidium bromide and acriflavine resistance were defined as MICs of more than 2 µg/ml. Isolates resistant to at least one fluoroquinolone and both dyes were defined as fluoroquinolone and dye resistant (FQDR). Isolates susceptible to dyes but resistant to one fluoroquinolone were designated as fluoroquinolone resistant (FQR), while isolates susceptible to fluoroquinolones but resistant to both dyes were defined as dye resistant (DR). Susceptible (S) isolates were defined as those that were susceptible to all four agents. Overall, 15 isolates were FQDR, 24 were FQR, 2 were DR and the remaining 16 isolates were S.

3.4 Analysis of *patA* and *patB* expression levels for each group

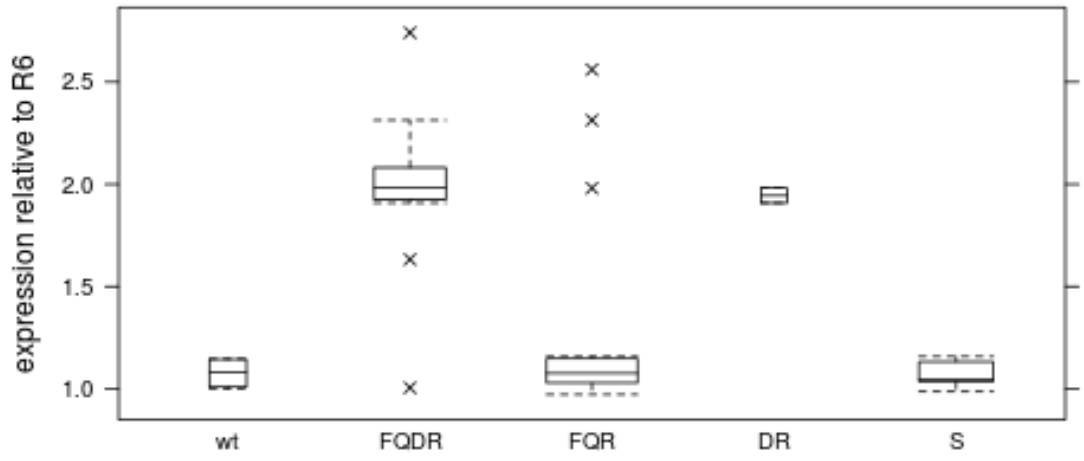
Expression of *patA* and *patB* in each of the clinical isolates was compared to that of the wild type strain R6 (Figure 3.1). In the group of 15 FQDR isolates, 14 isolates over-expressed *patA* and *patB* compared to R6 (>1.5 times the R6 expression level), while only 3 of the 24 FQR isolates did so. Both the DR strains also over-expressed *patA* and *patB*, while no over-expression was observed in any of the S isolates.

There was a significant association between resistance to both fluoroquinolones and dyes and high expression of *patAB* ($p < 0.001$, Fisher's exact test).

3.5 Effect of chemical and genetic inactivation of *patA* and *patB*

To determine the contribution of efflux mechanisms to the resistance phenotypes observed, the effect of efflux inhibitors on the MICs of ciprofloxacin and ethidium bromide was analysed for each isolate (Figure 3.2). Addition of reserpine (20 $\mu g/ml$) reduced the MICs of ethidium bromide by more than two dilutions for all 15 FQDR isolates (100%), both DR isolates (100%), 21 out of 24 FQR isolates (88%) and 16 out of 17 S isolates (94%). Ethidium bromide MICs for the remaining four isolates were reduced by one dilution. Identical results were observed for FQDR and DR isolates when sodium orthovanadate (50μM) was used, and for S isolates the proportion of isolates with MICs reduced by two or more dilutions was slightly reduced to 88% (15 out of 17). A greater difference between reserpine and sodium orthovanadate was observed in FQR isolates, where MICs were reduced by two or more dilutions in 63% of isolates by sodium

A



B

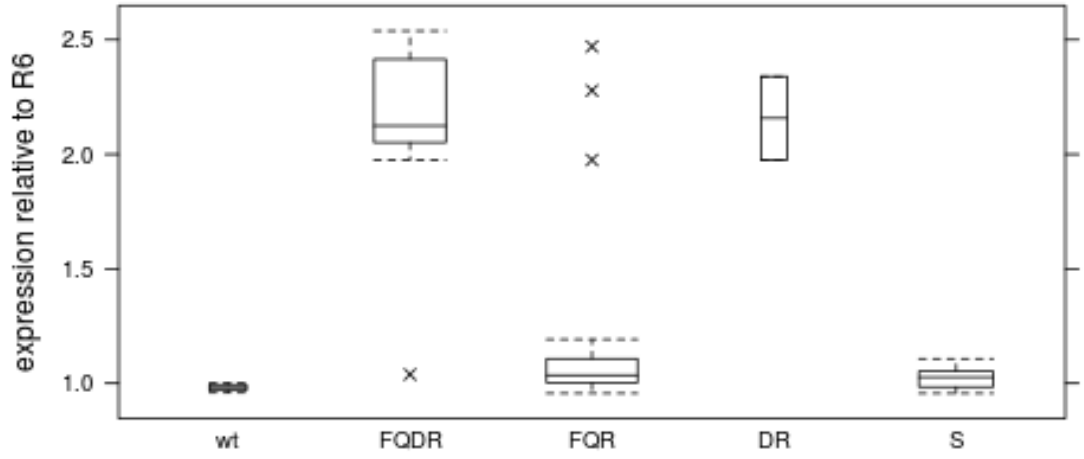


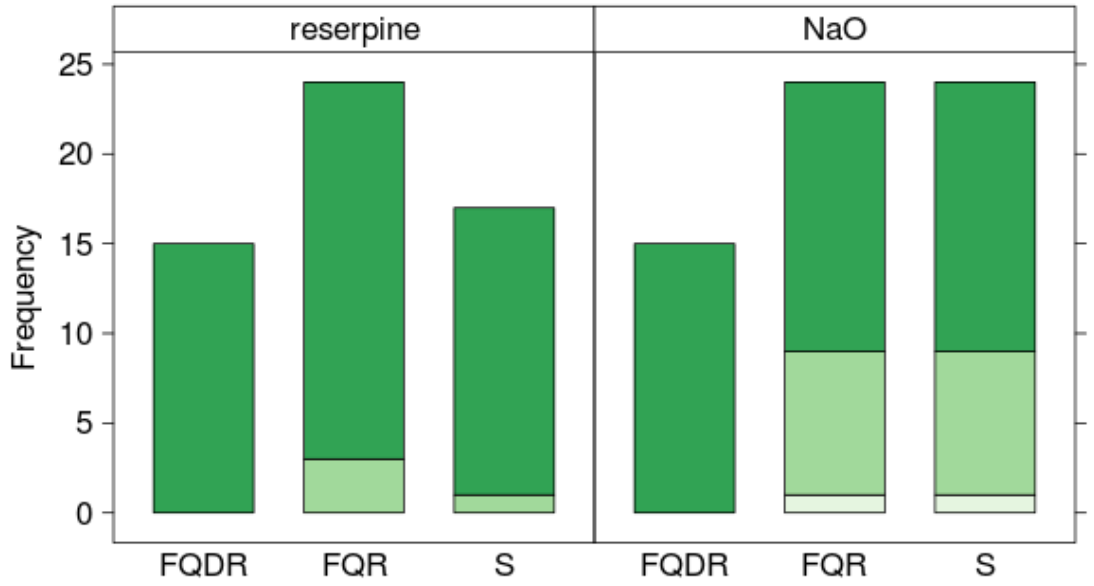
Figure 3.1. Expression of A *patA* and B *patB* in clinical isolates grouped by resistance profile. Boxes represent interquartile range (IQR) of expression values, and whiskers represent 1.5 IQR. Box widths are scaled by number of isolates in each group

orthovanadate.

The effect of efflux inhibitors on the MICs of ciprofloxacin was more variable (Figure 3.2B). In the presence of reserpine, ciprofloxacin MICs were reduced by two or more dilutions for eight FQDR isolates (53%), and MICs were reduced by one dilution for the remaining seven isolates. In the group of FQR isolates, three isolates (13%) had MICs reduced by two or more dilutions, 17 had MICs reduced by one dilution (71%) and MICs of the remaining four isolates were unaffected by reserpine. None of the S isolates had ciprofloxacin MICs reduced by two or more dilutions by reserpine, but six out of the 17 isolates (35%) had MICs reduced by one dilution. In the presence of sodium orthovanadate, ciprofloxacin MICs for FQDR isolates were again reduced by two or more dilutions for eight isolates, but only four isolates (23%) showed a reduction of one dilution, compared to 46% when reserpine was used. For FQR isolates, a greater proportion of the isolates had ciprofloxacin MICs reduced by two or more dilutions by sodium orthovanadate (38%, compared to 13% with reserpine), but there were also more isolates unaffected by sodium orthovanadate (33% compared to 16% with reserpine). A similar result was observed for S isolates, with four isolates (24%) having ciprofloxacin MICs reduced by two or more dilutions compared to 0% with reserpine. Five of the S isolates (29%) had ciprofloxacin MICs reduced by one dilution, while eight isolates (47%) were unaffected. Ciprofloxacin MICs in the two DR isolates were not affected by addition of either inhibitor.

Efflux inhibitors will inhibit any efflux pump to which they can bind. Therefore, to more specifically observe the effect of loss of PatAB, the distribution of MICs of ciprofloxacin and ethidium bromide was analysed in mutants made from eight FQDR and five FQR isolates in which either the *patA* or *patB* gene had been inactivated by insertion of a *magellan2* minitransposon (Figure 3.3). The two DR isolates and one of the S isolates were also tested. Inactivation of *patA* resulted in an increase in susceptibility to ethidium

A



B

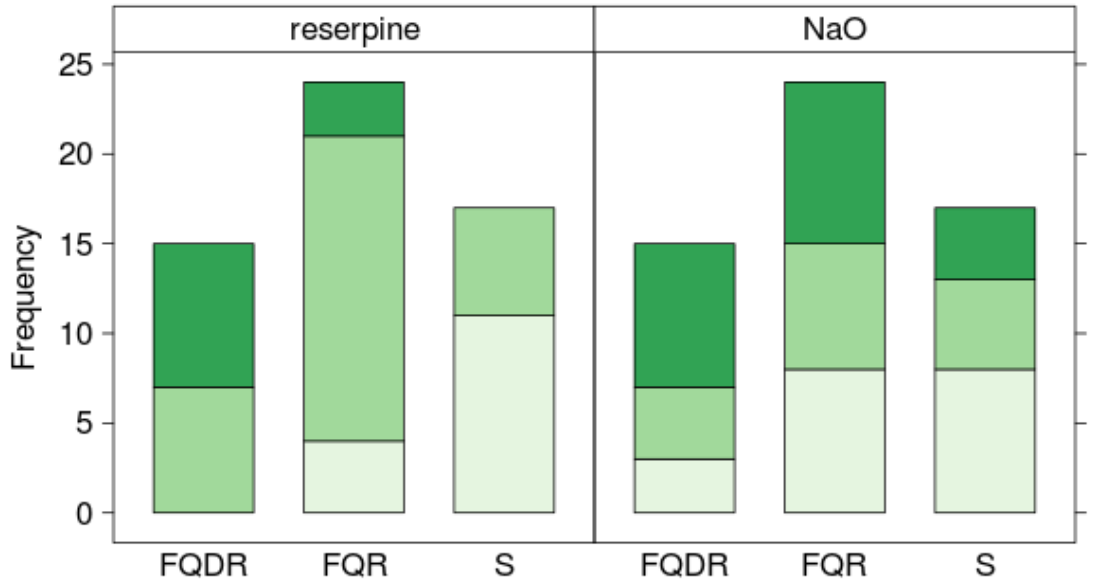


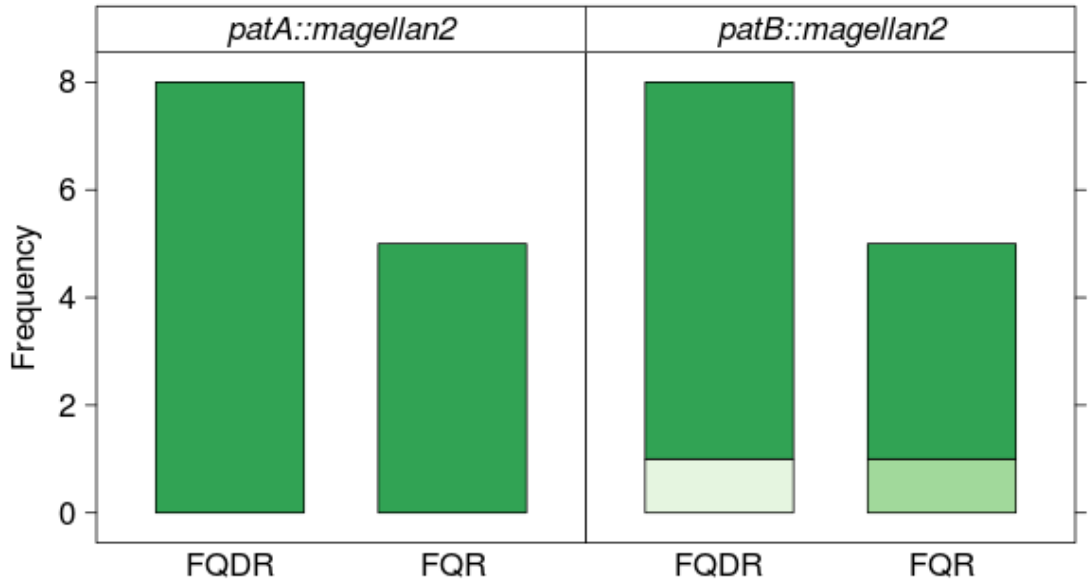
Figure 3.2. Number of isolates where MICs of (A) ciprofloxacin and (B) ethidium bromide were reduced by <1 (○), 1-2 (◔) or >2 (●) dilutions by addition of efflux inhibitors reserpine or sodium orthovanadate (NaO).

bromide of two or more dilutions in all isolates apart from the S isolate, in which the MIC of ethidium bromide was reduced by one dilution. When *patB* was inactivated, MICs of ethidium bromide were reduced by two or more dilutions for seven of the eight FQDR isolates and four of the five FQR isolates, and both DR isolates. The MIC of ethidium bromide for the S isolate was unaffected by *patB* inactivation. The effects of inactivating *patA* or *patB* on ciprofloxacin resistance were identical. MICs of ciprofloxacin were reduced by 2 or more dilutions in five FQDR isolates and four FQR isolates by inactivation of either *patA* or *patB*. The MICs of ciprofloxacin and ethidium bromide for the mutant isolates with inactivated *patA* or *patB* were not affected by more than one dilution by addition of reserpine or sodium orthovanadate.

3.6 Hoechst 33342 accumulation and relation to *patA* and *patB* expression level

To determine whether over-expression of *patA* and *patB* was linked to increased efflux, the steady-state intracellular accumulation of Hoechst 33342 was measured in a subset of isolates representing all combinations of MIC phenotype and *patAB* over-expression, and compared to that of the R6, M3 and M4 (Figure 3.4). Twenty-seven isolates were chosen in total, consisting of twelve FQDR, nine FQR, one DR, and five S isolates. Of these, eleven FQDR isolates, three FQR isolates and the DR isolate showed high expression of *patA* (defined as >1.5-fold higher expression of *patA* than in R6). Following addition of 2.5 μM Hoechst 33342, fluorescence increased rapidly for all tested isolates due to intracellular accumulation of the dye, and a steady state was reached in all cultures within four minutes. The average fluorescence of the control strains R6, M3 and M4 after ten minutes of measurement was was 123 arbitrary units (standard deviation = 3.4). Average steady-state accumulation of Hoechst 33342 in the group of isolates with

A



B

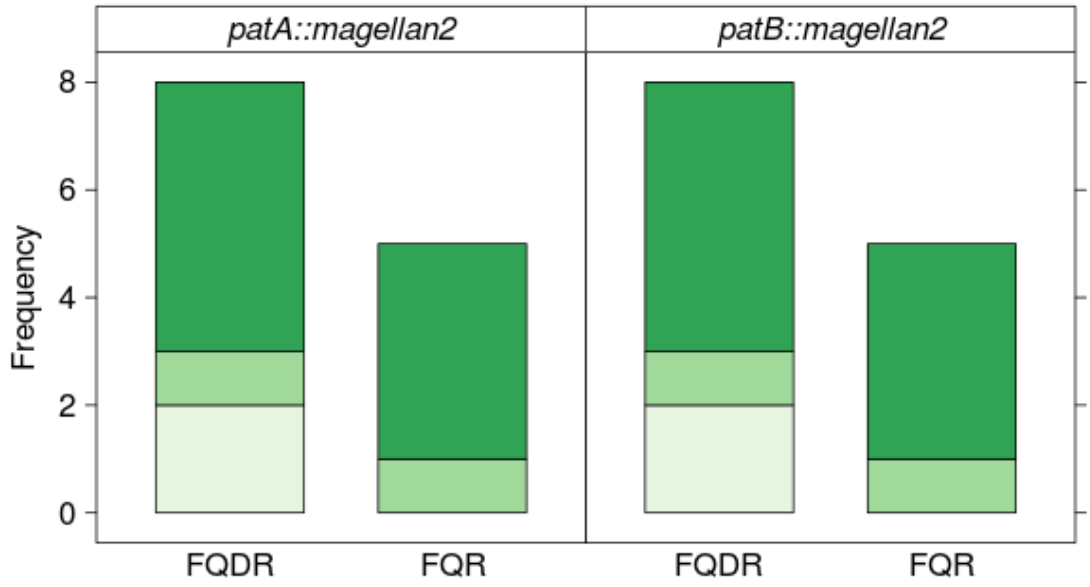


Figure 3.3. Number of isolates where MICs of (A) ciprofloxacin and (B) ethidium bromide were reduced by <1 (○), 1-2 (◔) or >2 (●) dilutions by inactivation of either *patA* or *patB* by insertion of a *magellan2* minitransposon.

high expression of *patAB* was 78 units (standard deviation = 23), which was significantly lower than the control strains ($p < 0.05$, one-tailed Student's t-test; Figure 3.4). There was no statistically significant difference between average accumulation in the control strains and the group of isolates not over-expressing *patAB* (124 units, standard deviation = 18; Figure 3.4). Fluorescence of all isolates was increased on addition of 20 μg/ml reserpine, but the mean fold increase observed for isolates with high expression of *patA* was significantly higher than that of isolates with low expression of *patA* (4.8 ± 0.7 and 3.4 ± 0.3 respectively, $p<0.01$, one-tailed Student's t-test; Figure 3.4).

In the groups of isolates over-expressing and not over-expressing *patAB* there were some outliers with Hoechst 33342 accumulation levels that did not fit the general pattern observed for the group. These were identified as isolates with fluorescence values lying outside a 95% confidence interval calculated for each group. For isolates not over-expressing *patAB*, 95% of the observed fluorescence values were between 112 and 136 units, with the exception of isolate M96, which had a fluorescence value of 68 units, similar to the majority of the *patAB* over-expressing isolates. Conversely, the 95% confidence interval for Hoechst accumulation in the isolates with high *patA* expression was 66 to 91 fluorescence units, but two *patAB* over-expressing isolates, M42 and M99, accumulated significantly more Hoechst than this (134 and 131 units, respectively).

3.7 Mutations in topoisomerase genes

The sequences of the QRDRs of the topoisomerase genes, *parC*, *parE*, *gyrA* and *gyrB* had been previously determined for 46 isolates by Piddock et al (2002). This dataset contained topoisomerase QRDR sequences from a further 11 isolates determined by Garvey and Wong (unpublished data). The sequences of the QRDRs from the clinical isolates were compared to those of R6 to identify base substitutions that caused an amino acid change in the protein sequence of the enzyme.

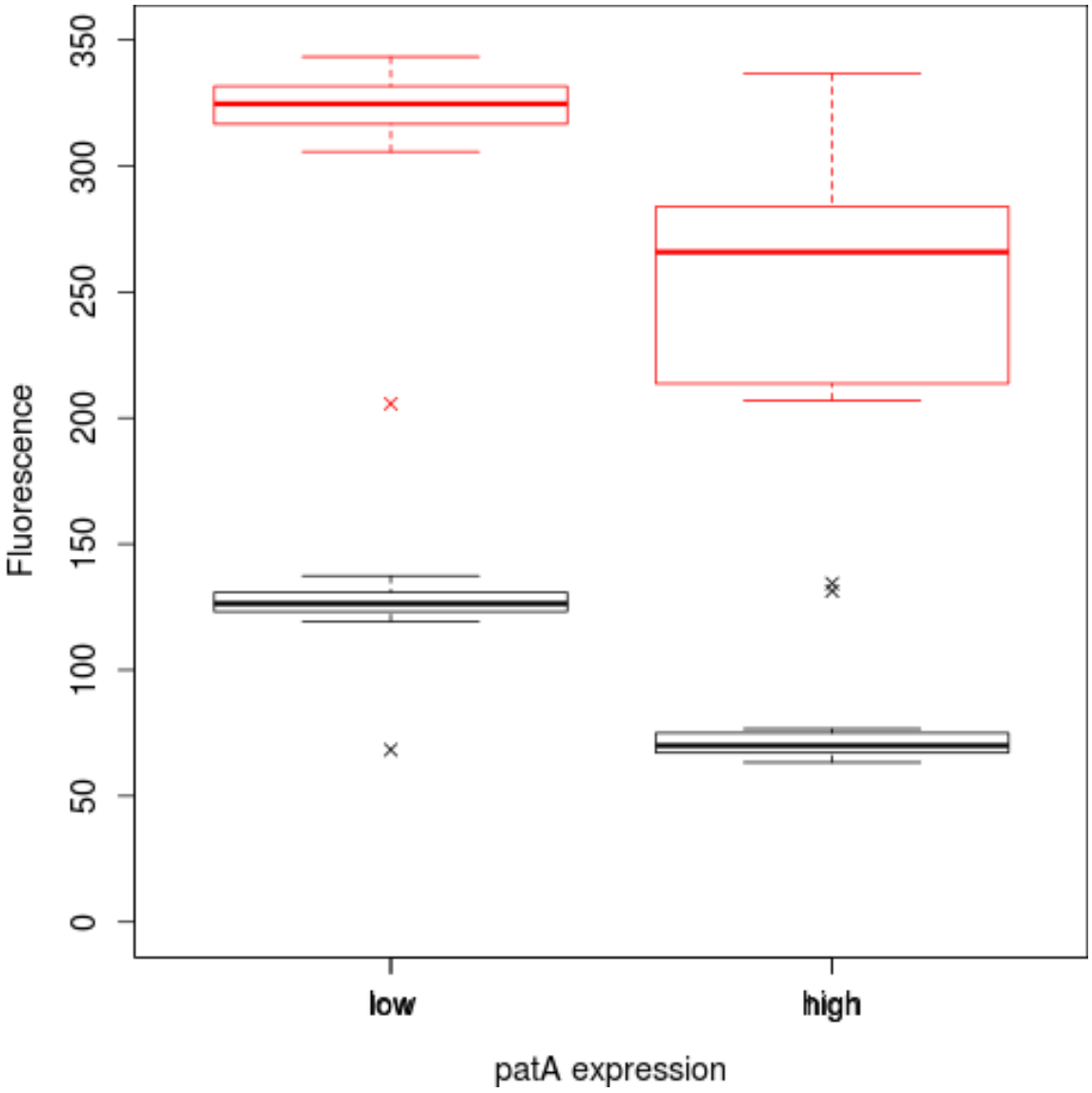


Figure 3.4. Fluorescence due to intracellular accumulation of Hoechst 33342 for isolates grouped by *patA* expression level. Distribution of fluorescence values for groups of isolates with high (>1.5 fold higher than R6) or low expression of *patA* with (red boxes) or without (black boxes) addition of 20 µg/ml reserpine.

In the 17 strains tested that were susceptible to all four agents, two were found to carry mutations in the topoisomerase genes. One isolate (M295) had an Ile460Val change in ParE, and the other (M77) carried a Lys137Asn mutation in ParC. These mutations were found in combination with other topoisomerase mutations in FQDR and FQR isolates. The ParE Ile460Val mutation was found in 12 isolates, and the Lys137Asn mutation in five. However, as these mutations did not cause fluoroquinolone resistance when present in isolation, they were assumed to not contribute significantly to fluoroquinolone resistance and excluded from further analysis. No topoisomerase mutations were found in the two DR isolates.

Following exclusion of non-functional topoisomerase mutations, the distribution of different genotypes in isolates with fluoroquionolone resistance was analysed. Overall, of the 15 FQDR and 24 FQR isolates, seven FQDR and 16 FQR isolates carried mutations in two or more topoisomerase genes (Table 3.1). A further five FQDR isolates and eight FQR isolates had carried a mutation in a single topoisomerase gene. There was no significant difference in the frequency of isolates carrying single topoisomerase mutations, or two or more topoisomerase mutations between the groups of isolates over-expressing or not over-expressing *patAB* (p=0.7, Fisher's exact test). Three FQDR isolates with increased *patAB* contained no functional topoisomerase mutations, while none of the FQR isolates showed this phenotype. This association was significant at the 10% level ($p < 0.1$), but a larger sample size would be needed to determine whether this is robust.

Mutations in ParC are the most common topoisomerase mutations observed in fluoroquinolone resistant *S. pneumoniae* (Muñoz and De La Campa, 1996). In this dataset, nine FQDR and 17 FQR isolates carried mutant *parC* alleles (60% and 70% respectively). Twenty-one of these encoded proteins where serine 79 was mutated to tyrosine or phenylalanine, which is well established as the most commonly observed ParC mutation causing fluoroquinolone resistance. Four other ParC mutations were observed in six isolates, either singly or in combination with serine 79 mutations.

Table 3.1. Mutations in topoisomerase genes detected in FQDR and FQR clinical isolates.

Strain	Group	parC	parE	gyrA	gyrB
M101	FQDR	-	-	S81F	-
M296	FQDR	-	I460V	-	-
M297	FQDR	S79Y	I460V	-	-
M42	FQDR	S79F	I460V	-	-
M44	FQDR	-	-	-	-
M45	FQDR	-	-	-	V163I, R169K, E181K
M50	FQDR	S79F	-	S81F	-
M70	FQDR	S79F	I460V	S81F	-
M74	FQDR	S79F	-	S81F	-
M86	FQDR	K137N	-	-	-
M87	FQDR	S79F, K137N	I460V	S81F	-
M89	FQDR	S79F	-	S81F	A639Q
M92	FQDR	S79F	I460V	S81F	-
M96	FQDR	K137N	I460V	S81F	-
M99	FQDR	S79F	I460V	S81F	-
M100	FQR	D78N, R95C	-	S81F	-
M37	FQR	S79F	-	-	V163I, R169K, E181K
M38	FQR	N91D	-	-	V163I, R169K, E181K
M40	FQR	S79F	-	-	-
M41	FQR	S79F	-	-	-
M49	FQR	S79F	-	S81F	-
M52	FQR	-	-	S81F	-
M67	FQR	-	-	-	V163I, R169K, E181K
M69	FQR	-	-	-	V163I, R169K, E181K
M75	FQR	S79F	-	-	-
M76	FQR	-	D435N	S81F	-
M78	FQR	S79F	-	-	-
M79	FQR	S79F	-	S81F	-
M80	FQR	-	D435N	S81F	-
M83	FQR	S80P, D83Y	-	S81F	-
M85	FQR	-	D435N	S81F	-
M88	FQR	S79F	-	S81F	T177I, E181K, I188T
M90	FQR	S79F, K137N	D435N, I460V	S81F	-
M91	FQR	S79F, N91D, E125D	-	-	T177I, E181K, I188T
M93	FQR	-	-	S81F	-
M94	FQR	S79F	-	S81F	-
M95	FQR	S79F	I460V	S81F	-
M97	FQR	N91D, K137N	I460V	S81F	-
M98	FQR	D83N	I460V	S81F	-

-, no mutation detected

GyrA mutations were found in nine FQDR and 15 FQR isolates, and all of these changed serine 81 to phenylalanine, which is the most commonly observed GyrA mutation associated with fluoroquinolone resistance (Janoir *et al.*, 1996; Pan *et al.*, 1996). In *S. pneumoniae*, *gyrA* mutations commonly arise secondarily to *parC* mutations, and in this dataset 70% of the isolates with a *gyrA* mutation also possessed a *parC* mutation (nine FQDR isolates and eight FQR isolates).

Five different *gyrB* alleles were observed in eight isolates (two FQDR and six FQR). In three isolates (one FQDR and two FQR) the change in GyrB was the only topoisomerase mutation present. A mutation of aspartate 435 to asparagine in ParE was found in three FQR isolates, but always in combination with a mutation in ParC and GyrA.

The possibility of an association between *patAB* over-expression and the number and type of topoisomerase mutations possessed by an isolate was also considered. A positive correlation might be observed if increased *patAB* expression promotes selection of further fluoroquinolone resistance mutations. Alternatively, selection of secondary topoisomerase mutations may be relatively less advantageous in an isolate over-expressing *patAB*, which could result in a negative correlation between *patAB* expression and presence of topoisomerase mutations. To investigate this, isolates with fluoroquinolone resistance were grouped by *patA* expression level as before and also by whether they possessed mutations in both *parC* and *gyrA*, *parC* only, *gyrA* only, or in neither gene. There was no significant difference in the distribution of *parC* and *gyrA* genotypes between isolates with high or low *patAB* expression ($p=0.2$, Fisher's Exact test). There was also no significant association between possession of a mutation in GyrB and high or low expression of *patAB* ($p=0.4$, Fisher's exact test).

3.8 Growth kinetics

In previous studies of laboratory selected mutants over-expressing *patA* and *patB*, no differences were observed in growth rate in broth culture between the mutant and parent strains (Garvey and Piddock, 2008). To confirm that this was also the case in clinical isolates, where in most cases the over-expression of *patA* and *patB* was combined with a mutation in one or more of the topoisomerase genes, the generation times of a randomly chosen set of 18 isolates were determined. In this set, nine isolates showed high expression of *patA* and *patB*.

Cultures were grown in triplicate in a microtitre tray on at least two occasions, and growth was monitored by measurement of optical density at 600 nm at ten minute intervals. The microtitre tray was incubated in the plate reader, without a CO_2-enriched atmosphere, and was shaken for five seconds before each absorbance reading. Under these conditions the generation times of the control strains R6 and M4 were 75 and 67 minutes, respectively. Generation times of the clinical isolates varied widely between 27 and 94 minutes (Figure 3.5) and there was no statistically significant difference between the groups of isolates with high and low *patAB* expression (p=0.65, two-tailed Student's t-test).

3.9 Discussion

This set of clinical isolates could be readily categorised into four groups based on resistance to the fluoroquinolone antibiotics ciprofloxacin and norfloxacin and the toxic dyes ethidium bromide and acriflavine. Decreased susceptibility to the ethidium bromide and acriflavine has been shown in previous studies to be indicative of increased efflux activity (Brenwald *et al.*, 2003; Pestova and Millichap, 2002; Piddock *et al.*, 2002).

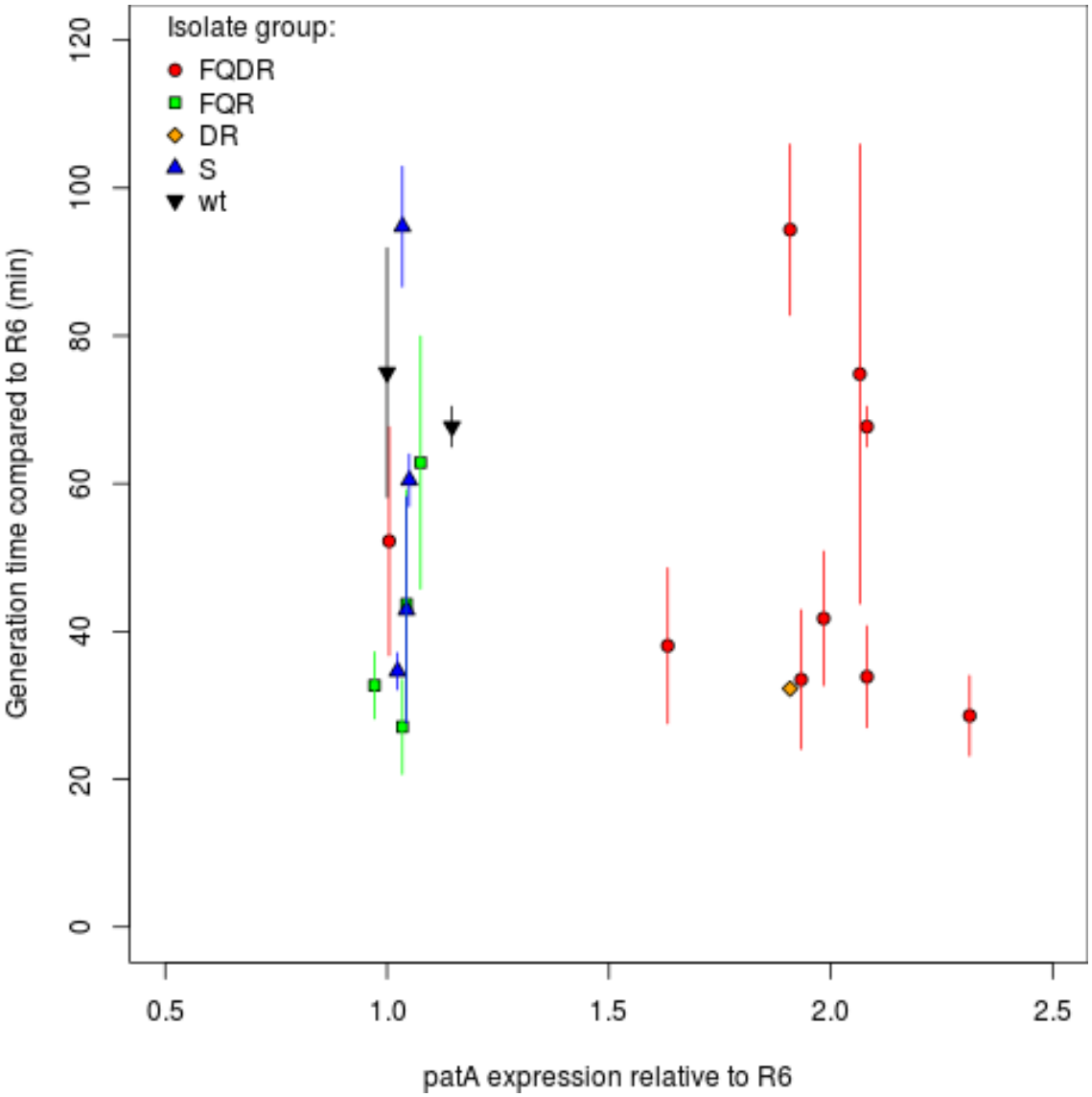


Figure 3.5. Generation times of 18 clinical isolates and the laboratory strains R6 and M4, plotted agains *patA* expression level.. Points are coloured according to isolate group. Error bars represent standard deviation of three biological replicates.

It was hypothesised that strains with the fluoroquinolone and dye resistant phenotype would be significantly associated with over-expression of *patA* and *patB*, since the PatAB transporter has been shown to confer resistance to all four of the tested agents in previous studies (Garvey and Piddock, 2008; Marrer *et al.*, 2006b). The results support this hypothesis, with 14 of the 15 FQDR isolates over-expressing *patA* and *patB*. In all cases, when *patA* and *patB* were over-expressed they were up-regulated together. This adds further support to previous reports suggesting that *patA* and *patB* are co-regulated (Garvey and Piddock, 2008).

The majority of the isolates investigated here were tested for *pmrA* in a previous study (Piddock *et al.*, 2002). No strong correlation was found between efflux activity and *pmrA* expression, which led to the conclusion that PmrA is not the predominant efflux pump associated with efflux-mediated fluoroquinolone resistance in clinical isolates of *S. pneumoniae*. Prior to this study, it had been hypothesised that this role was fulfilled by PatAB based on the phenotype of laboratory mutants (Garvey and Piddock, 2008; Marrer *et al.*, 2006b). The strong association between FQDR and *patAB* expression presented here provides support for this hypothesis.

However, some isolates showed phenotypes which did not fit the general pattern. Two isolates which over-expressed *patA* and *patB* had decreased susceptibility to dyes, but not to fluoroquinolones. Conversely, three isolates that were resistant to fluoroquinolones but not dyes nevertheless showed increased expression of *patA* and *patB*. The reasons for this are not clear, but could be due to differences in the genetic background of the isolates. One possibility is that an isolate could be naturally more susceptible to the effect of dyes than is standard for *S. pneumoniae* strains, such that increased expression of *patAB* is not sufficient to raise the MIC to the level above which isolates were designated dye resistant in this analysis. However, if this were true, it would be expected that FQR isolates over-expressing *patAB* would show particularly high susceptibility to dyes in the presence of an efflux inhibitor, which was not the case. An alternative explanation could

be that strain-specific differences in the PatAB transporter itself affect its efficiency at exporting different substrates.

Isolates over-expressing *patA* and *patB* accumulated significantly lower levels of the fluorescent dye Hoechst 33342 than wild-type strains and isolates not over-expressing these genes. This dye is frequently used as a marker of efflux activity (Garvey and Piddock, 2008; Velamakanni *et al.*, 2008; Venter *et al.*, 2008), so this result shows the expected link between increased expression of *patAB* and increased efflux. Again, however, there were outliers. Isolate M96, which had a FQDR phenotype but did not over-express *patAB*, accumulated low levels of Hoechst 33342. It is possible that Hoechst 33342 efflux in this isolate is mediated by over-expression of a different efflux pump. Although PmrA is not the primary transporter causing efflux mediated fluoroquinolone resistance in clinical isolates, over-expression of this pump has been observed in some isolates previously (Gill *et al.*, 1999), and has been shown to export ethidium bromide when over-expressed in R6 (Gill *et al.*, 1999; Pestova and Millichap, 2002). It is possible that the relative contributions of PatAB and PmrA to fluoroquinolone efflux differ between genetic backgounds. Additionally, the MATE transporter DinF has recently been identified as a third pneumococcal transporter capable of fluoroquinolone efflux (Tocci *et al.*, 2013). Inactivation of *dinF* in an antibiotic susceptible pneumococcal strain conferred hypersusceptibility to the fluoroquinolones ciprofloxacin, levofloxacin and moxifloxacin (Tocci *et al.*, 2013). PatAB does not affect moxifloxacin susceptibility (El Garch *et al.*, 2010), which suggests that PatAB and DinF have overlapping but non-identical substrate profiles. However, data are not yet available regarding the ability of DinF to transport dyes, or the phenotype of a mutant over-expressing *dinF*. Conversely, two of the FQDR isolates tested accumulated levels of Hoechst 33342 similar to wild type, despite over-expressing *patAB*. This could potentially be explained by strain-specific differences in the ability of the PatAB transporter to export Hoechst 33342.

The main difference between the results presented here, and previous studies of *patAB*

over-expression in laboratory mutants, is that the clinical isolates possessed topoisomerase mutations in addition to *patAB* over-expression. The role of PatAB in isolates with coexisting topoisomerase mutations had not previously been investigated. Three isolates which over-expressed *patAB* and had an FQDR phenotype did not contain functional topoisomerase mutations. This suggests that over-expression of *patA* and *patB* can cause fluoroquinolone resistance in clinical isolates in isolation. However, in the remaining 12 FQDR and 24 FQR isolates topoisomerase mutations were present. There was considerable variation in the combinations of topoisomerase gene alleles observed, but there was no correlation between any particular genotype and expression of *patAB*.

Genetic inactivation of *patA* or *patB*, or addition of efflux inhibitors, decreased MICs of ethidium bromide for the majority of the isolates by more than two dilutions, regardless of the initial level of resistance or expression of *patAB*. This implicates PatAB in intrinsic resistance to ethidium bromide, even in the absence of over-expression of the pump. For ciprofloxacin resistance, the picture was more complicated, but MICs were still reduced by one or more dilutions by either chemical or genetic inactivation of PatAB for the majority of the isolates tested. Ciprofloxacin MICs for a subset of FQR and S isolates were unaffected by addition of efflux inhibitors, and MICs of two FQDR isolates were unaffected by sodium orthovanadate. These results suggest that, in general, inactivation of PatAB reduces ciprofloxacin resistance, even in the presence of topoisomerase mutations. However, there is evidence of variation between strains, and PatAB may contribute to fluoroquinolone resistance to different degrees depending on the genetic background of the strain. Hoechst 33342 accumulation was also increased in all tested isolates by the addition of the efflux inhibitor reserpine, indicating that there is a basal level of efflux in all isolates, regardless of the expression level of *patAB*.

No significant difference in average generation time was observed between isolates that over-expressed *patA* and *patB* and those that did not. This indicates that over-expression of *patAB* does not confer a substantial fitness burden on pneumococcal isolates under

laboratory growth conditions, even when combined with topoisomerase mutations. This result complements previous work, which showed that over-expression of *patAB* did not affect growth of laboratory mutants (Garvey and Piddock, 2008). However, the generation times of the clinical isolates were highly variable, possibly because they were not grown under ideal *S. pneumoniae* culture conditions. Also, growth under laboratory conditions is only a crude measure of fitness. Further work would be needed to determine whether over-expression of *patAB* has an impact on pneumococcal growth and survival *in vivo*.

No mutations had been associated with the *patA* and *patB* genes in isolates that over-expressed *patAB* compared to isolates that did not, suggesting that the increased activity of *patA* and *patB* is not caused by a functional change in the transporter protein, or increased expression caused by an altered promoter sequence. This confirmed previous studies where no mutations had been found in *patA* and *patB* in laboratory mutants. This suggests that increased expression of these genes must be caused by a mutation elsewhere in the genome, possibly in a regulatory gene.

In conclusion, in this group of isolates over-expression of *patA* and *patB* was significantly associated with resistance to both fluoroquinolones and dyes. Inactivation of the PatAB pump caused reductions in resistance to these agents in the majority of the isolates, regardless of *patAB* expression level and possession of topoisomerase mutations, suggesting that PatAB contributes to intrinsic fluoroquinolone and dye resistance. These results have since been supported by other groups who have observed *patAB* over-expression in clinical isolates. El Garch *et al* (2010) investigated two clinical isolates with increased *patAB* expression, one of which possessed topoisomerase mutations and one of which did not, and observed similar reductions in fluoroquinolone and dye resistance by addition of reserpine or inactivation of *patA* or *patB*. Lupien and coworkers (Lupien *et al*., 2013) identified a ciprofloxacin resistant clinical isolate which over-expressed *patA* and *patB* and did not have any co-existing topoisomerase mutations, as was observed

in three of the isolates in this study.

3.9.1 Further work

The main limitation of this study is that it was carried out using a set of isolates that were collected over ten years ago. This means that the results may not reflect the clinical situation now. Additionally, the clinical history of the patients from which these isolates were taken is unknown, which means that correlations of expression of *patAB* with disease state or treatment using a particular antibiotic cannot be assessed. It would therefore be interesting to repeat this study using a modern set of clinical isolates collected systematically with details of the clinical context included. It has previously been shown that *patAB* over-expression can be selected by exposure to a variety of fluoroquinolones, not just those that are good substrates for the PatAB pump (El Garch *et al.*, 2010). It could therefore be hypothesised that a similar proportion of isolates would be found to over-express *patAB* despite declining use of ciprofloxacin. A mutation causing over-expression of *patA* and *patB* was found in an isolate from the cerebrospinal fluid of a meningitis patient, compared to an otherwise isogenic bloodstream isolate (Croucher *et al.*, 2013). It would therefore interesting to investigate whether there is a correlation between pneumococcal isolates causing meningitis and over-expression of *patAB*.

This study also identified some isolates with phenotypes that did not fit the expected pattern. Three isolates that over-expressed *patAB* accumulated wild-type levels of Hoechst 33342 and were susceptible to ethidium bromide. A further two isolates over-expressed *patAB* but showed wild-type susceptibility to fluoroquinolones. It would be interesting to compare the transmembrane domain sequences of PatA and PatB in these isolates with R6 to see if there are any differences that might alter the specificity of the transporter. Conversely, one isolate that was resistant to fluoroquinolones and dyes and accumulated

low levels of Hoechst 33342 did not over-express *patAB*. This isolate could be screened for increased expression of *pmrA* or *dinF* by qRT-PCR.

3.10 Key findings

- Over-expression of *patAB* was observed in 38% of the fluoroquinolone resistant clinical isolates in this set.
- Over-expression of *patAB* was significantly associated with resistance to both fluoroquinolones and dyes.
- Clinical isolates over-expressing *patAB* accumulated significantly lower levels of the fluorescent dye Hoechst 33342 than isolates not over-expressing the efflux pump.
- Chemical or genetic inactivation of PatAB reduced fluoroquinolone and dye resistant in isolates where the pump was not over-expressed, and in isolates with co-existing topoisomerase mutations, suggesting that PatAB is involved in intrinsic resistance to these agents.
- Over-expression of *patAB* did not affect growth of *S. pneumoniae* clinical isolates.

4 Using whole genome sequencing to identify mutations causing upregulation of *patA* and *patB* in M184

4.1 Background

Previous studies showed that mutants over-expressing *patA* and *patB* can be easily selected in the laboratory using ciprofloxacin (Marrer *et al.*, 2006b) and the efflux inhibitor reserpine (Garvey and Piddock, 2008). Results presented in chapter 3 suggested that over-expression of *patA* and *patB* contributing to fluoroquinolone resistance is common in clinical isolates (Garvey *et al.*, 2011). However, genetic changes responsible for the upregulation of *patA* and *patB* in these studies have not been identified.

M184 is a laboratory strain of *S. pneumoniae* that over-expresses *patA* and *patB* (Garvey and Piddock, 2008). M184 was derived from M169, a mutant of the unencapsulated laboratory strain R6 that was selected from increased resistance to the efflux inhibitor reserpine (Garvey and Piddock, 2008). Reserpine-resistant mutants, including M169, were selected from R6 following a single exposure to reserpine, at a frequency that suggested that reserpine resistance was conferred by a single point mutation (Garvey and Piddock, 2008). M184 was generated by transformation of genomic DNA from M169 into R6, followed by selection of successful transformants on 100 µg/ml reserpine (Garvey and Piddock, 2008). This was done to confirm that *patAB* over-expression was

conferred by a single genetic change, and to isolate this mutation in a clean genetic background.

M184 was previously shown to be less susceptible than R6 to fluoroquinolones (ciprofloxacin, norfloxacin and levofloxacin), dyes (ethidium bromide and acriflavine), cetrimide and reserpine (Garvey and Piddock, 2008). M184 was also shown to accumulate lower intracellular levels of ciprofloxacin and ethidium bromide than R6 (Garvey and Piddock, 2008). However, no mutations were found in the *patA* and *patB* genes or their upstream regions in M184 that might explain the observed phenotype, suggesting that a mutation elsewhere in the genome was responsible. Constitutive expression of efflux systems in several other bacteria has been shown to be caused by mutations in local or global transcriptional regulators (Abouzeed *et al.*, 2008; Webber *et al.*, 2005).

4.2 Aims and Hypothesis

The primary aim of this section was to determine the genetic change responsible for *patAB* over-expression in M184 using a whole genome sequencing approach. Following detection of a causative mutation, a secondary aim was to characterise the mutated gene(s) to begin to elucidate a regulatory pathway controlling expression of *patA* and *patB*.

Based on the mutation frequencies observed when strain M169 was selected, and the ease of transformation of the M169 phenotype into R6 (Garvey and Piddock, 2008), it was hypothesised that over-expression of *patA* and *patB* in M184 is conferred by a single point mutation. It was also hypothesised that the putative mutation would affect a transcriptional regulator.

4.3 Detection of mutations in M184 by whole genome resequencing

Whole genome resequencing of M184 and our current laboratory stock of R6 was carried out by GenePool (Edinburgh, UK) prior to the start of this project in 2009, as described in Materials and Methods (section 2.5.1). Briefly, 50 cycles of paired-end sequencing were carried out on an Illumina GAIIx system. This resulted in a set of 4,967,264 50bp read pairs for M184 and 2,043,226 for R6, giving theoretical coverages of the 2 Mb R6 genome of approximately 240x and 100x respectively. The quality of the M184 and R6 forward and reverse reads were assessed using FastQC. Average per-base quality scores were >20 along the entire length of the read for both M184 read sets and for the first read of each R6 pair. For the second reads from R6, a sizeable proportion showed a decline in quality in the final 10 bases.

4.3.1 Alignment

Sequencing reads obtained for each strain were mapped against the published R6 genome (Genbank accession number: NC_003098; Hoskins *et al.*, 2001) using the xBase next-generation sequencing pipeline (http://ng.xbase.ac.uk/my/). This is an online server allowing users to access the MAQ short-read alignment program (Li *et al.*, 2008) without the need to install and run the software locally. M184 and R6 reads were independently mapped against the reference sequence in paired-end mode, using an estimated insert size of 250bp. For M184, 9,161,550 out of 9,934,528 (92.2%) reads were successfully mapped, giving an average coverage of 225x, while 3,373,036 out of 4,086,452 (82.5%) R6 reads were mapped to give an overall coverage of 83x. In both cases, of the reads that were successfully mapped, more than 97% of these mapped as pairs. For M184, the majority of the genome positions were covered between 50 and 500 times, while for

R6 coverage mainly varied between 25 and 150 times (Figure 4.1).

4.3.2 Mutation detection

The xBase pipeline detects mutations in the aligned reads relative to the reference genome using the variation detection tools that are part of the MAQ program: cns2snp for calling SNPs and indelpe for detecting small indels (Li *et al.*, 2008). A set of SNPs and small indels was obtained for each strain relative to the published R6 reference genome, and these sets were compared to each other to find mutations unique to each strain and mutations shared between the strains.

4.3.2.1 Comparison of the R6 sequence to the published R6 reference genome

Sixty-eight mutations were detected by Maq in our R6 relative to the published R6 genome (Appendix 1). Six of these were unique to R6, but the remaining 62 were also found in the M184 sequence.

A cluster of ten mutations were predicted in the *pbpX* gene (spr0304). However, on closer inspection, coverage in this region was unusually low and dropped to zero in several places (figure 4.2). This read alignment pattern was also seen in the M184 data. This suggests that in the laboratory stock of R6 used in this study there has been a change in the *pbpX* region relative to the published R6 genome. This change could be a deletion of the *pbpX* gene, or a recombination with a *pbpX* variant sufficiently different such that reads generated from *pbpX* in the sequenced R6 cannot be aligned to *pbpX* in the published R6 genome. Mutation calls from the few reads that do align in this region are therefore unreliable, so mutations found in this region (302960 to 304013) were excluded from all following analyses.

Two further small clusters of five and four SNPs, respectively, were predicted in genes

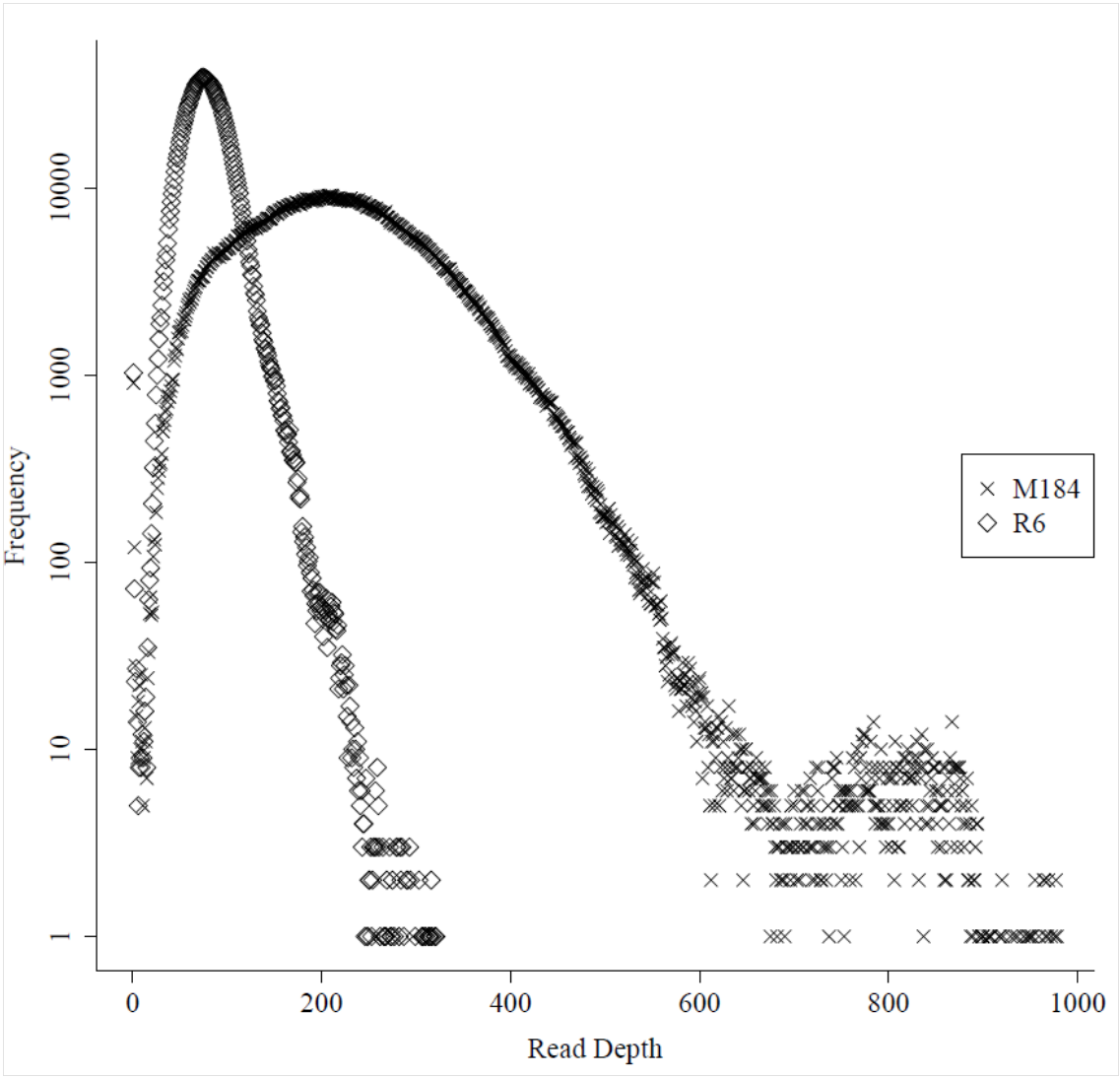


Figure 4.1. Comparison of M184 and R6 read depth distributions. Frequency distributions of read depths calculated for each genomic position when M184 and R6 Illumina reads were aligned against the published R6 genome

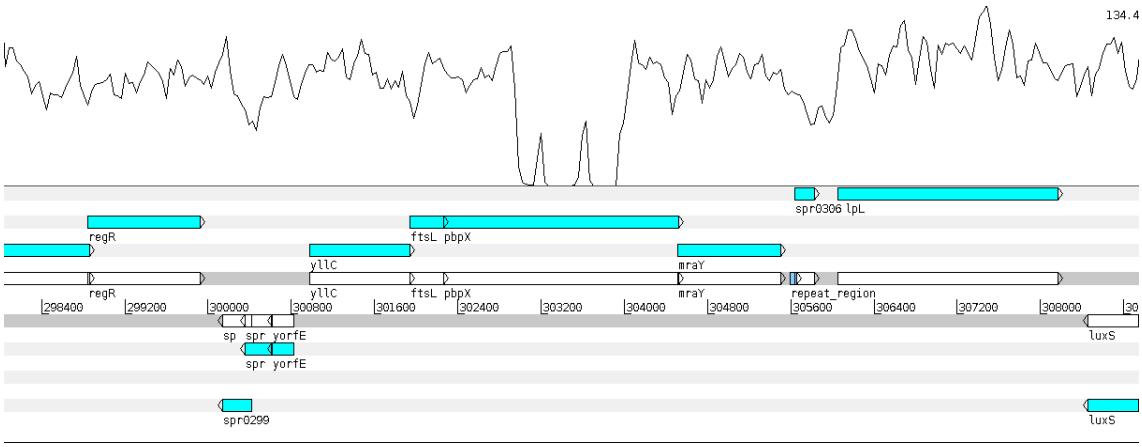


Figure 4.2. Read depth coverage of the *pbpX* region in R6. Viewed in Artemis

spr1675 and spr1721, both of which encode degenerate transposases. It is likely that these genes are non-functional, so not subjected to purifying selection, so a higher mutation rate than the surrounding genome would be expected. Alternatively, these genes could have been altered by recombination with another strain of *S. pneumoniae* at some point in the strain's evolutionary history. The remaining 47 mutations were distributed fairly evenly throughout the genome. Twenty-seven of these were predicted to affect the amino acid sequence of the respective gene product.

4.3.2.2 Identification of mutations in M184

Mutations found in both M184 and R6 genomes were excluded as these are likely to be pre-existing natural differences between the laboratory stock of R6 and the published R6 genome. In total, 36 SNPs and three small indels were found in M184 but not R6, and six SNPs were found in R6 but not M184. To confirm that the M184 mutations were real and not the result of sequencing error or mis-calling by the software, an area surrounding each mutation was amplified from M184 and R6 by PCR and resequenced by Sanger sequencing. The resulting sequences were examined for the presence or absence of the expected mutation. Of the original 39 mutations, 27 (24 SNPs and all three indels) were confirmed to be present in M184 and absent in R6 (Table 4.1). The remaining 12 mutations were not confirmed to be present in M184 and so were excluded from further analyses.

4.4 Investigating truncation of *hrcA*

Of the 27 confirmed SNPs found in M184 compared to R6, the most initially intriguing was a mutation in *hrcA*. HrcA is a negative regulator of the class I heat shock operons, *grpE-dnaK-dnaJ* and *groELS*, which are widely conserved in bacteria (Narberhaus, 1999;

Table 4.1. Mutations detected by Maq in M184 but not R6. Only mutations confirmed by targeted resequencing are shown.

Position[1]	Gene	Description	Mutation	Qual	Effect[2]	AA Change[3]
269365*	*sulB*	dihydrofolate synthetase	C-T	255	NS	S-L
263291	-	-	-8:ATATGTTG	31	I	-
387172	-	-	C-A	171	int	-
460764*	*hrcA*	heat-inducible transcription repressor	G-T	255	T	E-Stop
569420	-	-	G-R	76	int	-
634993*	*lctO*	lactate oxidase	G-T	242	NS	G-V
635851*	*lctO*	lactate oxidase	G-T	255	NS	G-V
808808	spr0807	hypothetical protein	T-C	48	NS	I-T
974512*	*murZ*	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	2:AT	155	I	trunc
1142521	*topA*	DNA topoisomerase I	G-T	196	NS	A-D
1167599*	*pyrP*	Uracil permease	G-A	255	NS	V-I
1327871*	*tufA*	elongation factor Tu	G-T	255	NS	T-N
1563763*	spr1582	hypothetical protein	C-A	173	NS	A-S
1570591*	spr1591	hypothetical protein	1:C	233	I	trunc
1595105	spr1620	ABC transporter SBP - sugar transport	G-T	250	T	Y-Stop
1670128*	*treP*	trehalose-specific PTS	G-A	106	NS	T-I
1845480*	spr1875	hypothetical protein	G-A	255	NS	A-V
1857615-8446*	*gltX*	glutamyl-tRNA synthetase	10 SNPs	211	-	various

[1] *, mutation chosen as candidate for transformation into R6, section 4.5; [2] NS, non-synonymous; S, synonymous; T, termination; int, intergenic; I, indel. [3] Stop, stop codon; trunc, mutation caused protein truncation. Qual, Phred-scaled quality score for variant call; AA, amino acid.

Schulz and Schumann, 1996). Loss-of-function mutations in negative regulators have been shown to cause derepression of efflux systems in other bacteria (Abouzeed *et al.*, 2008; Webber *et al.*, 2005), so the role of mutation of *hrcA* in the phenotype of M184 was further investigated. The mutation found in *hrcA* in M184 changes a glutamine codon to a stop codon, causing truncation of the HrcA protein. A comparative model of the HrcA protein structure was generated using SwissModel (Arnold *et al.*, 2006) using the crystal structure of *hrcA* from *Thermatoga maritima* as a template (Liu *et al.*, 2005). Based on this model, the truncation deletes part of a central, GAF-like domain that is thought to stabilise a dimerisation domain (Figure 4.3; Liu *et al.*, 2005). This seems likely to affect protein function.

4.4.1 Sequencing of *hrcA* gene from other *patAB* over-expressors

If HrcA is involved in regulation of *patAB*, it might be expected that other pneumococcal isolates that over-express *patAB* would also carry mutations in this gene. To investigate this, the sequence of the *hrcA* gene was determined from M168, a reserpine-selected mutant of M4 (Garvey and Piddock, 2008) that has the same phenotype as M184. The *hrcA* gene was also sequenced from eight of the *patAB* over-expressing clinical isolates described in Chapter 3. No *hrcA* mutations were found in the clinical isolates, but M168 contained a glycine to arginine mutation located in the putative DNA binding domain of HrcA (indicated in figure 4.3).

4.4.2 Transformation of R6 with mutated *hrcA*

To determine whether the truncation of *hrcA* in M184 is responsible for *patAB* over-expression, the *hrcA* gene containing the truncation mutation was amplified from M184 by PCR (primers 902 and 903), and the PCR amplimer was transformed into R6. Transformation of whole M184 genomic DNA into R6 was carried out in parallel as a positive

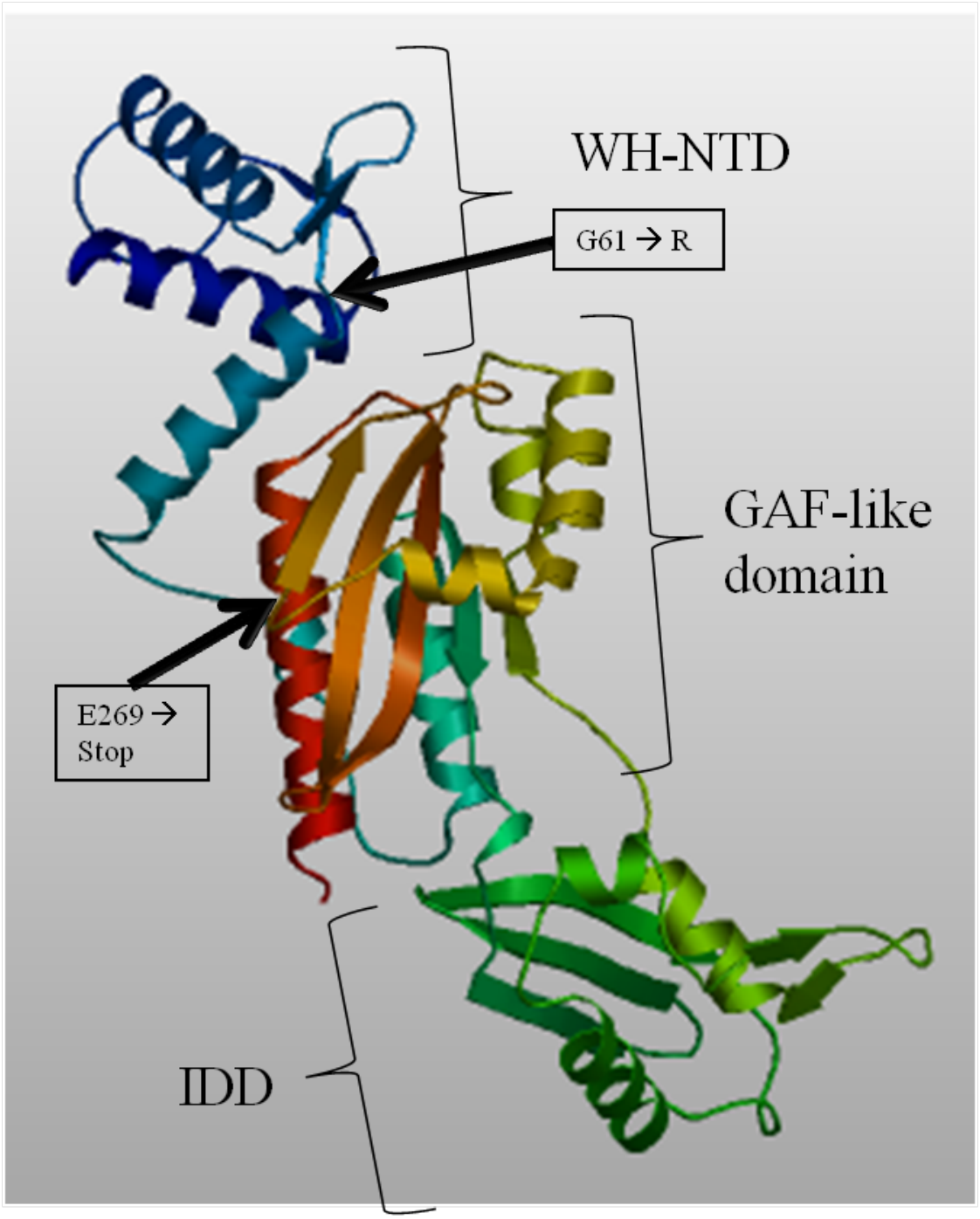


Figure 4.3. Model of Sp_HrcA produced by SwissModel. Positions of mutations found in M184 (E269→Stop) and M168 (G61→R) are indicated by arrows. WH-NTD: Winged helix-turn-helix N-terminal domain. IDD: inserted dimerisation domain

control. Transformation mixtures were spread onto agar plates containing 100 µg/ml reserpine. Two reserpine resistant transformants were recovered from transformation with the M184 *hrcA* amplimer and sequencing of the *hrcA* gene from these transformants confirmed that the truncation mutation had been transferred. One transformant was retained for further study and named M499. The MICs of ciprofloxacin, ethidium bromide and reserpine for M499, were measured and compared to R6 and M184 (Table 4.2). As observed previously (Garvey and Piddock, 2008), M184 shows decreased susceptibility to all three agents, and the resistance to ethidium bromide and ciprofloxacin was reversed by addition of 20 µg/ml reserpine. For M499, the MIC of reserpine was increased from 64 µg/ml to 128 µg/ml, as observed in M184, but the MICs of ciprofloxacin and ethidium bromide remained at the R6 levels of 2 and 4 µg/ml, respectively. This suggests that the *hrcA* mutation is unlikely to confer *patAB* over-expression in M184.

To confirm that M499 did not over-express *patA* and *patB*, the expression of these genes was measured by qRT-PCR from R6, M184 and M499 (Figure 4.4). Neither gene was expressed in M499 at a level significantly higher than R6. In contrast, in M184, both *patA* and *patB* were expressed at more that 100-fold higher levels than in R6.

4.5 Identifying mutations linked to the patAB over-expression phenotype

Since mutation of *hrcA* was not the cause of *patA* and *patB* over-expression, the involvement of the other mutations determined by the whole genome sequencing was assessed. A similar transformation process to that used to transfer the *hrcA* mutation was used. However, the resistance profile of M499 suggests that reserpine resistance can be conferred separately from ethidium bromide and fluoroquinolone resistance. Therefore, in the following experiments ethidium bromide and norfloxacin were used as selective agents

Table 4.2. MICs (μg/ml) of ciprofloxacin, ethidium bromide and reserpine for R6, M184 and M499, in the presence or absence of 20 μg/ml reserpine.

	Ciprofloxacin		Ethidium bromide		Reserpine
Reserpine (20 μg/ml)	-	+	-	+	
R6	2	1	4	1	64
M184	**8**	1	**32**	1	128
M499	2	1	4	1	128

Bold text indicates MIC values greater than those of R6 by two or more dilutions

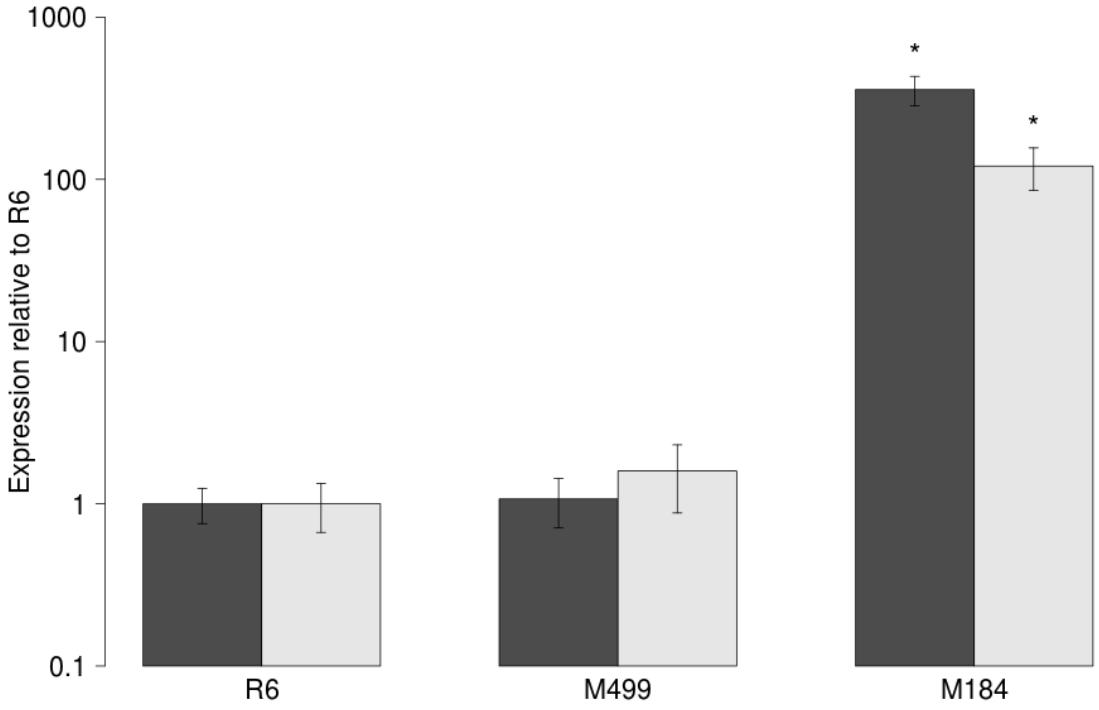


Figure 4.4. Expression of *patA* (dark bars) and *patB* (light bars) in R6, M499 and M184, determined by qRT-PCR. *, expression significantly higher than in R6, $p<0.05$, one-tailed Student's t-test on $\log_{10}$(expression).

as these provided a better indication of *patAB* over-expression.

4.5.1 Transformation with PCR amplimers

The remaining 26 candidate mutations were reduced to an initial test set of ten (indicated in table 4.1) based on the following criteria:

- the mutation occured within a gene and affected the protein coding sequence
- gene affected was part of the pneumococcal core genome, as defined by Hiller *et al.* (2007)
- the mutation was not found in any of the 17 published *S. pneumoniae* genomes described by Hiller *et al.* (2007). Conservation of a mutation in another genome implies that the mutation is a natural polymorphism unrelated to the M184 phenotype.

A cluster of ten mutations identified in the *gltX* gene was also tested. Due to their proximity, the ten mutations were counted as one genetic change, bringing the total set of mutations to test to eleven.

A region containing each candidate mutation was amplified by PCR and transformed into R6; the transformation mixtures were then cultured on agar plates containing 8 µg/ml of either norfloxacin or ethidium bromide. No transformants were recovered following selection on either agent for any of the PCR products, but a positive control transformation using M184 genomic DNA as the donor DNA consistently resulted in colonies resistant to both agents. The mutated PCR amplimers were also transformed into M499 to determine whether one of these mutations in combination with the HrcA truncation could confer the M184 phenotype but, again, no ethidium bromide or norfloxacin resistant transformants were recovered.

4.5.2 Transformation of M184 phenotype into R6 background using whole genomic DNA

None of the twelve mutations initially chosen for study could confer the M184 phenotype to R6 cells when transformed singly or in combination with the truncation mutation in *hrcA*. This implied that either one of the remaining five mutations not yet tested or a combination of more than one mutation was the cause of the phenotype. To assess the number of mutations required for over-expression of *patA* and *patB*, the frequency of transformation to ethidium bromide or norfloxacin resistance when R6 cells were transformed with whole M184 genomic DNA was measured and compared to that of a spectinomycin resistance gene.

R6 cells were transformed with 0.2 µg/ml M184 DNA on two separate occasions, and plated on either 8 µg/ml norfloxacin or 8 µg/ml ethidium bromide. In parallel, R6 cells were transformed with 0.2 µg/ml DNA from M240, an R6 mutants that carries a spectinomycin resistance cassette in the *patB* gene, and plated on 200 µg/ml spectinomycin. The number of colonies on each selective plate was counted to give the number of successful transformants per ml of culture. As a negative control to check the background rate of spontaneous mutations, R6 cells were transformed with R6 DNA and plated on spectinomycin, ethidium bromide and norfloxacin. For each transformation mixture the total number of bacteria per ml of culture, expressed as colony forming units (cfu) per ml, was determined by viable count, and transformation frequencies were calculated by dividing the number of transformants per ml obtained on selective media by the cfu/ml. The frequencies of transformation to ethidium bromide and norfloxacin resistance observed using M184 DNA were between 10^{-6} and 10^{-7}, while the frequency of transfer of the spectinomycin cassette from M240 was 10^{-5}. These values were compared to estimate how many mutations must be transferred to confer *patAB* over-expression. The average norfloxacin/spectinomycin ratio was 0.08 and the aver-

age ethidium bromide/spectinomycin ratio was 0.03. No spontaneous ethidium bromide or spectinomycin resistant mutants were observed in the negative control, but a background mutation frequency to norfloxacin resistance of 10^{-8} was observed, presumably due to mutations in topoisomerase genes. Three of the ethidium bromide-selected transformants (M501, M502 and M503) and one norfloxacin-selected transformant (M500) were randomly chosen and retained for further study.

4.5.3 Confirmation of successful transfer of *patAB* overexpression

To confirm that the norfloxacin and ethidium bromide resistance of the transformants was caused by *patAB* expression, the MIC profiles of the four retained transformants were determined in the presence and absence of reserpine and sodium orthovanadate (Table 4.3). All four transformants showed similar susceptibility to the two antibiotics and ethidium bromide as M184. The MICs of ciprofloxacin, norfloxacin and ethidium bromide were higher then those of R6 by at least two dilutions, and this was reversed by addition of 20 µg/ml reserpine or 50 µM sodium orthovanadate. However, norfloxacin MICs for three of the transformants (M500, M501 and M503) appeared to be affected to a greater extent by sodium orthovanadate than that of M184. Similarly, MICs of ethidium bromide for all four transformants were also reduced to a greater extent by sodium orthovanadate than for M184.

To confirm that the transformants showed increased efflux activity, as would be expected if they over-expressed *patA* and *patB*, intracellular accumulation of ethidium bromide was measured by fluorescence assay. All four transformants accumulated significantly lower levels of ethidium bromide after 10 minutes than wild type and M240 (R6 *patB::magellan2*; $p<0.05$, one-tailed Student's t-test; Figure 4.5). Interestingly, there was variation in the levels of accumulation between the transformants, with M500

Table 4.3. MICs of ciprofloxacin, norfloxacin and ethidium bromide for four R6^{M184} transformants in the presence and absence of efflux inhibitors reserpine and sodium orthovanadate.

	Cip	Cip + R	Cip + N	Nor	Nor + R	Nor + N	EtBr	EtBr + R	EtBr + N
R6	0.5	0.25	0.25	1	1	<1	2	<1	<1
M184	**2**	0.5	0.5	**8**	2	2	**16**	2	<1
M500	**2**	0.5	0.25	**8**	2	<1	**16**	2	<1
M501	**2**	0.25	0.25	**8**	2	<1	**16**	2	<1
M502	**2**	0.5	0.5	**16**	2	4	**32**	2	<1
M503	**2**	0.25	0.25	**8**	2	<1	**16**	2	<1

Cip, ciprofloxacin; Nor, norfloxacin; EtBr, ethidium bromide; R, 20 µg/ml reserpine; N, 50 µM sodium orthovanadate; Bold text indicates MIC values that are two or more dilutions greater than those of R6

and M503 accumulating significantly more ethidium bromide than M184, and M502 accumulating significantly less ($p<0.05$).

Finally, the expression of *patA* and *patB* was measured directly by qRT-PCR, using cDNA synthesised from total RNA extracted from mid-logarithmic phase cultures (OD_{660} = 0.4). Expression of *patA* and *patB* was determined relative to the control genes *rpoB* and 16S rRNA using the Pfaffl method (section 2.4. M184 and all four transformants expressed levels of *patA* and *patB* 100-1000-fold higher than those observed in R6 (Figure 4.6A). The levels of over-expression observed were surprisingly high, but were ascribed to very low levels of basal *patA* and *patB* expression observed in R6, that could be visualised by agarose gel electrophoresis of amplimers (Figure 4.6B).

Growth of M184 and the four $R6^{M184}$ transformants in BHI broth over seven hours was also measured and compared to that of R6 (Figure 4.7A). Generation times for each strain were calculated from exponential growth between two and four hours (Figure 4.7B). M184 grew significantly more slowly than R6 (generation time of 47±4 minutes compared to 35±2 minutes, $p<0.05$, One-tailed Student's t-test), but there was no significant difference between generation times of R6 and the four $R6^{M184}$ transformants.

4.5.4 Resequencing of known M184 mutations in $R6^{M184}$ transformants

To determine which, if any, of the M184 mutations detected by the genome resequencing were involved in *patAB* expression, regions surrounding each candidate mutation were amplified by PCR from all four of the transformants and sequenced by Sanger sequencing. The sequence of each PCR amplimer was then checked for the presence or absence of the candidate mutation. Surprisingly, none of the 27 candidate mutations detected in M184 were present in any of the transformants tested, which indicated that they were not the cause of *patAB* over-expression in these transformants. To check that the

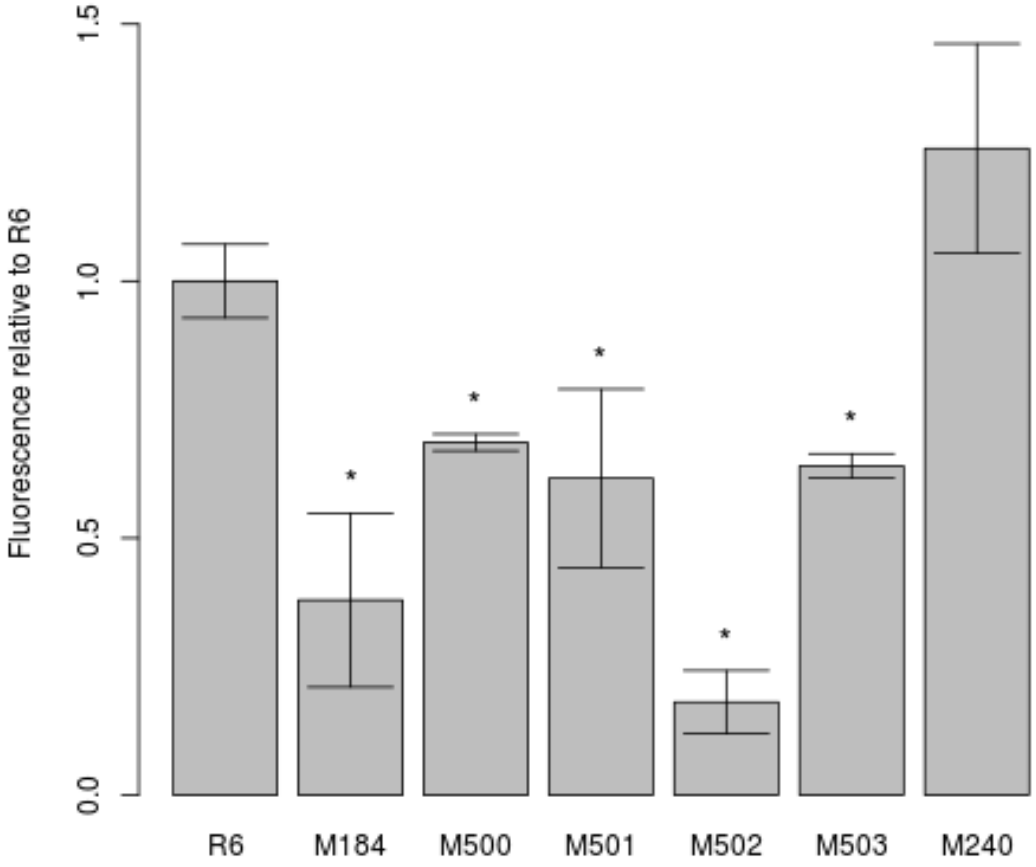


Figure 4.5. Ethidium bromide accumulation in M184 and four R6^{M184} transformants compared to R6. Bars represent average steady state fluorescence of three independent cell suspensions following 10 minutes of exposure to ethidium bromide (25 μM). Error bars represent standard deviation of three biological replicates. *, accumulates significantly lower levels of ethidium bromide than R6, $p<0.05$, one-tailed Student's t-test.

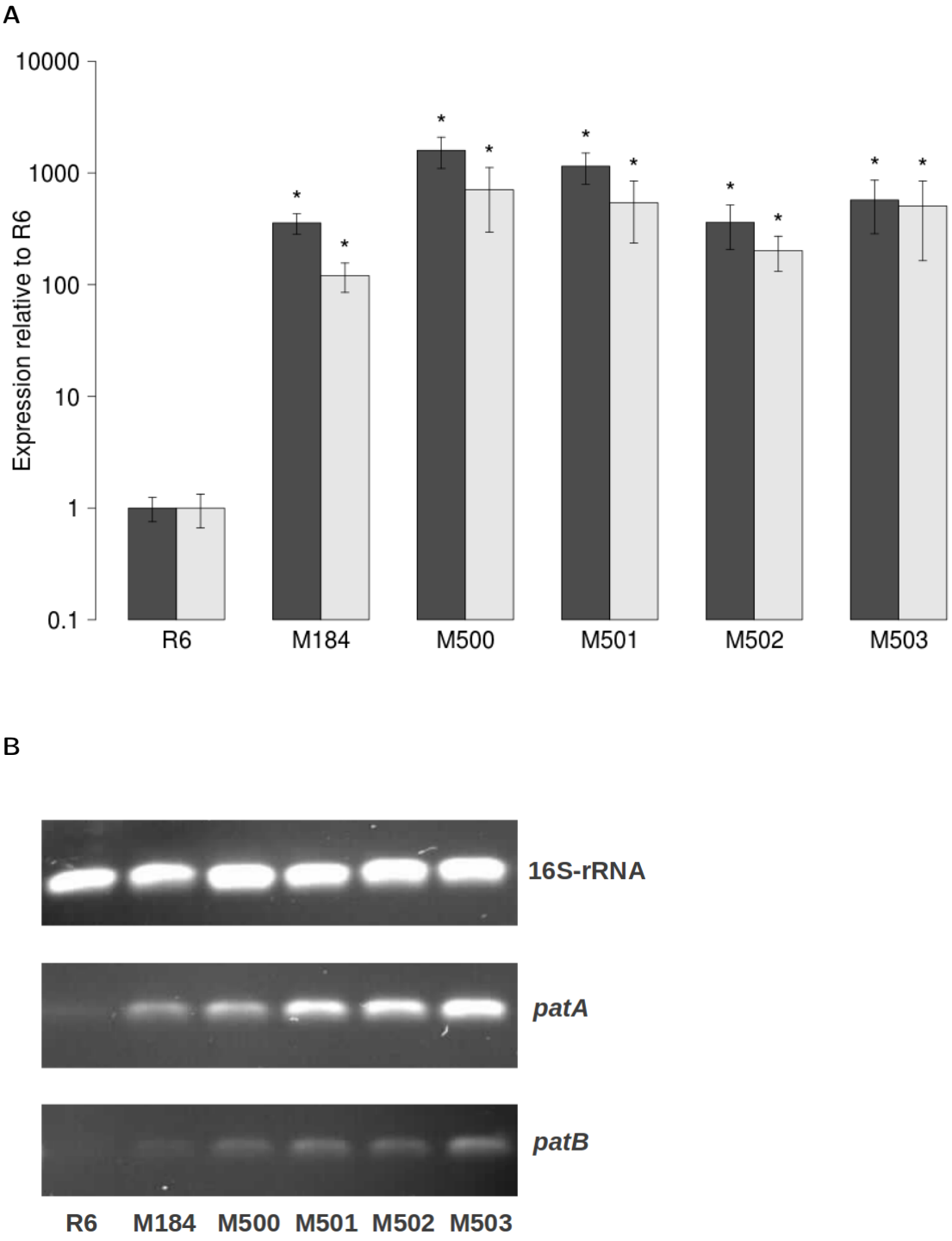


Figure 4.6. Expression of *patA* and *patB* from M184 and the four R6^{M184} transformants compared to R6 determined by qRT-PCR. (A) Fold expression of *patA* (dark bars) and *patB* (light bars) normalised to *rpoB*. *, Expression significantly higher than in R6, $p<0.05$, One-tailed Student's t-test on $\log_{10}$(Expression). **(B)** Amplimers obtained following 40 cycles of amplification of *patA*, *patB* and 16S rRNA from cDNA, visualised by gel electrophoresis.

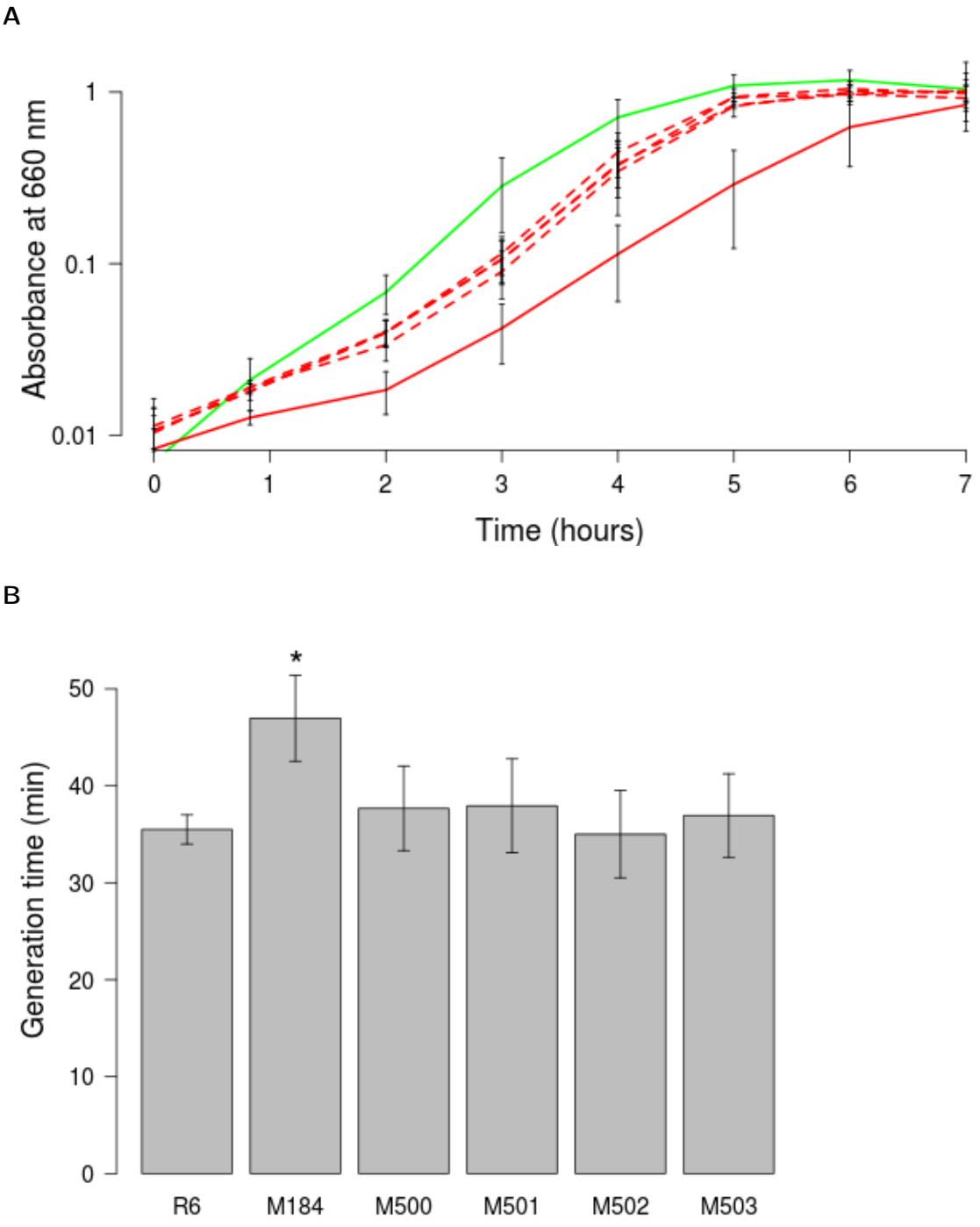


Figure 4.7. Growth of M184 and four R6^{M184} transformants in BHI compared to R6. (A) Absorbance (660 nm) of cultures of R6 (green line), M184 (solid red line) and four R6^{M184} transformants (dashed red lines). **(B)** Generation times per strain calculated from exponential phase growth between two and four hours. Error bars represent the standard deviation of three biological replicates. *, generation time significantly higher than that of R6, $p<0.05$, One-tailed Student's t-test.

causative mutation had not reverted during antibiotic-free culture of the transformants, genomic DNA from each transformant was transformed into R6 cells and transformants were selected on 8 µg/ml ethidium bromide. Ethidium bromide resistant colonies were recovered using DNA preparations from all four transformants, at frequencies equivalent to that obtained with M184 DNA (10^6 – 10^7), confirming that the transformants still carried the mutation(s) responsible for *patAB* over-expression.

4.6 Further analysis of genome sequencing data

As none of the mutations detected in M184 by the Maq analysis appeared to be responsible for the M184 phenotype, further analysis of the genome sequencing data was carried out to determine the genetic change causing upregulation of *patA* and *patB*. It was hypothesised that the causative mutation had been missed in the original SNP analysis, either due to a failure of the software, or due to the mutation being located in a region to which read mapping is complicated, such as a repetitive element.

4.6.1 Recalling SNPs using different software

Firstly, to rule out the possibility that the Maq software had failed to detect the causative SNP, the mutation detection analysis was repeated using the short-read aligner Bowtie2 (Langmead and Salzberg, 2012) to align reads against the R6 reference genome, and Samtools to identify variant genome positions. Identified variations were then filtered to keep only mutations with a Phred-scaled quality score of more than 50, resulting in 75 mutations in M184 and 52 in R6. These were compared with the mutations detected by Maq (figure 4.8). Of the 13 M184 mutations that had been detected by Maq that were subsequently found to be false positives by PCR and resequencing, only five were detected by the Bowtie2/Samtools approach. Visual inspection of read alignments

covering all SNPs detected by the two methods suggested that 17 of the mutations detected by Maq were false positives. Of these, only eight were also suggested to be SNPs by Bowtie2, suggesting that Bowtie2 detected fewer false positives. However, six of the mutations found by Maq, which appeared genuine on visual inspection, were not detected by Bowtie2. Five new mutations were detected by Bowtie2 that had not been found by Maq, four in both strains and one in M184 only. These new mutations were all small insertions and deletions, reflecting the superior indel detection algorithms in Bowtie2 compared to the older Maq program (Langmead and Salzberg, 2012).

One mutation was found in M184 by the Bowtie2 approach that had not previously been detected by Maq. This was a small insertion in the *pcpA* gene. A region surrounding this mutation was amplified from R6, M184 and the four $R6^{M184}$ transformants and sequenced by Sanger sequencing. The insertion was observed in M184, but not in R6 or any of the transformants. This suggests that the *pcpA* mutation is a genuine polymorphism in M184 relative to R6, but it is not the cause of over-expression of *patA* and *patB*.

4.6.2 Assembly of unaligned reads

A potential disadvantage of the genome resequencing approach used here is that it will not detect large insertions relative to the reference genome that are greater than the read length. Evidence of a potential recombination event affecting the *pbpX* gene had already been observed in the R6 and M184 sequences. It was therefore possible that other recombination events had occurred in the evolutionary history of R6 and/or M184 which could have led to insertion of extra DNA. This could mean that, either an insertion had occurred recently in M184 which caused *patAB* over-expression, or an insertion had occurred in the laboratory stock of R6 prior to selection of M184, and the mutation causing *patAB* over-expression was located in the inserted DNA. In support of these hypotheses, 8% of the M184 reads could not be aligned to the reference genome, which

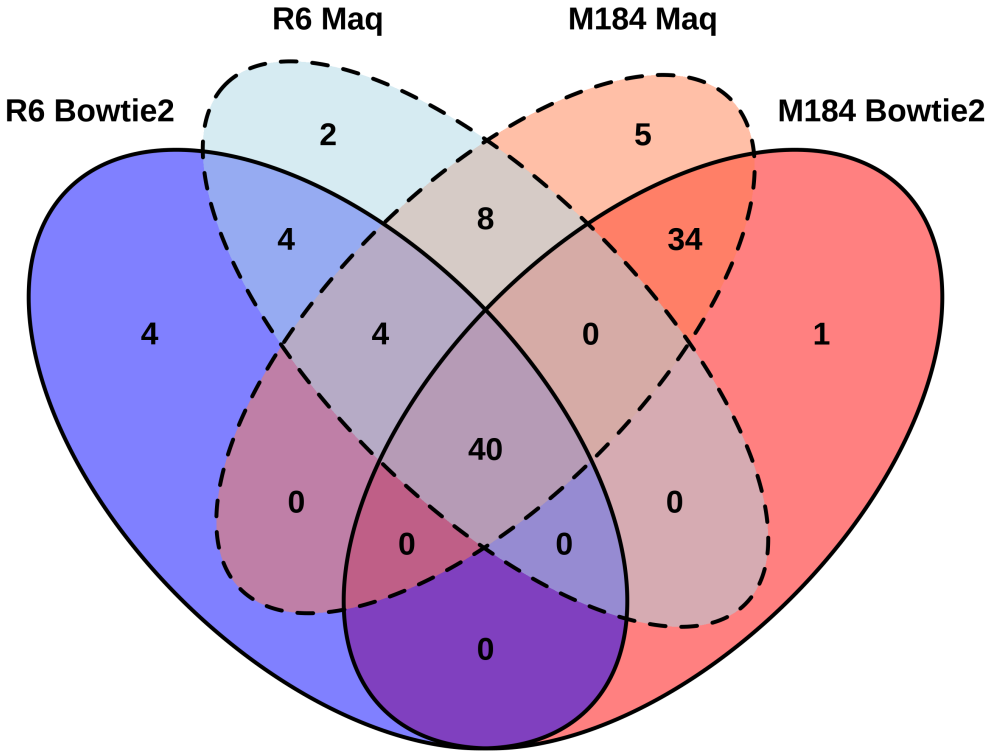


Figure 4.8. Comparison of numbers of mutations detected in M184 (red) and R6 (blue) from alignments of Illumina sequencing reads made by Maq (dotted border) and Bowtie2 (solid border).

equates to about 770,000 reads. Assuming the hypothetical extra DNA is covered to the same average depth as the rest of the genome, these extra reads could potentially represent 170 kb DNA. To investigate this, the M184 reads that did not align to the R6 reference were assembled using the *de novo* assembly software, Velvet (Zerbino and Birney, 2008). M184 reads were used for this analysis due to the higher overall quality of the reads.

Using a Velvet kmer value of 45, 24 contigs with length of 100 bp or greater were generated (Table 4.4). The 24 contigs were compared against the nr nucleotide database to determine the most similar sequence. One 1149 bp contig matched to an internal region of the *pbpX* gene from *S. pneumoniae* strain 670-6B (query coverage 100%, sequence identity 99%). This corresponded to the region of *pbpX* gene to which reads were not aligned in *S. pneumoniae* R6, implying that this region of the genome of the laboratory stock of R6 has undergone a recombination with a strain related to 670-6B. Twelve short contigs of median length 226 bp matched bovine microsatellite DNA, and one further contig of length 210 bp matched equine microsatellite DNA (median sequence identity 99%). Five contigs of median length 1129 bp matched transposase sequences from *Lactococcus lactis*, and three contigs of median length 554 matched sequences on lactococcal plasmids. Finally, a large 30 kb contig showed 79% sequence identity with the *Lactococcus lactis* lytic bacteriophage ASCC506, which was isolated from an Australian dairy factory.

Six sets of primers were designed to amplify regions of the putative bacteriophage sequence. However, no amplimers were observed when R6 or M184 genomic DNA was used as the template DNA. This suggested that the bacteriophage sequence was a contaminant of the original M184 genomic DNA preparation, and was not inserted into the R6 genome. The role of these extra sequences was therefore not investigated further. *S. pneumoniae* was routinely grown in rich medium derived from enzymatic digestion of beef products, meaning that the growth medium could have been the source of this

Table 4.4. Putative origins of contigs assembled from M184 reads that did not align to the published R6 genome.

Origin of sequence	Number of contigs	median length (bp)	median coverage
Bovine microsatellite DNA	12	226	273
Equine microsatellite DNA	1	210	145
Lactococcal phage	1	30774	301
Lactococcal plasmid	3	554	308
Lactococcal transposase	5	1129	181
Streptococcus pneumoniae pbpX	1	1149	229
Unknown	1	1083	192

extra DNA.

4.6.3 Sequencing of short repetitive regions

The *S. pneumoniae* genome contains many short interspersed sequences, such as BOX elements, repeat unit of pneumococcus (RUP) elements, insertion sequence (IS) elements and *S. pneumoniae* Rho-independent terminator-like elements (SPRITEs) (Croucher *et al.*, 2011b; Hoskins *et al.*, 2001). Altogether, these repetitive elements comprise more than 3% of the genome, which is an unusually high percentage compared to other sequenced bacteria (Hoskins *et al.*, 2001). It has been suggested that some pneumococcal short interspersed repeats can affect expression of downstream genes. For example, a BOX element upstream of the *com* operon was shown to activate gene expression (Knutsen *et al.*, 2006).

The positions of BOX, RUP and SPRITE elements in the R6 genome have been systematically determined by Croucher *et al* (2011b). Inspection of the alignment of R6 and M184 reads against the R6 reference genome showed that many of these regions have increased coverage relative to the surrounding genes due to reads from repetitive regions aligning to the genome multiple times. A SNP occurring in one of these repetitive regions could therefore be missed by the SNP calling software due to the mis-alignment of reads from multiple copies of the repeat. To confirm whether M184 and its transformants contained any mutations in short interspersed repeats, the sequences of 80 repeat regions were determined by PCR and Sanger sequencing from R6 and M184. Repeats were chosen based on their location and sequencing coverage, according to the following criteria. Candidate repeats were located between pairs of genes transcribed either in the same direction or divergently, while repeats located between convergently transcribed genes were excluded. Additionally, the average depth of reads mapped to candidate repeats was more than twice that of the flanking genes. For each candidate

repeat, primers were designed that bound in the unique sequence either side of the repeat (Appendix 2). Repeats were amplified by PCR from M184 and R6 DNA, and amplimers were sequenced by Sanger sequencing. No differences between R6 and M184 were found in any of the tested repeats.

4.7 Identification of mutations in R6^{M184} transformants by new whole genome sequencing

The results so far confirmed that transfer of genetic material from M184 to R6 confers *patAB* over-expression, but the SNPs and small indels identified in the M184 genome sequence are not the cause of this phenotype, and instead seem more likely to be compensatory mutations, or bystander mutations that are unrelated to the phenotype. This raised two hypotheses:

Hypothesis 1: As upregulation of *patAB* in the transformants is caused by incorporation of genetic material from M184, it was hypothesised that the R6^{M184} transformants and M184 should all contain the same genetic change relative to R6.

Hypothesis 2: If, as postulated, the mutations previously found in M184 have arisen to compensate for a fitness cost imposed by the over-expression of *patA* and *patB*, then the R6^{M184} should have also acquired mutations affecting similar systems.

To investigate these hypotheses, further genome sequencing was carried out on three of the four R6^{M184} transformants. Due to advances in Illumina technology since the original M184 sequence was obtained, longer read lengths and greater sequencing depth were possible.

4.7.1 Genome sequencing

Genomic DNA extracted from R6, M184 and three of the four R6^{M184} transformants (M500, M501 and M503) was sequenced by Genepool (Edinburgh, UK) using the Illumina HiSeq system. For each strain approximately 25 million 100 bp read pairs were obtained and quality was assessed using FastQC. For all five read sets, average per base sequence quality was good (>28) for the majority of the read length, but dropped to average (20 - 28) in the final 20 bp. The second reads for all five read sets all showed poor average quality (<20) in the final five bases of the read. To avoid poor quality bases confounding the SNP analysis, all reads were trimmed to 75 bp using a Python script written by Nick Loman (University of Birmingham), before aligning against the published R6 genome using Bowtie2. Variations were then identified using Samtools and, as before, mutations detected in the *pbpX* region were discarded. In contrast to the SNP sets identified by the Bowtie2 approach from the 50 bp sequencing, fewer than 80 mutations for each strain were identified with no filtering applied. Therefore, in the first instance, the raw SNP sets were used for comparison with the initial sequencing data.

4.7.2 Reanalysis of M184 and R6 mutations using new sequencing data

Sixty mutations were identified in the new R6 sequence relative to the published R6 genome sequence. This included 47 of the 58 mutations previously detected by Maq in the previous genome sequence (Section 4.3.2.1 and Figure 4.9A). Eleven mutations detected by Maq were not found in the new analysis, but thirteen additional mutations were found that had not been observed previously.

Seventy-nine mutations were detected in the new M184 genome sequence, of which 49 were also found in R6. These were excluded to give a set of 30 mutations unique

to M184, of which 25 had been detected in the previous analysis (Figure 4.9B). Five additional mutations were found that had not been detected previously, while fourteen of the mutations found by Maq were not detected in the new analysis.

4.7.3 Mutations detected in R6^{M184} transformants

The mutations detected in the three R6^{M184} transformants were compared with those found in R6 and M184 (figures 4.9C and 4.9D). Fifty-four of the 60 mutations found in R6 were also detected in all three of the transformants. Mutations were found in all three transformants that were not detected in R6 (six in M500, four in M501 and 20 in M503; summarised in table 4.5), but none of these were shared between all three transformants and M184 (figure 4.9D). The cluster of ten mutations previously found in *gltX* in M184 was also found in M503, suggesting that a recombination event may have occurred in this region. However, these mutations were not found in the other two transformants, and the lack of recovery of ethidium bromide-resistant transformants when R6 cells were transformed with a *gltX* PCR amplimer from M184 (section 4.5.1) suggests that this cluster of mutations is not responsible for *patAB* over-expression.

4.8 Detection of a novel gene duplication by analysis of improperly mapped read pairs

The previous analyses had suggested that over-expression of *patAB* in M184 and the R6^{M184} transformants was not caused by a point mutation in either the unique part of the genome or in a repetitive region, and could also not be explained by acquisition of foreign DNA. However, a possiblity not yet investigated was a large genomic rearrangement in M184 and the transformants relative to R6. Adaptive genomic rearrangements have been observed previously in bacteria. For example, inversions and duplications of parts

Table 4.5. Mutations detected in three sequenced R6^{M184} transformants

	Position	Base change	Mutation type	Gene affected			quality	depth
				Gene	*Description*	*protein change*		
M500	141189	C-A	int	-	-	-	222	1042
	541249	A-G	syn	*murM*	Ser/Ala adding enzyme, peptidoglycan biosynthesis	-	222	794
	792564	A-T	int	-	-	-	207	455
	1052448	C-A	syn	spr1057	hypothetical protein	-	222	570
	1282189	C-T	int	-	-	-	222	279
	1826248	G-T	non-syn	spr1850	hypothetical protein	Gln61Lys	222	1331
M501	599258	C-A	non-syn	spr0583	hypothetical protein	Ala78Glu	205	646
	635793	G-T	non-syn	*lctO*	Lactate oxidase	Val277Leu	222	2807
	969735	G-T	non-syn	*obgE*	GTPase implicated in regulation of chromosomal replication	Ala12Ser	222	1835
M503	391425	G-A	syn	*gatA*	aspartyl/glutamyl-tRNA amidotransferase subunit A	-	222	5815
	599258	C-A	non-syn	spr0583	hypothetical protein	Ala78Glu	222	2470
	635793	G-T	non-syn	*lctO*	Lactate oxidase	Val277Leu	222	5268
	969735	G-T	non-syn	*obgE*	GTPase implicated in regulation of chromosomal replication	Ala12Ser	222	4029
	1012631	C-T	non-syn	*glgC*	glucose-1-phosphate adenylyltransferase	Thr149Ile	222	3048
	1426032	C-T	syn	*oxlT*	oxalate:formate antiporter	-	222	2103
	1474250	C-G	non-syn	-	hypothetical protein	Gly284Arg	222	2834
	1550514	C-T	non-syn	*scrR*	sucrose operon repressor	Arg106Cys	222	4613
	1643773	G-A	non-syn	*adhB*	zinc-containing alcohol dehydrogenase	Pro84Leu	159	4012

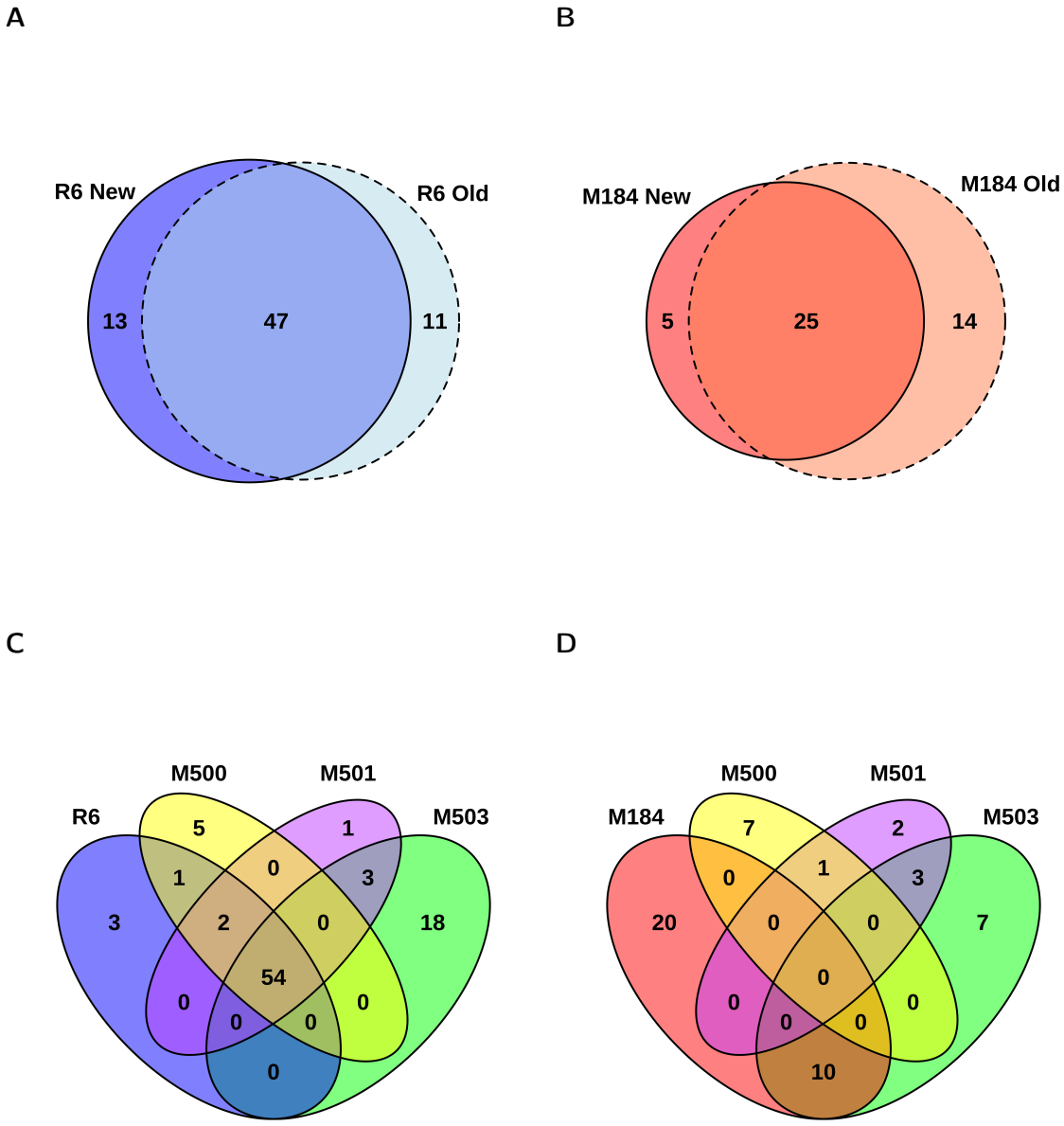


Figure 4.9. Comparison of mutations found in new and old sequences of R6 and M184, and three R6^{M184} transformants. **(A)** Mutations found in new and old sequences of R6. **(B)** Mutations found in the new and old sequences of M184. **(C)** Mutations in three R6^{M184} transformants compared to R6. **(D)** Mutations in three R6^{M184} transformants compared to R6 and M184.

of the *E. coli* chromosome were observed in mutants selected for suppression of a growth deficiency caused by dis-regulated transcription initiation (Skovgaard *et al.*, 2011).

To investigate this, the alignment of M184 reads against the reference genome was examined for a larger scale genomic rearrangement. Specialised software exists to detect genomic rearrangements, such as BreakDancer (Chen *et al.*, 2009) and CNVer (Medvedev *et al.*, 2010), but these are designed and optimised for detecting rearrangements in diploid eukaryotic genomes. Therefore, as the *S. pneumoniae* genome is relatively small, putative genomic rearrangements were identified more simply by filtering for "improperly" mapped read pairs. Illumina read pairs are derived from either end of a fragment of DNA of a known average size, in this case 250 bp. It was therefore expected that both members of a read pair should map to opposite strands of the genome within a defined distance of each other, facing each other in the 5' to 3' direction (Figure 4.10A). Improper pairs were defined as pairs of reads not meeting these criteria (Figure 4.10B-4.10E). To identify genomic regions with unusually high concentrations of improperly paired reads, read alignments were filtered to remove properly paired reads using Samtools, then filtered alignments were visually inspected using Artemis (Rutherford *et al.*, 2000).

All five read alignments showed regions containing high numbers of improperly paired reads corresponding to repetitive regions such as the rRNA loci, IS elements and degenerate transposases. However, two interesting regions of improperly mapped pairs were observed in the M184 alignment that were not seen in R6. The first was in two genes annotated as *hsdS*, spr0445 and spr0448, which encode Type I restriction-modification enzyme subunits. Although these *hsdS* genes were similar, they were not identical and the improper mapping of read pairs observed in M184 suggests that a recombination has occurred between these two genes in M184. However, improper mapping of reads to the *hsdS* genes was not observed in the $R6^{M184}$ transformants, so this putative recombination was not analysed further. The second cluster of improperly paired reads

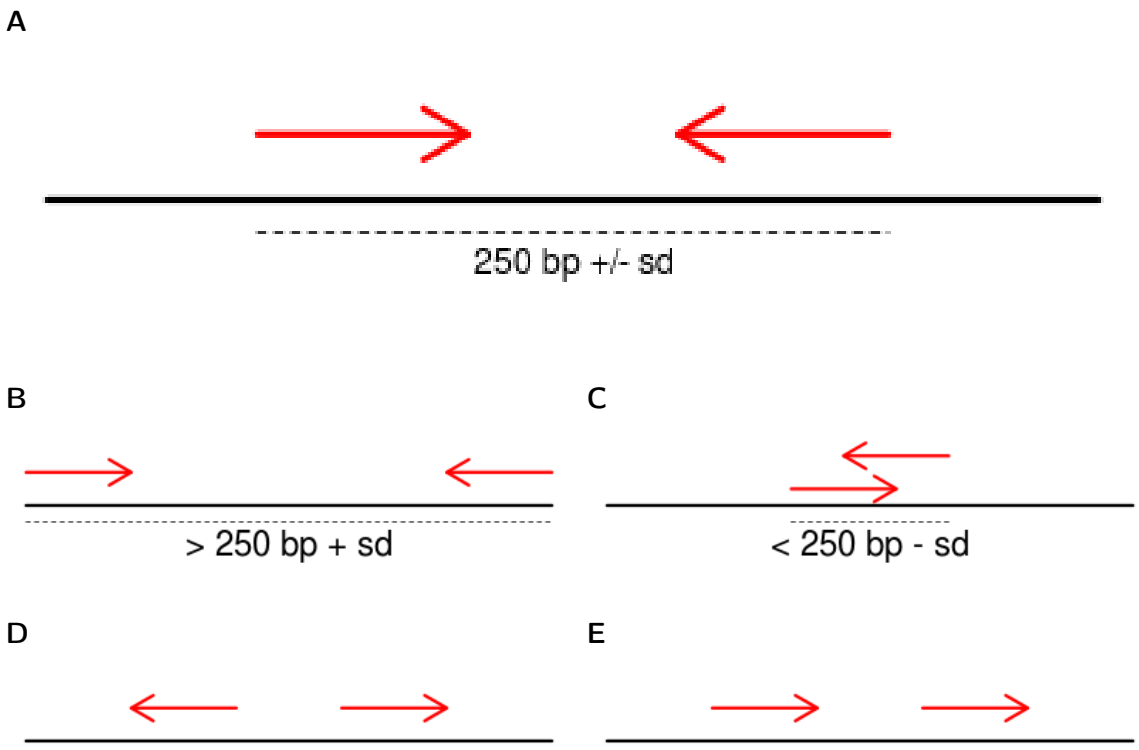


Figure 4.10. Possible mappings of Illumina read pairs (red arrows) to a reference genome (black line). (A) Read pairs are expected to map facing each other with an average insert size (dotted line) of 250 bp. Genomic rearrangements may cause improper mapping of read pairs by **(B)** increasing or **(C)** decreasing the insert size, or by causing reads to map **(D)** divergently or **(E)** to the same DNA strand. sd, standard deviation

mapped between *patA* and the upstream *hexA* gene in M184 (figure 4.11). This region of improperly paired reads was also observed in all three of the transformants. The absolute number of reads mapping to the *hexA-patA* intergenic region (genomic position 1865730-1865874) in M184, normalised for the total number of reads mapped, was 63 reads per million reads mapped, compared to 27 reads per million reads mapped in R6. This approximate doubling of normalised read depth suggests that the *hexA-patA* intergenic region may be duplicated in M184.

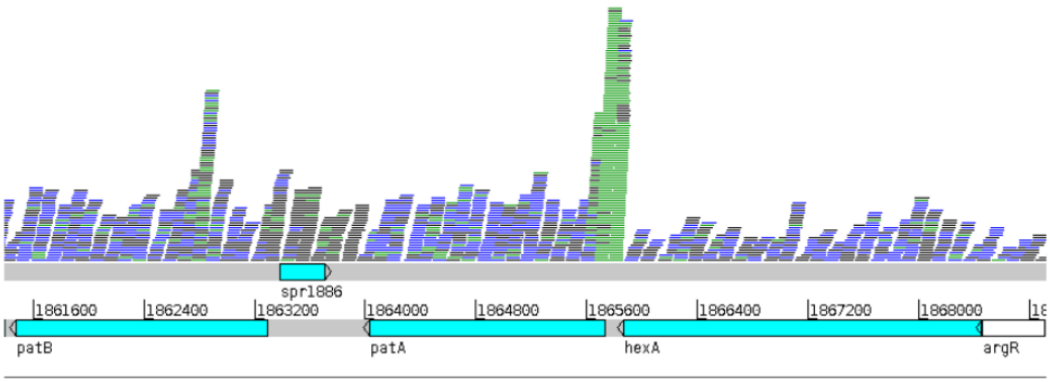
4.8.1 Investigation of putative rearrangement in the *patA* region of M184 and its transformants

The mapping of improperly paired reads in the intergenic region between *hexA* and *patA* suggests that a genomic rearrangement may have occurred involving the upstream region of *patA*. Three approaches were taken to investigate this further. Firstly, draft genomes were assembled for M184 and R6 and contigs containing *patA* and *hexA* were compared. Secondly, a local assembly was generated using the reads mapping to the intergenic region as improper pairs. Finally, the relative copy numbers of all genes in the R6 and M184 genomes were compared.

4.8.1.1 *De novo* assembly of R6 and M184 genomes

Draft genomes for R6 and M184 were assembled using the Velvet short-read assembly software (section 2.5.3). A range of Velvet kmer values were tested, and assemblies were ranked according to the size of the largest contig and the median length-weighted contig length (N50; Zerbino and Birney, 2008). Using a kmer size of 51, the R6 reads could be assembled into 1560 contigs, with an N50 of 5004 and maximum contig size of 22674, using 89% of the reads. The M184 reads were assembled into 508 contigs, with an N50 of 54708 and a maximum contig size of 168913, using 97% of the reads.

A



B

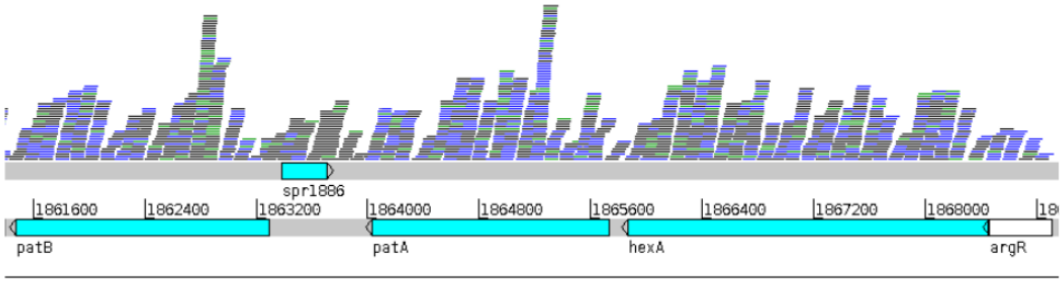


Figure 4.11. Alignment of improperly paired reads from (A) M184 and (B) R6 against the published R6 genome. Viewed in Artemis.

Megablast (Altschul *et al.*, 1997) was used to search both assemblies for contigs containing sequences matching *patA*, *hexA* and the intergenic region between the two. In M184, *patA* and *hexA* sequences were found on two contigs which overlapped in the intergenic region, while in R6 all three query sequences were only found once, on one contig spanning the whole *hexA*-*patA* region. The fact that the *patA* and *hexA* contigs could not be joined by Velvet in the M184 assembly suggests that a rearrangement has occurred in this genome, possibly involving duplication of the intergenic region between *patA* and *hexA*. However, it was not possible to determine the nature of the rearrangement from this assembly as sequences matching the intergenic region were not found in any other contigs in the assembly.

4.8.1.2 Realignment and local assembly of improperly mapped read pairs

Analysis of read depth in the intergenic region between *patA* and *hexA* suggested that this sequence had been duplicated elsewhere in the M184 genome. To identify the location of the postulated second copy, the improperly paired reads that aligned to the *hexA*-*patA* intergenic region were extracted from the read alignment using Samtools. The corresponding pair of each read was then extracted from the original sequencing data files using a custom Python script. These extracted reads were then re-aligned against the R6 reference genome using Bowtie2 in single-end mode. In the resulting alignment reads mapped to five distinct genomic regions. These corresponded to the intergenic region between *hexA* and *patA*, and regions about 400 bp upstream of each of the four copies of the 16S rRNA gene. The read extraction and alignment process was repeated for the three $R6^{M184}$ transformants and the same pattern was observed. This suggests that a rearrangement has occurred in M184 causing the upstream region of *patA* to be duplicated in close proximity to one or more of the rRNA loci.

To further determine the nature of this rearrangement, a local assembly was performed

with Velvet using the same set of extracted M184 reads. Using a kmer value of 69, a single contig of 608 bp was assembled using 88% of the reads. The contig was compared against the R6 genome using Megablast. As predicted, 389 bp of the 3' end of the contig matched the region upstream of *patA* and 373 bp at the 5' end matched regions upstream of the rRNA loci. The 5' end of the contig matched most closely to rRNA loci two and four (285 out of 286 bp identical; figure 4.12). When compared to the sequence upstream of copies one and three of the rRNA locus, the sequence identity was slightly lower (284 out of 286 bp). The predicted join between the two distinct parts of the contig occurred at 182 bp upstream of the *patA* start codon (36 bp from the end of *hexA*) and 453 bp upstream of the 16S rRNA genes.

A PCR assay was used to confirm the presence of the predicted genomic rearrangement in M184 and the R6^{M184} transformants. To determine wich rRNA locus was associated with the rearrangement, primers were designed which bound uniquely upstream of each rRNA locus (primers rRNA-1, rRNA-2, rRNA-3 and rRNA-4). These were combined with a primer that annealed to the 5' end of *patA* to attempt to amplify a region of DNA corresponding to the contig predicted by the local assembly (Figure 4.13A). No amplimers were obtained using any primer combination when R6 DNA was used as the template, but an amplimer was observed from the primer combination corresponding to rRNA locus four (primers 1772 and rRNA-4) when M184 DNA was used. Amplimers were also obtained when this PCR was repeated using DNA from the four R6^{M184} transformants (Figure 4.13B).

This, combined with previous results suggesting that the *hexA-patA* intergenic region is duplicated, suggests that a duplication of part of the genome has occurred in M184 and the transformants. The 5' extent of this duplication maps to 182 bp upstream of *patA*, and a region of unknown size has been duplicated between spr1880 and the fourth copy of the 16S rRNA gene.

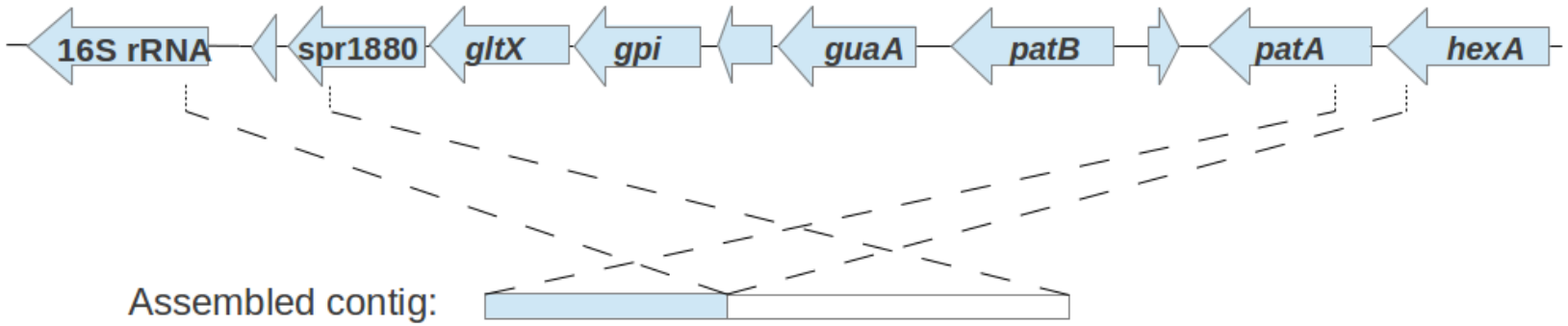


Figure 4.12. Comparison of contig assembled from M184 improperly mapped read pairs against the R6 genome. Regions of matching sequence (as determined by Blast) are indicated by dashed lines.

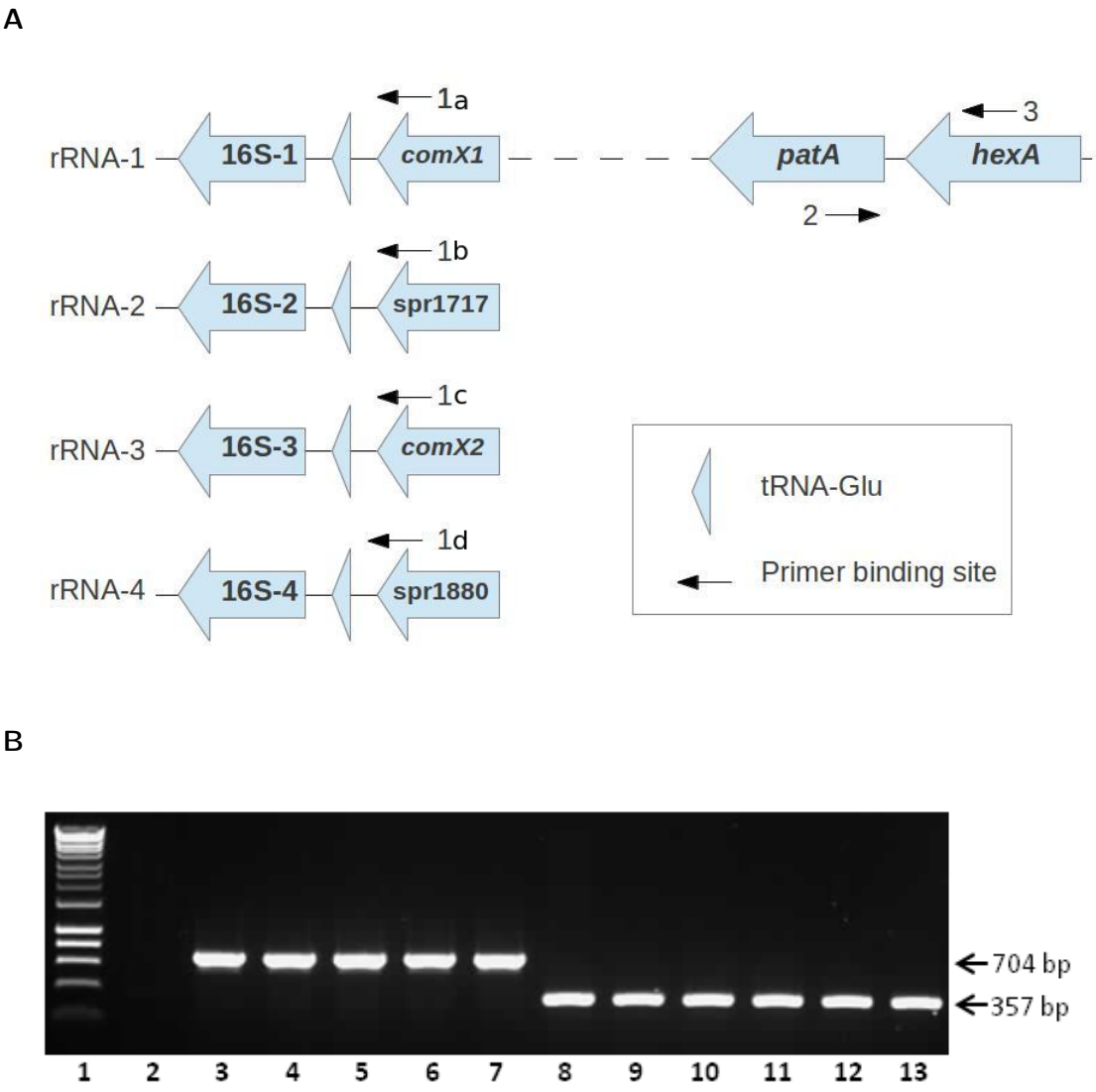


Figure 4.13. Investigation of a putative rearrangement in M184. (A) Location of primers in *patA* and rRNA loci. **(B)** amplification of rearranged region from R6, M184 and four $R6^{M184}$ transformants. Well 1: Hyperladder 1kb (Bioline). Wells 2-7: amplimers obtained from PCR using primers 1d and 2. Wells 8-13: amplimers obtained from control PCR using primers 2 and 3. Template DNA used was R6 (wells 2 and 8), M184 (wells 3 and 9), M500 (wells 4 and 10), M501 (wells 5 and 11), M502 (wells 6 and 12) or M503 (wells 7 and 13) genomic DNA preparations.

4.8.1.3 Comparative read depth analysis to determine extent of putative duplication

To determine the 3' extent of the putative duplication, the relative copy number of each gene in M184 was estimated compared to R6. To do this, normalised read depths, expressed as reads per kilobase per million reads mapped (RPKM), for each gene larger than 250 bp from M184 were plotted against those from R6 (figure 4.14). Interestingly, seven genes had RPKM values in M184 that were approximately two times higher in M184 than genes in R6. These genes were were co-located in a near-contiguous locus and consisted of *patA* (spr1887), spr1886, *patB* (spr1885), *guaA* (spr1884), *gpi* (spr1882), *gltX* (spr1881) and spr1880. Spr1883 was not included in the initial RPKM comparison as it is shorter than 250 bp in length. RPKM comparisons were repeated using per gene RPKM values from the three sequenced $R6^{M184}$ transformants and the same pattern was observed (figure 4.15). This suggests that the genetic change responsible for over-expression of *patA* and *patB* in M184 and its derivatives is a duplication of a 9.2 kb genomic region including all genes between *patA* and spr1880.

4.8.1.4 PCR analysis of putative duplication

The PCR analysis carried out previously demonstrated that there is a join present in M184 and its derivatives between the upstream region of *patA* and the region between spr1880 and rRNA locus 4. This suggests that the *patA*–spr1880 region has been tandemly duplicated in place, resulting in two copies of the locus side by side in the genome. To confirm whether this is the case, the whole duplicated region was amplified by PCR. To do this, a primer binding upstream of the duplicated region in *hexA* (primer 1771) was combined with a primer binding within *patA* (primer 1772) in a PCR cycle with an extension time of ten minutes (Figure 4.16). Using a DNA template containing the hypothesised duplication of the *patA* – spr1880 region, this reaction would result in two

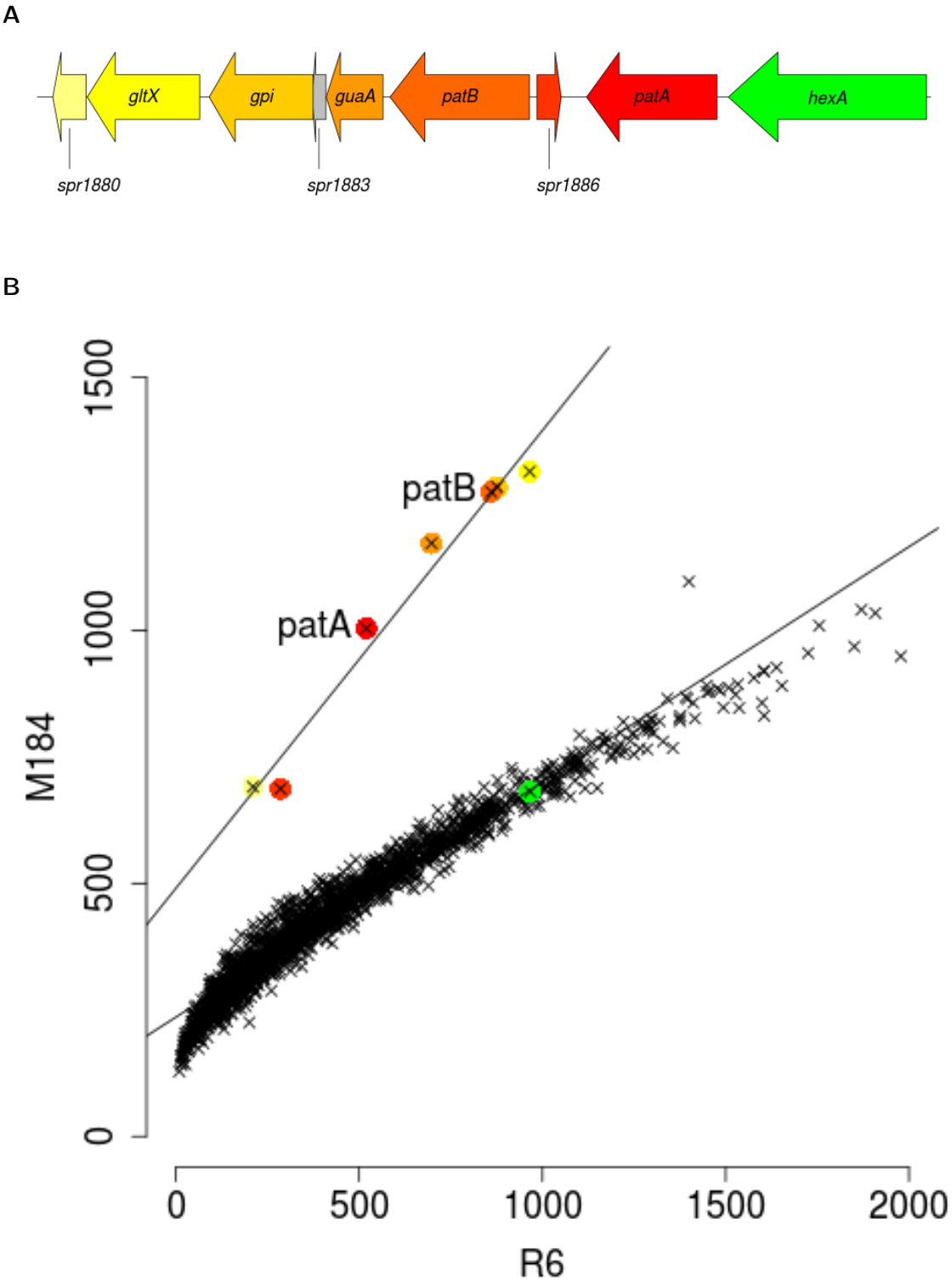


Figure 4.14. Comparisons of per gene RPKMs of M184 against R6. (A) Representation of genomic region around *patA* and *patB* **(B)** Per-gene RPKMs from M184 plotted against those of R6. Points representing genes included in *patAB* genomic region are coloured as in (A).

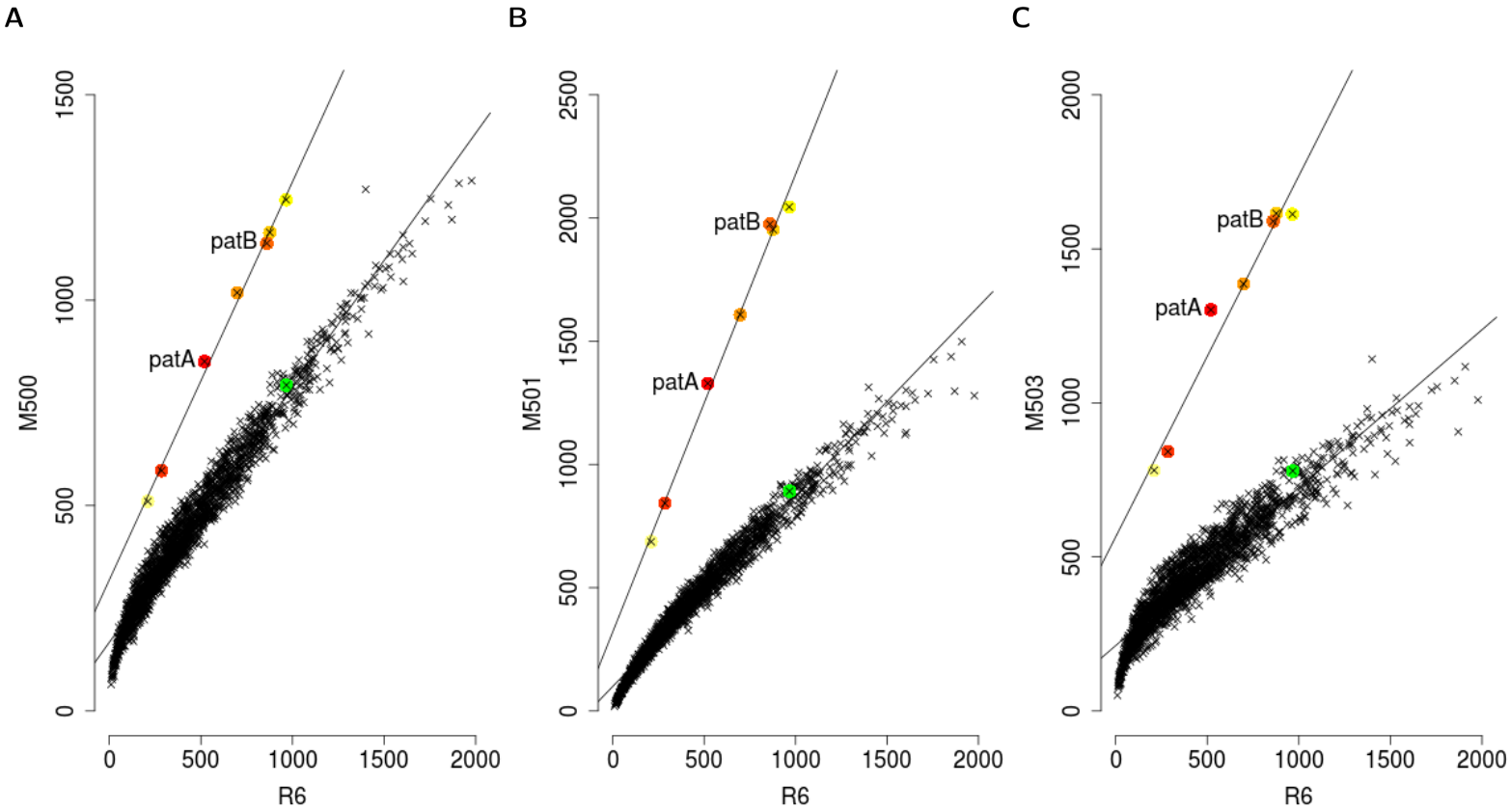


Figure 4.15. Comparisons of per gene RPKMs of (A) M500, (B) M501 and (C) M503 against R6. Points are coloured as in Figure 4.14

amplimers, one short amplimer corresponding to amplification between *hexA* and the first copy of *patA*, and one amplimer that is 9.2 kb larger corresponding to amplication from primers annealing to the second copy of *patA*. In contrast, from template DNA without the postulated duplication, only the short amplimer would be observed. As expected, only the short amplimer was observed when R6 DNA was used as a template. However, both short and long amplimers were obtained when DNA from M184 or the $R6^{M184}$ was used, confirming the hypothesis that the *patA* – spr1880 locus is tandemly duplicated in these strains. The large PCR amplimer from M184 was extracted from the agarose gel and sequenced from either end to confirm that it corresponded to the predicted sequence.

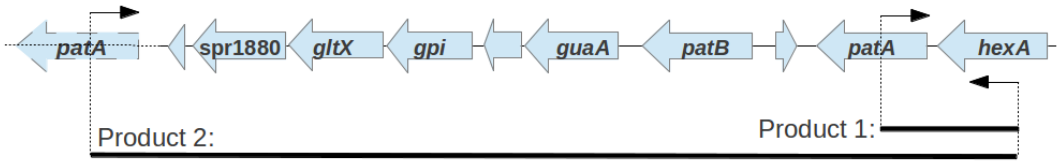
M168 and nine of the *patAB*-over-expressing clinical isolates described in chapter 3 were screened for the presence of the duplication by PCR for the duplication junction (primers rRNA-4 and 1772). No amplimers were obtained from any of the tested isolates.

4.8.2 Measuring segregation rate of duplication

Several examples of gene duplications have been observed in other bacteria (Sandegren and Andersson, 2009). In all cases, gene duplications were unstable and were lost from the genome at varying rates in the absence of selective pressure. This instability is thought to be caused by RecA-dependent homologous recombination between copies of repeated regions, which results in deletion of one copy (Andersson and Hughes, 2009).

To determine the stability of the *patA*-spr1880 duplication the rates of loss of norfloxacin resistance from populations of M184 and the $R6^{M184}$ transformant M502 were measured following growth in BHI broth in the absence of a selective agent. After zero and 12-14 hours of exponential growth, approximately 60 colonies per strain were tested for norfloxacin resistance. The proportions of the samples that were norfloxacin resistant before and after the growth period were compared to calculate percentage retention

A



B

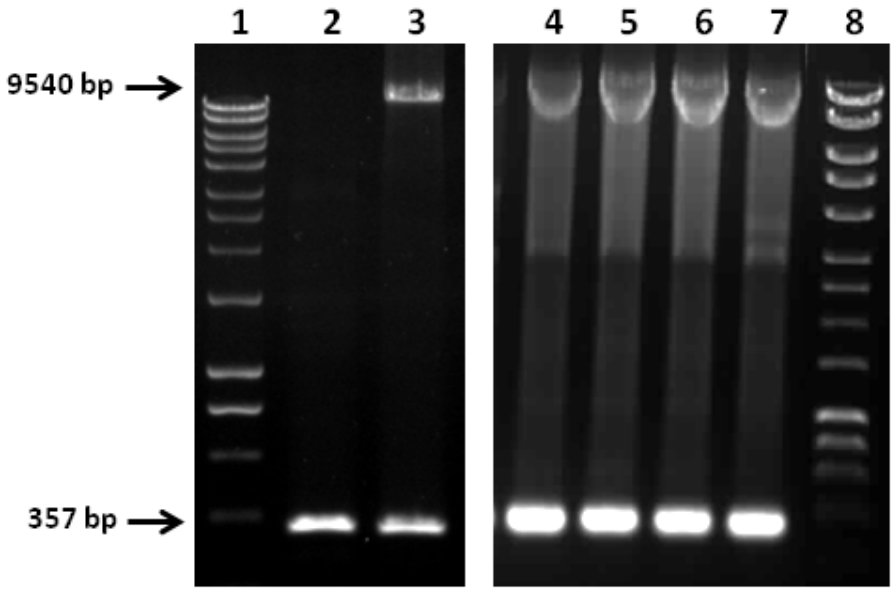


Figure 4.16. Investigation of putative patA-spr1880 tandem duplication by PCR. A Location of PCR primers in *patA* and *hexA*, assuming the existence of a tandem duplication of the *patA*-spr1880 region. **B** Amplification from R6 (well 1), M184 (well 2), M500 (well 3), M501 (well 4), M502 (well 5) and M503 (well 6) genomic DNA using the primers indicated in (A) and an extension time of ten minutes. Wells 1 and 8 contain Hyperladder 1kb (Bioline)

of norfloxacin resistance (P_{ret}). This was used to calculate the probability of loss of norfloxacin resistance per cell per generation, $p(loss)$, and the half life of the norfloxacin resistance within the population, using the equation $P_{ret} = (1-p)^n$ where n is the number of generations (section 2.7; Brochet *et al.*, 2008). Calculated values and 95% confidence intervals (CIs) are shown in Table 4.6.

Norfloxacin resistance was retained in approximately 76% of the M184 population (95% CI: 66-89%) and 71% (95% CI: 58-86%) of the M502 population, giving estimated probabilies of loss per cell per generation of 0.017$\pm$0.010 and 0.015$\pm$0.009, respectively. The half-life of the duplication within the populations was estimated to be between 17 hours and three days. The calculated confidence intervals of these measurements are large ($\pm$15 to 20% of $log(P_{ret})$) as 60 is a relatively small sample size. The dependence of confidence interval on sample size for different final proportions of norfloxacin-resistant cells is shown in Figure 4.17. To obtain a 95% confidence interval of $\pm$5%, using the formula $CI = \pm z_{\alpha/2}\sqrt{Var(log(P_{ret})}$ (section 2.7 of Materials and Methods), a sample size of about 1500 colonies would be needed, assuming that the percentage of norfloxacin resistant cells in the start population is about 100% and is $\geq$ 50% in the end population.

To confirm that loss of norfloxacin resistance correlated with loss of the *patA*-spr1880 duplication, seven norfloxacin resistant and seven norfloxacin sensitive isolates were tested for presence or absence of the duplication junction by PCR. PCR amplimers corresponding to the duplication junction were obtained for all of the resistant isolates, but not from the sensitive isolates.

Table 4.6. Retention of norfloxacin resistance by M184 and M502 in the absence of antibiotic selection

Strain	Num gens	Colony counts (norR/total)		% retention		p(loss)	Half life (days)	
		start	end	average	CI	average±CI	average	CI
M184	15	57/58	45/60	76	66-89	0.017±0.010	1.3	0.8-2.9
M502	22	56/60	39/59	71	58-86	0.015±0.009	1.2	0.7-2.7

CI, 95% confidence interval

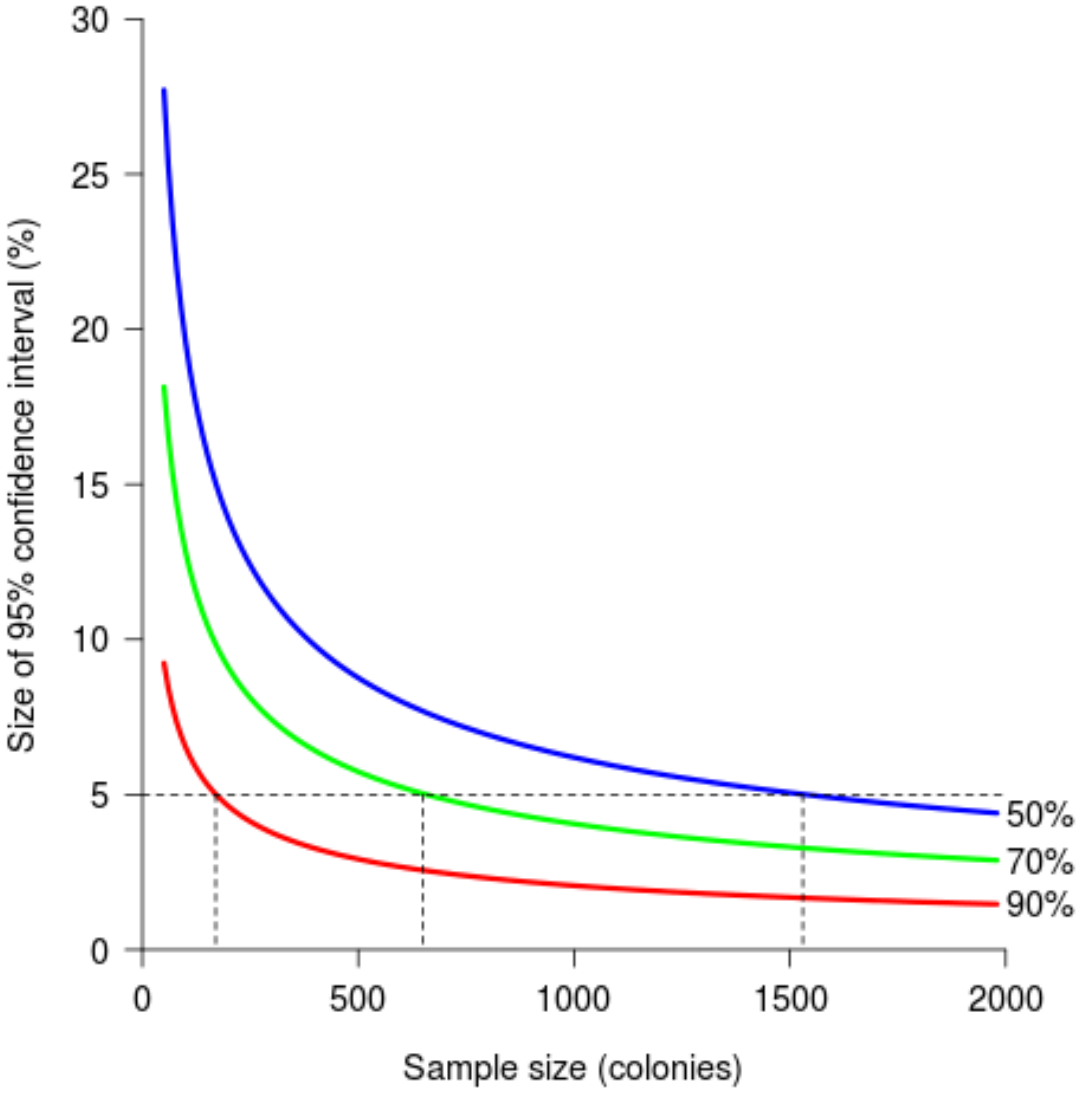


Figure 4.17. Dependence of the precision of measurement of the percentage retention of norfloxacin resistance in populations (P_{ret}) on sample size. Precision of measurement is expressed as ± a 95% confidence interval. The sample size required to measure P_{ret} to a precision of ± 5% (indicated by dashed lines) varies depending on the true value of P_{ret} in the population. Red line, P_{ret}=90%; Green line, P_{ret}=70%; Blue line, P_{ret}=50%

4.9 Effect of gene duplication on expression of *patA* and *patB*

It might be expected that duplication of the *patA* and *patB* genes would lead to a doubling of their expression. However, expression levels of *patA* and *patB* in M184 and the R6^{M184} transformants were determined previously (section 4.5.3) and found to be 100- to 1000-fold higher than in R6. This is higher than would be expected if the over-expression of these genes were caused by the duplication alone, and implies that the formation of the duplication must alter the regulation of *patA* and *patB* as well as changing their copy number. Three hypotheses could be suggested to explain this observation.

Hypothesis 1: The observed over-expression of *patAB* could be due to read-through from a highly expressed gene at the end of the first copy of the duplicated region into the second copy of the *patAB* locus.

Hypothesis 2: Although the duplicated region contains the intergenic region between *patA* and *hexA*, which includes the predicted *patA* promoter, there may be an important *cis*-regulatory sequence, such as a repressor binding site, located upstream of the 5' end of the duplicated region that is not included. Lack of this sequence upstream of the second copy of the duplication could cause unregulated expression of the second copy of *patAB*.

Hypothesis 3: Alternatively, the unexpectedly high levels of *patAB* expression observed could be due to a positive feedback loop activated by a small initial increase in the level of the PatAB transporter caused by the duplication itself.

To test these three hypotheses expression levels of other genes in the duplicated region were measured, and an attempt was made to inactivate individual copies of *patA*.

4.9.1 Measuring expression of genes contained within the duplication

As well as *patA* and *patB*, the duplicated region in M184 includes six other genes (figure 4.14A) and it is possible that transcription of these could also be changed by the duplication. Expression levels of five genes within the duplicated region, spr1886, spr1884 (*guaA*), spr1882 (*gpi*), spr1881 (*gltX*) and spr1880, were measured relative to *rpoB* as described for *patA* and *patB*. Expression of *hexA*, which is directly upstream of *patA* and outside of the duplicated region, was also measured. Transcription of spr1883, which is a small annotated open reading frame of unknown function that contains several regions of low-complexity sequence, could not be assessed as suitable qPCR primers could not be designed for this gene.

Measured expression levels for the tested genes are shown in Figure 4.18B. The gene immediately following *patB*, *guaA*, was significantly upregulated relative to R6 in M184 and the four R6^{M184} transformants ($p<0.05$, two-tailed Student's t-test of $log_{10}(expression)$ with Bonferroni correction for multiple testing). Observed levels of over-expression were 30$\pm$5-fold in M184, 59$\pm$18-fold in M500, 50$\pm$10-fold in M501, 26$\pm$8-fold in M502 and 69$\pm$44-fold in M503. Levels of cDNA corresponding to spr1886, which is encoded between *patA* and *patB* on the opposite DNA strand, were also significantly higher in transformants M500 and M501 than in R6 (21$\pm$8-fold and 12$\pm$2-fold respectively; $p<0.05$). The *gpi* gene was also slightly over-expressed in M500, M501 and M503 (4$\pm$1-fold, 4$\pm$1-fold and 5$\pm$2-fold, respectively; $p<0.05$). Expression of *hexA*, *gltX* and spr1880 was not significantly different from R6 in any of the tested strains.

GuaA is annotated as a glutamine amidotransferase enzyme in the published R6 genome sequence (Pfam domain: PF07722). While some glutamine amidotransferases exist as domains of larger proteins, such as CTP synthetase (Weng and Zalkin, 1987) and

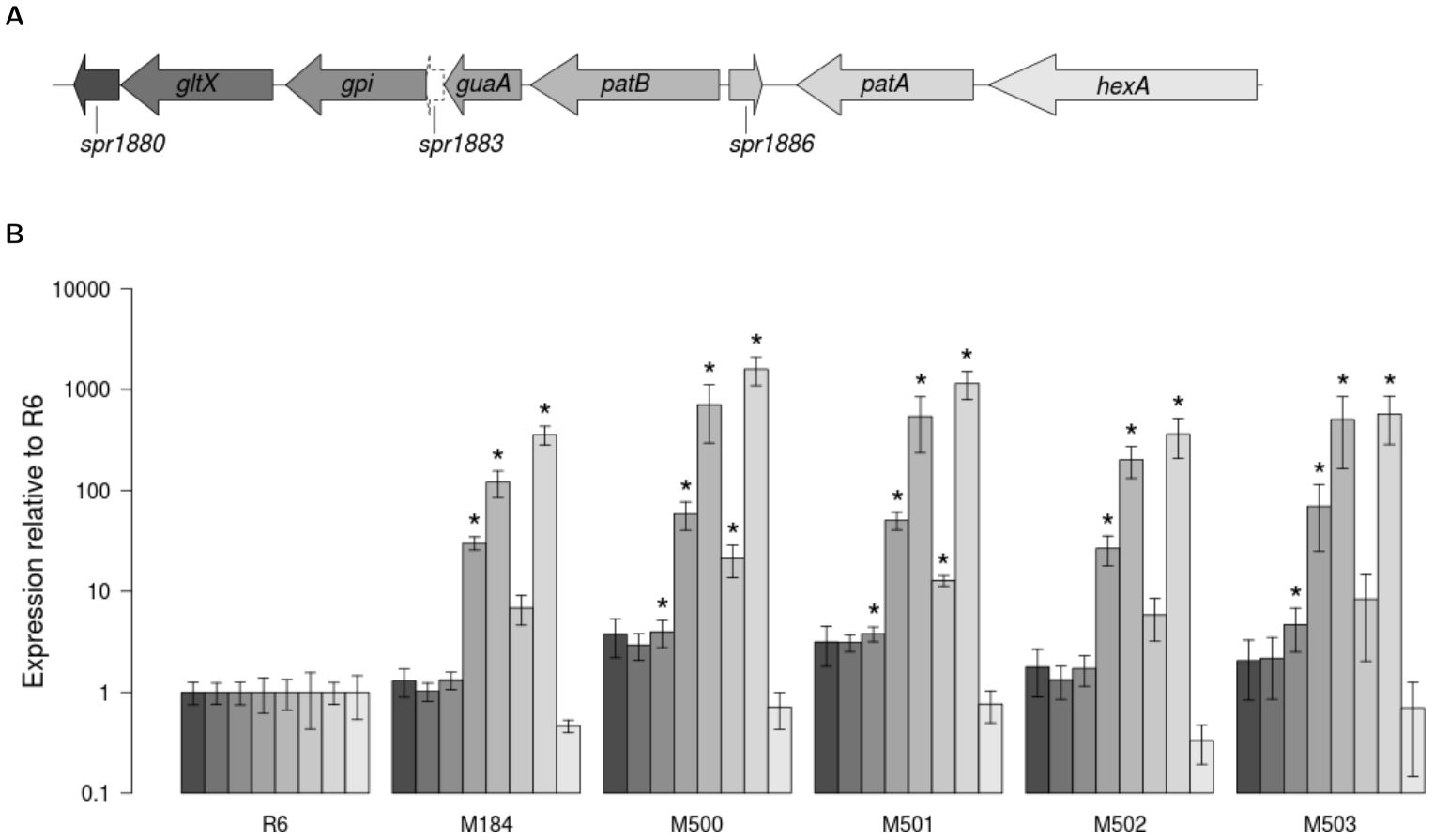


Figure 4.18. Expression of genes within the duplicated region in M184 and four R6^{M184} transformants, determined by qRT-PCR. **(A)** Representation of genomic region around *patA* and *patB*. **(B)** Gene expression relative to R6 (represented by bars shaded by gene as in panel A) determined by qRT-PCR in M184 and the four R6^{M184} transformants, M500-3. Error bars represent the standard deviation of three biological replicates. *, expression significantly different from that of the same gene in R6, $p<0.05$, two-tailed Student's t-test on $log_{10}(expression)$ with Bonferroni correction for multiple testing.

carbamoyl phosphate synthase (Raushel *et al.*, 1999), *guaA* appears to encode a single domain protein. Blast searches against the RefSeq and SwissProt protein databases did not reveal any further information about the function of GuaA in the pneumococcal cell, or its potential substrate.

The output of the TransTermHP program for *S. pneumoniae* R6 was examined for Rho-independent transcriptional terminator predictons in the intergenic region between *patB* and *guaA*. A terminator was predicted eight bp from the end of *patB*, but was located on the opposite DNA strand. Promoter predictions for this region were made using the Prokaryotic Promoter Prediction program, and no putative promoters upstream of *guaA* were found. It therefore looks likely that *guaA* forms an operon with *patA* and *patB* and is cotranscribed, although this would need to be confirmed experimentally as computational prediction of promoter and terminator sequences is not completely reliable.

4.9.2 Inactivation of individual copies of *patA*

Hypotheses one and two described above imply that over-expression of *patA* and *patB* is caused by loss of regulatory control of the second copy of the locus, while the expression of the first copy remains unchanged. If this is the case, it would be predicted that the antibiotic resistant phenotype of M184 would be reversed by inactivation of the second copy of *patA*, while inactivation of the first copy of *patA* would have no effect. In contrast, if the third hypothesis is true, the antibiotic resistant phenotype should be reversed by inactivation of either copy of *patA*.

To test this, an attempt was made to inactivate individual copies of *patA* in M184 by transformation with a PCR amplimer of *patA* from M246, an R6 mutant that contains a *magellan2* minitransposon insertion in *patA* (Garvey and Piddock, 2008). The PCR amplimer was generated using primers 608 and 1357, which amplify an internal region of

patA. As the *patA*-specific sequences flanking the minitransposon insertion were internal to the gene, and no upstream sequence was included, it was expected that the PCR amplimer should recombine with equal frequency with the first and second copies of *patA*. It is known that multiple recombination events can take place simultaneously in a pneumococcal cell (Croucher *et al.*, 2012). Therefore, to reduce the chances of recovering mutants with insertions in both copies of the gene, M184 cells were transformed with serial dilutions of the donor DNA (ranging from 1:10 to $1:10^6$). Candidate transformants were then picked from the lowest two dilutions for which spectinomycin-resistant colonies could be isolated. Recovered transformants were tested for the presence of the *magellan2* insertion by PCR using primer 1357, which anneals internally to *patA* downstream of the minitransposon insertion site, and primers 604 and rRNA-4, which anneal upstream of copy one and copy two of *patA*, respectively.

In total, 52 candidate transformants from three separate experiments were tested for minitransposon insertion. Unexpectedly, the minitransposon insertion was found in the first copy of *patA* in all tested candidates, and no mutants were obtained with an insertion in the second copy. A further ten transformants were tested from a transformation where the donor DNA was not diluted, but the same result was observed. Three M184 *patA1::magellan2* mutants were picked at random and retained for further experiments (strains M508, M509 and M510; Table 2.1).

To determine the effect of the inactivation of the first copy of *patA* on the M184 phenotype, MICs of ciprofloxacin, norfloxacin and ethidium bromide were measured for the three retained M184 *patA1::magellan2* mutants in the presence and absence of sodium orthovanadate (Table 4.7). All three mutants retained resistance to all three agents, and this resistance was reversible by addition of sodium orthovanadate. The resistance phenotype of the mutants was therefore identical to that of M184, which suggests that the first copy of the duplicated locus is not the major source of *patAB* over-expression. This observation supports hypotheses one and two, which both ascribe

over-expression of *patAB* to loss of regulatory control of the second copy.

4.9.3 Measuring readthrough from the promoter of a tRNA gene located upstream of patA copy 2

The first hypothesis suggested above suggests that duplication of the *patA*-spr1880 locus places the second copy of the locus under the control of the promoter of a highly-expressed upstream gene, resulting in over-expression of the second copy of *patAB*. The closest protein-coding gene upstream of the second copy of *patA* is the first copy of spr1880, but gene expression analysis in section 4.9.1 showed that this gene is not highly expressed. However, a glutamate-specific tRNA gene ($tRNA^{Glu}$) located between spr1880 and 16S rRNA in R6 is also included in the duplicated region in M184 and the $R6^{M184}$ transformants. A copy of this gene is therefore found 365 bp upstream of the second copy of *patA*. In its native location, this tRNA lies upstream of the rRNA gene locus that follows spr1880, and identical copies are also found in the same relative position upstream of the three other copies of the rRNA genes found in *S. pneumoniae*. There is also a fifth identical copy of $tRNA^{Glu}$ located between genes spr2040 and spr2041. As there are five identical copies of $tRNA^{Glu}$ in R6, (increasing to six in M184 and the $R6^{M184}$ transformants due to the duplicated genomic region), specific primers for direct measurement of the expression of the copy of $tRNA^{Glu}$ associated with the duplicated region by qRT-PCR could not be designed.

To identify putative promoters, the PPP webserver was used to scan the genomic region between $tRNA^{Glu}$ and spr1889 of R6 for putative lactococcal-like Sigma A binding sites, using Hidden Markov Models (HMMs) constructed from known lactococcal Sigma A binding sites (Zomer *et al.*, 2007). A promoter was predicted between 102 and 72 bp upstream of *patA* (-10 box: TAGAAT, -35 box TTGACA, separated by 17 bp, E-value=0.01), confirming previous predictions (Garvey and Piddock, 2008; Wasserscheid

Table 4.7. MICs of ciprofloxacin, norfloxacin and ethidium bromide for four R6^{M184} transformants in the presence and absence of efflux inhibitors reserpine and sodium orthovanadate.

	Cip	Cip + N	Nor	Nor + N	EtBr	EtBr + N
R6	1	0.5	4	1	2	1
M184	**4**	1	**32**	2	**32**	1
M508	**4**	0.5	**32**	2	**16**	1
M509	**4**	1	**32**	2	**32**	1
M510	**4**	1	**32**	2	**32**	1

Cip, ciprofloxacin; Nor, norfloxacin; EtBr, ethidium bromide; N, 50 μM sodium orthovanadate. Bold text indicates MIC values two or more dilutions higher than those of R6.

et al., 2009). A promoter sequence was also predicted between 53 and 24 bp upstream of tRNA^{Glu} (-10 box: TATAAT, -35 box TTGACA, separated by 16 bp, E-value = 0.0058).

To test whether expression of the second copy of *patA* could be increased by transcriptional read-through from tRNA^{Glu}, two promoter-GFP transcriptional fusions were constructed in a modified version of plasmid pBAV1K-T5-*gfp* (pBAV1K-*gfp2*). This plasmid was contructed by replacement of the T5 promoter-*gfp* cassette with a promoterless *gfp* allele encoding an unstable version of GFP (section 2.9.1). To make the first construct, a 144 bp synthesised DNA fragment (corresponding to genomic region 1865738 to 1865871 on the minus strand) that contained the predicted *patA* promoter region was cloned upstream of the promoterless GFP gene in pBAV1K-*gfp2* (p^{patA}; Figure 4.19A). In the second construct (tRNA-p^{patA}), the tRNA^{Glu} gene was inserted upstream of the cloned patA promoter (Figure 4.19A). This was done by amplifying the duplicated tRNA gene using primers 1892 and 1893, and cloning into the pBAV1K-*gfp2*-*p^{patA}* vector using an *EcoRI* site encoded by primer 1893 and a naturally occurring *SpeI* site found 134 bp upstream of the start of *patA*. Use of this naturally occuring restriction site meant that the sequence between the tRNA gene and the *patA* promoter in the plasmid construct was identical to that found between these two sequences in M184.

The two plasmid constructs and the empty pBAV1K-*gfp2* vector were transformed into R6, and cells containing plasmids were selected with 200 µg/ml kanamycin. These strains were grown in triplicate to mid-logarithmic phase (OD_{600} 0.5), and then one ml of each culture was resuspended in 200 µl PBS and fluorescence was measured (Figure 4.19B). Fluorescence from R6 cells containing pBAV1K-*gfp2*-*p^{patA}* plasmid was 2.4-fold higher than from cells containing empty pBAV1K-*gfp2* vector, which suggests that the upstream region of *patA* indeed contains a promoter sequence that mediates expression of downstream genes. Addition of the tRNA gene upstream of the *patA* promoter (pBAV1K-*gfp2*-tRNA-*p^{patA}*) resulted in a six-fold increase in fluorescence compared to pBAV1K-*gfp2*-*p^{patA}*, suggesting that transcriptional read-through from the tRNA gene

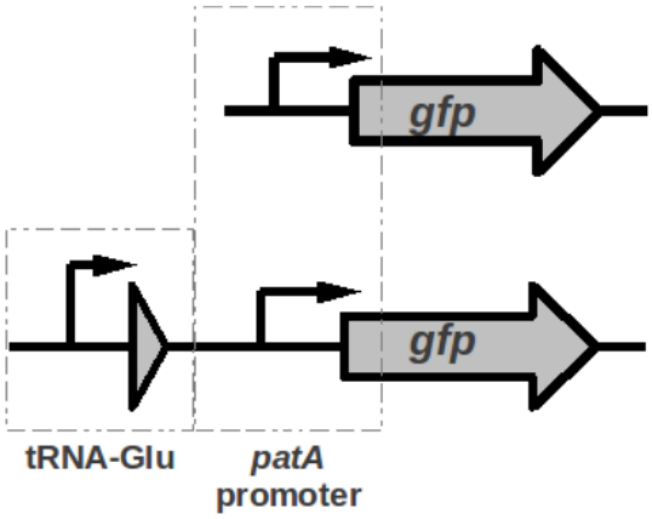
into the second copy of *patA* can occur. However, the strain containing the pBAV1K-*gfp2*-tRNA-p^{patA} construct showed a growth defect compared to R6 and strains containing the other tested constructs (generation time of 55±6 minutes for R6 containing pBAV1K-*gfp2*-tRNA-p^{patA}, compared to 34±1 minutes for R6 containing pBAV1K-*gfp2*-p^{patA}).

4.10 Discussion

The genetic change causing *patAB* over-expression in M184 was a duplication of a 9.2 kb region of the genome, which contained the *patA* and *patB* genes (Figure 4.20). This was unexpected, as it was originally hypothesised that *patAB* over-expression would be caused by mutation of a regulatory gene. The duplication in M184 and the $R6^{M184}$ transformants is novel, and represents the first time that a large adaptive gene duplication has been observed in *S. pneumoniae*.

Spontaneous duplication of genomic regions has been shown to occur frequently within bacterial populations (Anderson and Roth, 1981; Reams and Neidle, 2004; Sun *et al.*, 2012). Once formed, amplification of duplications can further increase gene copy number (Andersson and Hughes, 2009). Antibiotic resistance caused by increases in copy number of resistance determinants has been observed previously (reviewed in Sandegren and Andersson, 2009). Examples include five-fold amplification of the *blaA* gene in *Yersinia enterocolitica* causing β-lactam resistance (Seoane *et al.*, 2003), four-fold amplification of folate synthesis genes in *S. agalactiae* causing sulfonamide and trimethoprim resistance (Brochet *et al.*, 2008), and duplication of a large genomic region containing *acrAB* in *E. coli*, causing transient multidrug resistance (Nicoloff *et al.*, 2006, 2007). There is considerable variation between gene duplications and amplifications involved in antibiotic resistance in terms of the size of the region amplified, copy number, stability of the amplification and the proposed mechanism of formation (Sandegren and Andersson,

A



B

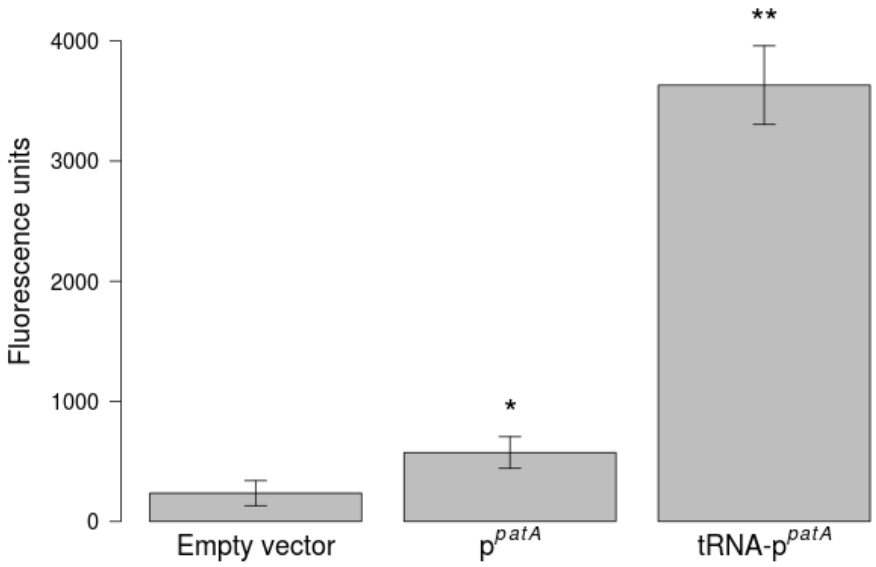


Figure 4.19. Measurement of transcriptional read-through from tRNA^{Glu} using a p^{patA}-*gfp* transcriptional fusion. (A) Organisation of promoter elements and tRNA in the *p^{patA}* (upper) and tRNA-*p^{patA}* (lower) constructs. (B) Fluorescence of 1 ml mid-logarithmic phase cells (resuspended in 200 µl PBS) containing empty vector, pBAV1K-*gfp*-*p^{patA}* or pBAV1K-*gfp*-tRNA-*p^{patA}* constructs. Error bars represent the standard deviation of three biological replicates. *, fluorescence significantly higher than from cells containing empty vector, $p<0.05$, one-tailed Student's t-test. **, fluorescence significantly higher than from cells containing pBAV1K-*gfp*-*p^{patA}*, $p<0.01$, one-tailed Student's t-test.

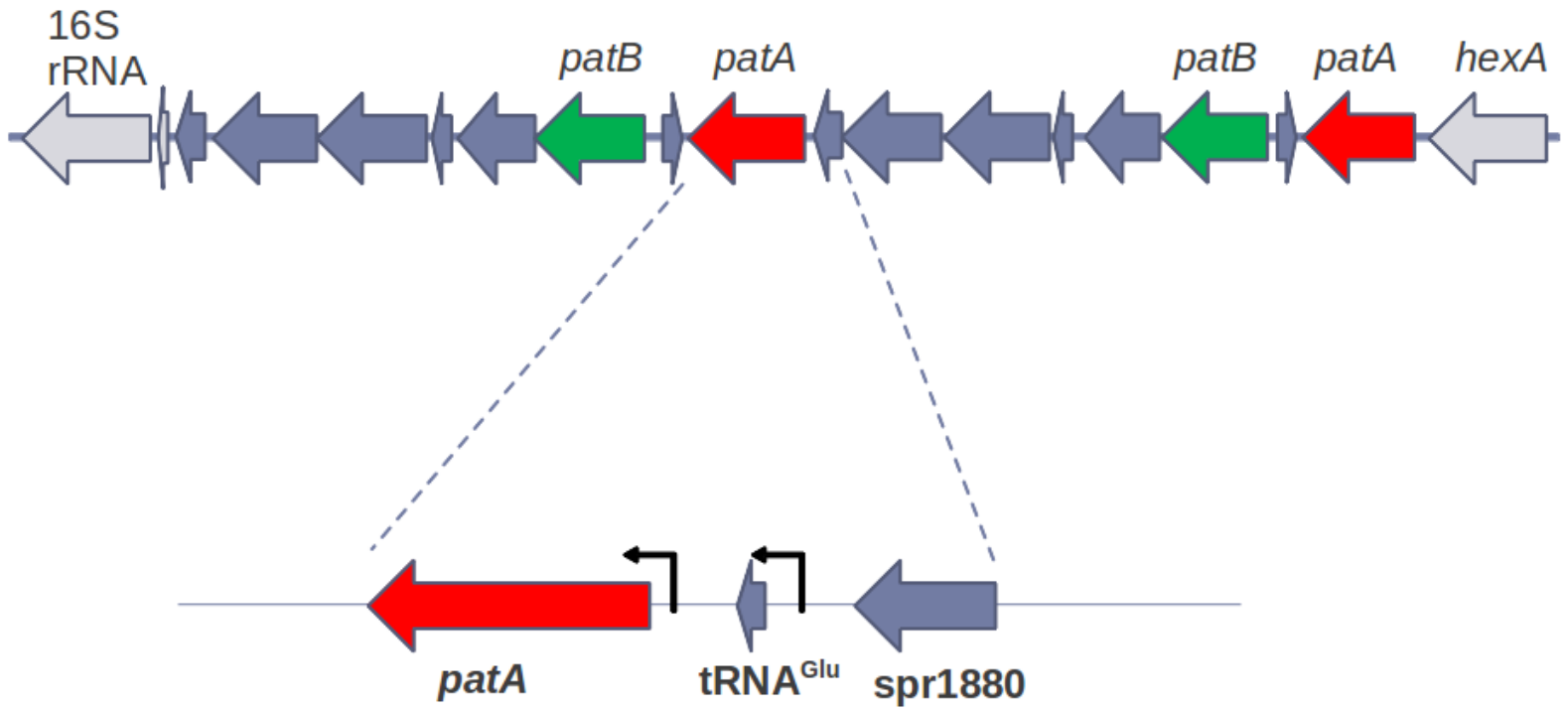


Figure 4.20. Representation of the genetic change causing *patAB* over-expression in M184. M184 contains a tandem duplication of a 9.2 kb region of the genome that includes the *patA* and *patB* genes, indicated in red and green respectively. This results in the second copy of *patA* being located directly downstream from a $tRNA^{Glu}$ gene, and work presented here suggests that high-level expression of the second copy of *patA* and *patB* may be mediated by transcriptional read-through from this tRNA gene.

2009).

Formation of duplications can be mediated by both RecA-dependent and RecA-independent mechanisms (Andersson and Hughes, 2009). RecA-mediated duplication formation requires regions of homology at either end of the duplicated region, such as directly repeated rRNA operons or IS elements (Anderson and Roth, 1981; Haack and Roth, 1995; Lehner and Hill, 1980). RecA-dependent recombination is therefore unlikely to be the mechanism of formation for the duplication in M184, as the duplicated region is not flanked by repetitive elements. Two mechanisms have been proposed to explain RecA-independent duplications (Andersson and Hughes, 2009). The first is strand-slippage during DNA replication, and requires short regions of homology at the duplication endpoints (Lovett *et al.*, 1993; Trinh and Sinden, 1993). The second is illegitimate ligation of DNA ends by DNA gyrase, which occurs by exchange of DNA gyrase subunits while they are bound to double-stranded breaks in the DNA (Ikeda *et al.*, 2004). This mechanism does not require homology at the duplication end points, so is therefore the most likely mechanism by which the duplication observed in M184 was formed. The sequence surrounding the duplication endpoints in M184 does not match an experimentally-determined consensus sequence of the DNA gyrase cleavage site (Leo *et al.*, 2005). However, DNA gyrase cleavage sites are degenerate, so this observation does not rule out involvement of DNA gyrase in formation of the M184 duplication (Leo *et al.*, 2005).

Following duplication of genomic regions, the copy number of duplicated regions can be further increased by amplification, either by RecA-mediated recombination between repeat copies, or by rolling circle replication (Andersson and Hughes, 2009). Under appropriate selective conditions, tandem arrays of up to 200 repeats can be formed by this mechanism, and the DNA content of the cell may be almost doubled (Andersson *et al.*, 1998; Tlsty *et al.*, 1984). More commonly, in examples where gene amplification has been linked to antibiotic resistance, tandem arrays of between two and 40 repeat

copies have been observed (Brochet *et al.*, 2008; Nilsson *et al.*, 2006; Seoane *et al.*, 2003; Yanai *et al.*, 2006). Further gene amplification does not appear to have occurred in M184, or in the R6^{M184} transformants. The slope of the regression line fitted to M184:R6 RPKM ratios for genes inside the proposed duplicated region was approximately double that for the genes outside of the duplicated region, and similar results were observed for all three sequenced transformants. As the number of reads mapping to a repeated region is expected to correlate with copy number, this strongly suggests that the copy number of the *patA*–spr1880 region is no more than two. Higher level amplifications in a small proportion of cells cannot be ruled out, however. The true distribution of copy numbers of the repeat could be confirmed by Southern blotting.

Expression of *patA* and *patB* in M184 and the four transformants was between 100- and 1000-fold higher than in R6. This is much higher than previously reported by Garvey and Piddock (2008). The methods used to measure *patAB* expression differed between the two studies. Garvey and Piddock used comparative reverse transcription PCR (C-RT-PCR) followed by quantification of amplimers using denaturing high-pressure liquid chromatography, whereas the more recently developed technique of qRT-PCR was used in the current study. Real-time PCR measures accumulation of PCR amplimers over the whole amplification cycle, and because of this has a much larger dynamic range than conventional PCR-based methods, such as cRT-PCR, which only quantify PCR amplimers at the end point of the RT-PCR cycle (Schmittgen *et al.*, 2000). This could explain the much larger expression difference detected in this study.

The high levels of *patAB* transcripts observed suggested that the doubling in copy number of these genes could not fully explain their over-expression. One proposed explanation for this was that the formation of the duplication caused dis-regulation of the expression of the second copy of the *patAB* genes due to the change in upstream sequence. Supporting this hypothesis, inactivation of the first copy of *patA* by insertion of the *magellan2* minitransposon did not affect resistance of M184 to ciprofloxacin, norfloxacin

or ethidium bromide. Transposon insertions in copy two of *patA* could not be obtained. This could be explained in two ways. The transposon insertion inactivates PatA by truncation of the protein. While transposon insertion has been previously used successfully to inactivate *patA* in wild-type strains (El Garch *et al.*, 2010; Garvey and Piddock, 2008), it is possible that insertion of the transposon into a gene that is being highly transcribed might result in toxic accumulation of truncated polypeptide. Alternatively, if the second copy of *patA* is being highly transcribed, then recombination of the *patA::magellan2* amplimer with this locus may be physically blocked by the presence of RNA polymerase complexes bound to the DNA. This would make recombination with the first copy of the locus more energetically favourable than recombination with the second.

The proposed mechanism for increased expression of the second copy of *patA* is transcriptional read-through from the tRNA^{Glu} gene located directly upstream. Expression of a GFP reporter was increased when the tRNA gene was inserted upstream of the *patA* promoter in plasmid pBAV1K-*gfp82*. This demonstrates that there are no termination signals in the sequence between tRNA and patA promoter, which implies that transcriptional read-through is possible. The increase in fluorescence when the tRNA gene was added also suggests that, as predicted, the promoter controlling expression of the tRNA is more transcriptionally active than the *patA* promoter. This is plausible because tRNAs have a housekeeping function, and in its native location the tRNA^{Glu} gene is closely associated with the highly expressed rRNA locus.

However, further work will be required to definitely show co-transcription of tRNA^{Glu} and the second copy of *patA*. In particular, the increase in fluorescence from the GFP reporter when the tRNA gene was inserted was much smaller than the observed over-expression of *patA* in M184. The GFP assay used has some limitations. Firstly, the unstable GFP used is not necessarily an ideal reporter gene, as it was taken from a plasmid used to measure promoter activity in *E. coli*, and has therefore not been validated for use in *S. pneumoniae*. It may have been preferable to use the GFP reporter gene used by Ruiz-

Cruz *et al* (Ruiz-Cruz *et al*., 2010) to measure expression from various pneumococcal promoters, as this version of GFP has been more extensively tested in *S. pneumoniae*. Secondly, the growth of R6 cells containing the pBAV1K-*gfp82*-tRNA-p^{patA} construct was impaired relative to growth of cells containing empty vector or the p^{patA} construct. This suggests that accumulation of GFP in cells contining the tRNA-p^{patA} construct is having toxic effects. These two factors may prevent the GFP fluorescence assay from being a quantitative measure of gene expression, and further optimisation of the assay is therefore required.

Finally, to definitively conclude that upregulation of *patAB* is caused by transcriptional read-through from $tRNA^{Glu}$, other mechanisms of upregulation related to the duplication of the tRNA gene would need to be ruled out. It is possible that duplication of the $tRNA^{Glu}$ gene, which increases the total number of $tRNA^{Glu}$ genes in the genome from five to six, could affect the cellular ratio of charged to uncharged tRNA. This might affect expression of genes controlled by T-box regulatory mechanisms, where interaction of uncharged tRNA with an RNA leader sequence affects transcription of the downstream gene (Henkin, 1994). Whether a mechanism such as this could be indirectly causing *patAB* upregulation in M184 remains to be investigated.

Expression of *guaA*, which directly follows *patB*, was also significantly increased in M184 and all four transformants. This over-expression has not been seen previously in studies where the expression of this gene was measured. Over-expression of *guaA* was not observed in microarray analyses of the *patAB* over-expressing mutant M22 (Marrer *et al*., 2006a). Feng *et al* (2009) measured *guaA* expression by qRT-PCR in a linezolid resistant mutant that over-expressed *patAB*, and found that *guaA* expression was not increased. Expression of *guaA* was not measured in studies by Garvey and Piddock (2008) or by El Garch *et al* (2010). It is therefore unclear whether the over-expression of *guaA* detected in this study represents true co-transcription of *patB* and *guaA*, or whether increased *guaA* transcript levels is due to leaky termination of *patB*. The expression of *patB* in

M184 was much higher than in the strain analysed by Feng *et al* (120-fold vs 11-fold), which could explain why leaky termination of *patB* was not observed in the latter study. The relative over-expression of *patA*, *patB* and *guaA* in M184 and the transformants decreased approximately exponentially across the locus, which may reflect a decrease in the effect of the upstream tRNA as intergenic distance increases.

If co-transcription of *patB* and *guaA* was genuine, this would be interesting. GuaA encodes a glutamine amidotransferase of the GATase II family (PFAM domain:), which catalyse transfer of an amide group from glutamine to a variety of recipient substrates (Raushel *et al.*, 1999). Glutamine amidotransferases are often domains within larger proteins, but can be encoded separately, as is the case for *guaA*. Glutamine amidotransferase domains are found in CTPase enzymes, which are involved in nucleotide biosynthesis (Weng and Zalkin, 1987). PatAB is upregulated in response to DNA damage, and co-transcription of *guaA* with *patB* could provide a causal link. There is also a small open reading frame directly downstream of *guaA*, spr1883. Expression of this gene was not measured in this study, and its function is unknown, but its location next to *guaA* suggests that it might be co-transcribed.

The levels of RNA corresponding to spr1886 were significantly higher than in R6 in two out of the four $R6^{M184}$ transformants. However, in both cases transcription was several hundred-fold lower than *patA* and *patB*, and spr1886 was not over-expressed in M184 and the remaining two $R6^{M184}$ transformants. It has been suggested several times previously that *patA* and *patB* are co-expressed (El Garch *et al.*, 2010; Garvey and Piddock, 2008; Marrer *et al.*, 2006b). This hypothesis is based on the lack of reports of strains that over-express one gene without over-expressing the other, and the observation that both genes are upregulated together in induction experiments (El Garch *et al.*, 2010; Marrer *et al.*, 2006b). However, it has not previously been shown whether *patA* and *patB* genes are co-transcribed, or whether they are transcribed separately but co-regulated. The low levels of transcript corresponding to spr1886 would suggest that the genes are not co-

transcribed. If co-transcription occurred, rRNA corresponding to the *patAB* intergenic region would be produced at levels equivalent to the flanking genes. Following conversion to double-stranded cDNA, this region would be recognised by the qRT-PCR primers for spr1886, since the qRT-PCR method used does not differentiate between strands. If *patA* and *patB* are not co-transcribed, this is in contrast to the situation in *S. suis* where co-transcription of the *patAB* homologues *satA* and *satB* has been shown (Escudero *et al.*, 2011, 2013; **?**). It is not clear whether the increased transcription of spr1886 observed in M500 and M501 is due to a genuine increase in expression of spr1886, or from leaky termination of transcription of *patA*. Determination of the direction of transcription in this region and accurate determination of transcription start sites would be required to distinguish these two scenarios.

Genomic duplication was shown to be the cause of *patAB* over-expression in M184, but was not found in another reserpine-selected mutant, M168, which was selected in the same study that gave rise to M184 (Garvey and Piddock, 2008). The duplication was also not found in nine tested clinical isolates that over-expressed *patAB*. This indicates that duplication of the *patAB* region is not the only mechanism that causes over-expression of *patA* and *patB*. Over-expression of *patAB* mediated by mutations in an upstream Rho-independent terminator has been observed in linezolid resistant mutants (Billal *et al.*, 2011; Feng *et al.*, 2009) and in a clinical isolate from a meningitis patient (Croucher *et al.*, 2013). It would be interesting to ascertain whether duplication-mediated *patAB* over-expression is prevalent among *patAB*-over-expressing isolates, or whether it is a rare event observed only in this particular strain. The duplication could be easily transformed into R6 recipient cells, was originally selected by a single exposure to reserpine, and does not appear to confer a fitness defect. These observations, taken together, imply that formation of the duplication is not genetically difficult or energetically unfavourable, which suggests that other examples may be found in a wider screen of *patAB* over-expressing isolates. The PCR assay used to detect duplication junctions in this study

could be expanded such that duplication junctions that are not identical to the example in M184 could be detected. As discussed above, the duplication in M184 is associated with very high levels of *patAB* expression. While no other incidences of this duplication have been reported to date, a ciprofloxacin resistant mutant was recently described by Lupien *et al* (2013) that had high *patAB* expression (60-fold increase compared to R6). It is therefore possible that a similar genomic duplication might be found in this strain.

Gene duplications are often unstable and can be lost from the genome by homologous recombination between copies of the repeats (Andersson and Hughes, 2009; Galitski and Roth, 1997). An estimate of the stability of the *patA*–spr1880 duplication was calculated by measuring the rate of loss of norfloxacin resistance from M184 and M502. The probability of loss of the duplication per cell per generation was estimated as being between 0.006 and 0.027 (Table 4.6). A range of stabilities of between 0.003 and 0.15 loss events per cell per generation have been observed for genomic duplications in other bacteria (Brochet *et al.*, 2008; Sandegren and Andersson, 2009). The duplication in M184 is therefore towards the more stable end of this range. In particular, the M184 duplication is much more stable than the duplication of *acrAB* in *E. coli*, the only previously observed example of duplication of an efflux pump (Nicoloff *et al.*, 2006). In this strain, the 149 kb duplication carrying the *acrAB* genes was lost following two passages on agar plates in the absence of naladixic acid selection. In contrast, the stability of the duplication in M184 in this study suggests that several days of continuous exponential growth would be required for complete loss of the duplication in the population. However, the precision of measurement of the duplication stability was quite low, and a larger sample size would be required to quantify duplication stability more accurately.

Before identification of the genomic duplication as the cause of *patAB* over-expression in M184, several other hypotheses were investigated and rejected. It was originally hypothesised that *patAB* over-expression would be caused by a point mutation in a regulatory gene, as upregulation of efflux pumps caused by mutation of transcriptional

regulators has been frequently observed in other bacteria. For example, mutations in the local repressor AcrR or the global regulators MarR, SoxS and Rob causes over-expression of AcrAB-TolC in clinical isolates of *E. coli* (Alekshun and Levy, 1997; Webber *et al.*, 2005).

Twenty-seven confirmed SNPs and small indels present in M184 relative to R6 were initially identified by whole genome resequencing of M184. Transformation of R6 cells with M184 DNA was carried out to distinguish mutations causing the phenotype from compensatory mutations and bystander mutations, and also to determine how many mutations might be required to confer the M184 phenotype. This whole genome transformation approach was also used successfully by Billal *et al* (Billal *et al.*, 2011) to identify mutations causing linezolid resistance. This approach allowed these authors to determine that linezolid resistance was conferred by 23S rRNA mutations, while a mutation causing *patAB* over-expression relieved a fitness cost but did not contribute towards resistance. In the study by Billal *et al* multiple rounds of transformation were required to fully reconstitute the linezolid resistant phenotype. In the current study, the genetic change conferring over-expression of *patAB*, which is now known to be the *patA*–spr1880 duplication, was transferred in a single step, but at a lower frequency than a spectinomycin resistance cassette. This suggested that the phenotype was not caused by a single transversion mutation, which are transformed with high efficiency (Claverys *et al.*, 1983), but could have been caused by a low efficiency mutation such as a transition, a combination of more than one point mutation or, as was eventually shown, a larger genomic rearrangement.

None of the 27 point mutations identified in M184 were shared by all four of the transformants, which showed that they were not directly involved in upregulation of *patAB*. The only M184 mutation that was transferred to any of the transformants was a cluster of ten mutations in *gltX* (spr1881), which was found in the genome sequence of M503. This gene is part of the duplicated region in M184 and the four transformants. The

transfer of the *gltX* mutations to M503 only suggests that formation of the duplication does not require transformation with a DNA fragment that covers the whole duplicated locus, as otherwise the *gltX* mutations would have been found in all four transformants. Reams and Neidle (Reams and Neidle, 2004) showed that transformation with DNA fragments corresponding to junctions between duplications could induce duplication formation in a wild-type *Acinetobacter* strain. It is likely that this is also the mechanism of transfer of the duplication from M184 to the four transformants. The transfer of the *gltX* mutations from M184 to M503 may have occurred through co-transfer of the duplication junction and the *gltX* gene on the same donor DNA fragment.

The roles of the 27 mutations found in M184 have not been determined. Although all four of the transformants over-expressed *patA* and *patB* their phenotype was not identical to that of M184. Ethidium bromide and norfloxacin resistance was reduced to a greater degree by sodium orthovanadate in the transformants than in M184. Three of the transformants (M500, M501 and M503) expressed higher levels of *patA* and *patB* than M184 (figure 4.6), but accumulated more ethidium bromide (figure 4.5). This suggests that some of the mutations detected in M184 may be involved in either modulating *patAB* expression and/or affecting ethidium bromide resistance. Growth of M184 was attenuated compared to R6, but this growth defect was not seen in the transformants. This implies that *patAB* over-expression does not confer a major fitness defect, and suggests that the growth defect in M184 is probably caused by one of the other mutations carried by this strain. The extra mutations possessed by M184 compared to the transformants may be a result of the difference in selective agents used. M184 was selected using reserpine, which has been shown to affect a variety of cellular processes (Marrer *et al.*, 2006b). The interaction of reserpine with the PatAB transporter is also ambiguous. Garvey and Piddock (2008) showed that reserpine resistance in M184 was reversed only by inactivation of *patA* and not *patB*, and in the current study it was shown that truncation of HrcA caused reserpine resistance but not *patAB* over-expression. In

contrast, the transformants generated in this study were selected using ethidium bromide and norfloxacin, which are known substrates of PatAB. Some of the mutations in M184 may have been selected to counteract the pleiotropic effect of reserpine, so are unrelated to over-expression of *patAB*.

The presence of an insertion in the genome of the laboratory stock of R6 relative to the published R6 genome was considered as a possible explanation for why mutations causing *patAB* over-expression were not initially discovered by the whole genome transformation approach. This was considered plausible because inserted prophages have been previously observed in *S. pneumoniae* (Romero *et al.*, 2009a,b), and the distributed genome hypothesis suggests that pneumococci can acquire accessory genes by transformation from a wider pool of pneumococcal genes (Hiller *et al.*, 2007). It has also been over ten years since the R6 genome sequence was published, and the laboratory stock of R6 may have evolved over that time. In particular, there was already evidence that a recombination had occurred in the *pbpX* gene. Assembly of unaligned reads from M184 using Velvet identified some non-pneumococcal sequences, including a large contig that resembled a lytic bacteriophage associated with the dairy industry (Castro-Nallar *et al.*, 2012). PCR using primers specific for this phage did not yield any amplimers. While not conclusively shown, it is likely that this phage was a contaminant of the M184 DNA preparation, possibly originating from the growth medium. Bovine and equine microsatellite sequences were also assembled, which are also likely to have come from the rich media and horse blood supplement used to grow the bacteria. This demonstrates the high sensitivity of Illumina genome sequencing.

Another hypothesis that was considered to explain why mutations causing *patAB* expression were not initially detected by the transformation approach was that the causative SNP was located in a repetitive region. Repetitive elements are particularly common in pneumococcal genomes, comprising over 3% of the R6 genome (Hoskins *et al.*, 2001). Mapping of reads from multiple genomic locations to each repeat sequence could have

prevented detection of a SNP. It was considered possible that mutation in a short interspersed repeat could cause *patAB* over-expression because some short repeat elements have been shown to affect expression of nearby genes. For example, expression of the competence operons *comAB* and *qsrAB* is modulated by upstream BOX elements (Knutsen *et al.*, 2006), and SPRITEs resemble transcriptional terminators (Croucher *et al.*, 2011b). However, a subset of short interspersed repeats was examined for the presence of mutations by PCR and Sanger sequencing approach and no mutations were found.

Overall, several hypotheses were considered to explain *patAB* over-expression in M184 and the four $R6^{M184}$ transformants. The genetic change causing *patAB* was ultimately identified as a tandem duplication of a 9.2 kb region of the genome containing the *patA* and *patB* genes. The results presented here suggest that the formation of this duplication leads to a high level of *patAB* over-expression via a promoter-switching mechanism, which has not been observed in previous examples of duplication-mediated antibiotic resistance.

4.11 Further work

The conclusion that over-expression of *patA* and *patB* came initially from the observation that insertional inactivation of the first copy of *patA* had no effect on resistance of M184 to ciprfloxacin, norfloxacin and ethidium bromide. However, inactivation of the second copy of *patA* would be required to fully confirm this hypothesis, and this could not be achieved by the transposon insertion method used here. This may be because transposon insertion into the second copy leads to accumulation of toxic levels of truncated PatA protein. If this is the case, inactivation *patA* copy two might be accomplished using a method that completely replaces the gene with an antibiotic resistance gene, meaning that no truncated PatA is produced. This approach was used by Boncoeur *et al* (2012), who completely replaced the *patA* gene in R6 with a chloramphenicol resistance gene.

Alternatively, recombination of the *patA::magellan2* PCR amplimer with the first copy of *patA* may have been favoured due to the presence of more transcription complexes bound to copy two, or a difference in tertiary DNA structure in the copy two region. If this is the case, recombination of a *patA::magellan2* amplimer could be encouraged by addition of copy two-specific upstream sequence to the donor PCR amplimer by overlap extension PCR (Horton *et al.*, 1989).

The proposed explanation for increased expression of the second copy of *patA* was transcriptional read-through from a $tRNA^{Glu}$ gene located at the 3' end of the duplicated region. The GFP assay carried out in this study suggests that read-through is possible. However, full confirmation of this hypothesis require demonstation that the tRNA and the second copy of *patA* are co-transcribed. This could be shown by Northern blotting, or by mapping of transcription start sites of *patA* using 5' rapid amplification of cDNA ends. Another way to investigate co-transcription of $tRNA^{Glu}$ and *patA* would be to determine the whole transcriptome of M184 using RNA-Seq. This would also allow investigation of whether or not *patA* and *patB* are encoded on the same transcript, and whether *guaA* and spr1883 are co-transcribed with *patB*.

The copy number of the duplication identified in M184 was determined to be two based on comparison of the slopes of regression lines fitted to M184:R6 RPKM ratios for genes located inside and outside the duplicated region. However, this does not rule out the possibility that a subpopulation of cells may have contained higher level amplifications of the *patA*–spr1880 region. The distribution of copy numbers of the *patA*–spr1880 region within a population of M184 cells could be determined by Southern blotting.

In a study investigating 72 gene amplification mutants of *Acinetobacter* with increased ability to grow on benzoate, certain duplication junctions were observed multiple times (Reams and Neidle, 2004). This indicated that formation of some amplifications was particularly favourable. Some of these common duplication junctions could be detected

at low levels in wild-type unselected populations of cells. This suggested that these duplications formed spontaneously within the population at a low frequency, and then cells containing these duplications were selected by exposure to the selective condition. The duplication in M184 could be a similar common spontaneous duplication in R6 populations that is selected by exposure to fluoroquinolones, ethidium bromide or reserpine. This possibility could be investigated by attempting to amplify the duplication junction by PCR from high concentrations of R6 genomic DNA, as was done by Reams and Neidle (2004), or by using the padlock probing method for validating low frequency genomic rearrangements that was used in a recent study by Sun *et al* (2012). Additionally, a larger collection of isolates or mutants over-expressing *patA* and *patB* could be screened for the presence of the duplication identified in M184 to determine whether formation of this duplication is a common mechanism of *patAB* over-expression.

The duplication identified in M184 and the $R6^{M184}$ transformants resulted in 100-1000-fold increases in transcription of *patA* and *patB*. However, changes in levels of PatAB protein were not measured in this study. It would be interesting to investigate the proportion of the *patA* and *patB* mRNA that is translated into functional protein, as it might be expected that a 1000-fold increase in production of a particular protein would be toxic to the cell. Protein production could be measured by Western blotting, either by raising antibodies that recognise PatA and PatB directly, or by adding HIS- or FLAG®-tags (Sigma Aldrich, UK) to the protein sequences. Tagging the proteins could also provide confirmation of the amount of protein originating from each copy of the *patAB* locus, as the two copies of *patA* could be modified to encode different tags.

Finally, an estimate of the stability of the *patA*–spr1880 duplication was obtained by monitoring the loss of norfloxacin resistance from a population. However, the 95% confidence interval of this estimate was large as the sample size tested was quite small (60 colonies). Repetition of this experiment using a larger number of colonies (up to 1500) would enable a more accurate determination of the half-life of the duplication

within the population.

4.12 Key Findings

- The *patAB* over-expression phenotype from M184 can be readily transferred into antibiotic susceptible R6 cells by genetic transformation
- The genetic change responsible for over-expression of *patA* and *patB* is not a point mutation
- Over-expression of *patA* and *patB* in M184 is caused by a duplication of a 9.2 kb region of the genome, which includes *patA* and *patB*
- The duplication confers 100-1000-fold over-expression of *patAB*, which is much higher than would be expected from the increase in gene copy number alone
- The duplication was lost from the population at a rate of approximately 0.01-0.03 loss events per cell per generation, but several days of exponential growth would be required for complete elimination from the population
- Inactivation of the first copy of *patA* did not affect antimicrobial resistance in M184, suggesting that the phenotype is mainly conferred by expression from the second copy of *patAB*.
- Formation of the duplication results in the location of the second copy of *patA* downstream of a tRNA^{Glu} gene
- Transcriptional read-through from the tRNA gene into the downstream *patA* gene is possible

5 Identification of mutations causing over-expression of *patAB* in different genetic backgrounds

5.1 Background

Duplication of a genomic region including the *patA* and *patB* genes was shown to be the mechanism for *patAB* over-expression in M184 (Chapter 4). However, evidence of this mechanism was not found in the set of clinical isolates screened (section 4.8.1.4) or in M168, a second reserpine-selected laboratory mutant selected from the serotype 1 strain M4 (NCTC7465; Garvey and Piddock, 2008). It is also unclear whether the gene amplification observed in M184 can explain the observation that *patAB* expression is inducible by subinhibitory concentrations of fluoroquinolones and mitomycin C (El Garch *et al.*, 2010; Marrer *et al.*, 2006b). Therefore, to identify other mechanisms of regulation of *patAB*, M168 and three clinical isolates (chapter 3 and Piddock *et al.*, 2002) were further investigated to identify mutations causing *patAB* up-regulation.

5.2 Aims and hypotheses

The aim of this study was determine the genetic cause of *patAB* over-expression in M168, and three clinical isolates, M101, M87 and M74, by whole genome sequencing

of R6 transformants of these strains. It was hypothesised that recombinations would be detected in or near genes involved in the regulation of *patAB* expression.

5.3 Transformation of *patAB* over-expression phenotype from M168 and three clinical isolates into R6

The non-susceptible parental strains of the three clinical isolates are not available, so the approach of resequencing and comparing mutant and parental strains could not be carried out due to the lack of a suitable reference genome. The parental strain of M168 (M4) was available, but both M4 and M168 produce a mucoid capsule and therefore extraction of sufficient good quality DNA for genome sequencing was difficult.

Instead, the natural transformation ability of the pneumococcus was exploited to transfer DNA regions containing causative mutations into the R6 genetic background. Pneumococci can take up DNA from the environment by the competence pathway, and integrate segments of DNA with a mean size of about four to six kb (Croucher *et al.*, 2012). Incorporated DNA of foreign origin can be seen as a region with a high density of SNPs compared to the recipient genome, named recombinant sequence segments (RSSs) by Croucher *et al* (2012).

This approach was chosen due to previous reports which found that there were no mutations in the *patA* promoter region in M168 or the clinical isolates (Garvey and Piddock, 2008, Garvey and Wong, unpublished data), suggesting that the mutation causing *patAB* over-expression were located elsewhere in the genome.

The three clinical isolates, chosen at random from the set of isolates resistant to both fluoroquinolones and dyes (Chapter 3), were M101, M87 and M74. The *patA* and *patB*

genes from M101 had been sequenced previously and no mutations were reported when compared against sequences from R6, M4 and M3 (Garvey and Wong, unpublished data). M87 and M74 had not been previously sequenced. The *patA* and *patB* sequences from M168 were previously reported to be identical to those from M4 (Garvey and Piddock, 2008).

Whole genomic DNA preparations extracted from M168, M101, M87 and M74 were used to transform CSP-treated R6 cultures, and transformants over-expressing *patA* and *patB* were selected on eight µg/ml ethidium bromide. No ethidium bromide resistant transformants were isolated from a control experiment where an R6 culture was transformed with R6 DNA, indicating that spontaneous ethidium bromide resistant transformants were not selected under the conditions used.

5.3.1 Confirmation of *patAB* over-expression in transformants

If the selected transformants have successfully acquired a genetic change that causes upregulation of *patA* and *patB* in the R6 genetic background, it would be expected that they should have a phenotype similar to M184. This means that they should have a reduced susceptibility to ciprofloxacin, norfloxacin and ethidium bromide, which is reversible by addition of the ABC transporter inhibitor sodium orthovanadate. They should also accumulate lower levels of ethidium bromide.

The minimum inhibitory concentrations of ciprofloxacin, norfloxacin and ethidium bromide were determined for 19 transformants (6x $R6^{M168}$, 5x $R6^{M101}$, 5x $R6^{M87}$ and 3x $R6^{M74}$) in the presence and absence of sodium orthovanadate, and compared to those of R6 and M184. The MICs of each agent for the tested transformants are summarised in Table 5.1.

The levels of intracellular accumulation of ethidium bromide were measured following incubation of the transformants for 14 minutes in the presence of 25 µM ethidium bromide.

Table 5.1. MICs of ciprofloxacin, norfloxacin and ethidium bromide for selected R6 transformants of M168, M101, M87 and M74 in the presence and absence of sodium orthovanadate.

	Transformant	Cip	Cip + N	Nor	Nor + N	EtBr	EtBr + N
R6	-	1	0.5	4	1	4	<1
M184	-	**4**	0.5	**32**	2	**64**	<1
$R6^{M168}$	1	**2**	0.5	**16**	1	**32**	<1
	2	**2**	0.5	**16**	2	**32**	<1
	3	**2**	0.5	**16**	2	**32**	<1
	4	**2**	0.5	**16**	2	**32**	<1
	5	**2**	0.5	**16**	2	**32**	<1
	6	**2**	0.5	**16**	2	**32**	<1
$R6^{M101}$	1	**4**	0.5	**32**	2	**64**	<1
	2	**4**	0.5	**32**	2	**64**	<1
	3	**4**	0.5	**32**	2	**64**	<1
	4	**4**	0.5	**32**	2	**64**	<1
	5	**4**	0.5	**32**	2	**64**	<1
$R6^{M87}$	1	**4**	0.5	**32**	2	**64**	<1
	2	**4**	0.5	**32**	2	**64**	<1
	3	**4**	0.5	**32**	2	**64**	<1
	4	**4**	0.5	**16**	2	**64**	<1
	5	**4**	0.5	**16**	2	**64**	<1
$R6^{M74}$	1	**4**	0.5	**32**	8	**16**	<1
	2	**16**	0.5	**128**	2	**16**	<1
	3	**4**	0.5	**32**	8	**16**	<1

Cip, ciprofloxacin; Nor, norfloxacin; EtBr, ethidium bromide; N, 50 μM sodium orthovanadate. Bold text indicates MIC values that are two or more dilutions higher than those of R6

This experiment was carried out on two separate occasions using the same transformants as for the MIC measurements. The level of accumulation of ethidium bromide differed within each group of transformants (Figure 5.1). Two of six $R6^{M168}$ transformants, one of five $R6^{M101}$ transformants, four of five $R6^{M87}$ transformants and one of three $R6^{M74}$ transformants accumulated significantly less ethidium bromide than R6 ($p<0.05$, pairwise Student's t-test with Benjamini-Hochberg correction to control the false discovery rate; Figure 5.1). All the tested transformants accumulated significantly more ethidium bromide than M184 ($p<0.05$). For each donor strain, the transformant that accumulated the lowest level of ethidium bromide was retained for further experiments; these were named M504 ($R6^{M168}$ transformant two), M505 ($R6^{M101}$ transformant four), M506 ($R6^{M87}$ transformant five), and M507 ($R6^{M74}$ transformant two).

Expression levels of *patA* and *patB* were measured by qRT-PCR (Figure 5.2) as described for M184 and its R6 transformants. As over-expression of *guaA* and spr1886 had also been observed in some of the $R6^{M184}$ transformants, the expression of these genes was also measured (Figure 5.2). In M504, M505 and M506, transcription of *patA*, *patB* and *guaA* was significantly increased compared to R6 ($p<0.05$, one-tailed pairwise Student's t-test on log(expression), with Bonferroni correction for multiple testing). Levels of transcript corresponding to spr1886 were also significantly increased in M505 and M506 ($p<0.05$). Over-expression of *patA* varied from 3.3 ± 0.5-fold in M504 to 6.4 ± 1.4-fold in M506, while over-expression of *patB* ranged between 3.5 ± 0.7-fold in M504 to 8.2 ± 1.9-fold in M506. Unexpectedly, neither *patA* or *patB* was significantly over-expressed in M507, the $R6^{M74}$ transformant. However, *guaA* was significantly over-expressed in this strain (4.2 ± 1.6-fold, $p<0.05$).

Growth of the four chosen transformants in BHI broth was measured over six hours and no significant differences in generation time were observed compared to R6.

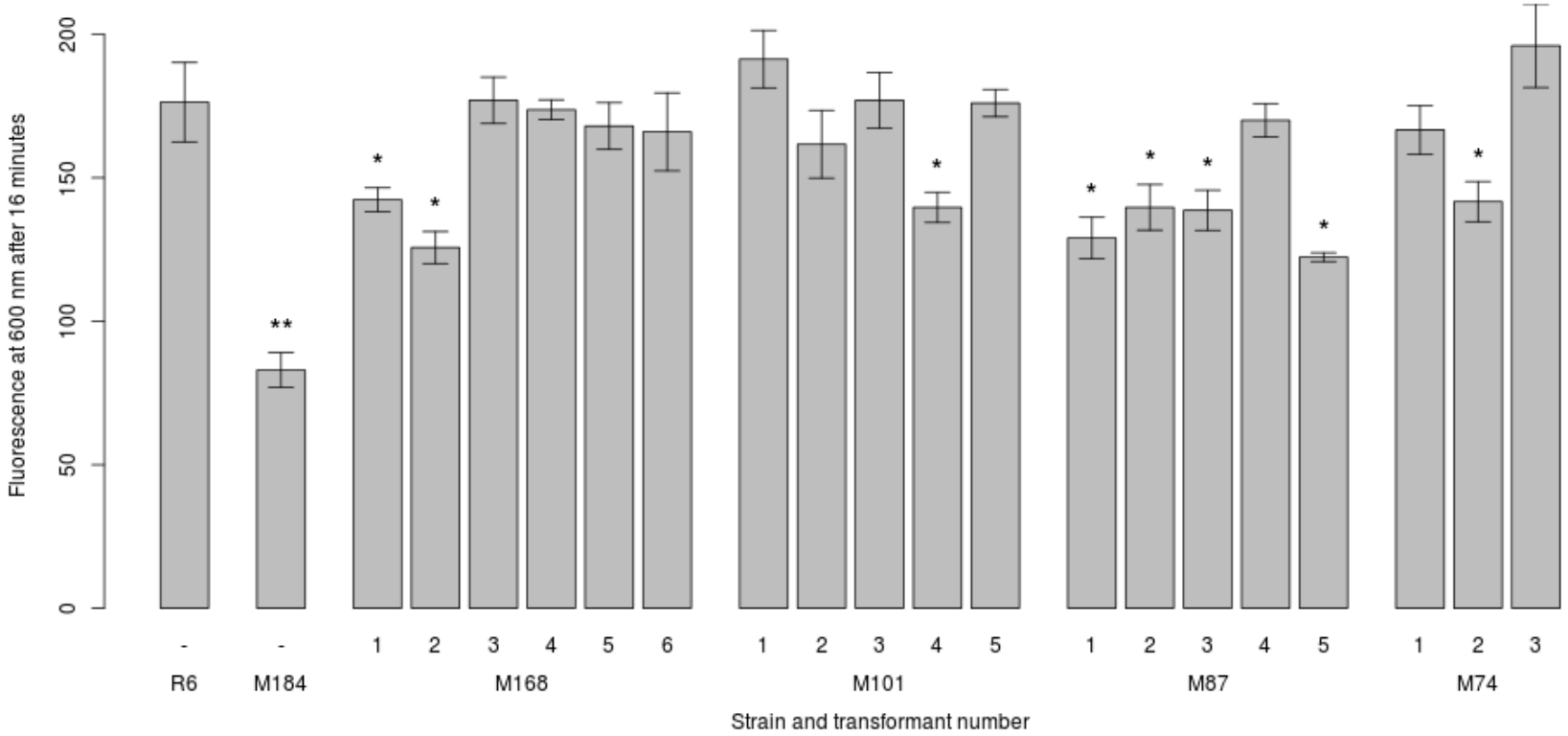


Figure 5.1. Fluorescence of R6 clinical isolate transformants due to intracellular accumulation of ethidium bromide 14 minutes after exogenous addition of 25 μM ethidium bromide. Error bars represent standard deviation of three biological replicates. *, fluorescence significantly lower than R6, $p<0.05$, one-tailed Student's t-test with Benjamini-Hochberg correction to control false discovery rate. **, fluorescence significantly lower than R6, $p<0.01$, tested as above.

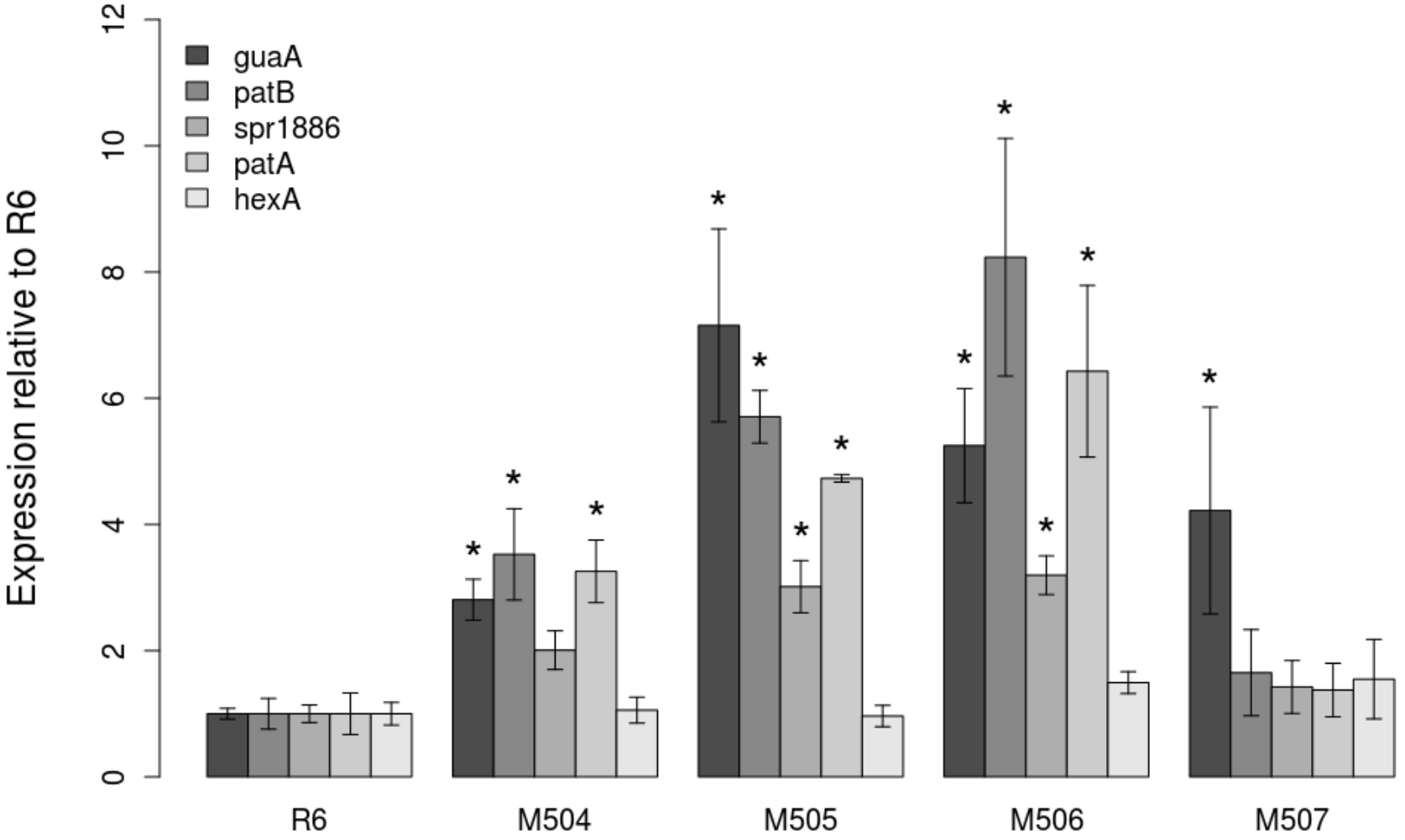


Figure 5.2. Expression of *hexA*, *patA*, spr1886, *patB* and *guaA* in R6 transformants of M168 and three clinical isolates, determined by qRT-PCR. Error bars represent standard deviation of three biological replicates. *, Expression is significantly higher than that of R6, $p<0.05$, One-tailed Student's t-test with Bonferroni correction for multiple testing.

5.4 Whole-genome sequencing of R6 transformants of clinical isolates

Whole genomic DNA of the transformants from the three clinical isolates and M168 was sequenced using by Illumina sequencing as described in Materials and Methods (section 2.5.1) to give sets of 100 bp paired end reads. Theoretical coverage levels of the 2.0 Mb R6 genome of 2856, 1756, 1453 and 2057 were achieved for $R6^{M74}$, $R6^{M101}$, $R6^{M87}$ and $R6^{M168}$ respectively.

Read sets were assessed for quality using FastQC. All read sets had high average quality per read. However, all reads were trimmed to 75 bp, using a Python script written by Nick Loman, as in some of the read sets per-base quality scores dropped below 20 at the 3' end of the read.

Reads were mapped against the R6 genome using Bowtie2, and SNPs and small indels relative to the reference sequence were identified using Samtools mpileup and BCFtools (Materials and Methods, section 2.5.2). Identified SNPs were filtered to retain only those with a Phred scaled quality score greater than 100 and a covering read depth greater than 50. Polymorphisms that were found in the sequence of the R6 laboratory stock (Appendix 1) were also discarded.

The total numbers of polymorphisms identified in $R6^{M74}$, $R6^{M101}$, $R6^{M87}$ and $R6^{M168}$ were 95, 668, 631 and 189 respectively, excluding SNPs shared with R6. As expected, the polymorphisms were found in clusters, identifying regions of recombinant DNA originating from the donor strain. In all four strains, more than one distinct cluster of polymorphisms was observed, suggesting that more than one recombination event had occurred in each strain (figure 5.3). Clusters were defined as genomic regions for which the average number of SNPs in a 5 kb sliding window was consistently greater than

two. Using this definition, four clusters were identified in $R6^{M74}$, 14 in $R6^{M101}$, 24 in $R6^{M87}$ and one in $R6^{M168}$. Comparison of the locations of clusters between the four strains revealed that all four contained a large cluster mapping between chromosomal positions 1855000 and 1875000. This region is roughly centred around the *patA* gene (Figure 5.4). As these overlapping clusters were found in all four transformants, it was hypothesised that these represent the primary recombination event that conferred *patAB* over-expression in the transformants. Furthermore it was postulated that the other observed clusters of mutations were secondary recombinations that do not contribute to the observed phenotype.

5.4.1 Determination of the extent of primary recombination events

The full extent of the polymorphic clusters in the *patA* region, as measured from the first polymorphism in the cluster to the last, ranged from 9.7 kb in $R6^{M74}$ to 12.2 kb in $R6^{M168}$. This is larger than the average recombination sizes that have been previously reported, which range between 2 and 6 kb (Croucher *et al.*, 2012; Feil *et al.*, 2000; Fox and Allen, 1964; Gurney Jr. and Fox, 1968; Lacks, 1966). However, it has been shown that a high density of secondary recombination events can occur around a primary selected recombination due to multiple, non-contiguous recombinations occurring simultaneously from the same molecule of donor DNA (Croucher *et al.*, 2012).

To investigate this, the polymorphic clusters were resolved into regions representing putative recombination events using an approach similar to that of Croucher *et al* (2012). The *patA* and *patB* nucleotide sequences from each donor strain were compared against the RefSeq genome database using MegaBlast to determine the most closely related genome. M101 and M74 most closely matched strains JJA and ATCC700669 (sequence identity: 99% *patA* and 100% *patB*), M87 was closest to strain OXC141 (sequence

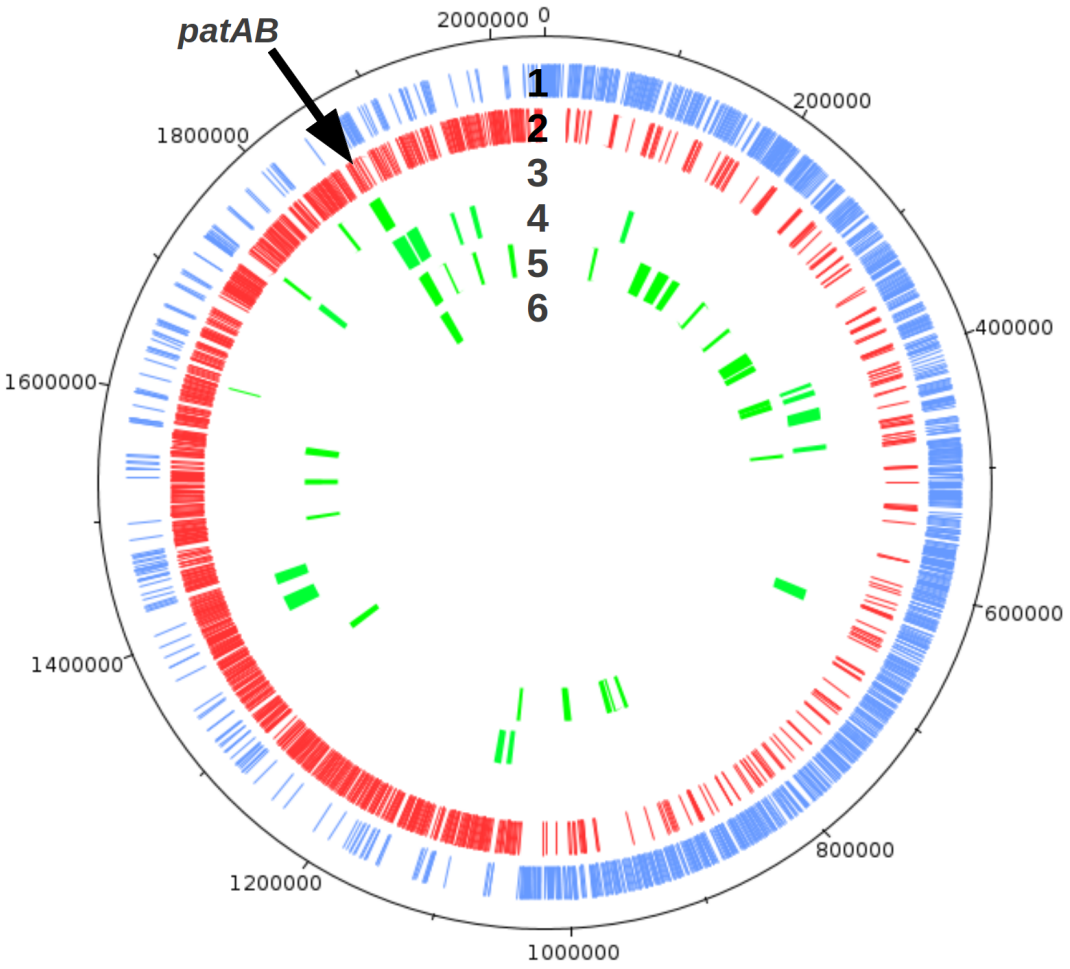


Figure 5.3. Positions of putative recombination events in genome sequences of R6^{M168}, R6^{M101}, R6^{M87} and R6^{M74}. Rings 1 and 2 represent positions of coding sequences on the forward and reverse strands of the R6 genome, respectively. Positions SNP clusters representing putative recombinations are indicated by green segments for R6^{M74} (ring 3), R6^{M101} (ring 4), R6^{M87} (ring 5) and R6^{M168} (ring 6). A recombination had occurred in the *patAB* region (indicated by the black arrow) in all four transformants.

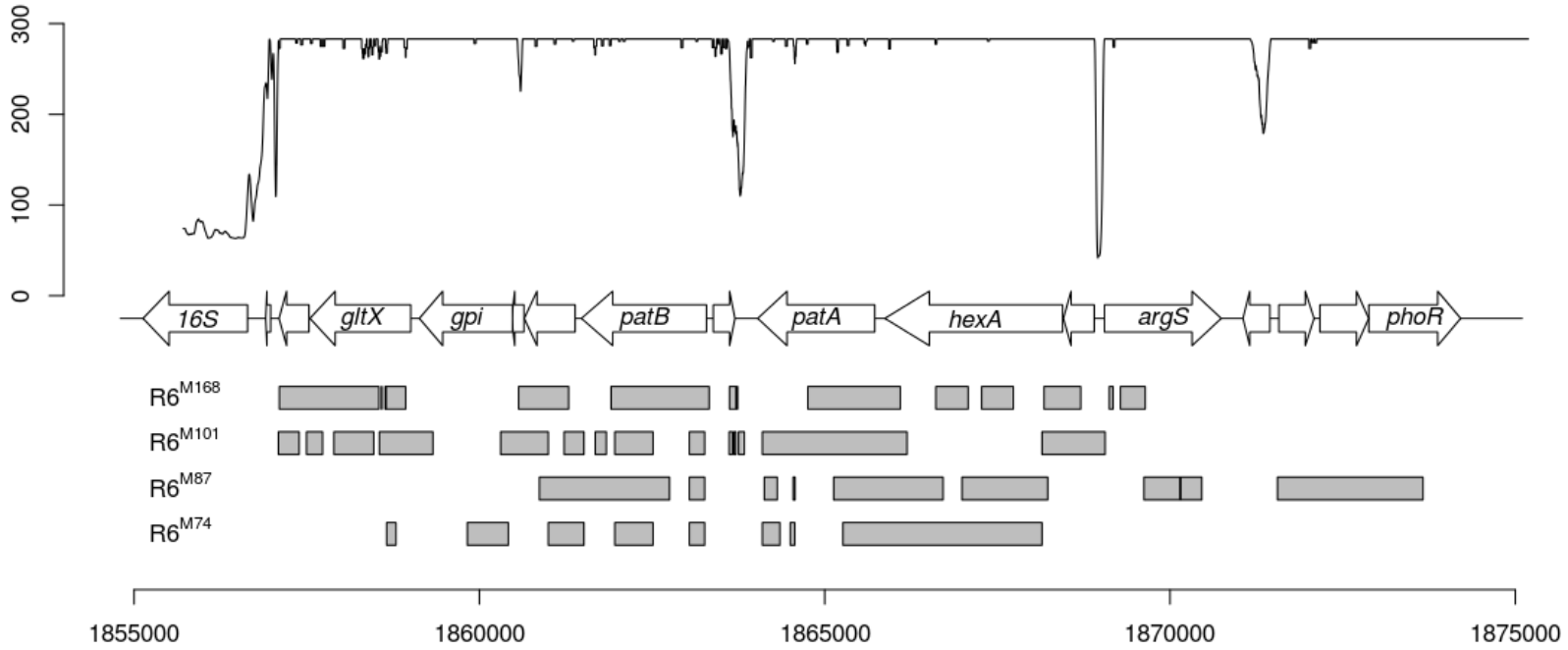


Figure 5.4. Extent of recombinant sequence segments in the *patAB* region of R6^{M74}, R6^{M101}, R6^{M87} and R6^{M168}. Boxes represent recombinant sequence segments defined as the median distance the last donor allele of a segment and the following recipient allele. Line represents average per base quality score across the region, smoothed over a 25 bp window.

identity 100% for *patA* and *patB*), and M168 most closely matched 670-6B (sequence identity: 100% for *patA* and *patB*). The regions corresponding to R6 region 1855700 to 1875200 were extracted from each of these genome sequences and aligned with the R6 sequence and that of the corresponding transformant. At each polymorphic site in the alignment the allele present in the transformant sequence was classified as "donor" or "R6" based on whether it matched the donor strain sequence of the R6 sequence. A recombinant donor segment was defined as an unbroken series of donor alleles with no intervening R6 alleles, and the recombination sites were tentatively defined as the median distance between the first or last donor allele of the segment and the next recipient allele. By this method, between eight and thirteen putative recombinant segments were identified for each transformant (Figure 5.4). However, it is possible that the sizes of some segments may be under- or overestimated due to mis-calling of bases in low quality regions, particularly around the spr1886 gene and in the intergenic region between spr1889 and 1890 (mapping quality across the region is depicted in figure 5.4).

5.5 The region upstream of *patA* contains genetic changes responsible for patAB over-expression

Examination of predicted primary recombinations (Figure 5.4) revealed overlapping segments in all four transformants, which mapped to a region spanning the 5' end of the *patA* gene and the 3' end of the *hexA* gene. Additionally, smaller overlapping segments mapped to two intergenic regions of the *patB* gene. There was also some overlap between segments mapping to *guaA*. This suggests that the genetic change responsible for *patAB* over-expression in the four donor isolates is most likely to be localised in one of these areas.

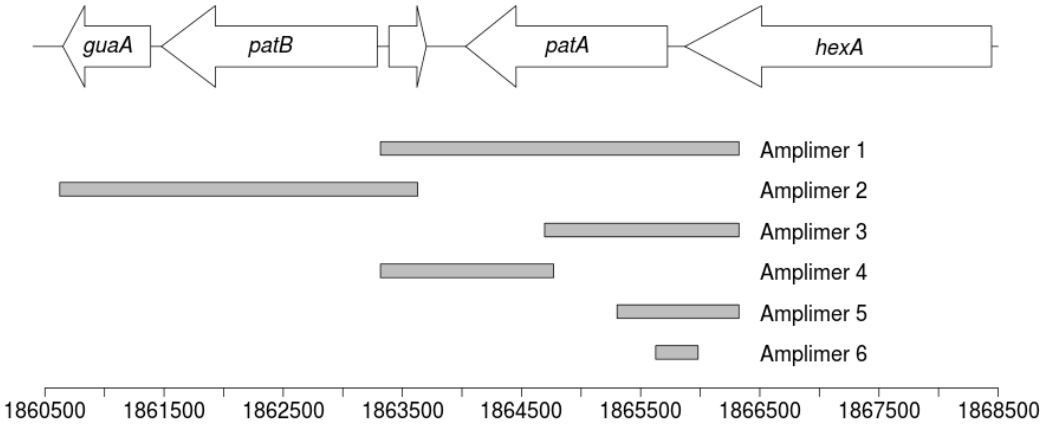
To investigate this, 3009 bp and 3005 bp regions encompassing *patA* and *patB*, respec-

tively, were amplified from each transformant using primers 604 and 605 for the *patA* amplimer and 606 and 607 for *patB*. These PCR amplimers were used to transform R6 cells, and transformants were selected on eight µg/ml ethidium bromide. Ethidium bromide resistant transformants were recovered from the transformation with the *patA* amplimer, but not when the *patB* amplimer was used.

Further transformations with PCR amplimers corresponding to progressively smaller regions were carried out to more precisely localise the *patA* mutation responsible for *patAB* over-expression. The extents of these amplimers and the outcomes of transformation experiments are summarised in Figure 5.5. Transformants were obtained with an amplimer covering the 5' half of the gene, including 600 bp upstream, generated with primers 604 and 1357, but not with an amplimer made using primers 605 and 1358 that covers the 3' half of the genome. This was narrowed down with further transformation experiments to a region between 1865302 and 1866325 (primers 604 and 609). For $R6^{M101}$, $R6^{M87}$ and $R6^{M168}$, transformants were obtained using a PCR amplimer for a smaller region, 1865626 to 1865982, generated using primers 1771 and 1772. This region encompasses the intergenic region between *hexA* and *patA*, suggesting that the cause of *patAB* over-expression was a mutation in a *cis*-regulatory sequence. However, no transformants were obtained using this PCR amplimer from $R6^{M74}$, suggesting that the causative mutation in this isolate lies outside of this region.

In $R6^{M101}$ and $R6^{M87}$, the DNA region identified as the source of *patAB* over-expression contained only one mutation relative to R6 in each strain, at genomic positions 1865772 and 1865771 respectively. No mutation was initially found in this region in $R6^{M168}$, but when the raw, unfiltered SNPs were examined, a single mutation was found with a quality score of 96 at genomic position 1865758. The mutations in all three isolates were confirmed by PCR and Sanger sequencing of this region. The three mutations were closely clustered within a 15 bp region, which suggests that they may disrupt a single *cis*-regulatory element. A 100 bp DNA region containing these mutations was anal-

A



B

	$R6^{M168}$	$R6^{M101}$	$R6^{M87}$	$R6^{M74}$
Amplimer 1	+	+	+	+
Amplimer 2	-	-	-	-
Amplimer 3	+	+	+	+
Amplimer 4	-	-	-	-
Amplimer 5	+	+	+	+
Amplimer 6	+	+	+	-

Figure 5.5. Determination of the location of polymorphisms causing *patAB* overexpression in $R6^{M74}$, $R6^{M101}$, $R6^{M87}$ and $R6^{M168}$. (A) Extent of PCR amplimers used to transform R6 cells **(B)** Results of transformation experiments using different transformant DNA as the template for PCR. +, ethidium bromide resistant transformants isolated; -, no ethidium bromide resistant colonies isolated.

ysed bioinformatically to identified putative regulatory structures. The palindrome and equicktandem tools from EMBOSS explorer (Rice *et al.*, 2000) were used to search for inverted or direct repeat sequences that might indicate a protein binding site. No convincing palindromic sequences (minimum length of ten bases, less than three mismatches), or direct repeats (score >10) were found. A search against known riboswitch elements using RibEx (Abreu-Goodger and Merino, 2005) revealed no significant matches, but did predict a transcriptional attenuator structure between genomic positions 1865743 and 1865827 (18 to 102 bp upstream of *patA*), which encompasses all three identified mutations. This prediction was supported by the independent prediction of a Rho-independent terminator located between positions 1865753 and 1865774 in a set of pre-computed Rho-independent terminator predictions for the *S. pneumoniae* R6 genome made by TransTermHP (Kingsford *et al.*, 2007).

A transcriptional attenuator is a sequence located upstream of a gene that, when transcribed into RNA, can fold into two or more competing secondary structures that modulate the expression of the downstream gene by affecting the association of RNA polymerase with the DNA (Henkin and Yanofsky, 2002). The putative transcriptional attenuator upstream of *patA* was predicted by RibEx to consist of three overlapping parts: a terminator, an anti-terminator and an anti-anti-terminator. Unexpectedly, however, the *patA* promoter sequence that was predicted earlier (section 4.9.3) overlapped the attenuator between the anti-anti-terminator loops. The mutations found in this region in transformants $R6^{M101}$, $R6^{M87}$ and $R6^{M168}$ all mapped to the predicted terminator structure. To predict the effect that these mutations might have, the RNA secondary structure of the terminator region from R6 was predicted using RNAfold, which is part of the Vienna RNA package (Lorenz *et al.*, 2011). A single stem-loop structure with a 3' U-rich tail was predicted, which is typical for a Rho-independent terminator (Henkin and Yanofsky, 2002). This is depicted in Figure 5.6A, along with the positions of the mutations found in $R6^{M101}$, $R6^{M87}$ and $R6^{M168}$. All three mutations mapped to predicted regions of base

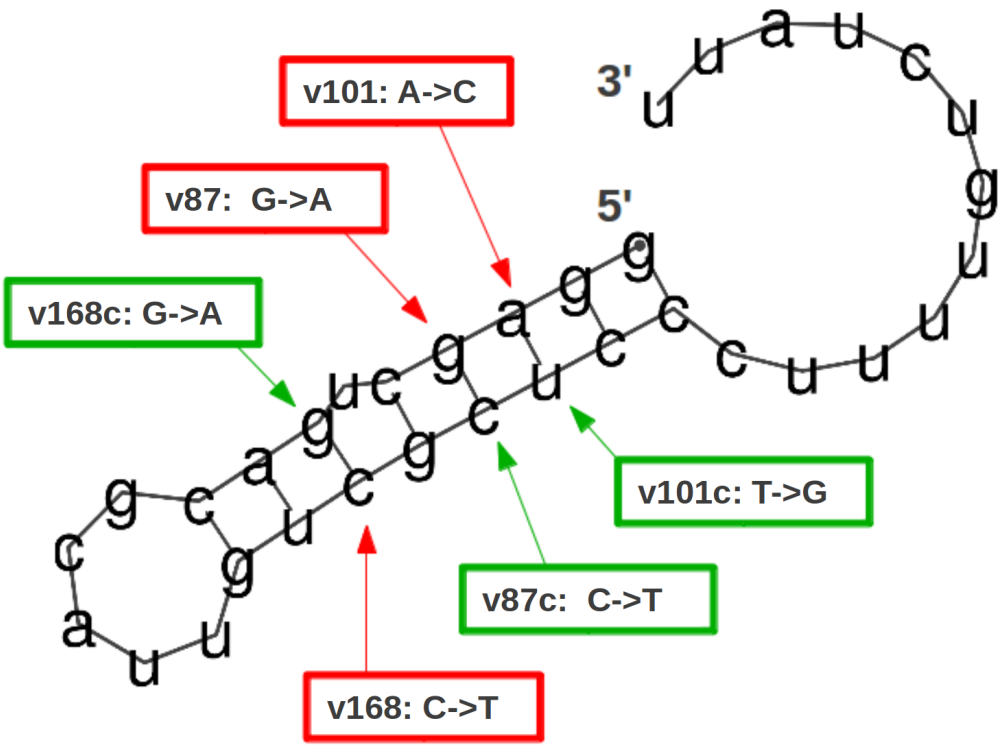
pairing in the terminator stem-loop. It was therefore hypothesised that the mutations in R6^{M101}, R6^{M87} and R6^{M168} cause *patAB* over-expression by disrupting the transcriptional terminator, causing *patA* and *patB* to be constitutively expressed. When the modelled RNA region was extended to include the whole attenuator, a second stem-loop structure was predicted that was identical in structure to the anti-antiterminator loop suggested by the RibEx server. Predicted energies of the wild type and mutated RNA structures were calculated using RNAeval and are shown in Figure 5.6B.

Ethidium bromide resistance in R6^{M74} is not conferred by the same mechanism as in R6^{M101}, R6^{M87} and R6^{M168} as no mutations were found in the terminator region in this strain. The narrowed-down region containing the causative mutation in this strain was 1864694 to 1866325, but excluding the internal region 1865626 to 1865982. Polymorphisms in these regions were compared against the sequence of the same region in the closely related strain JJA. High-quality SNPs found at positions 1865339, 1865583 and 1865599 were found in both R6^{M74} and JJA, indicating that they are natural polymorphisms and are unlikely to be involved in the R6^{M74} phenotype. However, examination of the raw, unfiltered SNPs in this region revealed a possible G:C to A:T mutation at position 1865940. This had previously been filtered out as the SNP quality score was only 25, but was confirmed to be a real polymorphism in R6^{M74} by PCR and sequencing. The mutated base is in codon 836 of the *hexA* gene, but does not cause a change in the protein sequence. A 600 bp region upstream of the start of *patA* was compared against the Rfam database to search for matches to known functional RNA structures. No significant matches were found.

5.6 Construction of a promoter probe plasmid

A plasmid vector was constructed to measure activity of cloned promoters by modification of the broad host-range plasmid pBAV1K-T5-*gfp* (Bryksin and Matsumura, 2010). This

A



B

	wt	v101	v101c	v87	v87c	v168	v168c
Folding Energy	-10.6	-6.1	-11.9	-3.2	-8.2	-7.4	-8.8

Figure 5.6. Prediction of structure of the Rho-independent terminator upstream of *patA* using RNAfold. (A) Predicted structure of the terminator. Mutations found in $R6^{M101}$, $R6^{M87}$ and $R6^{M168}$ sequences are indicated in red. Bases mutated to restore base pairing in promoter probe constructs v101c, v87c and v168c are indicated in green. **(B)** RNAeval prediction of free energy of folding for each terminator structure.

vector contains a stable GFP reporter gene controlled by a constitutive T5 promoter. This makes it ideal for applications such as monitoring the localisation of bacteria in a host, where stable, continuous expression of GFP is required for imaging. However, stable GFP is not suitable for promoter activity measurements as accumulation of active GFP prevents subtle trends in GFP expression throughout the growth cycle from being observed. Furthermore, in a promoter probe plasmid, GFP expression should be solely under the control of the tested promoter, meaning that the T5 promoter in pBAV1K-T5-*gfp* should be removed.

To fulfil these requirements, the T5-*gfp* cassette was replaced with a promoterless *gfp* gene from plasmid pMW82 (Lawler *et al.*, 2013), as described in section 2.9.1. The replacement *gfp* gene encodes an unstable version of GFP, and includes a 5' *EcoRI-XbaI* cloning site into which test promoters can be inserted. The new vector was named pBAV1K-*gfp82*.

To confirm that GFP was not expressed from pBAV1K-*gfp82* in the absence of a cloned test promoter, for example due to read-through from the kanamycin resistance gene into the *gfp* gene, cultures of R6 and R6 containing pBAV1K-*gfp82* were grown and fluorescence of a cell sample was measured at early and late logarithmic phase and at stationary phase. No significant differences in fluorescence were observed between R6 with and without pBAV1K-*gfp82* at any stage of growth.

5.7 Measuring GFP expression from *patA* promoter regions containing mutated and complemented transcriptional attenuators

To test whether mutations in the terminator region found upstream of *patA* lead to increased gene expression, the *patA* promoter and attenuator sequences were cloned into pBAV1K-*gfp82* and the effect of the terminator mutations found in R6^{M101}, R6^{M87} and R6^{M168} on GFP expression was measured. To do this, seven variant DNA oligonucleotides were synthesised by GeneArt that corresponded to a 134 bp region of the upstream sequence of *patA*, excluding the Shine-Dalgarno sequence (positions -6 to -11 from the *patA* start codon), with added restriction sites (5' *EcoRI* site and 3' *XbaI* site) for cloning. The first synthesised oligonucleotide corresponded to the sequence of this region in R6 and the vector containing this sequence was named pBAV1K-*gfp82*wt throughout the following experiments. The remaining six oligonucleotides were mutated as follows (depicted in Figure 5.6A). Oligonucleotides v101, v87 and v168 contained mutations corresponding to the mutations identified in the terminator structure in R6^{M101}, R6^{M87} and R6^{M168}, respectively. Oligonucleotides v101c, v87c and v168c were constructed by mutating the base that is predicted to pair with the mutated base in each strain, as determined by the RNAfold model, to the appropriate base such that base pairing in the stem-loop should be restored. Predicted energies of the RNA structures with restored base pairing are shown in Figure 5.6B.

The resulting pBAV1K-*gfp82* constructs were transformed into R6 cells. A preliminary attempt was made to measure fluorescence from one ml samples of logarithmic phase cells resuspended in 200 µl PBS, as described for p^{patA} and tRNA-p^{patA} constructs in section 4.9.3. To assess the degree of variation in replicate fluorescence measurements in the FluoStar Optima plate reader, resuspended cells for three biological replicates

of each culture were added to a microtitre tray, and fluorescence was measured from the same samples three times in quick succession, with five seconds of shaking between each reading. Considerable variation was observed between replicate measurements of the same sample (Figure 5.7). This implied that there was background variation in the fluorescence measurements made by the FluoStar Optima that was large enough to obscure putative differences in fluorescence between samples. To allow smoothing of fluorescence measurements to distinguish genuine signal from background noise, it was decided to measure fluorescence continuously throughout growth of strains containing each construct. This also allowed investigation of differences in relative expression of GFP from the *patA* promoter in different phases of growth. To do this, cultures containing pBAV1K-*gfp82*wt, v101, v101c, v87, v87c, v168 or v168c were grown in triplicate in BHI broth in a microtitre tray and simultaneous fluorescence and absorbance measurements were taken using a Fluostar Optima plate reader. No significant differences in growth kinetics were identified between any of the strains (Figures 5.8, 5.9 and 5.10, dashed lines).

Cells containing the pBAV1K-*gfp82*v101 construct showed a consistently higher level of fluorescence than that of the pBAV1K-*gfp82*wt construct from two hours of growth (mid-logarithmic phase) onwards (Figure 5.8). From the pBAV1K-*gfp82*v101c construct, where the corresponding base in the terminator structure had been mutated to restore base pairing, fluorescence was consistently lower than that of the wild type construct (Figure 5.8). Cells containing the $R6^{M87}$ and $R6^{M168}$ promoter constructs also showed higher levels of fluorescence than wild type from two hours onwards (Figures 5.9 and 5.10). However, for these attenuator variants, mutation of the opposing base in the predicted terminator structure did not result in reduced fluorescence. In fact, the pBAV1K-*gfp82*v87c and pBAV1K-*gfp82*v168c constructs produced consistently higher levels of fluorescence than pBAV1K-*gfp82*v87 and pBAV1K-*gfp82*v168, respectively (Figures 5.9 and 5.10).

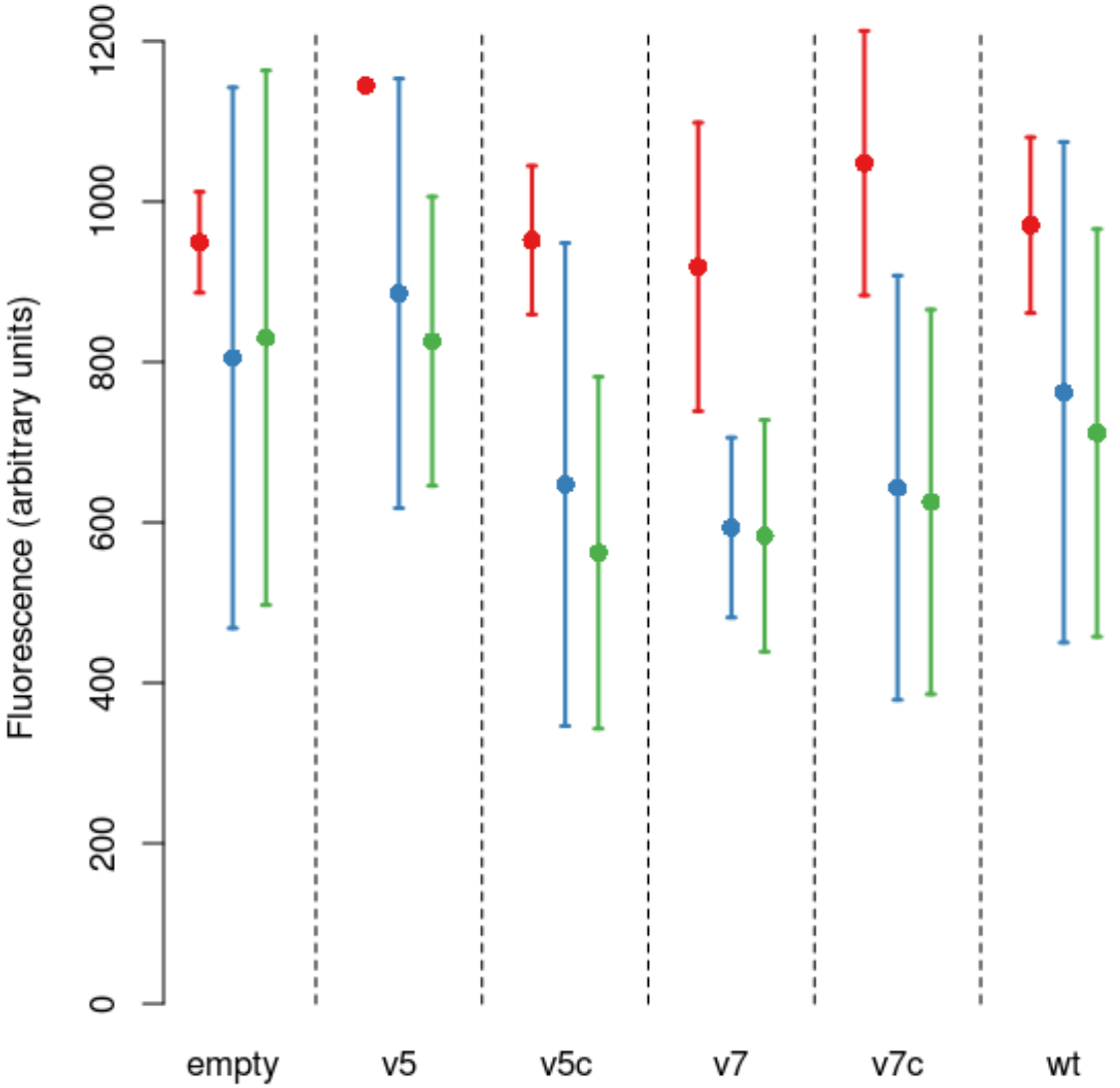


Figure 5.7. Variation in endpoint fluorecence measurements from 1 ml R6 cells containing pBAV1K-*gfp2*-p^{patA} constructs resuspended in 200 µl PBS. Coloured points represent independent biological replicate measurements for each construct. Error bars represent the standard deviation of three technical replicate measurements.

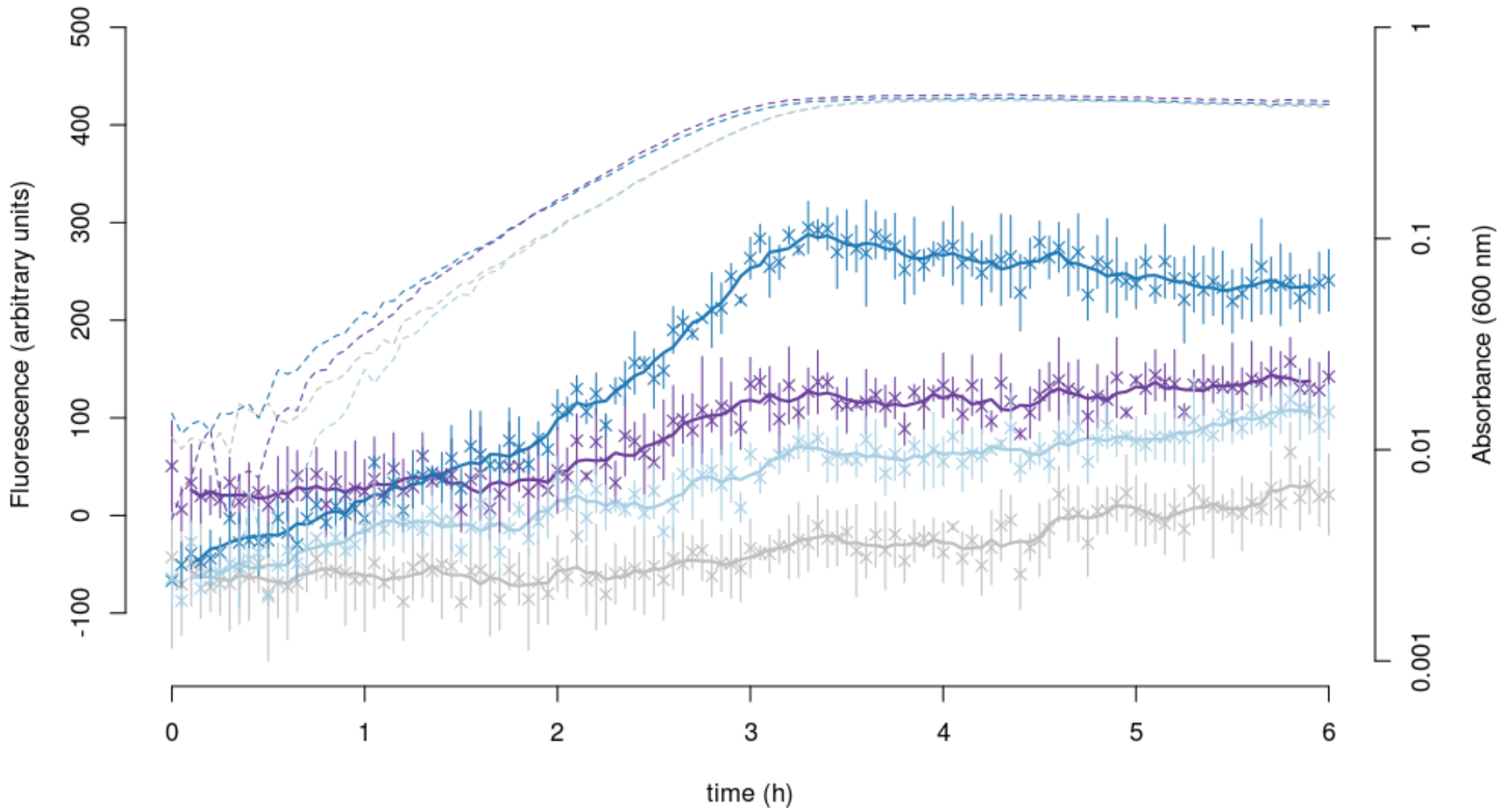


Figure 5.8. Fluorescence from pBAV1K-*gfp*-v101 and pBAV1K-*gfp*-v101c transcriptional fusions during growth in BHI broth. Fluorescence (solid lines) and absorbance at 600 nm (dashed lines) were measured simultaneously from growing cultures of R6 containing empty pBAV1K-*gfp82* vector ●, or the P^{patA} promoter constructs wt ●, v101 ● or v101c ●. Points and error bars represent average and standard deviation of fluorescence from three biological replicates. Solid lines represent a smoothed average of fluorescence readings over a 15 minute window.

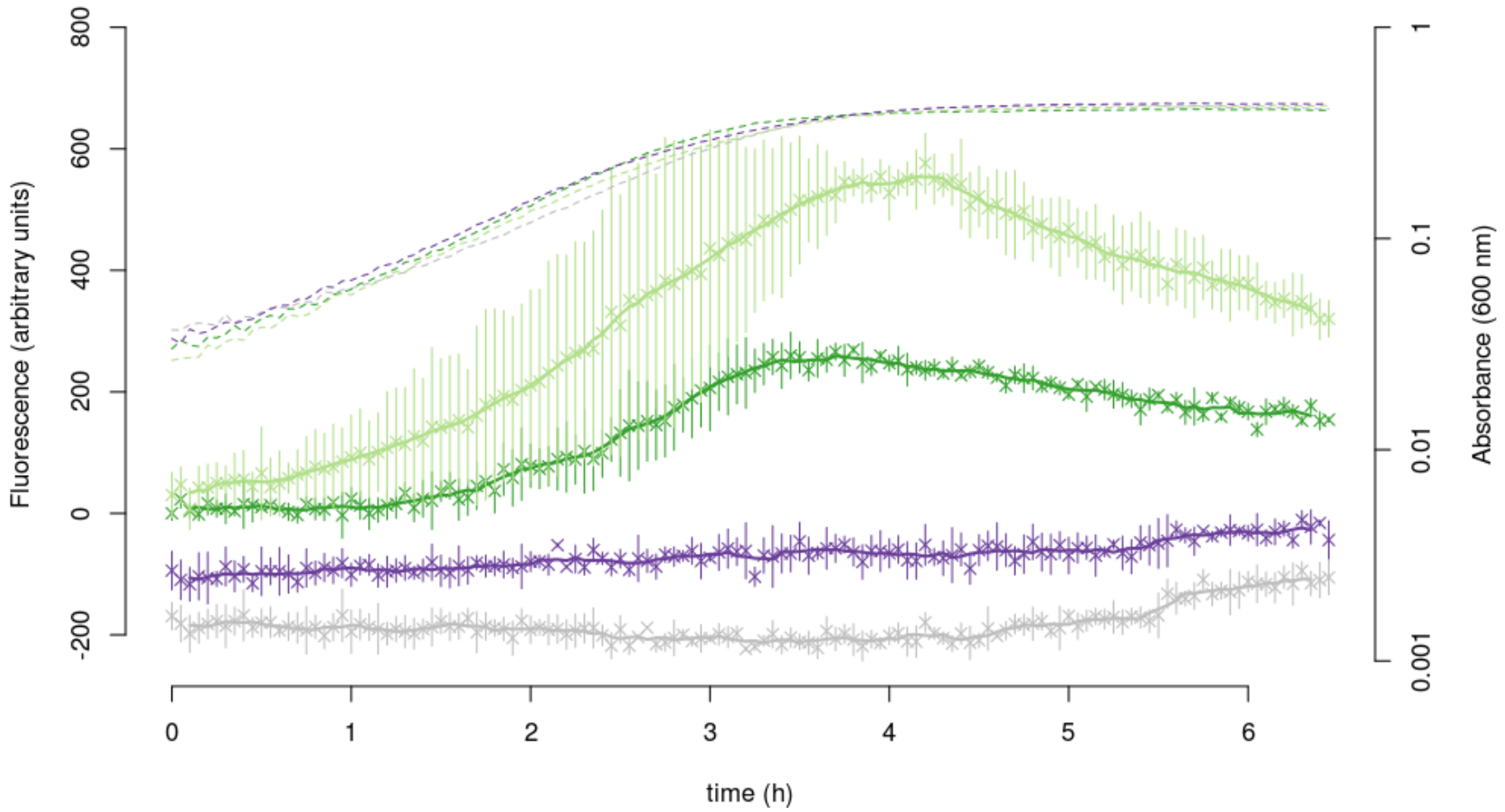


Figure 5.9. Fluorescence from pBAV1K-*gfp*-v87 and pBAV1K-*gfp*-v87c transcriptional fusions during growth in BHI broth. Fluorescence (solid lines) and absorbance at 600 nm (dashed lines) were measured simultaneously from growing cultures of R6 containing empty pBAV1K-*gfp82* vector ●, or the $\int^{patA}$ promoter constructs wt ●, v87 ● or v87c ●. Points and error bars represent average and standard deviation of fluorescence from three biological replicates. Solid lines represent a smoothed average of fluorescence readings over a 15 minute window.

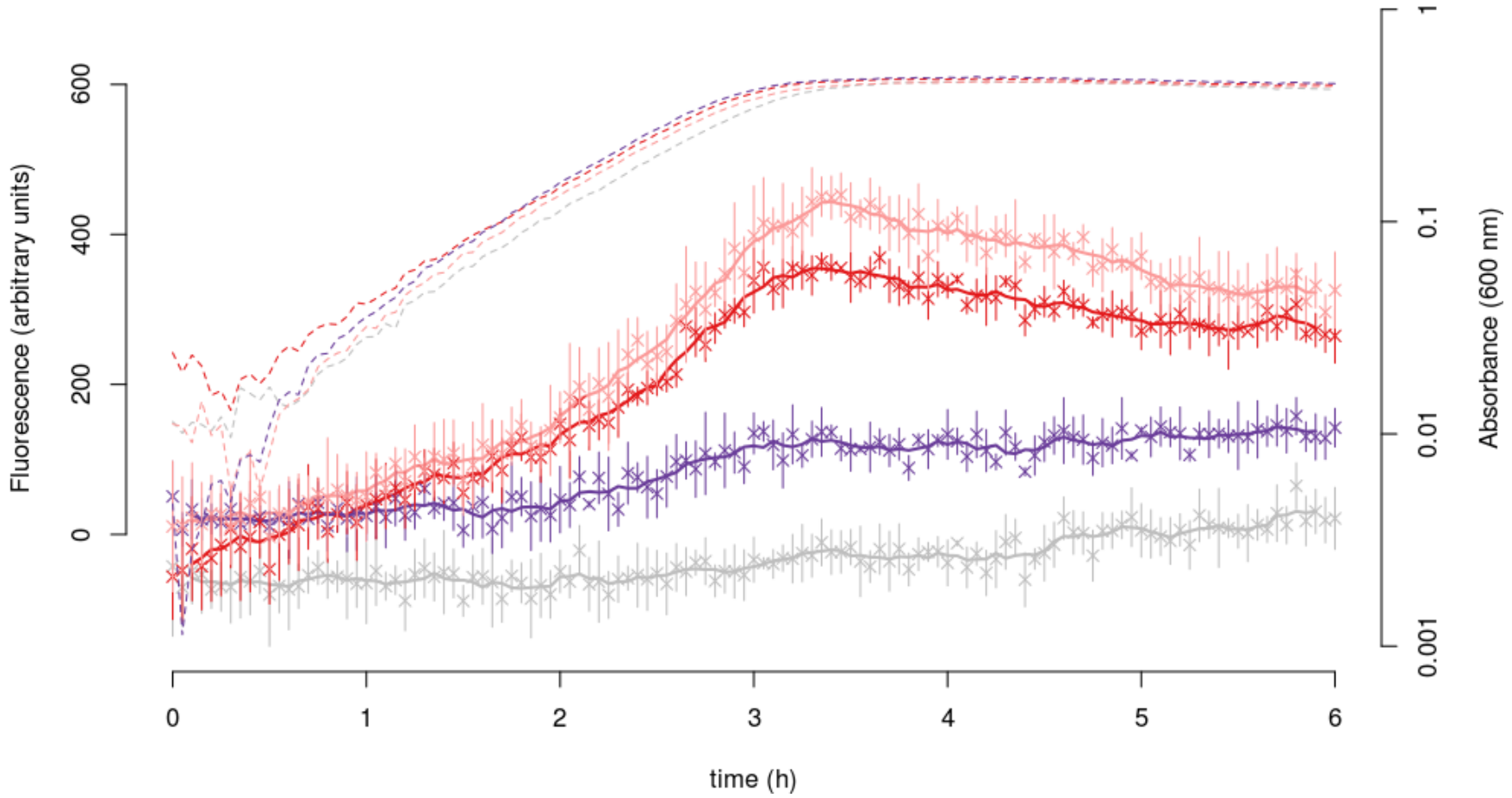


Figure 5.10. Fluorescence from pBAV1K-*gfp*-v168 and pBAV1K-*gfp*-v168c transcriptional fusions during growth in BHI broth. Fluorescence (solid lines) and absorbance at 600 nm (dashed lines) were measured simultaneously from growing cultures of R6 containing empty pBAV1K-*gfp82* vector ●, or the P^{patA} promoter constructs wt ●, v168 ● or v168c ●. Points and error bars represent average and standard deviation of fluorescence from three biological replicates. Solid lines represent a smoothed average of fluorescence readings over a 15 minute window.

To determine whether relative expression of GFP per cell differed at different stages of growth, values of fluorescence divided by optical density were averaged over 20 minute windows centered on two hours (mid-logarithmic phase), three hours (late-logarithmic phase) and five hours (stationary phase) (Figure 5.11). The fluorescence of each strain relative to its optical density was consistent across these different stages of growth. This suggests that GFP production from these constructs occurs at a consistent rate and is not up- or downregulated in particular growth phases, meaning that fluorescence is proportional to the cell density. This analysis also showed that fluorescence from construct pBAV1K-*gfp82*v101 was significantly higher than that of wild type construct in late-logarithmic, stationary and late-stationary phases of growth ($p<0.05$, one-tailed Student's t-test with Bonferroni correction for multiple testing). The v168 and v168c constructs both produced higher levels of fluorescence than wild type in all the tested growth phases ($p<0.05$). The v101c constructed fluoresced at a significantly lower level than R6 in all tested growth phases ($p<0.05$).

5.8 Discussion

Genetic changes causing over-expression of *patA* and *patB* in three clinical isolates and the laboratory mutant M168 were successfully transferred into the R6 genetic background by whole genome transformation. In all four of the tested transformants a recombination had occurred that affected the genomic region encompassing *patA* and *patB*. In three of the transformants the mutation causing *patAB* over-expression was localised to a Rho-independent terminator sequence directly upstream of *patA*. Different mutations affecting this structure have been previously observed in other isolates over-expressing *patAB* (Billal *et al.*, 2011; Croucher *et al.*, 2013)

In the four isolates investigated here, mutations causing upregulation of *patAB* were identified by sequencing R6 transformed with genomic DNA from the isolates, rather

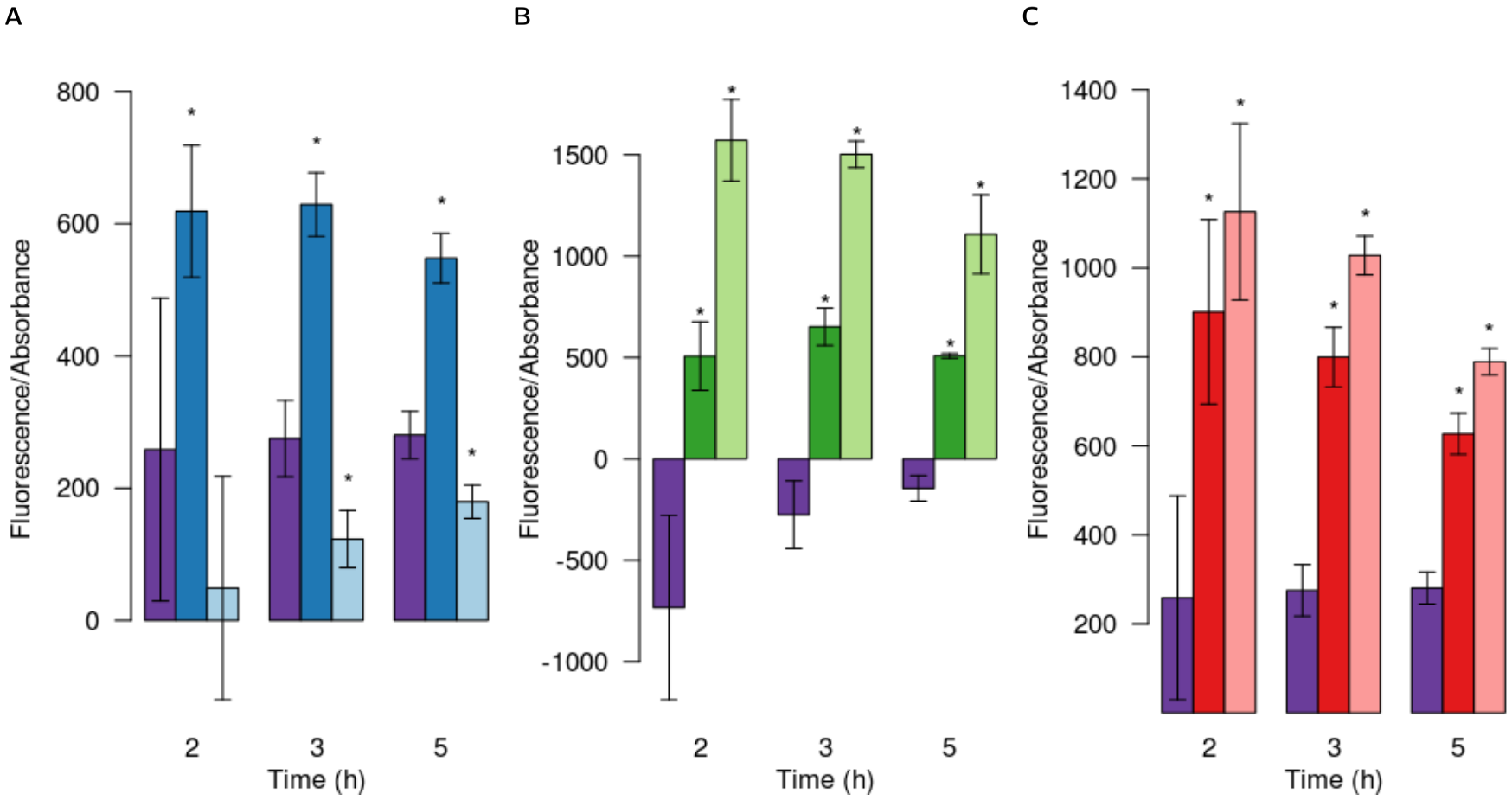


Figure 5.11. Fluorescence per OD_{660} unit from p^{patA}-*gfp* transcriptional fusions during three phases of growth Fluorescence/Absorbance for R6 containing p^{patA} variants v101 ●, v101c ●, v87 ● or v87c ●, v168 ● or v168c ● was compared to that of p^{patA}-wt ●, at two, three and five hours of growth (mid-logarithmic, late logarithmic and stationary phases of growth, respectively). Bars represent fluorescence per OD unit smoothed over a twenty minute window. Error bars represent standard deviation of three biological replicates. *, fluorescence per OD unit is significantly different from wt, two-tailed Student's t-test, with Bonferroni correction for multiple testing.

than sequencing the isolates directly. The main advantage of this approach was that all the transformants could be compared to a common reference genome and regions where recombination had occurred could be readily identified due to the presence of clusters of SNPs (Croucher *et al.*, 2012). For the three clinical isolates, isogenic drug susceptible parental strains were not available for comparison. Unless the isolates were very closely related to a previously sequenced genome, sequencing directly could have resulted in identifying large of polymorphisms, which would have obscured SNPs related to the investigated phenotype. Identifying recombinant regions in R6 transformants significantly narrowed down mutations causing the phenotype to a smaller area of the genome. Although the eventual conclusion of this analysis was that the causative mutations were associated with the *patAB* locus itself, if this had not been the case, the method used here could have identified recombinant DNA segments anywhere in the genome.

The *patAB* over-expression phenotype was transferred from all four isolates into R6 by a single round of transformation. However, variation was observed in the phenotypes of the selected ethidium bromide resistant transformants. Several transformants did not accumulate significantly lower levels of ethidium bromide than R6, although their ethidium bromide MICs were elevated. Also, the MIC of norfloxacin for the $R6^{M74}$ that was picked as M507 was two dilutions higher than for the other two tested transformants. Transformation with DNA of foreign origin may result in a higher frequency of selection of spontaneous ethidium bromide resistant mutants than transformation with R6 DNA, due to secondary recombinations elsewhere in the genome. This could have been tested by using M4 DNA as the negative control for transformation experiments instead of R6 DNA. The frequency of occurence of secondary recombinations could also be reduced by transforming with a lower concentration of DNA (Croucher *et al.*, 2012).

Over-expression of *patAB* was confirmed in three of the four transformants. The $R6^{M74}$ transformant M507 did not significantly over-express *patA* and *patB*, but transcription of *guaA* was significantly increased. The reasons for this are unclear. However, the MICs

of norfloxacin and ciprofloxacin were unusually high for M507 compared to all the other tested transformants (Table 5.1). It is therefore possible that this particular transformant has acquired a secondary mutation in addition to the during the transformation process that modulates *patAB* expression and increases norfloxacin resistance. Alternatively, it may be a spurious multidrug resistant mutant selected by exposure to ethidium bromide. In the other three tested transformants, *patA* and *patB* were expressed at levels three- to eight-fold higher than in R6. This over-expression is much lower than that observed in $R6^{M184}$ transformants (section 4.9.1).

Three out of four of the transformants contained more than one recombinant region. Four recombinations were observed in $R6^{M74}$, 14 in $R6^{M101}$ and 24 in $R6^{M87}$. This observation is consistent with previous work by Croucher *et al* (2012), which showed that several recombination events can occur simultaneously in the same cell. In the $R6^{M168}$ transformant, the single observed recombination affected a genomic region from approximately 1856000 to 1868000, which includes the *patAB* genes. The remaining three transformants all contained a recombination that overlapped with this region, which suggested that this was the primary recombination causing the ethidium bromide resistance phenotype in all four transformants. This was confirmed by the ability of PCR amplimers from this region to confer ethidium bromide when transformed into R6. Interestingly, transformation of R6 with *patA* amplimers from M507 ($R6^{M74}$ produced ethidium bromide resistant transformants, despite qRT-PCR results that suggested that M507 did not over-express *patA* and *patB*. This could indicate that a secondary mutation has occurred in M507 that modulates *patAB* expression, but the original mutation still exists and can cause *patAB* upregulation when transferred to another strain.

The mutations causing *patAB* over-expression in $R6^{M168}$, $R6^{M101}$ and $R6^{M87}$ were localised to a putative Rho-independent terminator upstream of *patA* by transformation of R6 with PCR amplimers progressively decreasing in size. Rho independent transcriptional terminators consist of an RNA stem-loop structure that arrests the transcription

complex, followed by a uracil-rich tail that promotes dissociation of the polymerase from the DNA (Gusarov and Nudler, 1999). Rho-independent termination is predominant over Rho-dependent termination in Gram-positive bacteria (Washburn *et al.*, 2001). This is particularly true in *S. pneumoniae*, which does not contain a homologue of the Rho protein (Opperman and Richardson, 1994). The terminator directly upstream of *patA* may act to prevent transcriptional read-through into *patA* from the upstream *hexA* gene. Alternatively, it could form part of a regulated transcriptional attenuator structure.

The effect of terminator mutation on downstream gene expression were investigated by transcriptional fusion of the predicted *patA* promoter, combined with wild-type or mutated terminator sequences, with GFP in the pBAV1K-*gfp82* plasmid. All three mutant terminators allowed higher levels of expression of GFP than the wild type terminator during all phases of growth. However, when the predicted base pairing of the terminator was restored by mutation of the opposing base in the stem-loop, GFP expression was not reduced from two of the three constructs (v168c and v87c). In contrast, GFP expression from the v101c construct was significantly lower in all growth phases than from the wild type construct, suggesting that repression was restored to a greater extent than in the wild type attenuator. These observations could be explained by the difference in strength between AT and GC base pairs. In the complemented $R6^{M101}$ attenuator, the original AT base pairing interaction was replaced with a stronger CG base pair, which could result in higher stability. In contrast, in the complemented $R6^{M87}$ and $R6^{M168}$ attenuators, strong CG base pairs were replaced by weaker TA base pairs, which may not have been sufficient to restore the terminator structure. Also, the mutated base pair in the v168 and v168c terminators was located adjacent to an unpaired base, whereas the mutated base pair in v101 and v101c was in the centre of a helix. The presence of unpaired bases next to the mutated bases in v168c could mean that the base pairing interactions formed are different from those naively predicted from the wild type terminator structure.

The GFP fusion vector used to monitor expression from the *patA* promoter in combina-

tion was the same as that used to measure read-through from $tRNA^{Glu}$ into *patA* in M184 (section 4.9.3). As discussed in Chapter 4, there are some limitations to this approach. The main limitation is that the GFP allele from pMW82 has not been validated for use in *S. pneumoniae*. The maximum fluorescence observed from the constructs in this study was approximately 4,500 fluorescence units (measured from broth without blank correction), whereas fluorescence values up to ten-fold higher have been observed using the same reporter gene in *S. Typhimurium* in the same laboratory (Lawler *et al.*, 2013). This suggests that this GFP may not function optimally in *S. pneumoniae*, limiting the sensitivity of the assay.

Fluorescence from the terminator constructs was measured during growth in BHI broth over several hours, instead of taking single endpoint fluorescence measurements as was done for the p^{patA} and tRNA-pp^{patA} constructs in the previous chapter (section 4.9.3). This approach was taken as a preliminary experiment suggested that there was considerable variation in three fluorescence readings taken in quick succession from the same sample of cells (Figure 5.7). The differences in fluorescence between wild-type and mutant terminator constructs were less pronounced than between p^{patA} and tRNA-p^{patA} (section 4.9.3), which meant that the large signal to noise ratio had a greater effect on comparisons between constructs. Measuring fluorescence at regular intervals over the whole growth cycle allowed smoothing of the data, allowing trends representing genuine differences in fluorescence between constructs to be distinguished from the background variation. No differences in growth were observed between strains carrying any of the tested constructs, suggesting that the levels of GFP produced by these constructs were not toxic.

The Rho-independent terminator upstream of *patA* was predicted by RibEx (Abreu-Goodger and Merino, 2005) to form part of a transcriptional attenuator. Transcriptional attenuation, defined as the premature termination of the elongation phase of transcription, is an important regulatory strategy in bacteria. Attenuators are conserved RNA

structures in the leader sequence of regulated genes that can fold into at least two different confirmations, one of which is a Rho-independent terminator (Henkin and Yanofsky, 2002). The simplest, and most common, attenuator structure is an RNA sequence that can fold into two mutually exclusive stem-loops, one of which is a terminator. The alternative stem-loop of which is an anti-terminator that permits passage of RNA polymerase and prevents formation of the terminator structure (Henkin and Yanofsky, 2002). The attenuator structure predicted upstream of *patA* consisted of this basic structure, with the addition of a third, anti-anti-terminator stem-loop.

Transcriptional attenuation is conserved throughout all bacteria, but is particularly common in the Firmicutes, where 2.6% of genes are controlled by transcriptional attenuation, compared to 1.6% bacteria in general (Naville and Gautheret, 2010a). Expression of a gene downstream of a transcriptional attenuator is regulated by cellular signals that selectively stabilise or destabilise one of the conformations of the attenuator (Henkin and Yanofsky, 2002). A variety of different sensory mechanisms have been identified. For example, riboswitches sense small metabolites using a conserved sensory domain that folds into a complex secondary structure (Coppins *et al.*, 2007; Dambach and Winkler, 2009; Vitreschak *et al.*, 2004). A variety of metabolites can be sensed by riboswitches, including purines, glycine, flavin mononucleotide and S-adenosyl methionine (Epshtein *et al.*, 2003; Mandal and Breaker, 2004; Mandal *et al.*, 2004; McDaniel *et al.*, 2003; Mironov *et al.*, 2002; Winkler *et al.*, 2002, 2003). Similarly, T-boxes encode structured RNA leader sequences, and these fold into a secondary structure that responds to uncharged tRNAs (Grundy and Henkin, 1993; Grundy *et al.*, 2002). Some attenuators are regulated *in trans* by an anti-termination protein, which either stabilises an antiterminator structure or prevents its formation (Amster-Choder, 2005; Bonner *et al.*, 2001). Transcriptional attenuation can also be mediated by a ribosome-dependent mechanism, where a ribosome will stall at a leader sequence enriched in codons for a particular amino acid, when levels of that amino acid are low, and prevent terminator formation

(Vitreschak *et al.*, 2004).

In the last decade there have been several studies that have attempted to computationally detect the full complement of genes controlled by transcriptional attenuation in order to identify regulatory signals. The two most comprehensive were performed by Merino and Yanofsky (2005) and Naville and Gautheret (2010b). Merino and Yanofsky searched 180 completed bacterial genomes for sequences containing mutually exclusive RNA stem-loop structures, and grouped the downstream genes based on their classification in the clusters of orthologous genes (COG) database. In the COG1132 family of ABC-type multidrug transport systems, putative transcriptional attenuators were found upstream of 153 of 302 members, including *patA* in *S. pneumoniae* R6. The predicted attenuator sequence was identical to that predicted in this study, although this is unsurprising as the algorithm used to detect transcriptional attenuators is identical to that implemented in the RibEx webserver (Abreu-Goodger and Merino, 2005; Merino and Yanofsky, 2005). However, no predictions were made by Merino and Yanofsky as to what might regulate the attenuator. In a similar study by Naville and Gautheret the *patA* transcriptional attenuator was not identified. A different algorithm was used to detect attenuators in this study, which was based on sequence similarity to a pre-computed profile of a Rho-independent terminator, instead of prediction of RNA secondary structure.

Riboswitches and T-boxes have conserved structures due to structural constraints required to bind metabolites and tRNAs (Vitreschak *et al.*, 2003, 2008). They are also relatively large (100-200 nt for riboswitches and 300 nt for T-boxes), so are usually characterised by long 5' untranslated regions upstream of genes (Barrick *et al.*, 2004; Grundy and Henkin, 1993). The upstream sequence of *patA* did not match any known RNA structures in the Rfam database (Griffiths-Jones *et al.*, 2005), and the intergenic region between *hexA* and *patA* is only 147 nt. Taken together, these observations suggest that this transcriptional attenuator is not a riboswitch or T-box. If the putative attenuator controls *patAB* expression, it could be regulated *in trans* by a protein factor that binds

to the anti-anti-terminator. This has been observed previously for the *pyr* operon in *B. subtilis*, which encodes genes involved in uracil uptake and biosynthesis (Bonner *et al.*, 2001). When cellular levels of UMP are high, the PyrR anti-termination factor represses the operon by binding to the anti-anti-terminator loops of three attenuators distributed throughout the operon. PyrR binding is reduced when UMP levels decrease, promoting formation of the anti-terminator, which is thermodynamically more stable than the terminator loop.

However, the hypothesis that expression of *patA* is regulated by a transcriptional attenuator is not supported by the predicted position of the *patA* promoter, which is postulated to be located within the predicted attenuator overlapping the anti-anti-terminator and anti-terminator loops. Transcription from this promoter precludes the attenuator from regulating transcription as the anti-anti-terminator and anti-terminator loops will not be transcribed.

There are several possible explanations for this observation:

1. The *patA* promoter prediction may be inaccurate. However, the prediction made by PPP (section 4.9.3) is supported by predictions from two other bacterial promoter prediction webservers, BPROM (Softberry Inc., NY, USA) and PePPER (de Jong *et al.*, 2012). This gives weight to the hypothesis that it is the most likely promoter sequence and not a spurious prediction from a single piece of software. Furthermore, there must be a valid promoter in the DNA regions cloned into pBAV1K-*gfp82*, otherwise GFP expression would not have been observed, regardless of the structure of the transcriptional terminator.

2. Alternatively, the attenuator prediction may be a false positive. The presence of the Rho-independent terminator was detected by RibEx and TransTermHP, which use the same underlying algorithm for prediction (Abreu-Goodger and Merino, 2005). However, the attenuator was not predicted by the method used in the

genome-wide study be Naville and Gautheret which uses the algorithms from the attenuator prediction program ARNold (Naville *et al.*, 2011).

3. Another possibility is that there might be a second *patA* promoter further upstream of the *patA* gene. The promoter located close to the *patA* start codon may control basal expression, while expression from the hypothetical alternative promoter could be inducible via the transcriptional attenuator. The nearest promoter detected in the PPP analysis was 3.2 kb upstream of *patA*. However, Merino and Yanofsky predicted a possible promoter sequence 232 bp upstream of *patA*, within the *hexA* gene (Merino and Yanofsky, 2005, supplementary data). Presumably, this prediction was made for the DNA sequence within a certain distance of the predicted attenuator, with the attenuator structure itself being excluded. However, the authors do not provide details of the promoter prediction algorithm used. This predicted promoter had a -35 box sequence of TTGATA and a -10 box sequence of TATCCT, separated by 18 bp. Inducible transcription from an alternative promoter located within an upstream gene was recently observed for the *cbpA* gene in *E. coli* (Chintakayala *et al.*, 2013). In this case, expression of *cbpA* was repressed during exponential phase growth by binding of the nucleoid protein Fis, which blocked transcription from the upstream promoter. Expression of *cbpA* was then increased in stationary phase, due to reduction in concentration of Fis. It is possible that a similar mechanism may operate to control expression of *patA*, mediated by attenuation of transcription from the predicted alternative promoter within *hexA*.

4. Alternatively, the attenuator may regulate read-through into the *patA* gene from the *hexA* gene. If this were true, basal transcription of *patA* could be mediated transcription from the predicted *patA* promoter and inducible expression could be mediated by repression of termination between *hexA* and *patA* mediated by the transcriptional attenuator. This hypothesis may also explain why the mutation

found in $R6^{M74}$, a synonymous SNP towards the 3' end of *hexA*, potentially affects expression of *patA*, if it alters transcription from the end of the *hexA* gene.

5. Finally, the predicted attenuator may be real, but could be a non-functional evolutionary relic. The genomic context of the *patA* gene is not conserved in other *Streptococci* (Escudero *et al.*, 2011; Garvey and Piddock, 2008), and the presence of the spr1886 degenerate transposase gene suggests that a genomic rearrangement has occurred in this region. It is possible that the attenuator conferred inducible expression of a downstream gene in a previous context, but is now non-functional due to loss of an upstream promoter. Evidence of acquisition, loss and switching of attenuators upstream of some families of genes throughout evolution was reported in both previously discussed genome-wide studies of transcriptional attenuation (Merino and Yanofsky, 2005; Naville and Gautheret, 2010b).

However, to distinguish between these possibilities, further work will be required to accurately map *patA* transcriptional start sites.

5.9 Further Work

The sensitivity of the GFP assay used in this work was limited by the low fluorescence in *S. pneumoniae* of the version of GFP used. Better differentiation between levels of reporter expression from the contructs tested here could be achieved by optimisation of this assay. This could provide a high-throughput method to measure inducible expression of *patA*, as has been done for *ramA* expression using the pMW82 reporter in *Salmonella* (Lawler *et al.*, 2013). Optimisation of *gfp* fluorescence in *S. pneumoniae* could be achieved by using an alternative GFP allele, which may produce higher fluorescence, such as the GFP variant used by Ruiz-Cruz *et al* (Ruiz-Cruz *et al.*, 2010) in a different *S. pneumoniae* reporter plasmid. Alternatively, the *gfp* gene could be replaced with a

luciferase operon, for example from pBAV1K-T5-*lux* (Bryksin and Matsumura, 2010). BHI broth emits a high level of background fluorescence at the wavelengths used to detect GFP fluorescence, but does not auto-luminesce. Using luciferase as a reporter instead of GFP would therefore be an advantage for assays involving growth of strains in broth, as it would reduce background fluorescence.

A putative transcriptional attenuator was detected upstream of *patA* but it has not been shown whether this can control inducible expression of *patA* in response to fluoroquinolones. This could be tested by cloning larger fragments of the upstream sequence of *patA* into an improved version of the reporter plasmid, and measuring changes in GFP/luciferase production in response to sub-MIC concentrations of fluoroquinolones. Cloned fragments should include either the putative alternative *patA* promoter located within the *hexA* gene, or the whole *hexA* gene sequence, to determine whether regulated expression of *patA* is mediated by increased read-through from *hexA* or by expression from an alternative promoter.

To further investigate inducible expression of *patA* potentially mediated by the transcriptional attenuator, the exact location of transcription start sites for *patA* need to be experimentally verified. Transcription start sites could be determined using RNAseq. The advantages of RNAseq are that multiple transcription start sites can be visualised, transcription from each DNA strand can be investigated separately, and data relating to the rest of the transcriptome collected during the experiment can be used to answer other biological questions. Alternatively, *patA* transcription start sites could be investigated in isolation by carrying out 5' rapid amplification of cDNA ends, mRNA primer extension assays (Lloyd *et al.*, 2008) or DNA footprinting of RNA polymerase binding sites (Singh and Grainger, 2013).

The origins of the putative transcriptional attenuator upstream of *patA* could also be further investigated by determining the phylogeny of the *patA* and *patB* genes, as it has

been previously observed that the genomic context of these genes is different in other streptococci (Escudero *et al.*, 2011; Garvey and Piddock, 2008).

Finally, it was observed that, when R6 was transformed with DNA from clinical isolates that over-expressed *patAB*, not all of the selected ethidium bromide resistant transformants showed a phenotype typical of *patAB* over-expression. It would be interesting to determine the genome sequences of these transformants to find the cause of their ethidium bromide resistance.

5.10 Key Findings

- Over-expression of *patAB* could be transferred to R6 by transformation with whole genomic DNA from three clinical isolates and a laboratory mutant from a non-R6 genetic background
- Expression of *patAB* in the transformants was up to eight-fold higher than R6, which was much lower than that observed in M184
- Regions where recombination had occurred between recipient and donor DNA in the transformants could be readily identified from Illumina sequencing data by defining clusters of SNPs
- In all four investigated transformants a primary recombination had occurred in the *patAB* region of the genome
- In three of the transformants, other, secondary recombination events had occurred throughout the genome
- Mutations in a Rho-independent transcriptional terminator upstream of *patA* caused over-expression of *patA* and *patB* in three out of four transformants
- This terminator structure may be part of a transcriptional attenuator that could

explain inducible expression of *patAB*

6 Overall Discussion

The two main aims of this study were to establish the role of over-expression of *patAB* in fluoroquinolone resistance in clinical isolates, and to identify how expression of *patA* and *patB* is controlled.

The first of these aims was accomplished, with a strong association shown between over-expression of *patAB* and cross-resistance to both fluoroquinolones and dyes in clinical isolates (Chapter 3). Similar results have been observed in fluoroquinolone resistant clinical isolates investigated in other studies, two by El Garch *et al* (2010) and one by Lupien *et al* (2013). However, to date, *patAB* over-expression has not been investigated in any larger screening studies. Some isolates in the set investigated had phenotypes that did not fit the general pattern. In particular, one isolate was observed which expressed low levels of *patAB*, yet was resistant to fluoroquinolones and dyes and accumulated low levels of Hoechst 33342. As discussed previously (section 3.4), this could be explained by expression of a different efflux pump in this isolate, such as PmrA or DinF (Gill *et al.*, 1999; Tocci *et al.*, 2013).

The second aim of this study was to identify how transcription of *patAB* is controlled. It was postulated that there is a transcriptional regulator of *patAB* expression because induction of transcription of *patAB* had been observed following exposure to fluoroquinolone antibiotics and mitomycin C (El Garch *et al.*, 2010; Marrer *et al.*, 2006a). This led to the hypothesis that *patAB* is upregulated as part of a response to DNA damage (El Garch *et al.*, 2010); subsequent studies have also implicated PatAB in tol-

erance of high pH (van Opijnen and Camilli, 2012), meningitis (Croucher *et al.*, 2013) and alleviating the fitness cost of linezolid resistance (Billal *et al.*, 2011). It has been well established that multidrug efflux pumps in other bacteria have diverse functions beyond antibiotic transport (Piddock, 2006b), and for several, particularly in *E. coli* and *Salmonella spp.*, complex regulatory networks controlling efflux pump expression have been elucidated (Nishino *et al.*, 2009). Knowledge of the regulatory network controlling an efflux pump can provide insight into its physiological role. For instance, the MacAB ABC transporter in *Salmonella* is co-regulated with many other genes involved in virulence under the control of the master regulator PhoPQ (Zwir *et al.*, 2005). However, no direct regulators of *patAB* expression had been identified prior to this study. Several global regulators have been identifed in *S. pneumoniae* such as StkP (Saskova *et al.*, 2007), CiaR (Halfmann *et al.*, 2007) and CodY (Hendriksen *et al.*, 2008). The effects of inactivation of these regulators on the pneumococcal transcriptome have been investigated, but *patAB* was not represented in any of the indentified putative regulons.

In this study, the approach used to identify putative regulators of *patAB* was to sequence the genomes of strains with constitutive over-expression of *patAB* and identify causative mutations. Constitutive over-expression of efflux pumps is commonly associated with mutations in their regulators. There are many examples of this in both laboratory selected mutants and clinical isolates. For example, loss of function mutations in AcrR and RamR, and gain of function mutations in RamA have been observed which lead to over-expression of AcrAB-TolC in *Salmonella enterica* (Abouzeed *et al.*, 2008; Chiu *et al.*, 2005; Zheng *et al.*, 2009), and mutations in the AdeRS two-component system were recently associated with over-expression of the AdeABC RND pump in clinical isolates of *Acinetobacter baumanii* (Yoon *et al.*, 2013).

The results obtained in this study were therefore surprising, given previous successes at finding mutated regulators in mutants with constitutive upregulation of efflux pumps. Two mechanisms were found causing constitutive expression of *patAB*, (1) duplication

of a genomic region containing *patAB*, and (2) mutation of a transcriptional attenuator upstream of *patA*. However, neither of these allowed regulatory pathways controlling inducible expression to be deduced. A third mechanism involving a synonymous mutation in *hexA* was tentatively identified in one strain, but was not investigated further.

The first mechanism, found in the laboratory mutant M184, involved duplication of a chromosomal region containing *patA* and *patB* and subsequent "capture" of the second copy of *patAB* by a highly active tRNA promoter. This resulted in a very high level of over-expression of *patAB*. This is the first report of a gene duplication causing antibiotic resistance in *S. pneumoniae*, and is also the first report of a gene duplication causing antibiotic resistance by expression of a resistance determinant from an alternative promoter. Although several examples of antibiotic resistance caused by gene duplication and amplification have been reported previously (reviewed by Sandegren and Andersson, 2009), the level of antibiotic resistance observed was directly related to the copy number of the amplified gene, rather than major changes in its transcription. The duplication observed in M184 was not found in any of the other mutants or clinical isolates used in this study. However, a previous study of gene amplification in *Acinetobacter* species found that identical duplications were observed independently multiple times when a selective pressure was applied (Reams and Neidle, 2004). Gene duplications are difficult to observe as they are often unstable, and the nucleotide sequences of duplicated copies are identical (Andersson and Hughes, 2009). Therefore, it is unclear whether the duplication found in M184 has not been observed previously because it is rare or unique, or because gene duplications have not been investigated. It would be interesting to carry out a larger screening study to specifically look for duplications of the *patAB* locus in fluoroquinolone resistant pneumococci to ascertain the prevalence of this phenomenon. Interestingly, a recent study by Lupien *et al* (2013) described a laboratory mutant with high expression of *patAB*, with no identified cause. It would be interesting to examine this mutant for the presence of a *patAB* duplicate.

Although induction of *patAB* expression in response to DNA damage has been hypothesised to be mediated by a classical regulatory interaction, this has not definitively been shown in this study or by others. Presence of the gene duplication in M184 and its derivatives was associated with very high over-expression of *patAB*. In previous studies, induction of *patAB* expression in response to DNA damaging agents was measured using qRT-PCR and microarrays (El Garch *et al.*, 2010; Marrer *et al.*, 2006a); however, neither of these techniques can differentiate between a regulatory interaction affecting *patAB* expression in the whole population, or very high expression of *patAB* from a small subpopulation. Therefore, the possibility that *patAB* induction in these studies was caused by transient gene duplication and subsequent high *patAB* expression in a small proportion of the population cannot be ruled out. Illegitimate joining of DNA ends by DNA gyrase has been suggested as a possible mechanism for formation of gene duplications without homology at the duplication junction, as was observed in M184 (Seoane *et al.*, 2003). This process could be promoted by the sub-inihbitory concentrations of DNA damaging agents used in the induction assays, either by increasing the amount of DNA gyrase bound to DNA or, in the case of fluoroquinolones, reducing the accuracy of double-stranded DNA break religation. However, there are problems with this hypothesis. Firstly, the kinetics of induction observed by El Garch *et al* (2010) suggested that *patAB* expression dropped to wild-type levels within two hours when selection was removed, whereas the half-life of the duplication measured in this study was much longer. Furthermore, if the duplication observed in M184 could form with a high enough frequency for *patAB* over-expression to be observed within the timescale of an induction experiment, it would not explain why R6 cultures are antibiotic susceptible since a subpopulation of cells would be expected to develop *patAB*-mediated fluoroquinolone resistance rapidly following fluoroquinolone exposure. Further work is needed to establish whether duplication of *patA* and *patB* can be induced by fluoroquinolone exposure and, if so, to determine the frequency and kinetics of this.

The second mechanism of constitutive over-expression of *patAB* identified in this project was mutation of a predicted Rho-independent transcriptional terminator directly upstream of *patA*. Mutations affecting this structure have also been shown to cause *patAB* over-expression in linezolid resistant laboratory-selected mutants (Billal *et al.*, 2011) and in a clinical isolate isolated from a meningitis patient (Croucher *et al.*, 2013). This terminator forms part of a putative transcriptional attenuator, consisting of a terminator, anti-terminator and anti-anti-terminator. Transcriptional attenuators control expression of downstream genes by forming mutually exclusive RNA secondary structures that affect transcription elongation (Henkin and Yanofsky, 2002). Folding of attenuators, and therefore gene transcription, can be controlled by multiple factors, including tRNA levels, small metabolites and anti-terminator proteins (Green *et al.*, 2010; Henkin, 2008; Turner *et al.*, 1994). Examples are common throughout bacteria, but particularly in the Firmicutes (Naville and Gautheret, 2010a). It is therefore possible that inducible expression of *patAB* could be mediated via this transcriptional attenuator, in conjunction with a yet to be identified regulatory factor. Inducible attenuation of expression was not investigated in this study, however the placement of the predicted *patA* promoter does not support this hypothesis. An alternative hypothesis is that the predicted attenuator is simply just a terminator for the upstream *hexA* gene that has somehow been co-opted for regulation of *patA*. To determine the role of this attenuator in inducible *patAB* expression further work will be required to properly define the transcription start site of *patA*.

A more detailed study of the phylogeny of *patAB* could provide some insight. In *S. suis*, *satA* and *satB*, which are homologues of *patA* and *patB*, are located directly adjacent to each other, transcribed as an operon and regulated by a local regulator SatR, encoded directly upstream of *satA* (Escudero *et al.*, 2011, 2013). A similar gene organisation is also observed for *lmrCD* in *L. lactis*, which are also *patAB* homologues, although the local regulator belongs to a different gene family (Lubelski *et al.*, 2006). The genes directly surrounding *patAB* homologues also differ between species. In *S. pneumoniae*, *patA* is

preceded by *hexA*, which encodes a component of the Hex mismatch repair system, and *patB* is followed by the glutamine amidotransferase gene *guaA* which, based on qRT-PCR results obtained here, may be co-transcribed with *patB*. In *S. suis*, the *satRAB* operon is flanked on the 5' side by a gene encoding a NAD(P)H-dependent glycerol-3-phosphate dehydrogenase, and on the 3' side by genes encoding a dUTPase enzyme and fructose-2,6-bisphosphatase (Zhang *et al.*, 2011). In *L. lactis* the gene organisation is different again, with the *lmrRCD* operon flanked by genes encoding a cation transporter and UDP-N-acetylglucosamine 1-carboxyvinyltransferase (Makarova *et al.*, 2006). Examination of the genomic contexts of *patAB* homologues in published genome sequences of other *Streptococci* suggests that at least three other gene organisations exist (Figure 6.1). This, combined with the presence of a degenerate transposase gene between *patA* and *patB* in *S. pneumoniae*, implies that several genomic rearrangements have occurred in the evolutionary history of the *Streptococci* that have altered the location of the *patAB* genes. Mapping these events may provide insight into the origin and function of the putative transcriptional attenuator in *S. pneumoniae*, and could distinguish between alternative evolutionary scenarios. For example, the attenuator may be linked to *patA* and co-locate with *patA* homologues in other *Streptococci*, or it could have regulated a different gene that was once located downstream of *hexA* and which has since been replaced by *patA*. It has been previously shown that acquisition, loss and switching of attenuators upstream of several gene families has occurred frequently throughout bacterial evolution (Naville and Gautheret, 2010b).

The two mechanisms found in this study differ in the level of *patAB* expression produced. The gene duplication in M184 and its derivatives resulted in *patAB* expression that was several hundred-fold higher than in R6, while mutation of the upstream terminator increased expression by only two- to ten-fold. This led to phenotypic differences between strains with the duplication and strains with the terminator mutation, although these were not as large as might be expected from the magnitude of the difference in

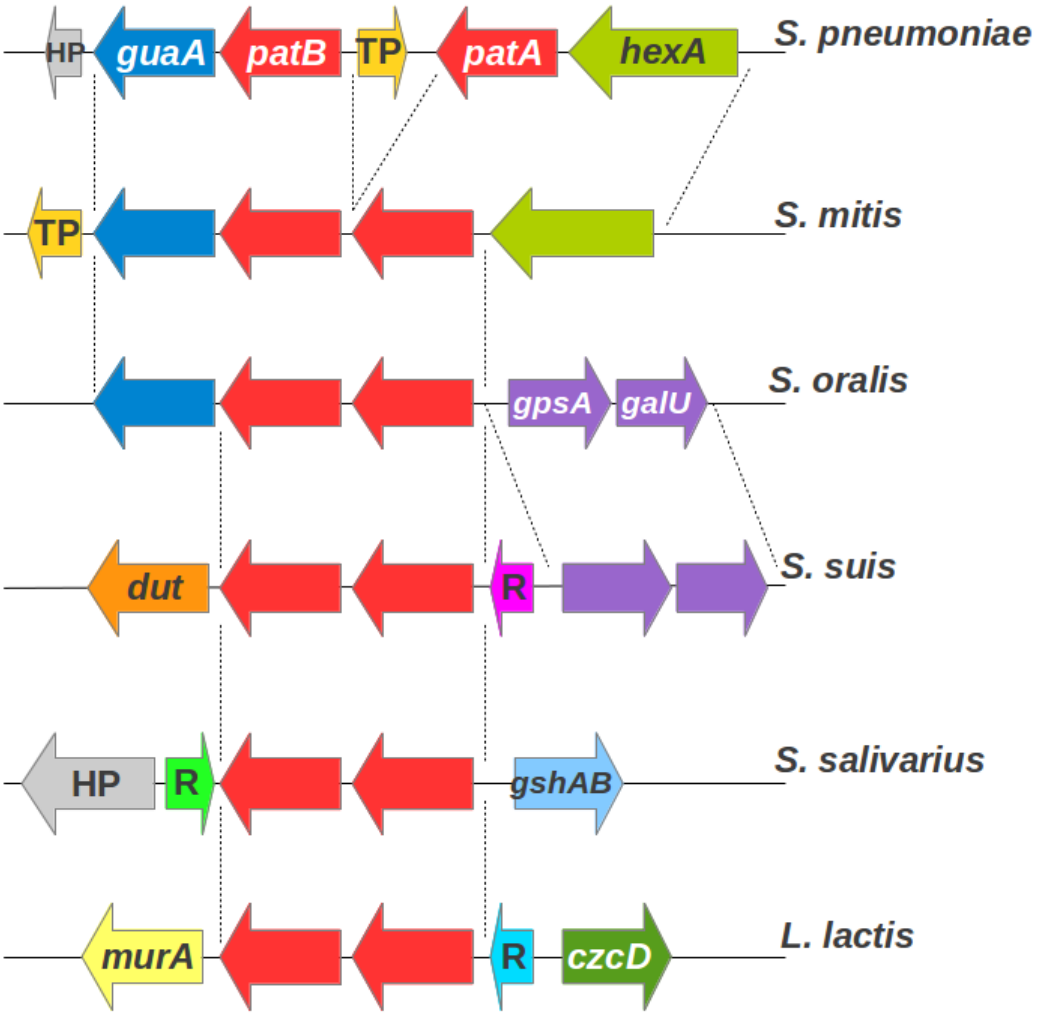


Figure 6.1. Gene organisation surrounding *patAB* homologues in several *Streptococcus* species and *Lactococcus lactis*.. Homologues of *patA* and *patB* are indicated in red throughout. Regions represented with identical colouring between species are homologous, and homology is also indicated by dotted lines. HP, hypothetical protein; TP, degerate transposase; R, regulator; *guaA*, glutamine amidotransferase; *gpsA*, NAD(P)H-dependent glycerol-3-phosphate dehydrogenase; *galU*, glucose-1-phosphate-uridylyltransferase; *dut*, dUTPase; *gshAB*, bifunctional glutathione biosynthesis protein; *murA*, UDP-N-acetylglucosamine 1-carboxyvinyltransferase; *czcD*, cation transporter

patAB transcription. Strains carrying the duplication accumulated approximately half the amount of ethidium bromide accumulated by strains with the terminator mutation. However, there was little to no difference in the increase in ciprofloxacin, norfloxacin and ethidium bromide MICs compared to R6 for strains with either mechanism. In this study, only the transcription of *patAB* was measured, and not the production of functional PatAB transporter proteins. The discrepancy between transcription level and observed phenotype may be explained by incomplete conversion of the over-produced *patAB* transcript to functional protein in strains with the duplication, either due to mRNA degradation or incomplete translation, or by degradation of excess protein. Additionally, MICs are an insensitive measure of the efficiency of efflux of a compound (Lim and Nikaido, 2010), which may explain why the difference in ethidium bromide accumulation did not result in a larger difference in MICs of ethidium bromide conferred by the two mechanisms.

Neither mechanism of *patAB* over-expression had an impact on the growth of R6. The only mutant with a reduced growth rate compared to R6 was M184, but since none of the $R6^{M184}$ transformants had similar alterations in growth, it seems likely that the growth defect was caused by one or more of the 27 point mutations carried by this mutant, instead of the presence of the duplication. This has implications for stability of resistance caused by each mechanism. The duplication was lost from cells over time at a measurable rate and, since the duplication has no effect on growth, this is due to recombination between duplication copies instead of competition between variants with and without the duplication. The stability of the mutations in the transcriptional terminator was not measured, but the persistence of point mutations in a population of cells tends to depend on the degree to which a strain carrying a mutation is outcompeted by a strain lacking the mutation, since true reversion of a specific mutated base is rare. The two mechanisms therefore differ in that the duplication transiently produces very high levels of *patAB* expression, while mutation of the transcriptional terminator results

in lower levels of *patAB* production which are likely to persist for a longer period of time.

In conclusion, two mechanisms conferring constitutive over-expression of *patA* and *patB* have been identified from four genetic backgrounds. One of these, the duplication of a genomic region containing *patAB*, is novel as antibiotic resistance conferred by this mechanism has not been previously observed in *S. pneumoniae*. Efflux-mediated fluoroquinolone resistance has been observed in many laboratory mutants and clinical isolates in previous studies, and the results presented here suggest simple, PCR-based assays that could be carried out to determine the cause of *patAB* expression in these isolates, and therefore estimate the prevalence of each mechanism in the clinical setting. However, the diversity of the physiological roles postulated for PatAB suggests that further layers of regulation remain to be identified.

References

Abeyta, M., Hardy, G., and Yother, J. (2003), Genetic alteration of capsule type but not PspA type affects accessibility of surface-bound complement and surface antigens of *Streptococcus pneumoniae*. *Infection and Immunity*, 71(1):218–225.

Abouzeed, Y. M., Baucheron, S., and Cloeckaert, A. (2008), ramR mutations involved in efflux-mediated multidrug resistance in Salmonella enterica serovar Typhimurium. *Antimicrobial Agents and Chemotherapy*, 52(7):2428–34.

Abreu-Goodger, C. and Merino, E. (2005), RibEx: a web server for locating riboswitches and other conserved bacterial regulatory elements. *Nucleic Acids Research*, 33(Web Server issue):W690–2.

Achard-Joris, M., van den Berg van Saparoea, H. B., Driessen, A. J. M., and Bourdineaud, J.-P. (2005), Heterologously expressed bacterial and human multidrug resistance proteins confer cadmium resistance to *Escherichia coli*. *Biochemistry*, 44(15): 5916–22.

Adam, H. J., Schurek, K. N., Nichol, K. a., Hoban, C. J., Baudry, T. J., Laing, N. M., Hoban, D. J., and Zhanel, G. G. (2007), Molecular characterization of increasing fluoroquinolone resistance in *Streptococcus pneumoniae* isolates in Canada, 1997 to 2005. *Antimicrobial Agents and Chemotherapy*, 51(1):198–207.

Adams, D. E., Shekhtman, E. M., Zechiedrich, E. L., Schmid, M. B., and Cozzarelli, N. R. (1992), The role of topoisomerase IV in partitioning bacterial replicons and the structure of catenated intermediates in DNA replication. *Cell*, 71(2):277–88.

Agustiandari, H., Lubelski, J., van den Berg van Saparoea, H. B., Kuipers, O. P., and Driessen, A. J. M. (2008), LmrR is a transcriptional repressor of expression of the multidrug ABC transporter LmrCD in *Lactococcus lactis*. *Journal of Bacteriology*, 190(2):759–63.

Ahmed, M., Borsch, C. M., Neyfakh, A. A., and Schuldiner, S. (1993), Mutants of the Bacillus subtilis multidrug transporter Bmr with altered sensitivity to the antihypertensive alkaloid reserpine. *The Journal of biological chemistry*, 268(15):11086–9.

Ahmed, M., Lyass, L., Markham, P. N., Taylor, S. S., Vázquez-Laslop, N., and Neyfakh, A. A. (1995), Two highly similar multidrug transporters of *Bacillus subtilis* whose expression is differentially regulated. *Journal of Bacteriology*, 177(14):3904–10.

Alekshun, M. N. and Levy, S. B. (1997), Regulation of chromosomally mediated multiple antibiotic resistance: the mar regulon. *Antimicrobial Agents and Chemotherapy*, 41 (10):2067–75.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997), Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17):3389–402.

Ambler, R. P. (1980), The structure of beta-lactamases. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 289(1036):321–31.

Amster-Choder, O. (2005), The bgl sensory system: a transmembrane signaling pathway controlling transcriptional antitermination. *Current Opinion in Microbiology*, 8(2): 127–34.

Anderson, P. and Roth, J. (1981), Spontaneous tandem genetic duplications in *Salmonella typhimurium* arise by unequal recombination between rRNA (rrn) cistrons. *Proceedings of the National Academy of Sciences of the United States of America*, 78(5):3113–7.

Andersson, D. I., Slechta, E. S., and Roth, J. R. (1998), Evidence that gene amplification underlies adaptive mutability of the bacterial lac operon. *Science (New York, N.Y.)*, 282(5391):1133–5.

Andersson, D. I. and Hughes, D. (2009), Gene amplification and adaptive evolution in bacteria. *Annual Review of Genetics*, 43:167–95.

Aniansson, G., Alm, B., Andersson, B., Larsson, P., Nylén, O., Peterson, H., Rignér, P., Svanborg, M., and Svanborg, C. (1992), Nasopharyngeal colonization during the first year of life. *The Journal of Infectious Diseases*, 165 Suppl:S38–42.

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006), The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics (Oxford, England)*, 22(2):195–201.

Avery, O. T., Macleod, C. M., and McCarty, M. (1944), Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *The Journal of Experimental Medicine*, 79(2):137–58.

Avrain, L., Garvey, M., Mesaros, N., Glupczynski, Y., Mingeot-Leclercq, M.-P., Piddock, L. J. V., Tulkens, P. M., Vanhoof, R., and Van Bambeke, F. (2007), Selection of quinolone resistance in Streptococcus pneumoniae exposed in vitro to subinhibitory drug concentrations. *The Journal of Antimicrobial Chemotherapy*, 60(5):965–72.

Ball, P. (2000), Quinolone generations: natural history or natural selection? *The Journal of Antimicrobial Chemotherapy*, 46 Suppl T:17–24.

Balsalobre, L. and De La Campa, A. (2008), Fitness of *Streptococcus pneumoniae* fluoroquinolone-resistant strains with topoisomerase IV recombinant genes. *Antimicrobial Agents and Chemotherapy*, 52(3):822.

Balsalobre, L., Ferrándiz, M. J., Liñares, J., Tubau, F., and de la Campa, A. G. (2003), Viridans group streptococci are donors in horizontal transfer of topoisomerase IV genes to *Streptococcus pneumoniae*. *Antimicrobial Agents and Chemotherapy*, 47(7):2072–81.

Balsalobre, L., Ferrándiz, M. J., de Alba, G., and de la Campa, A. G. (2011), Nonoptimal DNA topoisomerases allow maintenance of supercoiling levels and improve fitness of *Streptococcus pneumoniae*. *Antimicrobial Agents and Chemotherapy*, 55(3):1097–105.

Baranova, N. N. and Neyfakh, a. a. (1997), Apparent involvement of a multidrug transporter in the fluoroquinolone resistance of *Streptococcus pneumoniae*. *Antimicrobial agents and chemotherapy*, 41(6):1396–8.

Barbosa, T. M. and Levy, S. B. (2000), Differential expression of over 60 chromosomal genes in *Escherichia coli* by constitutive expression of MarA. *Journal of Bacteriology*, 182(12):3467–74.

Barrett, J. F. (2000), Moxifloxacin Bayer. *Current opinion in investigational drugs (London, England : 2000)*, 1(1):45–51.

Barrick, J. E., Corbino, K. a., Winkler, W. C., Nahvi, A., Mandal, M., Collins, J., Lee, M., Roth, A., Sudarsan, N., Jona, I., Wickiser, J. K., and Breaker, R. R. (2004), New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proceedings of the National Academy of Sciences of the United States of America*, 101(17):6421–6.

Bartilson, M., Marra, A., Christine, J., Asundi, J., Schneider, W., and Hromockyj, A. (2001), Differential fluorescence induction reveals *Streptococcus pneumoniae* loci regulated by competence stimulatory peptide. *Molecular Microbiology*, 39(1):126–135.

Baucheron, S., Imberechts, H., Chaslus-Dancla, E., and Cloeckaert, A. (2002), The AcrB multidrug transporter plays a major role in high-level fluoroquinolone resistance in Salmonella enterica serovar typhimurium phage type DT204. *Microbial drug resistance (Larchmont, N.Y.)*, 8(4):281–9.

Bay, D. C., Rommens, K. L., and Turner, R. J. (2008), Small multidrug resistance proteins: a multidrug transporter family that continues to grow. *Biochimica et biophysica acta*, 1778(9):1814–38.

Behlau, I. and Miller, S. I. (1993), A PhoP-repressed gene promotes Salmonella typhimurium invasion of epithelial cells. *Journal of bacteriology*, 175(14):4475–84.

Bentley, S. D., Aanensen, D. M., Mavroidi, A., Saunders, D., Rabbinowitsch, E., Collins, M., Donohoe, K., Harris, D., Murphy, L., Quail, M. A., Samuel, G., Skovsted, I. C., Kaltoft, M. S., Barrell, B., Reeves, P. R., Parkhill, J., and Spratt, B. G. (2006), Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genetics*, 2(3):e31.

Bergé, M., Mortier-Barrière, I., Martin, B., and Claverys, J.-P. (2003), Transformation of Streptococcus pneumoniae relies on DprA- and RecA-dependent protection of incoming DNA single strands. *Molecular Microbiology*, 50(2):527–536.

Berry, A. M. and Paton, J. C. (1996), Sequence heterogeneity of PsaA, a 37-kilodalton putative adhesin essential for virulence of Streptococcus pneumoniae. *Infection and immunity*, 64(12):5255–62.

Billal, D. S., Feng, J., Leprohon, P., Légaré, D., and Ouellette, M. (2011), Whole genome analysis of linezolid resistance in *Streptococcus pneumoniae* reveals resistance and compensatory mutations. *BMC Genomics*, 12(1):512.

Bina, J. E. and Mekalanos, J. J. (2001), Vibrio cholerae tolC is required for bile resistance and colonization. *Infection and immunity*, 69(7):4681–5.

Blondeau, J. M. and Missaghi, B. (2004), Gemifloxacin: a new fluoroquinolone. *Expert opinion on pharmacotherapy*, 5(5):1117–52.

Bogaert, D., Engelen, M. N., Timmers-Reker, A. J. M., Elzenaar, K. P., Peerbooms, P. G. H., Coutinho, R. A., De Groot, R., and Hermans, P. W. M. (2001), Pneumococcal carriage in children in the Netherlands: A molecular epidemiological study. *Journal of Clinical Microbiology*, 39(9):3316–3320.

Bogaert, D., Hermans, P. W. M., Grivea, I. N., Katopodis, G. S., Mitchell, T. J., Sluijter, M., De Groot, R., Beratis, N. G., and Syrogiannopoulos, G. A. (2003), Molecular epidemiology of penicillin-susceptible non-beta-lactam-resistant *Streptococcus pneumoniae* isolates from Greek children. *Journal of Clinical Microbiology*, 41(12):5633–9.

Bogaert, D., Groot, R. D., and Hermans, P. (2004), *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. *The Lancet Infectious Diseases*, 4(March).

Boncoeur, E., Durmort, C., Bernay, B., Ebel, C., Di Guilmi, A. M., Croizé, J., Vernet, T., and Jault, J.-M. (2012), PatA and PatB form a functional heterodimeric ABC multidrug efflux transporter responsible for the resistance of *Streptococcus pneumoniae* to fluoroquinolones. *Biochemistry*, 51(39):7755–65.

Bonner, E. R., D'Elia, J. N., Billips, B. K., and Switzer, R. L. (2001), Molecular recognition of pyr mRNA by the *Bacillus subtilis* attenuation regulatory protein PyrR. *Nucleic acids research*, 29(23):4851–65.

Bouige, P., Laurent, D., Piloyan, L., and Dassa, E. (2002), Phylogenetic and functional classification of ATP-binding cassette (ABC) systems. *Current protein & peptide science*, 3(5):541–59.

Bourdineaud, J.-P., Nehmé, B., Tesse, S., and Lonvaud-Funel, A. (2004), A bacterial gene homologous to ABC transporters protect Oenococcus oeni from ethanol and other stress factors in wine. *International journal of food microbiology*, 92(1):1–14.

Brenwald, N. P., Gill, M. J., and Wise, R. (1997), The effect of reserpine, an inhibitor of multi-drug efflux pumps, on the in-vitro susceptibilities of fluoroquinolone-resistant strains of Streptococcus pneumoniae to norfloxacin. *The Journal of antimicrobial chemotherapy*, 40(3):458–60.

Brenwald, N. P., Gill, M. J., and Wise, R. (1998), Prevalence of a putative efflux mechanism among fluoroquinolone-resistant clinical isolates of *Streptococcus pneumoniae*. *Antimicrobial Agents and Chemotherapy*, 42(8):2032–5.

Brenwald, N. P., Appelbaum, P., Davies, T., and Gill, M. J. (2003), Evidence for efflux pumps, other than PmrA, associated with fluoroquinolone resistance in *Streptococcus pneumoniae*. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 9(2):140–3.

Briles, D. E., Crain, M. J., Gray, B. M., Forman, C., and Yother, J. (1992), Strong Association between Capsular Type and Virulence for Mice. *Infection and Immunity*, 60(1):111–116.

Brochet, M., Couvé, E., Zouine, M., Poyart, C., and Glaser, P. (2008), A naturally occurring gene amplification leading to sulfonamide and trimethoprim resistance in *Streptococcus agalactiae*. *Journal of Bacteriology*, 190(2):672–80.

Brown, M. B. and Roberts, M. C. (1991), Tetracycline resistance determinants in streptococcal species isolated from the bovine mammary gland. *Veterinary microbiology*, 29(2):173–80.

Brown, M. H., Paulsen, I. T., and Skurray, R. A. (1999), The multidrug efflux protein NorM is a prototype of a new family of transporters. *Molecular microbiology*, 31(1): 394–5.

Brueggemann, A. B., Peto, T. E. a., Crook, D. W., Butler, J. C., Kristinsson, K. G., and Spratt, B. G. (2004), Temporal and geographic stability of the serogroup-specific invasive disease potential of Streptococcus pneumoniae in children. *The Journal of infectious diseases*, 190(7):1203–11.

Brundish, D. E. and Baddiley, J. (1968), Pneumococcal C-substance, a ribitol teichoic acid containing choline phosphate. *The Biochemical journal*, 110(3):573–82.

Bryksin, A. V. and Matsumura, I. (2010), Rational design of a plasmid origin that replicates efficiently in both gram-positive and gram-negative bacteria. *PloS One*, 5 (10):e13244.

Buckley, A. M., Webber, M. A., Cooles, S., Randall, L. P., La Ragione, R. M., Woodward, M. J., and Piddock, L. J. V. (2006), The AcrAB-TolC efflux system of Salmonella

enterica serovar Typhimurium plays a role in pathogenesis. *Cellular microbiology*, 8 (5):847–56.

Bush, L. M., Calmon, J., and Johnson, C. C. (1995), Newer penicillins and beta-lactamase inhibitors. *Infectious Disease Clinics of North America*, 9(3):653–86.

Butler, J. C. Epidemiology of Pneumococcal Disease. In Tuomanen, E. I., editor, *The Pneumococcus*, chapter 10, pages 148–168. ASM Press, (2004).

Cámara, M., Boulnois, G. J., Andrew, P. W., and Mitchell, T. J. (1994), A neuraminidase from Streptococcus pneumoniae has the features of a surface protein. *Infection and immunity*, 62(9):3688–95.

Campbell, E. a., Choi, S. Y., and Masure, H. R. (1998), A competence regulon in *Streptococcus pneumoniae* revealed by genomic analysis. *Molecular Microbiology*, 27 (5):929–39.

Canton, R., Morosini, M., Enright, M. C., and Morrissey, I. (2003), Worldwide incidence, molecular epidemiology and mutations implicated in fluoroquinolone-resistant Streptococcus pneumoniae: data from the global PROTEKT surveillance programme. *The Journal of antimicrobial chemotherapy*, 52(6):944–52.

Carattoli, A. (2001), Importance of integrons in the diffusion of resistance. *Veterinary Research*, 32(3-4):243–59.

Carver, T., Harris, S. R., Berriman, M., Parkhill, J., and McQuillan, J. A. (2012), Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics*, 28(4):464–9.

Castro-Nallar, E., Chen, H., Gladman, S., Moore, S. C., Seemann, T., Powell, I. B., Hillier, A., Crandall, K. a., and Chandry, P. S. (2012), Population Genomics and Phylogeography of an Australian Dairy Factory Derived Lytic Bacteriophage. *Genome Biology and Evolution*, 4(3):382–393.

Chang, G. (2003), Multidrug resistance ABC transporters. *FEBS Letters*, 555(1):102–105.

Chen, K., Wallis, J. W., McLellan, M. D., Larson, D. E., Kalicki, J. M., Pohl, C. S., McGrath, S. D., Wendl, M. C., Zhang, Q., Locke, D. P., Shi, X., Fulton, R. S., Ley, T. J., Wilson, R. K., Ding, L., and Mardis, E. R. (2009), BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nature Methods*, 6(9): 677–81.

Chien, Y.-W., Klugman, K. P., and Morens, D. M. (2009), Bacterial pathogens and death during the 1918 influenza pandemic. *New England Journal of Medicine*, 361 (26):2582–2583.

Chintakayala, K., Singh, S. S., Rossiter, A. E., Shahapure, R., Dame, R. T., and Grainger, D. C. (2013), E. coli Fis protein insulates the cbpA gene from uncontrolled transcription. *PLoS genetics*, 9(1):e1003152.

Chiu, C.-H., Tang, P., Chu, C., Hu, S., Bao, Q., Yu, J., Chou, Y.-Y., Wang, H.-S., and Lee, Y.-S. (2005), The genome sequence of Salmonella enterica serovar Choleraesuis, a highly invasive and resistant zoonotic pathogen. *Nucleic acids research*, 33(5): 1690–8.

Ciftçi, E., DoÄ§ru, U., Aysev, D., Ince, E., Güriz, H., and Aysev, U. D. (2001), Investigation of risk factors for penicillin-resistant Streptococcus pneumoniae carriage in Turkish children. *Pediatrics international : official journal of the Japan Pediatric Society*, 43(4):385–90.

Clancy, J., Petitpas, J., Dib-Hajj, F., Yuan, W., Cronan, M., Kamath, A. V., Bergeron, J., and Retsema, J. A. (1996), Molecular cloning and functional analysis of a novel macrolide-resistance determinant, mefA, from Streptococcus pyogenes. *Molecular microbiology*, 22(5):867–79.

Claverys, J. P., Méjean, V., Gasc, a. M., and Sicard, a. M. (1983), Mismatch repair in *Streptococcus pneumoniae*: relationship between base mismatches and transformation efficiencies. *Proceedings of the National Academy of Sciences of the United States of America*, 80(19):5956–60.

Claverys, J.-P., Prudhomme, M., and Martin, B. (2006), Induction of competence regulons as a general response to stress in gram-positive bacteria. *Annual Review of Microbiology*, 60:451–75.

Claverys, J.-P., Martin, B., and Polard, P. (2009), The genetic transformation machinery: composition, localization, and mechanism. *FEMS microbiology reviews*, 33(3):643–56.

Colomer-Lluch, M., Imamovic, L., Jofre, J., and Muniesa, M. (2011), Bacteriophages carrying antibiotic resistance genes in fecal waste from cattle, pigs, and poultry. *Antimicrobial Agents and Chemotherapy*, 55(10):4908–11.

Coppins, R. L., Hall, K. B., and Groisman, E. A. (2007), The intricate world of riboswitches. *Current opinion in microbiology*, 10(2):176–81.

Corso, A., Severina, E. P., Petruk, V. F., Mauriz, Y. R., and Tomasz, A. (1998), Molecular characterization of penicillin-resistant Streptococcus pneumoniae isolates causing respiratory disease in the United States. *Microbial drug resistance (Larchmont, N.Y.)*, 4(4):325–37.

Crook, D. W., Brueggemann, A. B., Sleeman, K. L., and Peto, T. E. A. Pneumococcal Carriage. In Tuomanen, E. I., editor, *The Pneumococcus*, chapter 9, pages 136–147. ASM Press, (2004).

Croucher, N. J., Harris, S. R., Fraser, C., Quail, M. a., Burton, J., van der Linden, M., McGee, L., von Gottberg, A., Song, J. H., Ko, K. S., Pichon, B., Baker, S., Parry, C. M., Lambertsen, L. M., Shahinas, D., Pillai, D. R., Mitchell, T. J., Dougan, G., Tomasz, A., Klugman, K. P., Parkhill, J., Hanage, W. P., and Bentley, S. D. (2011)a, Rapid pneumococcal evolution in response to clinical interventions. *Science*, 331(6016):430–4.

Croucher, N. J., Vernikos, G. S., Parkhill, J., and Bentley, S. D. (2011)b, Identification, variation and transcription of pneumococcal repeat sequences. *BMC Genomics*, 12 (1):120.

Croucher, N. J., Harris, S. R., Barquist, L., Parkhill, J., and Bentley, S. D. (2012), A high-resolution view of genome-wide pneumococcal transformation. *PLoS Pathogens*, 8(6):e1002745.

Croucher, N. J., Mitchell, A. M., Gould, K. a., Inverarity, D., Barquist, L., Feltwell, T., Fookes, M. C., Harris, S. R., Dordel, J., Salter, S. J., Browall, S., Zemlickova, H., Parkhill, J., Normark, S., Henriques-Normark, B., Hinds, J., Mitchell, T. J., and Bentley, S. D. (2013), Dominant role of nucleotide substitution in the diversification of serotype 3 pneumococci over decades and during a single infection. *PLoS Genetics*, 9(10):e1003868.

Cundell, D. R. and Tuomanen, E. I. (1994), Receptor specificity of adherence of Streptococcus pneumoniae to human type-II pneumocytes and vascular endothelial cells in vitro. *Microbial pathogenesis*, 17(6):361–74.

Cuthbertson, L., Kos, V., and Whitfield, C. (2010), ABC transporters involved in export of cell surface glycoconjugates. *Microbiology and Molecular Biology Reviews*, 74(3): 341–62.

Dagan, R., Leibovitz, E., Greenberg, D., Yagupsky, P., Fliss, D. M., and Leiberman, A. (1998), Dynamics of pneumococcal nasopharyngeal colonization during the first days of antibiotic treatment in pediatric patients. *The Pediatric infectious disease journal*, 17(10):880–5.

Dambach, M. D. and Winkler, W. C. (2009), Expanding roles for metabolite-sensing regulatory RNAs. *Current opinion in microbiology*, 12(2):161–9.

Dantas, G., Sommer, M. O. A., Oluwasegun, R. D., and Church, G. M. (2008), Bacteria subsisting on antibiotics. *Science*, 320(5872):100–3.

Dassa, E. and Bouige, P. (2001), The ABC of ABCs: a phylogenetic and functional classification of ABC systems in living organisms. *Research in Microbiology*, 152(3-4): 211–29.

Davies, T. a., Pankuch, G. a., Dewasse, B. E., Jacobs, M. R., and Appelbaum, P. C. (1999), In vitro development of resistance to five quinolones and amoxicillin-clavulanate in *Streptococcus pneumoniae*. *Antimicrobial Agents and Chemotherapy*, 43(5):1177–82.

Daw, N. C., Wilimas, J. A., Wang, W. C., Presbury, G. J., Joyner, R. E., Harris, S. C., Davis, Y., Chen, G., and Chesney, P. J. (1997), Nasopharyngeal carriage of penicillin-resistant Streptococcus pneumoniae in children with sickle cell disease. *Pediatrics*, 99 (4):E7.

Dawson, R. J. P. and Locher, K. P. (2006), Structure of a bacterial multidrug ABC transporter. *Nature*, 443(7108):180–5.

Dawson, R. J. P., Hollenstein, K., and Locher, K. P. (2007), Uptake or extrusion: crystal structures of full ABC transporters suggest a common mechanism. *Molecular microbiology*, 65(2):250–7.

D'Costa, V. M., McGrann, K. M., Hughes, D. W., and Wright, G. D. (2006), Sampling the antibiotic resistome. *Science*, 311(5759):374–7.

de Galan, B. E., van Tilburg, P. M., Sluijter, M., Mol, S. J., de Groot, R., Hermans, P. W., and Jansz, A. R. (1999), Hospital-related outbreak of infection with multidrug-resistant Streptococcus pneumoniae in the Netherlands. *The Journal of hospital infection*, 42(3):185–92.

de Jong, A., Pietersma, H., Cordes, M., Kuipers, O. P., and Kok, J. (2012), PePPER: a webserver for prediction of prokaryote promoter elements and regulons. *BMC genomics*, 13(1):299.

de la Campa, A. G., Balsalobre, L., Ardanuy, C., Fenoll, A., Pérez-Trallero, E., and Liñares, J. (2004), Fluoroquinolone resistance in penicillin-resistant Streptococcus pneumoniae clones, Spain. *Emerging infectious diseases*, 10(10):1751–9.

Dockrell, D. H., Whyte, M. K. B., and Mitchell, T. J. (2012), Pneumococcal pneumonia: mechanisms of infection and resolution. *Chest*, 142(2):482–91.

Dowell, S. F., Butler, J. C., Giebink, G. S., Jacobs, M. R., Jernigan, D., Musher, D. M., Rakowsky, A., and Schwartz, B. (1999), Acute otitis media: management and surveillance in an era of pneumococcal resistance–a report from the Drug-resistant Streptococcus pneumoniae Therapeutic Working Group. *The Pediatric infectious disease journal*, 18(1):1–9.

Dowson, C. G., Coffey, T. J., Kell, C., and Whiley, R. A. (1993), Evolution of penicillin resistance in Streptococcus pneumoniae; the role of Streptococcus mitis in the formation of a low affinity PBP2B in S. pneumoniae. *Molecular microbiology*, 9(3): 635–43.

Drlica, K. and Zhao, X. (1997), DNA gyrase, topoisomerase IV, and the 4-quinolones. *Microbiology and molecular biology reviews : MMBR*, 61(3):377–92.

Drlica, K., Malik, M., Kerns, R. J., and Zhao, X. (2008), Quinolone-mediated bacterial death. *Antimicrobial agents and chemotherapy*, 52(2):385–92.

Dunais, B., Bruno-Bazureault, P., Carsenti-Dellamonica, H., Touboul, P., and Pradier, C. (2011), A decade-long surveillance of nasopharyngeal colonization with *Streptococcus pneumoniae* among children attending day-Care centres in south-Eastern France: 1999-2008. *European Journal of Clinical Microbiology and Infectious Diseases*, 30(7): 837–843.

Dunais, B., Pradier, C., Carsenti, H., Sabah, M., Mancini, G., Fontas, E., and Dellamonica, P. (2003), Influence of child care on nasopharyngeal carriage of Streptococcus pneumoniae and Haemophilus influenzae. *The Pediatric infectious disease journal*, 22 (7):589–92.

Edlund, T., Grundström, T., and Normark, S. (1979), Isolation and characterization of DNA repetitions carrying the chromosomal beta-lactamase gene of Escherichia coli K-12. *Molecular & general genetics : MGG*, 173(2):115–25.

Ekdahl, K., Ahlinder, I., Hansson, H. B., Melander, E., Mölstad, S., Söderström, M., and Persson, K. (1997), Duration of nasopharyngeal carriage of penicillin-resistant Streptococcus pneumoniae: experiences from the South Swedish Pneumococcal Intervention Project. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 25(5):1113–7.

El Garch, F., Lismond, A., Piddock, L. J. V., Courvalin, P., Tulkens, P. M., and Van Bambeke, F. (2010), Fluoroquinolones induce the expression of patA and patB, which encode ABC efflux pumps in *Streptococcus pneumoniae*. *The Journal of Antimicrobial Chemotherapy*, 65(10):2076–82.

Epshtein, V., Mironov, A. S., and Nudler, E. (2003), The riboswitch-mediated control of sulfur metabolism in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 100(9):5052–6.

Escudero, J. A., San Millan, A., Gutierrez, B., Hidalgo, L., La Ragione, R. M., AbuOun, M., Galimand, M., Ferrándiz, M. J., Domínguez, L., de la Campa, A. G., and Gonzalez-Zorn, B. (2011), Fluoroquinolone efflux in *Streptococcus suis* is mediated by SatAB and not by SmrA. *Antimicrobial Agents and Chemotherapy*, 55(12):5850–60.

Escudero, J. A., San Millan, A., Montero, N., Gutierrez, B., Ovejero, C. M., Carrilero, L., and González-Zorn, B. (2013), SatR is a repressor of fluoroquinolone efflux pump SatAB. *Antimicrobial agents and chemotherapy*, 57(7):3430–3.

Facklam, R. (2002), What happened to the streptococci: overview of taxonomic and nomenclature changes. *Clinical microbiology reviews*, 15(4):613–630.

Feil, E. J., Smith, J. M., Enright, M. C., and Spratt, B. G. (2000), Estimating recombinational parameters in Streptococcus pneumoniae from multilocus sequence typing data. *Genetics*, 154(4):1439–1450.

Felmingham, D. and Grüneberg, R. N. (2000), The Alexander Project 1996-1997: latest susceptibility data from this international study of bacterial pathogens from community-acquired lower respiratory tract infections. *The Journal of antimicrobial chemotherapy*, 45(2):191–203.

Felmingham, D., Reinert, R. R., Hirakata, Y., and Rodloff, A. (2002), Increasing prevalence of antimicrobial resistance among isolates of Streptococcus pneumoniae from the PROTEKT surveillance study, and compatative in vitro activity of the ketolide, telithromycin. *The Journal of antimicrobial chemotherapy*, 50 Suppl S:25–37.

Feng, J., Lupien, A., Gingras, H., Wasserscheid, J., Dewar, K., Légaré, D., and Ouellette, M. (2009), Genome sequencing of linezolid-resistant Streptococcus pneumoniae mutants reveals novel mechanisms of resistance. *Genome research*, 19(7):1214–23.

Ferrándiz, M.-J., Martín-Galiano, A. J., Schvartzman, J. B., and de la Campa, A. G. (2010), The genome of *Streptococcus pneumoniae* is organized in topology-reacting gene clusters. *Nucleic Acids Research*, 38(11):3570–81.

Fischer, W., Behr, T., Hartmann, R., Peter-Katalinić, J., and Egge, H. (1993), Teichoic acid and lipoteichoic acid of Streptococcus pneumoniae possess identical chain structures. A reinvestigation of teichoid acid (C polysaccharide). *European journal of biochemistry / FEBS*, 215(3):851–7.

Floyd, J. L., Smith, K. P., Kumar, S. H., Floyd, J. T., and Varela, M. F. (2010), LmrS is a multidrug efflux pump of the major facilitator superfamily from Staphylococcus aureus. *Antimicrobial Agents and Chemotherapy*, 54(12):5406–12.

Fox, M. S. and Allen, M. K. (1964), On the Mechanism of Deoxyribonucleate Integration in Pneumococcal Transformation. *Proceedings of the National Academy of Sciences of the United States of*, 52:412–419.

Fraud, S. and Poole, K. (2011), Oxidative stress induction of the MexXY multidrug efflux genes and promotion of aminoglycoside resistance development in *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy*, 55(3):1068–74.

Fraud, S., Campigotto, A. J., Chen, Z., and Poole, K. (2008), MexCD-OprJ multidrug efflux system of Pseudomonas aeruginosa: involvement in chlorhexidine resistance and induction by membrane-damaging agents dependent upon the AlgU stress response sigma factor. *Antimicrobial agents and chemotherapy*, 52(12):4478–82.

Fuda, C., Suvorov, M., Vakulenko, S. B., and Mobashery, S. (2004), The basis for resistance to beta-lactam antibiotics by penicillin-binding protein 2a of methicillin-resistant *Staphylococcus aureus*. *The Journal of Biological Chemistry*, 279(39):40802–6.

Fuller, J. D. and Low, D. E. (2005), BRIEF REPORT A Review of *Streptococcus pneumoniae* Infection Treatment Failures Associated with Fluoroquinolone Resistance. 41(table 1).

Galitski, T. and Roth, J. R. (1997), Pathways for homologous recombination between chromosomal direct repeats in *Salmonella typhimurium*. *Genetics*, 146(3):751–67.

García-Rey, C., Aquilar, L., and Baquero, F. (2000), Influences of different factors on prevalence of ciprofloxacin resistance in Streptococcus pneumoniae in Spain. *Antimicrobial agents and chemotherapy*, 44(12):3481–2.

Garvey, M. I. and Piddock, L. J. V. (2008), The efflux pump inhibitor reserpine selects multidrug-resistant *Streptococcus pneumoniae* strains that overexpress the ABC

transporters PatA and PatB. *Antimicrobial Agents and Chemotherapy*, 52(5):1677–85.

Garvey, M. I., Baylay, A. J., Wong, R. L., and Piddock, L. J. V. (2011), Overexpression of patA and patB, which encode ABC transporters, is associated with fluoroquinolone resistance in clinical isolates of *Streptococcus pneumoniae*. *Antimicrobial Agents and Chemotherapy*, 55(1):190–6.

Gay, K. and Stephens, D. S. (2001), Structure and dissemination of a chromosomal insertion element encoding macrolide efflux in *Streptococcus pneumoniae*. *The Journal of Infectious Diseases*, 184(1):56–65.

Gay, K., Baughman, W., Miller, Y., Jackson, D., Whitney, C. G., Schuchat, A., Farley, M. M., Tenover, F., and Stephens, D. S. (2000), The emergence of Streptococcus pneumoniae resistant to macrolide antimicrobial agents: a 6-year population-based assessment. *The Journal of infectious diseases*, 182(5):1417–24.

Geddes, A. M. (2009), Influenza and bacterial pneumonia. *International journal of antimicrobial agents*, 34(4):293–4.

Gellert, M., Mizuuchi, K., O'Dea, M. H., and Nash, H. A. (1976), DNA gyrase: an enzyme that introduces superhelical turns into DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 73(11):3872–6.

Gill, M. J., Brenwald, N. P., and Wise, R. (1999), Identification of an efflux pump gene, *pmrA*, associated with fluoroquinolone resistance in *Streptococcus pneumoniae*. *Antimicrobial agents and chemotherapy*, 43(1):187–9.

Gillespie, S. H. (2002), Evolution of drug resistance in Mycobacterium tuberculosis: clinical and molecular perspective. *Antimicrobial agents and chemotherapy*, 46(2): 267–74.

Gillespie, S. H., Voelker, L. L., Ambler, J. E., Traini, C., and Dickens, A. (2003), Fluoroquinolone resistance in Streptococcus pneumoniae: evidence that gyrA mutations arise at a lower rate and that mutation in gyrA or parC predisposes to further mutation. *Microbial drug resistance (Larchmont, N.Y.)*, 9(1):17–24.

Gladstone, R. A., Jefferies, J. M., Faust, S. N., and Clarke, S. C. (2011), Continued control of pneumococcal disease in the UK - the impact of vaccination. *Journal of medical microbiology*, 60(Pt 1):1–8.

Green, N. J., Grundy, F. J., and Henkin, T. M. (2010), The T box mechanism: tRNA as a regulatory molecule. *FEBS Letters*, 584(2):318–24.

Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S. R., and Bateman, A. (2005), Rfam: annotating non-coding RNAs in complete genomes. *Nucleic acids research*, 33(Database issue):D121–4.

Grinius, L., Dreguniene, G., Goldberg, E. B., Liao, C. H., and Projan, S. J. (1992), A staphylococcal multidrug resistance gene product is a member of a new protein family. *Plasmid*, 27(2):119–29.

Grundy, F. J. and Henkin, T. M. (1993), tRNA as a positive regulator of transcription antitermination in B. subtilis. *Cell*, 74(3):475–82.

Grundy, F. J., Winkler, W. C., and Henkin, T. M. (2002), tRNA-mediated transcription antitermination in vitro: codon-anticodon pairing independent of the ribosome. *Proceedings of the National Academy of Sciences of the United States of America*, 99 (17):11121–6.

Guiral, S., Mitchell, T. J., Martin, B., and Claverys, J.-P. (2005), Competence-programmed predation of noncompetent cells in the human pathogen Streptococcus pneumoniae: genetic requirements. *Proceedings of the National Academy of Sciences of the United States of America*, 102(24):8710–5.

Gurney Jr., T. and Fox, M. S. (1968), Physical and genetic hybrids formed in bacterial transformation. *Journal of Molecular Biology*, 32(1):83–100.

Gusarov, I. and Nudler, E. (1999), The mechanism of intrinsic transcription termination. *Molecular cell*, 3(4):495–504.

Haack, K. R. and Roth, J. R. (1995), Recombination between chromosomal IS200 elements supports frequent duplication formation in *Salmonella typhimurium*. *Genetics*, 141(4):1245–52.

Håvarstein, L. S., Coomaraswamy, G., and Morrison, D. a. (1995), An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae*. *Proceedings of the National Academy of Sciences of the United States of America*, 92(24):11140–4.

Håvarstein, L. S., Gaustad, P., Nes, I. F., and Morrison, D. A. (1996), Identification of the streptococcal competence-pheromone receptor. *Molecular microbiology*, 21(4): 863–9.

Halfmann, A., Hakenbeck, R., and Brückner, R. (2007), A new integrative reporter plasmid for *Streptococcus pneumoniae*. *FEMS Microbiology Letters*, 268(2):217–24.

Hammerschmidt, S., Talay, S. R., Brandtzaeg, P., and Chhatwal, G. S. (1997), SpsA, a novel pneumococcal surface protein with specific binding to secretory immunoglobulin A and secretory component. *Molecular microbiology*, 25(6):1113–24.

Harvey, H. A., Swords, W. E., and Apicella, M. A. (2001), The mimicry of human glycolipids and glycosphingolipids by the lipooligosaccharides of pathogenic neisseria and haemophilus. *Journal of autoimmunity*, 16(3):257–62.

Hausdorff, W. P., Bryant, J., Paradiso, P. R., and Siber, G. R. (2000), Which pneumococcal serogroups cause the most invasive disease: implications for conjugate vaccine formulation and use, part I. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 30(1):100–21.

Hava, D. L. and Camilli, A. (2002), Large-scale identification of serotype 4 Streptococcus pneumoniae virulence factors. *Molecular microbiology*, 45(5):1389–406.

Hendley, J. O., Sande, M. A., Stewart, P. M., and Gwaltney, J. M. (1975), Spread of Streptococcus pneumoniae in families. I. Carriage rates and distribution of types. *The Journal of infectious diseases*, 132(1):55–61.

Hendriksen, W. T., Bootsma, H. J., Estevão, S., Hoogenboezem, T., de Jong, A., de Groot, R., Kuipers, O. P., and Hermans, P. W. M. (2008), CodY of *Streptococcus pneumoniae*: link between nutritional gene regulation and colonization. *Journal of Bacteriology*, 190(2):590–601.

Henkin, T. M. (1994), tRNA-directed transcription antitermination. *Molecular microbiology*, 13(3):381–7.

Henkin, T. M. (2008), Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes & Development*, 22(24):3383–90.

Henkin, T. M. and Yanofsky, C. (2002), Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 24(8):700–7.

Higgins, C. (2001), ABC transporters: physiology, structure and mechanismâĂŞan overview. *Research in Microbiology*, 152(3-4):205–10.

Higgins, C., Hiles, I., and Salmond, G. (1986), A family of related ATP-binding subunits coupled to many distinct biological processes in bacteria. *Nature*, 323(6087):448–50.

Hiller, N. L., Janto, B., Hogg, J. S., Boissy, R., Yu, S., Powell, E., Keefe, R., Ehrlich, N. E., Shen, K., Hayes, J., Barbadora, K., Klimke, W., Dernovoy, D., Tatusova, T., Parkhill, J., Bentley, S. D., Post, J. C., Ehrlich, G. D., and Hu, F. Z. (2007), Comparative genomic analyses of seventeen *Streptococcus pneumoniae* strains: insights into the pneumococcal supragenome. *Journal of Bacteriology*, 189(22):8186–95.

Hiller, N. L., Eutsey, R. A., Powell, E., Earl, J. P., Janto, B., Martin, D. P., Dawid, S., Ahmed, A., Longwell, M. J., Dahlgren, M. E., Ezzo, S., Tettelin, H., Daugherty, S. C., Mitchell, T. J., Hillman, T. A., Buchinsky, F. J., Tomasz, A., de Lencastre, H., Sá-Leão, R., Post, J. C., Hu, F. Z., and Ehrlich, G. D. (2011), Differences in genotype and virulence among four multidrug-resistant Streptococcus pneumoniae isolates belonging to the PMEN1 clone. *PloS one*, 6(12):e28850.

Hirakata, Y., Srikumar, R., Poole, K., Gotoh, N., Suematsu, T., Kohno, S., Kamihira, S., Hancock, R. E. W., and Speert, D. P. (2002), Multidrug efflux systems play an important role in the invasiveness of Pseudomonas aeruginosa. *The Journal of Experimental Medicine*, 196(1):109–18.

Ho, P. L., Yung, R. W., Tsang, D. N., Que, T. L., Ho, M., Seto, W. H., Ng, T. K., Yam, W. C., and Ng, W. W. (2001), Increasing resistance of Streptococcus pneumoniae to fluoroquinolones: results of a Hong Kong multicentre study in 2000. *The Journal of antimicrobial chemotherapy*, 48(5):659–65.

Högberg, L., Geli, P., Ringberg, H. k., Melander, E., Lipsitch, M., and Ekdahl, K. (2007), Age- and serogroup-related differences in observed durations of nasopharyngeal carriage of penicillin-resistant pneumococci. *Journal of clinical microbiology*, 45 (3):948–52.

Hoge, C. W., Reichler, M. R., Dominguez, E. A., Bremer, J. C., Mastro, T. D., Hendricks, K. A., Musher, D. M., Elliott, J. A., Facklam, R. R., and Breiman, R. F. (1994), An epidemic of pneumococcal disease in an overcrowded, inadequately ventilated jail. *New England Journal of Medicine*, 331(10):643–648.

Hopfner, K. P., Karcher, A., Shin, D. S., Craig, L., Arthur, L. M., Carney, J. P., and Tainer, J. A. (2000), Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell*, 101(7):789–800.

Horne, D. S. and Tomasz, a. (1993), Possible role of a choline-containing teichoic acid in the maintenance of normal cell shape and physiology in Streptococcus oralis. *Journal of bacteriology*, 175(6):1717–22.

Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K., and Pease, L. R. (1989), Engineering hybrid genes without the use of restriction enzymes : gene splicing by overlap extension sequences ; frequency of errors ; exon ; intron ; mosaic fusion protein ; mouse histocompatibility genes). *Science*, 77:61–68.

Hoskins, J., Alborn, W. E., Arnold, J., Blaszczak, L. C., Burgett, S., Hoff, B. S. D. E., Estrem, S. T., Fritz, L., Fu, D.-j., Fuller, W., Geringer, C., Gilmour, R., Glass, J. S., Khoja, H., Kraft, A. R., Lagace, R. E., Blanc, D. J. L. E., Lee, L. N., Lefkowitz, E. J., Lu, J. I. N., Matsushima, P., Ahren, S. M. M. C., Henney, M. M. C., Leaster, K. M. C., Mundy, C. W., Nicas, T. I., Norris, F. H., Gara, M. O., Peery, R. B., Robertson, G. T., Rockey, P., Sun, P.-m., Winkler, M. E., Yang, Y., Young-bellido, M., Zhao, G., Zook, C. A., Baltz, R. H., Jaskunas, S. R., Rosteck, P. R., Skatrud, P. L., and Glass, J. I. (2001), Genome of the Bacterium *Streptococcus pneumoniae* Strain R6. *Society*, 183 (19):5709–5717.

Hossain, A., Ferraro, M. J., Pino, R. M., Dew, R. B., Moland, E. S., Lockhart, T. J., Thomson, K. S., Goering, R. V., and Hanson, N. D. (2004), Plasmid-mediated carbapenem-hydrolyzing enzyme KPC-2 in an *Enterobacter* sp. *Antimicrobial Agents and Chemotherapy*, 48(11):4438–40.

Hotchkiss, R. D. (1954), Cyclical Behavior in Pneumococcal Growth and Transformability Occasioned by Environmental Changes. *Proceedings of the National Academy of Sciences of the United States of America*, 40(2):49–55.

Huda, M. N., Chen, J., Morita, Y., Kuroda, T., Mizushima, T., and Tsuchiya, T. (2003), Gene cloning and characterization of VcrM, a Na+-coupled multidrug efflux pump, from *Vibrio cholerae* non-O1. *Microbiology and Immunology*, 47(6):419–27.

Hui, F. M. and Morrison, D. A. (1991), Genetic transformation in Streptococcus pneumoniae: nucleotide sequence analysis shows comA, a gene required for competence induction, to be a member of the bacterial ATP-dependent transport protein family. *Journal of bacteriology*, 173(1):372–81.

Humbert, O., Prudhomme, M., Hakenbeck, R., Dowson, C. G., and Claverys, J. P. (1995), Homeologous recombination and mismatch repair during transformation in *Streptococcus pneumoniae*: saturation of the Hex mismatch repair system. *Proceedings of the National Academy of Sciences of the United States of America*, 92(20): 9052–6.

Huovinen, P. (2001), Resistance to trimethoprim-sulfamethoxazole. *Clinical Infectious Diseases*, 32(11):1608–14.

Ikeda, H., Shiraishi, K., and Ogata, Y. (2004), Illegitimate recombination mediated by double-strand break and end-joining in *Escherichia coli*. *Advances in biophysics*, 38 (Complete):3–20.

Janoir, C., Zeller, V., Kitzis, M. D., Moreau, N. J., and Gutmann, L. (1996), High-level fluoroquinolone resistance in *Streptococcus pneumoniae* requires mutations in *parC* and *gyrA*. *Antimicrobial agents and chemotherapy*, 40(12):2760–4.

Jefferies, J. and Smith, A. (2004), Genetic analysis of diverse disease-causing pneumococci indicates high levels of diversity within serotypes and capsule switching. *Journal of Clinical Microbiology*, 42(12):5681–5688.

Jefferies, J. M. C., Macdonald, E., Faust, S. N., and Clarke, S. C. (2011), 13-valent pneumococcal conjugate vaccine (PCV13). *Human vaccines*, 7(10):1012–8.

Jessop, A. P. and Clugston, C. (1985), Amplification of the ArgF region in strain HfrP4X of E. coli K-12. *Molecular & general genetics : MGG*, 201(2):347–50.

Johnsborg, O. and Håvarstein, L. S. (2009), Regulation of natural genetic transformation and acquisition of transforming DNA in *Streptococcus pneumoniae*. *FEMS Microbiology Reviews*, 33(3):627–642.

Johnsborg, O., Eldholm, V., and Håvarstein, L. S. (2007), Natural genetic transformation: prevalence, mechanisms and function. *Research in Microbiology*, 158(10): 767–78.

Johnson, C. N., Briles, D. E., Benjamin, W. H., Hollingshead, S. K., and Waites, K. B. (2005), Relative fitness of fluoroquinolone-resistant Streptococcus pneumoniae. *Emerging infectious diseases*, 11(6):814–20.

Jones, R. N. and Pfaller, M. a. (2000), Macrolide and fluoroquinolone (levofloxacin) resistances among *Streptococcus pneumoniae* strains: significant trends from the SENTRY Antimicrobial Surveillance Program (North America, 1997-1999). *Journal of clinical microbiology*, 38(11):4298–9.

Jones, R. N., Jacobs, M. R., and Sader, H. S. (2010), Evolving trends in *Streptococcus pneumoniae* resistance: implications for therapy of community-acquired bacterial pneumonia. *International Journal of Antimicrobial Agents*, 36(3):197–204.

Jorgensen, J. H., Weigel, L. M., Ferraro, M. J., Swenson, J. M., and Tenover, F. C. (1999), Activities of newer fluoroquinolones against *Streptococcus pneumoniae* clinical isolates including those with mutations in the *gyrA*, *parC*, and *parE* loci. *Antimicrobial agents and chemotherapy*, 43(2):329–34.

Jumbe, N. L., Louie, A., Miller, M. H., Liu, W., Deziel, M. R., Tam, V. H., Bachhawat, R., and Drusano, G. L. (2006), Quinolone Efflux Pumps Play a Central Role in Emergence of Fluoroquinolone Resistance in *Streptococcus pneumoniae*. 50(1):310–317.

Kaatz, G. W., Seo, S. M., O'Brien, L., Wahiduzzaman, M., and Foster, T. J. (2000), Evidence for the existence of a multidrug efflux transporter distinct from NorA in Staphylococcus aureus. *Antimicrobial agents and chemotherapy*, 44(5):1404–6.

Kaatz, G. W., Moudgal, V. V., Seo, S. M., Kristiansen, E., Kaatz, G. W., Moudgal, V. V., Seo, S. M., and Kristiansen, J. E. (2003), Phenothiazines and Thioxanthenes Inhibit Multidrug Efflux Pump Activity in Staphylococcus aureus Phenothiazines and Thioxanthenes Inhibit Multidrug Efflux Pump Activity in Staphylococcus aureus.

Kato, J., Nishimura, Y., Imamura, R., Niki, H., Hiraga, S., and Suzuki, H. (1990), New topoisomerase essential for chromosome segregation in E. coli. *Cell*, 63(2):393–404.

Kawamura, Y., Hou, X. G., Sultana, F., Miura, H., and Ezaki, T. (1995), Determination of 16S rRNA sequences of *Streptococcus mitis* and *Streptococcus gordonii* and phylogenetic relationships among members of the genus *Streptococcus*. *International Journal of Systematic Bacteriology*, 45(2):406–8.

Kaye, K. S., Cosgrove, S., Harris, A., Eliopoulos, G. M., and Carmeli, Y. (2001), Risk factors for emergence of resistance to broad-spectrum cephalosporins among *Enterobacter* spp. *Antimicrobial Agents and Chemotherapy*, 45(9):2628–30.

Kim, P. E., Musher, D. M., Glezen, W. P., Rodriguez-Barradas, M. C., Nahm, W. K., and Wright, C. E. (1996), Association of invasive pneumococcal disease with season, atmospheric conditions, air pollution, and the isolation of respiratory viruses. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 22(1):100–6.

Kingsford, C. L., Ayanbule, K., and Salzberg, S. L. (2007), Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome biology*, 8(2):R22.

Kislak, J. W., Razavi, L. M., Daly, A. K., and Finland, M. (1965), Susceptibility of pneumococci to nine antibiotics. *The American journal of the medical sciences*, 250 (3):261–8.

Kloosterman, T. G., van der Kooi-Pol, M. M., Bijlsma, J. J. E., and Kuipers, O. P. (2007), The novel transcriptional regulator SczA mediates protection against Zn2+ stress by activation of the Zn2+-resistance gene *czcD* in *Streptococcus pneumoniae*. *Molecular Microbiology*, 65(4):1049–63.

Kloosterman, T. G., Witwicki, R. M., van der Kooi-Pol, M. M., Bijlsma, J. J. E., and Kuipers, O. P. (2008), Opposite effects of Mn2+ and Zn2+ on PsaR-mediated expression of the virulence genes *pcpA*, *prtA*, and *psaBCA* of *Streptococcus pneumoniae*. *Journal of Bacteriology*, 190(15):5382–93.

Knutsen, E., Johnsborg, O., and Quentin, Y. (2006), BOX-elements modulate gene expression in *Streptococcus pneumoniae*: impact on the fine-tuning of competence development. *Journal of Bacteriology*, 188(23):8307–8312.

Kobayashi, N., Nishino, K., and Yamaguchi, A. (2001), Novel macrolide-specific ABC-type efflux transporter in Escherichia coli. *Journal of bacteriology*, 183(19):5639–44.

Kuroda, T. and Tsuchiya, T. (2009), Multidrug efflux transporters in the MATE family. *Biochimica et biophysica acta*, 1794(5):763–8.

Kyaw, M. H., Lynfield, R., Schaffner, W., Craig, A. S., Hadler, J., Reingold, A., Thomas, A. R., Harrison, L. H., Bennett, N. M., Farley, M. M., Facklam, R. R., Jorgensen, J. H., Besser, J., Zell, E. R., Schuchat, A., and Whitney, C. G. (2006), Effect of introduction of the pneumococcal conjugate vaccine on drug-resistant Streptococcus pneumoniae. *The New England journal of medicine*, 354(14):1455–63.

Lacks, S. (1966), Integration efficiency and genetic recombination in pneumococcal transformation. *Genetics*, 53(1):207–235.

Lacks, S. A. Transformation. In Tuomanen, E. I., editor, *The Pneumococcus*, chapter 7, pages 89–115. ASM Press, (2004).

Laible, G. and Hakenbeck, R. (1987), Penicillin-binding proteins in beta-lactam-resistant laboratory mutants of Streptococcus pneumoniae. *Molecular microbiology*, 1(3):355–63.

Laible, G., Spratt, B. G., and Hakenbeck, R. (1991), Interspecies recombinational events during the evolution of altered PBP 2x genes in penicillin-resistant clinical isolates of Streptococcus pneumoniae. *Molecular microbiology*, 5(8):1993–2002.

Langmead, B. and Salzberg, S. L. (2012), Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–9.

Lawler, a. J., Ricci, V., Busby, S. J. W., and Piddock, L. J. V. (2013), Genetic inactivation of acrAB or inhibition of efflux induces expression of ramA. *The Journal of antimicrobial chemotherapy*, 68(7):1551–7.

Leclercq, R., Dutka-Malen, S., Duval, J., and Courvalin, P. (1992), Vancomycin resistance gene vanC is specific to *Enterococcus gallinarum*. *Antimicrobial Agents and Chemotherapy*, 36(9):2005–8.

Lee, E.-W., Huda, M. N., Kuroda, T., Mizushima, T., and Tsuchiya, T. (2003), EfrAB, an ABC multidrug efflux pump in Enterococcus faecalis. *Antimicrobial agents and chemotherapy*, 47(12):3733–8.

Lee, M. S. and Morrison, D. a. (1999), Identification of a new regulator in *Streptococcus pneumoniae* linking quorum sensing to competence for genetic transformation. *Journal of Bacteriology*, 181(16):5004–16.

Lehner, A. F. and Hill, C. W. (1980), Involvement of Ribosomal Ribonucleic Acid Operons in *Salmonella typhimurium* Chromosomal Rearrangements. 143(1):492–498.

Leino, T., Auranen, K., Jokinen, J., Leinonen, M., Tervonen, P., and Takala, A. K. (2001), Pneumococcal carriage in children during their first two years: important role of family exposure. *The Pediatric infectious disease journal*, 20(11):1022–7.

Leitner, I., Nemeth, J., Feurstein, T., Abrahim, A., Matzneller, P., Lagler, H., Erker, T., Langer, O., and Zeitlinger, M. (2011), The third-generation P-glycoprotein inhibitor tariquidar may overcome bacterial multidrug resistance by increasing intracellular drug concentration. *The Journal of Antimicrobial Chemotherapy*, 66(4):834–9.

Leo, E., Gould, K. a., Pan, X.-S., Capranico, G., Sanderson, M. R., Palumbo, M., and Fisher, L. M. (2005), Novel symmetric and asymmetric DNA scission determinants for *Streptococcus pneumoniae* topoisomerase IV and gyrase are clustered at the DNA breakage site. *The Journal of Biological Chemistry*, 280(14):14252–63.

Lesher, G. Y., Froelich, E. J., Gruett, M. D., Bailey, J. H., and Brundage, R. P. (1962), 1,8-Naphthyridine Derivatives. A New Class of Chemotherapeutic Agents. *Journal of medicinal and pharmaceutical chemistry*, 91:1063–5.

Lewis, K. (1994), Multidrug resistance pumps in bacteria: variations on a theme. *Trends in biochemical sciences*, 19(3):119–23.

Lewis, K. (2000), Programmed death in bacteria. *Microbiology and Molecular Biology Reviews*, 64(3):503–14.

Li, H. and Durbin, R. (2009), Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14):1754–60.

Li, H., Ruan, J., and Durbin, R. (2008), Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Research*, 18(11):1851–8.

Lim, S. P. and Nikaido, H. (2010), Kinetic parameters of efflux of penicillins by the multidrug efflux transporter AcrAB-TolC of *Escherichia coli*. *Antimicrobial Agents and Chemotherapy*, 54(5):1800–6.

Lin, H. T., Bavro, V. N., Barrera, N. P., Frankish, H. M., Velamakanni, S., van Veen, H. W., Robinson, C. V., Borges-Walmsley, M. I., and Walmsley, A. R. (2009), MacB ABC transporter is a dimer whose ATPase activity and macrolide-binding capacity are regulated by the membrane fusion protein MacA. *The Journal of Biological Chemistry*, 284(2):1145–54.

Lin, J., Sahin, O., Michel, L. O., and Zhang, Q. (2003), Critical role of multidrug efflux pump CmeABC in bile resistance and in vivo colonization of Campylobacter jejuni. *Infection and immunity*, 71(8):4250–9.

Lin, R. J., Capage, M., and Hill, C. W. (1984), A repetitive DNA sequence, rhs, responsible for duplications within the Escherichia coli K-12 chromosome. *Journal of molecular biology*, 177(1):1–18.

Lindberg, F., Lindquist, S., and Normark, S. (1987), Inactivation of the *ampD* gene causes semiconstitutive overproduction of the inducible *Citrobacter freundii* beta-lactamase. *Journal of Bacteriology*, 169(5):1923–8.

Little, J. W. and Mount, D. W. (1982), The SOS regulatory system of Escherichia coli. *Cell*, 29(1):11–22.

Littlejohn, T. G., Paulsen, I. T., Gillespie, M. T., Tennent, J. M., Midgley, M., Jones, I. G., Purewal, A. S., and Skurray, R. A. (1992), Substrate specificity and energetics of antiseptic and disinfectant resistance in Staphylococcus aureus. *FEMS microbiology letters*, 74(2-3):259–65.

Liu, J., Huang, C., Shin, D.-H., Yokota, H., Jancarik, J., Kim, J.-S., Adams, P. D., Kim, R., and Kim, S.-H. (2005), Crystal structure of a heat-inducible transcriptional repressor HrcA from *Thermotoga maritima*: structural insight into DNA binding and dimerization. *Journal of Molecular Biology*, 350(5):987–96.

Livermore, D. M. (2001), Of *Pseudomonas*, porins, pumps and carbapenems. *The Journal of Antimicrobial Chemotherapy*, 47(3):247–50.

Lloyd, G. S., Hollands, K., Godfrey, R. E., and Busby, S. J. W. (2008), Transcription initiation in the Escherichia coli K-12 malI-malX intergenic region and the role of the cyclic AMP receptor protein. *FEMS microbiology letters*, 288(2):250–7.

Lomovskaya, O. and Watkins, W. (2001), Inhibition of efflux pumps as a novel approach to combat drug resistance in bacteria. *Journal of molecular microbiology and biotechnology*, 3(2):225–36.

Lorenz, R., Bernhart, S. H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., and Hofacker, I. L. (2011), ViennaRNA Package 2.0. *Algorithms for molecular biology : AMB*, 6:26.

Lovett, S. T., Drapkin, P. T., Sutera, V. A., and Gluckman-peskind, T. J. (1993), A Sister-Strand Exchange Mechanism For recA-Independent Deletion of Repeated DNA Sequences in *Escherichia coli*. 642(Cm).

Lubelski, J., Mazurkiewicz, P., van Merkerk, R., Konings, W. N., and Driessen, A. J. M. (2004), ydaG and ydbA of Lactococcus lactis encode a heterodimeric ATP-binding cassette-type multidrug transporter. *The Journal of biological chemistry*, 279(33): 34449–55.

Lubelski, J., de Jong, A., van Merkerk, R., Agustiandari, H., Kuipers, O. P., Kok, J., and Driessen, A. J. M. (2006), LmrCD is a major multidrug resistance transporter in *Lactococcus lactis*. *Molecular microbiology*, 61(3):771–81.

Lubelski, J., Konings, W. N., and Driessen, A. J. M. (2007), Distribution and physiology of ABC-type transporters contributing to multidrug resistance in bacteria. *Microbiology and Molecular Biology Reviews*, 71(3):463–76.

Luo, N., Sahin, O., Lin, J., Michel, L. O., and Zhang, Q. (2003), In vivo selection of Campylobacter isolates with high levels of fluoroquinolone resistance associated with gyrA mutations and the function of the CmeABC efflux pump. *Antimicrobial agents and chemotherapy*, 47(1):390–4.

Lupien, A., Billal, D. S., Fani, F., Soualhine, H., Zhanel, G. G., Leprohon, P., and Ouellette, M. (2013), Genomic characterisation of ciprofloxacin resistance in a laboratory derived mutant and clinical isolate of *Streptococcus pneumoniae*. *Antimicrobial agents and chemotherapy*.

Ma, D., Cook, D. N., Alberti, M., Pon, N. G., Nikaido, H., and Hearst, J. E. (1995), Genes acrA and acrB encode a stress-induced efflux system of Escherichia coli. *Molecular microbiology*, 16(1):45–55.

Magee, A. D. and Yother, J. (2001), Requirement for capsule in colonization by Streptococcus pneumoniae. *Infection and immunity*, 69(6):3755–61.

Makarova, K., Slesarev, a., Wolf, Y., Sorokin, a., Mirkin, B., Koonin, E., Pavlov, a., Pavlova, N., Karamychev, V., Polouchine, N., Shakhova, V., Grigoriev, I., Lou, Y., Rohksar, D., Lucas, S., Huang, K., Goodstein, D. M., Hawkins, T., Plengvidhya, V., Welker, D., Hughes, J., Goh, Y., Benson, a., Baldwin, K., Lee, J.-H., Díaz-Muñiz, I., Dosti, B., Smeianov, V., Wechter, W., Barabote, R., Lorca, G., Altermann, E., Barrangou, R., Ganesan, B., Xie, Y., Rawsthorne, H., Tamir, D., Parker, C., Breidt, F., Broadbent, J., Hutkins, R., O'Sullivan, D., Steele, J., Unlu, G., Saier, M., Klaenhammer, T., Richardson, P., Kozyavkin, S., Weimer, B., and Mills, D. (2006), Comparative genomics of the lactic acid bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 103(42):15611–6.

Mandal, M. and Breaker, R. R. (2004), Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nature structural & molecular biology*, 11 (1):29–35.

Mandal, M., Lee, M., Barrick, J. E., Weinberg, Z., Emilsson, G. M., Ruzzo, W. L., and Breaker, R. R. (2004), A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science (New York, N.Y.)*, 306(5694):275–9.

Mandell, L. a., Wunderink, R. G., Anzueto, A., Bartlett, J. G., Campbell, G. D., Dean, N. C., Dowell, S. F., File, T. M., Musher, D. M., Niederman, M. S., Torres, A., and Whitney, C. G. (2007), Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 44 Suppl 2(Suppl 2):S27–72.

Markham, P. N. and Neyfakh, A. A. (1996), Inhibition of the multidrug transporter NorA prevents emergence of norfloxacin resistance in Staphylococcus aureus. *Antimicrobial agents and chemotherapy*, 40(11):2673–4.

Marra, A., Lawson, S., Asundi, J. S., Brigham, D., and Hromockyj, A. E. (2002), In vivo characterization of the psa genes from Streptococcus pneumoniae in multiple models of infection. *Microbiology (Reading, England)*, 148(Pt 5):1483–91.

Marrer, E., Satoh, A., Johnson, M., Piddock, L., and Page, M. (2006)a, Global transcriptome analysis of the responses of a fluoroquinolone-resistant *Streptococcus pneumoniae* mutant and its parent to ciprofloxacin. *Antimicrobial Agents and Chemotherapy*, 50(1):269.

Marrer, E., Schad, K., Satoh, A., Page, M., Johnson, M., and Piddock, L. (2006)b, Involvement of the putative ATP-dependent efflux proteins PatA and PatB in fluoroquinolone resistance of a multidrug-resistant mutant of *Streptococcus pneumoniae*. *Antimicrobial Agents and Chemotherapy*, 50(2):685.

Marsh, R., Smith-Vaughan, H., Hare, K. M., Binks, M., Kong, F., Warning, J., Gilbert, G. L., Morris, P., and Leach, a. J. (2010), The nonserotypeable pneumococcus: phenotypic dynamics in the era of anticapsular vaccines. *Journal of Clinical Microbiology*, 48(3):831–5.

Martínez-Martínez, L., Pascual, A., and Jacoby, G. A. (1998), Quinolone resistance from a transferable plasmid. *Lancet*, 351(9105):797–9.

Matsuoka, M., Endou, K., Kobayashi, H., Inoue, M., and Nakajima, Y. (1998), A plasmid that encodes three genes for resistance to macrolide antibiotics in *Staphylococcus aureus*. *FEMS Microbiology Letters*, 167(2):221–7.

Matthews, P. R. and Stewart, P. R. (1988), Amplification of a section of chromosomal DNA in methicillin-resistant Staphylococcus aureus following growth in high concentrations of methicillin. *Journal of general microbiology*, 134(6):1455–64.

McAleese, F., Petersen, P., Ruzin, A., Dunman, P. M., Murphy, E., Projan, S. J., and Bradford, P. A. (2005), A novel MATE family efflux pump contributes to the reduced susceptibility of laboratory-derived Staphylococcus aureus mutants to tigecycline. *Antimicrobial agents and chemotherapy*, 49(5):1865–71.

McCullers, J. A. and Bartmess, K. C. (2003), Role of neuraminidase in lethal synergism between influenza virus and *Streptococcus pneumoniae*. *Journal of Infectious Diseases*, 187(6):1000–1009.

McDaniel, B. A. M., Grundy, F. J., Artsimovitch, I., and Henkin, T. M. (2003), Transcription termination control of the S box system: direct measurement of S-adenosylmethionine by the leader RNA. *Proceedings of the National Academy of Sciences of the United States of America*, 100(6):3083–8.

McGee, L., McDougal, L., Zhou, J., Spratt, B. G., Tenover, F. C., George, R., Hakenbeck, R., Hryniewicz, W., Lefévre, J. C., Tomasz, A., and Klugman, K. P. (2001), Nomenclature of major antimicrobial-resistant clones of Streptococcus pneumoniae defined by the pneumococcal molecular epidemiology network. *Journal of clinical microbiology*, 39(7):2565–71.

Medvedev, P., Fiume, M., Dzamba, M., Smith, T., and Brudno, M. (2010), Detecting copy number variation with mated short reads. *Genome Research*, 20(11):1613–22.

Merino, E. and Yanofsky, C. (2005), Transcription attenuation: a highly conserved regulatory strategy used by bacteria. *Trends in Genetics*, 21(5):260–4.

Millar, M. R., Brown, N. M., Tobin, G. W., Murphy, P. J., Windsor, A. C., and Speller, D. C. (1994), Outbreak of infection with penicillin-resistant Streptococcus pneumoniae in a hospital for the elderly. *The Journal of hospital infection*, 27(2):99–104.

Mironov, A. S., Gusarov, I., Rafikov, R., Lopez, L. E., Shatalin, K., Kreneva, R. A., Perumov, D. A., and Nudler, E. (2002), Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell*, 111(5):747–56.

Modi, S. R., Lee, H. H., Spina, C. S., and Collins, J. J. (2013), Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature*, 499(7457):219–222.

Mook-Kanamori, B. B., Geldhoff, M., van der Poll, T., and van de Beek, D. (2011), Pathogenesis and pathophysiology of pneumococcal meningitis. *Clinical Microbiology Reviews*, 24(3):557–91.

Morgan, P. J., Hyman, S. C., Rowe, a. J., Mitchell, T. J., Andrew, P. W., and Saibil, H. R. (1995), Subunit organisation and symmetry of pore-forming, oligomeric pneumolysin. *FEBS letters*, 371(1):77–80.

Morita, Y., Kodama, K., Shiota, S., Mine, T., Kataoka, A., Mizushima, T., and Tsuchiya, T. (1998), NorM, a putative multidrug efflux protein, of Vibrio parahaemolyticus and

its homolog in Escherichia coli. *Antimicrobial agents and chemotherapy*, 42(7):1778–82.

Morita, Y., Sobel, M. L., and Poole, K. (2006), Antibiotic inducibility of the MexXY multidrug efflux system of Pseudomonas aeruginosa: involvement of the antibiotic-inducible PA5471 gene product. *Journal of bacteriology*, 188(5):1847–55.

Morrison, D. A. and Guild, W. R. (1972), Transformation and deoxyribonucleic acid size: extent of degradation on entry varies with size of donor. *Journal of bacteriology*, 112(3):1157–68.

Morrison, D. A., Mortier-Barrière, I., Attaiech, L., and Claverys, J.-P. (2007), Identification of the major protein component of the pneumococcal eclipse complex. *Journal of bacteriology*, 189(17):6497–500.

Mortier-Barrière, I., De Saizieu, A., Claverys, J.-P., and Martin, B. (1998), Competence-specific induction of *recA* is required for full recombination proficiency during transformation in *Streptococcus pneumoniae*. *Molecular Microbiology*, 27(1):159–170.

Mosser, J. L. and Tomasz, A. (1970), Choline-containing teichoic acid as a structural component of pneumococcal cell wall and its role in sensitivity to lysis by an autolytic enzyme. *The Journal of biological chemistry*, 245(2):287–98.

Muñoz, R. and De La Campa, a. G. (1996), ParC subunit of DNA topoisomerase IV of *Streptococcus pneumoniae* is a primary target of fluoroquinolones and cooperates with DNA gyrase A subunit in forming resistance phenotype. *Antimicrobial agents and chemotherapy*, 40(10):2252–7.

Mühlemann, K., Matter, H. C., Täuber, M. G., and Bodmer, T. (2003), Nationwide surveillance of nasopharyngeal Streptococcus pneumoniae isolates from children with respiratory infection, Switzerland, 1998-1999. *The Journal of infectious diseases*, 187 (4):589–96.

Musher, D. M. (2003), How contagious are common respiratory tract infections? *The New England journal of medicine*, 348(13):1256–66.

Musher, D. M. A Pathogenic Characterisation of Clinical Symptoms caused by *Streptococcus pneumoniae*. In Tuomanen, E. I., editor, *The Pneumococcus*, pages 211–220. ASM Press, (2004).

Nakamura, A., Miyakozawa, I., Nakazawa, K., O-Hara, K., and Sawai, T. (2000), Detection and characterization of a macrolide 2'-phosphotransferase from a Pseudomonas aeruginosa clinical isolate. *Antimicrobial Agents and Chemotherapy*, 44(11):3241–2.

Narberhaus, F. (1999), Negative regulation of bacterial heat shock genes. *Molecular Microbiology*, 31(1):1–8.

Navarro, F. and Courvalin, P. (1994), Analysis of genes encoding D-alanine-D-alanine ligase-related enzymes in Enterococcus casseliflavus and Enterococcus flavescens. *Antimicrobial Agents and Chemotherapy*, 38(8):1788–93.

Naville, M. and Gautheret, D. (2010)a, Transcription attenuation in bacteria: theme and variations. *Briefings in Functional Genomics*, 9(2):178–189.

Naville, M. and Gautheret, D. (2010)b, Premature terminator analysis sheds light on a hidden world of bacterial transcriptional attenuation. *Genome biology*, 11(9):R97.

Naville, M., Ghuillot-Gaudeffroy, A., Marchais, A., and Gautheret, D. (2011), ARNold: a web tool for the prediction of Rho-independent transcription terminators. *RNA biology*, 8(1):11–3.

Nelson, A. L., Roche, A. M., Gould, J. M., Chim, K., Ratner, A. J., and Weiser, J. N. (2007), Capsule enhances pneumococcal colonization by limiting mucus-mediated clearance. *Infection and immunity*, 75(1):83–90.

Neyfakh, A. A., Bidnenko, V. E., and Chen, L. B. (1991), Efflux-mediated multidrug resistance in Bacillus subtilis: similarities and dissimilarities with the mammalian system. *Proceedings of the National Academy of Sciences of the United States of America*, 88(11):4781–5.

Neyfakh, A. A., Borsch, C. M., and Kaatz, G. W. (1993), Fluoroquinolone resistance protein NorA of Staphylococcus aureus is a multidrug efflux transporter. *Antimicrobial agents and chemotherapy*, 37(1):128–9.

Nicoloff, H., Perreten, V., McMurry, L. M., and Levy, S. B. (2006), Role for tandem duplication and lon protease in AcrAB-TolC- dependent multiple antibiotic resistance (Mar) in an Escherichia coli mutant without mutations in marRAB or acrRAB. *Journal of bacteriology*, 188(12):4413–23.

Nicoloff, H., Perreten, V., and Levy, S. B. (2007), Increased genome instability in Escherichia coli lon mutants: relation to emergence of multiple-antibiotic-resistant (Mar) mutants caused by insertion sequence elements and large tandem genomic amplifications. *Antimicrobial Agents and Chemotherapy*, 51(4):1293–303.

Nikaido, E., Yamaguchi, A., and Nishino, K. (2008), AcrAB multidrug efflux pump regulation in Salmonella enterica serovar Typhimurium by RamA in response to environmental signals. *The Journal of biological chemistry*, 283(35):24245–53.

Nikaido, H. (2003), Molecular basis of bacterial outer membrane permeability revisited. *Microbiology and Molecular Biology Reviews*, 67(4):593–656.

Nilsson, A. I., Zorzet, A., Kanth, A., Dahlström, S., Berg, O. G., and Andersson, D. I. (2006), Reducing the fitness cost of antibiotic resistance by amplification of initiator tRNA genes. *Proceedings of the National Academy of Sciences of the United States of America*, 103(18):6976–81.

Nishino, K., Latifi, T., and Groisman, E. A. (2006), Virulence and drug resistance roles of multidrug efflux systems of Salmonella enterica serovar Typhimurium. *Molecular microbiology*, 59(1):126–41.

Nishino, K., Nikaido, E., and Yamaguchi, A. (2009), Regulation and physiological function of multidrug efflux pumps in Escherichia coli and Salmonella. *Biochimica et Biophysica Acta*, 1794(5):834–43.

Noguchi, N., Emura, A., Matsuyama, H., O'Hara, K., Sasatsu, M., and Kono, M. (1995), Nucleotide sequence and characterization of erythromycin resistance determinant that encodes macrolide 2'-phosphotransferase I in Escherichia coli. *Antimicrobial Agents and Chemotherapy*, 39(10):2359–63.

Nordmann, P., Poirel, L., Walsh, T. R., and Livermore, D. M. (2011), The emerging NDM carbapenemases. *Trends in Microbiology*, 19(12):588–95.

Normark, S., Edlund, T., Grundström, T., Bergström, S., and Wolf-Watz, H. (1977), Escherichia coli K-12 mutants hyperproducing chromosomal beta-lactamase by gene repetitions. *Journal of bacteriology*, 132(3):912–22.

Novak, R., Cauwels, A., Charpentier, E., and Tuomanen, E. (1999), Identification of a Streptococcus pneumoniae gene locus encoding proteins of an ABC phosphate transporter and a two-component regulatory system. *Journal of Bacteriology*, 181(4): 1126–33.

Ochs, M. M., McCusker, M. P., Bains, M., and Hancock, R. E. (1999), Negative regulation of the Pseudomonas aeruginosa outer membrane porin OprD selective for imipenem and basic amino acids. *Antimicrobial Agents and Chemotherapy*, 43(5): 1085–90.

Opperman, T. and Richardson, J. P. (1994), Phylogenetic analysis of sequences from diverse bacteria with homology to the Escherichia coli rho gene. *Journal of bacteriology*, 176(16):5033–43.

Pan, X. S., Ambler, J., Mehtar, S., and Fisher, L. M. (1996), Involvement of topoisomerase IV and DNA gyrase as ciprofloxacin targets in *Streptococcus pneumoniae*. *Antimicrobial Agents and Chemotherapy*, 40(10):2321–6.

Pao, S. S., Paulsen, I. T., and Saier, M. H. (1998), Major facilitator superfamily. *Microbiology and molecular biology reviews : MMBR*, 62(1):1–34.

Parsley, L. C., Consuegra, E. J., Kakirde, K. S., Land, A. M., Harper, W. F., and Liles, M. R. (2010), Identification of diverse antimicrobial resistance determinants carried on bacterial, plasmid, or viral metagenomes from an activated sludge microbial assemblage. *Applied and Environmental Microbiology*, 76(11):3753–7.

Patel, S. N., McGeer, A., Melano, R., Tyrrell, G. J., Green, K., Pillai, D. R., and Low, D. E. (2011), Susceptibility of Streptococcus pneumoniae to fluoroquinolones in Canada. *Antimicrobial agents and chemotherapy*, 55(8):3703–8.

Paton, J. H. and Reeves, D. S. (1988), Fluoroquinolone antibiotics. Microbiology, pharmacokinetics and clinical use. *Drugs*, 36(2):193–228.

Paulsen, I. T., Brown, M. H., and Skurray, R. A. (1996), Proton-dependent multidrug efflux systems. *Microbiological Reviews*, 60(4):575–608.

Paulsen, I. T., Sliwinski, M. K., and Saier, M. H. (1998), Microbial genome analyses: global comparisons of transport capabilities based on phylogenies, bioenergetics and substrate specificities. *Journal of molecular biology*, 277(3):573–92.

Paulsen, I. T., Nguyen, L., Sliwinski, M. K., Rabus, R., and Saier, M. H. (2000), Microbial genome analyses: comparative transport capabilities in eighteen prokaryotes. *Journal of molecular biology*, 301(1):75–100.

Peltola, V. T. and McCullers, J. A. (2004), Respiratory viruses predisposing to bacterial infections: role of neuraminidase. *The Pediatric infectious disease journal*, 23(1 Suppl): S87–97.

Pelton, S. I., Loughlin, A. M., and Marchant, C. D. (2004), Seven valent pneumococcal conjugate vaccine immunization in two Boston communities: changes in serotypes and antimicrobial susceptibility among Streptococcus pneumoniae isolates. *The Pediatric infectious disease journal*, 23(11):1015–22.

Perry, C. M., Barman Balfour, J. A., and Lamb, H. M. (1999), Gatifloxacin. *Drugs*, 58 (4):683–96; discussion 697–8.

Pestova, E. V., Havarstein, L. S., and Morrison, D. A. (1996), Regulation of competence for genetic transformation in *Streptococcus pneumoniae* by an auto-induced peptide pheromone and a two-component regulatory system. *Molecular Microbiology*, 21(4): 853–862.

Pestova, E. and Millichap, J. (2002), Non-PmrA-mediated multidrug resistance in *Streptococcus pneumoniae*. *Journal of Antimicrobial Chemotherapy . . .* , 49:553–556.

Peterson, S., Cline, R. T., Tettelin, H., Sharov, V., and Morrison, D. A. (2000), Gene expression analysis of the Streptococcus pneumoniae competence regulons by use of DNA microarrays. *Journal of Bacteriology*, 182(21):6192–202.

Peterson, S. N., Sung, C. K., Cline, R., Desai, B. V., Snesrud, E. C., Luo, P., Walling, J., Li, H., Mintz, M., Tsegaye, G., Burr, P. C., Do, Y., Ahn, S., Gilbert, J., Fleischmann, R. D., and Morrison, D. a. (2004), Identification of competence pheromone responsive genes in Streptococcus pneumoniae by use of DNA microarrays. *Molecular Microbiology*, 51(4):1051–1070.

Pfaffl, M. W. (2001), A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*, 29(9):e45.

Piddock, L. J., Johnson, M., Ricci, V., and Hill, S. L. (1998), Activities of new fluoroquinolones against fluoroquinolone-resistant pathogens of the lower respiratory tract. *Antimicrobial agents and chemotherapy*, 42(11):2956–60.

Piddock, L. J. V., Jin, Y.-F., and Everett, M. J. (1997), Non-*gyrA*-mediated ciprofloxacin resistance in laboratory mutants of *Streptococcus pneumoniae*. *Journal of Antimicrobial Chemotherapy*, pages 609–615.

Piddock, L. J. V. (2006)a, Clinically Relevant Chromosomally Encoded Multidrug Resistance Efflux Pumps in Bacteria. *Society*, 19(2):382–402.

Piddock, L. J. V. and Johnson, M. M. (2002), Accumulation of 10 Fluoroquinolones by Wild-Type or Efflux Mutant *Streptococcus pneumoniae*. 46(3):813–820.

Piddock, L. J. V., Johnson, M. M., Simjee, S., and Pumbwe, L. (2002), Expression of Efflux Pump Gene *pmrA* in Fluoroquinolone-Resistant and -Susceptible Clinical Isolates of *Streptococcus pneumoniae*. *Society*, 46(3):808–812.

Piddock, L. (2006)b, Multidrug-resistance efflux pumps? not just for resistance. *Nature Reviews Microbiology*, 4(8):629–36.

Pinho, M. G., de Lencastre, H., and Tomasz, A. (2001), An acquired and a native penicillin-binding protein cooperate in building the cell wall of drug-resistant staphylococci. *Proceedings of the National Academy of Sciences of the United States of America*, 98(19):10886–91.

Pletz, M. W. R., Mcgee, L., Jorgensen, J., Beall, B., Facklam, R. R., Whitney, C. G., and Klugman, K. P. (2004), Levofloxacin-Resistant Invasive Streptococcus pneumoniae in the United States : Evidence for Clonal Spread and the Impact of Conjugate Pneumococcal Vaccine. *Society*, 48(9):3491–3497.

Poisson, J., Le Hir, A., Goutarel, R., and Janot, M. M. (1954), [Isolation of reserpine from roots of Rauwolfia vomitoria Afz]. *Comptes rendus hebdomadaires des séances de l'Académie des sciences*, 238(15):1607–9.

Pomposiello, P. J., Bennik, M. H., and Demple, B. (2001), Genome-wide transcriptional profiling of the Escherichia coli responses to superoxide stress and sodium salicylate. *Journal of bacteriology*, 183(13):3890–902.

Poole, K. (2005), Efflux-mediated antimicrobial resistance. *The Journal of Antimicrobial Chemotherapy*, 56(1):20–51.

Prystowsky, J., Siddiqui, F., Chosay, J., Shinabarger, D. L., Millichap, J., Peterson, L. R., and Noskin, G. A. (2001), Resistance to linezolid: characterization of mutations in rRNA and comparison of their occurrences in vancomycin-resistant enterococci. *Antimicrobial agents and chemotherapy*, 45(7):2154–6.

Quagliarello, V. J. and Scheld, W. M. (1997), Treatment of bacterial meningitis. *The New England journal of medicine*, 336(10):708–16.

Ramirez, M. S. and Tolmasky, M. E. (2010), Aminoglycoside modifying enzymes. *Drug Resistance Updates*, 13(6):151–71.

Rane, L. and Subbarow, Y. (1940), Nutritional Requirements of the Pneumococcus: I. Growth Factors for Types I, II, V, VII, VIII. *Journal of bacteriology*, 40(5):695–704.

Raushel, F. M., Thoden, J. B., and Holden, H. M. (1999), The amidotransferase family of enzymes: molecular machines for the production and delivery of ammonia. *Biochemistry*, 38(25):7891–9.

Raymond, J., Le Thomas, I., Moulin, F., Commeau, A., Gendrel, D., and Berche, P. (2000), Sequential colonization by Streptococcus pneumoniae of healthy children living in an orphanage. *The Journal of infectious diseases*, 181(6):1983–8.

Reams, A. B. and Neidle, E. L. (2004), Gene amplification involves site-specific short homology-independent illegitimate recombination in Acinetobacter sp. strain ADP1. *Journal of Molecular Biology*, 338(4):643–56.

Reinert, S., Linden, M. V. D., Cil, M. Y., Al-lahham, A., and Appelbaum, P. (2005), Antimicrobial Susceptibility of *Streptococcus pneumoniae* in Eight European Countries from 2001 to 2003. 49(7):2903–2913.

Rice, P., Longden, I., and Bleasby, A. (2000), EMBOSS: the European Molecular Biology Open Software Suite. *Trends in genetics : TIG*, 16(6):276–7.

Rimini, R., Jansson, B., Feger, G., Roberts, T. C., de Francesco, M., Gozzi, A., Faggioni, F., Domenici, E., Wallace, D. M., Frandsen, N., and Polissi, A. (2000), Global analysis of transcription kinetics during competence development in *Streptococcus pneumoniae* using high density DNA arrays. *Molecular microbiology*, 36(6):1279–92.

Robertson, G., Doyle, T., and Lynch, A. (2005), Use of an efflux-deficient *Streptococcus pneumoniae* strain panel to identify ABC-class multidrug transporters involved in intrinsic resistance to antimicrobial agents. *Antimicrobial Agents and Chemotherapy*, 49(11):4781–4783.

Rodrigues, F., Nunes, S., Sá-Leão, R., Gonçalves, G., Lemos, L., and Lencastre, H. D. (2009), *Streptococcus pneumoniae* nasopharyngeal carriage in children attending day-care centers in the central region of Portugal, in the era of 7-valent pneumococcal conjugate vaccine. *Microbial Drug Resistance*, 15(4):269–277.

Romero, P., Croucher, N. J., Hiller, N. L., Hu, F. Z., Ehrlich, G. D., Bentley, S. D., García, E., and Mitchell, T. J. (2009)a, Comparative genomic analysis of ten Streptococcus pneumoniae temperate bacteriophages. *Journal of Bacteriology*, 191(15): 4854–62.

Romero, P., García, E., and Mitchell, T. J. (2009)b, Development of a prophage typing system and analysis of prophage carriage in Streptococcus pneumoniae. *Applied and Environmental Microbiology*, 75(6):1642–9.

Rouquette-Loughlin, C., Dunham, S. A., Kuhn, M., Balthazar, J. T., and Shafer, W. M. (2003), The NorM efflux pump of Neisseria gonorrhoeae and Neisseria meningitidis recognizes antimicrobial cationic compounds. *Journal of bacteriology*, 185(3):1101–6.

Rouquette-Loughlin, C. E., Balthazar, J. T., and Shafer, W. M. (2005), Characterization of the MacA-MacB efflux system in Neisseria gonorrhoeae. *The Journal of antimicrobial chemotherapy*, 56(5):856–60.

Rozen, D. E., McGee, L., Levin, B. R., and Klugman, K. P. (2007), Fitness costs of fluoroquinolone resistance in Streptococcus pneumoniae. *Antimicrobial Agents and Chemotherapy*, 51(2):412–6.

Rozen, S. and Skaletsky, H. (2000), Primer3 on the WWW for general users and for biologist programmers. *Methods in molecular biology (Clifton, N.J.)*, 132:365–86.

Ruiz, J. (2003), Mechanisms of resistance to quinolones: target alterations, decreased accumulation and DNA gyrase protection. *The Journal of Antimicrobial Chemotherapy*, 51(5):1109–17.

Ruiz-Cruz, S., Solano-Collado, V., Espinosa, M., and Bravo, A. (2010), Novel plasmid-based genetic tools for the study of promoters and terminators in Streptococcus pneumoniae and Enterococcus faecalis. *Journal of Microbiological Methods*, 83(2):156–63.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. a., and Barrell, B. (2000), Artemis: sequence visualization and annotation. *Bioinformatics*, 16(10):944–5.

Sá-Leão, R., Nunes, S., Brito-Avô, A., Alves, C. R., Carriço, J. a. A., Saldanha, J., Almeida, J. S., Santos-Sanches, I., and de Lencastre, H. (2008), High rates of transmission of and colonization by Streptococcus pneumoniae and Haemophilus influenzae within a day care center revealed in a longitudinal study. *Journal of clinical microbiology*, 46(1):225–34.

Sadowy, E., Kuch, A., Gniadkowski, M., and Hryniewicz, W. (2010), Expansion and evolution of the Streptococcus pneumoniae Spain9V-ST156 clonal complex in Poland. *Antimicrobial agents and chemotherapy*, 54(5):1720–7.

Sakamoto, K., Margolles, A., van Veen, H. W., and Konings, W. N. (2001), Hop resistance in the beer spoilage bacterium Lactobacillus brevis is mediated by the ATP-binding cassette multidrug transporter HorA. *Journal of bacteriology*, 183(18):5371–5.

Salter, S. J., Hinds, J., Gould, K. a., Lambertsen, L., Hanage, W. P., Antonio, M., Turner, P., Hermans, P. W. M., Bootsma, H. J., O'Brien, K. L., and Bentley, S. D. (2012), Variation at the capsule locus, *cps*, of mistyped and non-typable *Streptococcus pneumoniae* isolates. *Microbiology (Reading, England)*, 158(Pt 6):1560–9.

Salyers, a. a., Shoemaker, N. B., Stevens, a. M., and Li, L. Y. (1995), Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. *Microbiological Reviews*, 59(4):579–90.

Sandegren, L. and Andersson, D. I. (2009), Bacterial gene amplification: implications for the evolution of antibiotic resistance. *Nature reviews Microbiology*, 7(8):578–88.

Santagati, M., Iannelli, F., Oggioni, M. R., Stefani, S., and Pozzi, G. (2000), Characterization of a genetic element carrying the macrolide efflux gene mef(A) in Streptococcus pneumoniae. *Antimicrobial agents and chemotherapy*, 44(9):2585–7.

Saskova, L., Novakova, L., Basler, M., and Branny, P. (2007), Eukaryotic-Type Serine/Threonine Protein Kinase StkP Is a Global Regulator of Gene Expression in Streptococcus pneumoniae. *Journal of Bacteriology*, 189(11):4168–4179.

SáâĂŘLeão, R. and Nunes, S. (2009), Changes in pneumococcal serotypes and antibiotypes carried by vaccinated and unvaccinated day-care centre attendees in Portugal, a country with widespread use of the seven-valent pneumococcal conjugate vaccine. *Clinical Microbiology . . .*, 15(11):1002–1007.

Schito, G. C. and Felmingham, D. (2005), Susceptibility of Streptococcus pneumoniae to penicillin, azithromycin and telithromycin (PROTEKT 1999-2003). *International journal of antimicrobial agents*, 26(6):479–85.

Schmittgen, T. D., Zakrajsek, B. a., Mills, a. G., Gorn, V., Singer, M. J., and Reed, M. W. (2000), Quantitative reverse transcription-polymerase chain reaction to study mRNA decay: comparison of endpoint and real-time methods. *Analytical biochemistry*, 285(2):194–204.

Schmitz, F.-J., Higgins, P. G., Mayer, S., Fluit, A. C., and Dalhoff, A. (2002), Activity of quinolones against gram-positive cocci: mechanisms of drug action and bacterial resistance. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology*, 21(9):647–59.

Schulz, A. and Schumann, W. (1996), hrcA, the first gene of the Bacillus subtilis dnaK operon encodes a negative regulator of class I heat shock genes. *Journal of Bacteriology*, 178(4):1088–93.

Schwarz, S., Kehrenberg, C., Doublet, B., and Cloeckaert, A. (2004), Molecular basis of bacterial resistance to chloramphenicol and florfenicol. *FEMS Microbiology Reviews*, 28(5):519–42.

Seeger, M. a. and van Veen, H. W. (2009), Molecular basis of multidrug transport by ABC transporters. *Biochimica et biophysica acta*, 1794(5):725–37.

Semchyshyn, H., Bagnyukova, T., Storey, K., and Lushchak, V. (2005), Hydrogen peroxide increases the activities of soxRS regulon enzymes and the levels of oxidized proteins and lipids in Escherichia coli. *Cell biology international*, 29(11):898–902.

Seoane, A., Sánchez, E., and García-Lobo, J. M. (2003), Tandem amplification of a 28-kilobase region from the Yersinia enterocolitica chromosome containing the blaA gene. *Antimicrobial Agents and Chemotherapy*, 47(2):682–8.

Shakhnovich, E. A., King, S. J., and Weiser, J. N. (2002), Neuraminidase expressed by Streptococcus pneumoniae desialylates the lipopolysaccharide of Neisseria meningitidis and Haemophilus influenzae: a paradigm for interbacterial competition among pathogens of the human respiratory tract. *Infection and immunity*, 70(12):7161–4.

Shen, P. and Huang, H. V. (1986), Homologous recombination in Escherichia coli: dependence on substrate length and homology. *Genetics*, 112(3):441–57.

Shi, J., Blundell, T. L., and Mizuguchi, K. (2001), FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *Journal of molecular biology*, 310(1):243–57.

Singh, S. S. and Grainger, D. C. (2013), H-NS can facilitate specific DNA-binding by RNA polymerase in AT-rich gene regulatory regions. *PLoS genetics*, 9(6):e1003589.

Sjöström, K., Spindler, C., Ortqvist, a., Kalin, M., Sandgren, a., Kühlmann-Berenzon, S., and Henriques-Normark, B. (2006), Clonal and capsular types decide whether pneumococci will act as a primary or opportunistic pathogen. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 42(4): 451–9.

Sjöström, K., Blomberg, C., Fernebro, J., Dagerhamn, J., Morfeldt, E., Barocchi, M. a., Browall, S., Moschioni, M., Andersson, M., Henriques, F., Albiger, B., Rappuoli, R., Normark, S., and Henriques-Normark, B. (2007), Clonal success of piliated penicillin nonsusceptible pneumococci. *Proceedings of the National Academy of Sciences of the United States of America*, 104(31):12907–12.

Skovgaard, O., Bak, M., Lø bner Olesen, A., and Tommerup, N. (2011), Genome-wide detection of chromosomal rearrangements, indels, and mutations in circular chromosomes by short read sequencing. *Genome Research*, 21(8):1388–93.

Slonczewski, J. L. and Foster, J. W. *Microbiology*. Norton, 2 edition, (2009).

Stanhope, M. J., Walsh, S. L., Becker, J. A., Italia, M. J., Ingraham, K. A., Gwynn, M. N., Mathie, T., Poupard, J. A., Miller, L. A., Brown, J. R., and Amrine-Madsen, H. (2005), Molecular evolution perspectives on intraspecific lateral DNA transfer of topoisomerase and gyrase loci in Streptococcus pneumoniae, with implications for fluoroquinolone resistance development and spread. *Antimicrobial agents and chemotherapy*, 49(10):4315–26.

Stavri, M., Piddock, L. J. V., and Gibbons, S. (2007), Bacterial efflux pump inhibitors from natural sources. *The Journal of antimicrobial chemotherapy*, 59(6):1247–60.

Steinmoen, H., Knutsen, E., and Håvarstein, L. S. (2002), Induction of natural competence in Streptococcus pneumoniae triggers lysis and DNA release from a subfraction of the cell population. *Proceedings of the National Academy of Sciences of the United States of America*, 99(11):7681–6.

Sugawara, E., Nestorovich, E. M., Bezrukov, S. M., and Nikaido, H. (2006), Pseudomonas aeruginosa porin OprF exists in two different conformations. *The Journal of Biological Chemistry*, 281(24):16220–9.

Sun, S., Ke, R., Hughes, D., Nilsson, M., and Andersson, D. I. (2012), Genome-Wide Detection of Spontaneous Chromosomal Rearrangements in Bacteria. 7(8).

Swartz, M. N. (2002), Attacking the pneumococcus – a hundred years' war. *The New England journal of medicine*, 346(10):722.

Swiatlo, E., McDaniel, L. S., and Briles, D. E. Choline-binding proteins. In Tuomanen, E. I., editor, *The Pneumococcus*, chapter 4, pages 49–60. ASM Press, (2004).

Talbot, U. M., Paton, a. W., and Paton, J. C. (1996), Uptake of Streptococcus pneumoniae by respiratory epithelial cells. *Infection and immunity*, 64(9):3772–7.

Thaker, M., Spanogiannopoulos, P., and Wright, G. D. (2010), The tetracycline resistome. *Cellular and molecular life sciences*, 67(3):419–31.

Thanassi, D. G., Cheng, L. W., and Nikaido, H. Active efflux of bile salts by *Escherichia coli<\i>*., url = http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=178997&tool=pmcentrez&rendertype= volume = 179, year = 1997. *Journal of bacteriology*, (8):2512–8.

Tlsty, T. D., Albertini, A. M., and Miller, J. H. (1984), Gene amplification in the lac region of E. coli. *Cell*, 37(1):217–24.

Tocci, N., Iannelli, F., Bidossi, A., Ciusa, M. L., Decorosi, F., Viti, C., Pozzi, G., Ricci, S., and Oggioni, M. R. (2013), Functional analysis of pneumococcal drug efflux pumps associates the MATE DinF transporter with quinolone susceptibility. *Antimicrobial agents and chemotherapy*, 57(1):248–53.

Tomasz, A. (1967), Choline in the cell wall of a bacterium: novel type of polymer-linked choline in Pneumococcus. *Science (New York, N.Y.)*, 157(3789):694–7.

Tong, H. H., Blue, L. E., James, M. A., and DeMaria, T. F. (2000), Evaluation of the virulence of a Streptococcus pneumoniae neuraminidase-deficient mutant in nasopharyngeal colonization and development of otitis media in the chinchilla model. *Infection and immunity*, 68(2):921–4.

Tran, J. H. and Jacoby, G. a. (2002), Mechanism of plasmid-mediated quinolone resistance. *Proceedings of the National Academy of Sciences of the United States of America*, 99(8):5638–42.

Trieu-Cuot, P., Poyart-Salmeron, C., Carlier, C., and Courvalin, P. (1990), Nucleotide sequence of the erythromycin resistance gene of the conjugative transposon Tn1545. *Nucleic acids research*, 18(12):3660.

Trinh, T. Q. and Sinden, R. R. (1993), The influence of primary and secondary DNA structure in deletion and duplication between direct repeats in Escherichia coli. *Genetics*, 134(2):409–22.

Turner, R. J., Lu, Y., and Switzer, R. L. (1994), Regulation of the *Bacillus subtilis* pyrimidine biosynthetic (*pyr*) gene cluster by an autogenous transcriptional attenuation mechanism. *Journal of bacteriology*, 176(12):3708–22.

Tyrrell, G. J., Lovgren, M., Chui, N., Minion, J., Garg, S., Kellner, J. D., and Marrie, T. J. (2009), Serotypes and antimicrobial susceptibilities of invasive Streptococcus pneumoniae pre- and post-seven valent pneumococcal conjugate vaccine introduction in Alberta, Canada, 2000-2006. *Vaccine*, 27(27):3553–60.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J. A. M. (2007), Primer3Plus, an enhanced web interface to Primer3. *Nucleic acids research*, 35(Web Server issue):W71–4.

van Opijnen, T. and Camilli, A. (2012), A fine scale phenotype-genotype virulence map of a bacterial pathogen. *Genome Research*, 22(12):2541–51.

van Opijnen, T., Bodi, K. L., and Camilli, A. (2009), Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature Methods*, 6(10):767–72.

van Veen, H. W., Venema, K., Bolhuis, H., Oussenko, I., Kok, J., Poolman, B., Driessen, A. J., and Konings, W. N. (1996), Multidrug resistance mediated by a bacterial homolog of the human multidrug transporter MDR1. *Proceedings of the National Academy of Sciences of the United States of America*, 93(20):10668–72.

van Velkinburgh, J. C. and Gunn, J. S. (1999), PhoP-PhoQ-regulated loci are required for enhanced bile resistance in Salmonella spp. *Infection and immunity*, 67(4):1614–22.

Vanderkooi, O. G., Low, D. E., Green, K., Powis, J. E., and McGeer, A. (2005), Predicting antimicrobial resistance in invasive pneumococcal infections. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 40(9): 1288–97.

Velamakanni, S., Yao, Y., Gutmann, D. A. P., and van Veen, H. W. (2008), Multidrug transport by the ABC transporter Sav1866 from Staphylococcus aureus. *Biochemistry*, 47(35):9300–8.

Venter, H., Velamakanni, S., Balakrishnan, L., and van Veen, H. W. (2008), On the energy-dependence of Hoechst 33342 transport by the ABC transporter LmrA. *Biochemical pharmacology*, 75(4):866–74.

Vitreschak, A. G., Rodionov, D. A., Mironov, A. A., and Gelfand, M. S. (2003), Regulation of the vitamin B12 metabolism and transport in bacteria by a conserved RNA structural element. *RNA (New York, N.Y.)*, 9(9):1084–97.

Vitreschak, A. G., Lyubetskaya, E. V., Shirshin, M. A., Gelfand, M. S., and Lyubetsky, V. A. (2004), Attenuation regulation of amino acid biosynthetic operons in proteobacteria: comparative genomics analysis. *FEMS microbiology letters*, 234(2):357–70.

Vitreschak, A. G., Mironov, A. A., Lyubetsky, V. A., and Gelfand, M. S. (2008), Comparative genomic analysis of T-box regulatory systems in bacteria. *RNA (New York, N.Y.)*, 14(4):717–35.

Washburn, R. S., Marra, A., Bryant, A. P., Rosenberg, M., and Gentry, D. R. (2001), rho is not essential for viability or virulence in Staphylococcus aureus. *Antimicrobial Agents and Chemotherapy*, 45(4):1099–103.

Wasserscheid, J., Dewar, K., Feng, J., Ouellette, M., and Le, D. (2009), Genome sequencing of linezolid-resistant Streptococcus pneumoniae mutants reveals novel mechanisms of resistance. *Genome Research*, pages 1214–1223.

Webber, M. A., Talukder, A., and Piddock, L. J. V. (2005), Contribution of mutation at amino acid 45 of AcrR to acrB expression and ciprofloxacin resistance in clinical and veterinary Escherichia coli isolates. *Antimicrobial Agents and Chemotherapy*, 49 (10):4390–2.

Webber, M. a., Bailey, A. M., Blair, J. M. a., Morgan, E., Stevens, M. P., Hinton, J. C. D., Ivens, A., Wain, J., and Piddock, L. J. V. (2009), The global consequence of disruption of the AcrAB-TolC efflux pump in *Salmonella enterica* includes reduced expression of SPI-1 and other attributes required to infect the host. *Journal of Bacteriology*, 191(13):4276–85.

Weigel, L. M., Anderson, G. J., Facklam, R. R., and Tenover, F. C. (2001), Genetic Analyses of Mutations Contributing to Fluoroquinolone Resistance in Clinical Isolates of *Streptococcus pneumoniae*. *Antimicrobial Agents and Chemotherapy*, 45(12):3517–3523.

Weil-Olivier, C., van der Linden, M., de Schutter, I., Dagan, R., and Mantovani, L. (2012), Prevention of pneumococcal diseases in the post-seven valent vaccine era: a European perspective. *BMC infectious diseases*, 12:207.

Weisblum, B. (1995)a, Erythromycin resistance by ribosome modification. *Antimicrobial Agents and Chemotherapy*, 39(3):577–85.

Weisblum, B. (1995)b, Insights into erythromycin action from studies of its activity as inducer of resistance. *Antimicrobial agents and chemotherapy*, 39(4):797–805.

Weiser, J. N., Austrian, R., Sreenivasan, P. K., and Masure, H. R. (1994), Phase variation in pneumococcal opacity: relationship between colonial morphology and nasopharyngeal colonization. *Infection and immunity*, 62(6):2582–9.

Weng, M. L. and Zalkin, H. (1987), Structural role for a conserved region in the CTP synthetase glutamine amide transfer domain. *Journal of bacteriology*, 169(7):3023–8.

Whoriskey, S. K., Nghiem, V. H., and Leong, P. M. (1987), Genetic rearrangements and gene amplification in Escherichia coli: DNA sequences at the junctures of amplified gene fusions. *Genes & Development*, pages 227–237.

Winkler, W. C., Cohen-Chalamish, S., and Breaker, R. R. (2002), An mRNA structure that controls gene expression by binding FMN. *Proceedings of the National Academy of Sciences of the United States of America*, 99(25):15908–13.

Winkler, W. C., Nahvi, A., Sudarsan, N., Barrick, J. E., and Breaker, R. R. (2003), An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nature structural biology*, 10(9):701–7.

Woodford, N. and Ellington, M. J. (2007), The emergence of antibiotic resistance by mutation. *Clinical Microbiology and Infection*, 13(1):5–18.

Woodford, N., Reddy, S., Fagan, E. J., Hill, R. L. R., Hopkins, K. L., Kaufmann, M. E., Kistler, J., Palepou, M.-F. I., Pike, R., Ward, M. E., Cheesbrough, J., and Livermore, D. M. (2007), Wide geographic spread of diverse acquired AmpC beta-lactamases among Escherichia coli and Klebsiella spp. in the UK and Ireland. *The Journal of Antimicrobial Chemotherapy*, 59(1):102–5.

Yanai, K., Murakami, T., and Bibb, M. (2006), Amplification of the entire kanamycin biosynthetic gene cluster during empirical strain improvement of Streptomyces kanamyceticus. *Proceedings of the National Academy of Sciences of the United States of America*, 103(25):9661–6.

Yerushalmi, H., Lebendiker, M., and Schuldiner, S. (1995), EmrE, an Escherichia coli 12-kDa multidrug transporter, exchanges toxic cations and H+ and is soluble in organic solvents. *The Journal of Biological Chemistry*, 270(12):6856–63.

Yong, D., Toleman, M. A., Giske, C. G., Cho, H. S., Sundman, K., Lee, K., and Walsh, T. R. (2009), Characterization of a new metallo-beta-lactamase gene, bla(NDM-1), and a novel erythromycin esterase gene carried on a unique genetic structure in *Klebsiella pneumoniae* sequence type 14 from India. *Antimicrobial Agents and Chemotherapy*, 53(12):5046–54.

Yoon, E.-J., Courvalin, P., and Grillot-Courvalin, C. (2013), RND-type efflux pumps in multidrug-resistant clinical isolates of *Acinetobacter baumannii*: major role for AdeABC overexpression and AdeRS mutations. *Antimicrobial Agents and Chemotherapy*, 57(7):2989–95.

Yoshida, H., Bogaki, M., Nakamura, S., Ubukata, K., and Konno, M. (1990), Nucleotide sequence and characterization of the Staphylococcus aureus norA gene, which confers resistance to quinolones. *Journal of Bacteriology*, 172(12):6942–9.

Zerbino, D. R. and Birney, E. (2008), Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, 18(5):821–9.

Zhang, A., Yang, M., Hu, P., Wu, J., Chen, B., Hua, Y., Yu, J., Chen, H., Xiao, J., and Jin, M. (2011), Comparative genomic analysis of *Streptococcus suis* reveals significant genomic diversity among different serotypes. *BMC Genomics*, 12(1):523.

Zhang, J. R., Mostov, K. E., Lamm, M. E., Nanno, M., Shimida, S., Ohwaki, M., and Tuomanen, E. (2000), The polymeric immunoglobulin receptor translocates pneumococci across human nasopharyngeal epithelial cells. *Cell*, 102(6):827–37.

Zheng, J., Cui, S., and Meng, J. (2009), Effect of transcriptional activators RamA and SoxS on expression of multidrug efflux pumps AcrAB and AcrEF in fluoroquinolone-resistant Salmonella Typhimurium. *The Journal of antimicrobial chemotherapy*, 63 (1):95–102.

Zomer, A. L., Buist, G., Larsen, R., Kok, J., and Kuipers, O. P. (2007), Time-resolved determination of the CcpA regulon of Lactococcus lactis subsp. cremoris MG1363. *Journal of bacteriology*, 189(4):1366–81.

Zwir, I., Shin, D., Kato, A., Nishino, K., Latifi, T., Solomon, F., Hare, J. M., Huang, H., and Groisman, E. A. (2005), Dissecting the PhoP regulatory network of Escherichia coli and Salmonella enterica. *Proceedings of the National Academy of Sciences of the United States of America*, 102(8):2862–7.