Pathogenesis of *Streptococcus pneumoniae*

Lauren Joy M^{c}Allister, BSc (Mol. Biol.), BSc (Hons)

A thesis submitted in fulfillment of the requirements for the degree of

Doctor of Philosophy from the University of Adelaide

May, 2010

Research Centre for Infectious Diseases,

School of Molecular & Biomedical Sciences,

University of Adelaide,

Adelaide, S.A., Australia

This thesis is dedicated to my mother, and sisters Janine and Suzanne.

Table of Contents

Chapter 3: Initial Characterisation of Serotype 3 and Serogroup 11 Clinical Isolates

Chapter 4: DNA Microarray Analysis and PCR-based Subtractive Hybridisation

Chapter 5: Identification of a Genomic Island Carrying a Putative Cellobiose PTS

Abstract

Streptococcus pneumoniae is the principal bacterial cause of otitis media (OM). While OM does not have the high rates of mortality associated with invasive pneumococcal diseases, such as meningitis, it has an extremely high rate of morbidity, with most children suffering at least one episode of OM during their early childhood years. Despite being a major public health burden worldwide, certain communities are more impacted than others. In Australia, remote Aboriginal communities have the highest rates of OM in the world with absolutely no improvement over the last few decades. Therefore, understanding the pathogenesis of pneumococcal OM is of vital importance.

At the start of this project, serotype 3 middle ear and nasopharyngeal isolates were obtained from both remote Aboriginal communities and the general Australian population from the Menzies School of Health Research (MSHR) in Darwin and the Women's and Children's Hospital (WCH) in Adelaide, respectively. This serotype is an important middle ear pathogen, including in remote Aboriginal communities. In addition, serogroup 11 nasopharyngeal and sinusitis isolates were received from the same sources. Serogroup 11 is rarely reported in OM in remote Australian Aboriginal communities and is generally not as prevalent as serotype 3 in other communities. Furthermore, the serogroup 11 isolates tested were avirulent in mouse invasive disease models, while the serotype 3 isolates tested were virulent.

A serotype 11A nasopharyngeal isolate from a remote Aboriginal community was capsule switched to serotype 3. However, in a pneumonia/sepsis mouse model, it was less fit in the middle ears compared to the middle ear isolate from which the serotype 3 capsule

locus was derived. This indicated the serotype 3 isolates of this study possess genetic factors apart from serotype that influence OM.

Multilocus sequence typing identified two STs in the serotype 3 middle ear isolates (ST180 and ST458). Generally, ST180 is the dominant ST within serotype 3 worldwide, but interestingly, all the isolates typed from MSHR belonged to ST458. The genomes of ST180 and ST458 isolates were analysed using a variety of molecular biological techniques, including a comparison with serogroup 11 isolates by DNA microarray analysis. However, no gene common to one serotype/group but not the other was identified and therefore, the capsule loci remained the only distinguishing feature between serotype 3 and serogroup 11. In order to determine if there were any genes present in the genomes of ST180 and ST458 isolates which were not represented on the microarray slide, PCR-based subtractive hybridisation was employed, and a putative cellobiose phosphotransferase system (PTS) was identified. This PTS is part of a 10 kb island, which includes a sulfatase and ROK family protein. However, there is a large deletion in ST458, which includes the sulfatase and part of the putative cellobiose PTS operon. A large number of strains from a variety of serotype/groups, which included serogroup 11, were found to carry the island through PCR analysis or bioinformatic searches, with most possessing the full island. In ST180, mutagenesis of the island and subsequent virulence studies revealed that the island confers a competitive advantage in a variety of niches, including the ear. Unfortunately, mutagenesis in ST458 was not possible despite numerous attempts.

Finally, the availability of next generation sequencing and a fully sequenced genome of a serotype 3 ST180 strain (OXC141), allowed the investigation for genes which had been missed by PCR-based subtractive hybridisation due to limitations with this technique. The genome of an ST458 middle ear isolate (MSHR17) was analysed, along with an unrelated serotype 3 strain (WU2) as a comparison. Both OXC141 and MSHR17

had regions not represented on the microarray slide, but these regions had not been detected by PCR-based subtractive hybridisation due to their absence in the other ST. Therefore, the PTS island remained the only region common to both STs. Nonetheless, the repertoire of regions possessed by ST180 may explain the dominance of this ST in serotype 3 worldwide, while the repertoire possessed by ST458 may mean it is better adapted to the microenvironmental niches encountered in remote Aboriginal communities.

Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution to Lauren Joy McAllister, and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying, subject to the provisions of the Copyright Act 1968. I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue, the Australasian Digital Theses Program (ADTP) and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

……

Lauren Joy M^{c}Allister

…./…./…..…

Acknowledgements

I really could not have completed my PhD without the support of many others, and it is difficult to find the words to express the extent of my gratitude to the people below.

First of all, thank you to Professor James Paton for being willing to show me around the lab and introduce me to his work all those years ago when I approached him as a Science Scholars student. You have really fostered a love for microbiology in me, and I appreciate the belief you have shown in me and the opportunity to have done a summer scholarship, Honours and PhD with you. I would also like to thank my other two supervisors, Dr Uwe Stroeher and Dr David Ogunniyi, for their tremendous support and encouragement and for always being available, except Thursday lunchtimes! Special thanks to Uwe for helping me with the tedious task of making files for all the poor alignments and gaps in the second generation sequencing data, and Dave for checking on the mice in the middle of the night for me. A huge thanks to the former lab manager Jan Cook for keeping the lab in order and for always willing to help sort out any issues I may have had, as well as all the help you gave me for my mouse work. Thanks to Dr Tony Focareta for always taking an interest in what I was doing (as well as my family). With comments such as “Did the hairdresser make a mistake with your hair dye?”, at least you have ensured I will not leave my PhD with a big ego. Thanks also to Dr Kerrie Grabowski

for always being over-the-top considerate in the lab, especially worrying about distracting me or making too much noise while I wrote.

To Richard Harvey, thanks for your great company throughout my PhD and always lending an ear when I have wanted someone to talk about my project with. Also, for the timeless advice of not eating M^{c}Donald's before running. Thanks to everyone else I have shared an office with throughout my PhD no matter how long or short: Dr Kim LeMessurier, James Byrne, Charlie Plumptre, Brock Herdman, Daphne Mermans and Nadine Verhoeven. I have enjoyed everyone's company, as well as thankful for you putting up with my mess. I would like to thank the other members of the Paton lab both past and present: Dr Austin Chen, Dr Damien Chong, Xiao Hui Gao, Dr Sylvia Herold, Dr Layla Mahdi, Dr Judy Morona, Dr Adrienne Paton, Stephanie Philp, Dr Adam Potter, Dr Trish Rogers, Dr Alistair Standish, Ursula Talbot, Dr Claudia Trappetti, Dr Hui Wang, as well as Dr Chris McDevitt. There are too many people to acknowledge individually, but you have all helped make my PhD an enjoyable and rewarding experience. Thanks to everyone from the Morona lab for their friendliness and who like all good neighbours are always willing to lend something when needed. Thanks also to the CSU, LAS, and Molecular Pathology Unit of the IMVS, as well as Geneworks, for their services.

A huge thanks to my family, especially Mum, for always being there for me. I couldn't ask for a more supportive family. Thanks also to my friends, especially the girls at Craft Saturday for giving me the chance to think about something other than my thesis, Sophia Tan for organising lots of trips to Chocolate Bean, and Keiko, whose huge work ethic and fighting spirit inspires me when I am having a bad day. Also, to my former sharemate Anna and current sharemate Jenny for not only being willing to share a house with someone doing a PhD, but always being there to share good and bad days with.

Special thanks to Jenny for all her support, understanding and encouragement while I finished writing up.

Finally, I would like to thank Dr Amanda Leach from the Menzies School of Health Research in Darwin and Andrew Lawrence from the Women's and Children's Hospital in Adelaide for providing strains used in this study, as well as acknowledge the Australian indigenous communities who allowed the use of the bacterial isolates from the Menzies School of Health Research in this study.

List of Abbreviations

This thesis uses abbreviations that are acceptable to the American Society for Microbiology without definition. Listed below are additional abbreviations, which are defined when first used in the text.

$A_{520,540,570,600}$	Absorbance at 520, 540, 570, 600 nm, respectively
Amp	Ampicillin
AOM	Acute otitis media
AP	Alkaline phophatase
AR	Accessory region
BA	Blood agar
BHI	Brain heart infusion
bp	Base pair(s)
BSA	Bovine serum albumin
CBP	Choline binding protein
CFU	Colony forming unit
CSOM	Chronic supparative otitis media
CPS	Capsule
CSP	Competence stimulating peptide
C-terminus	Carboxy terminus
DIG	Digoxigenin
DOC	Sodium deoxycholate
EDTA	Ethylene diamine tetraacetic acid
Ery	Erythromycin
g	Gravity units
GAG	Glycoaminoglycan
Gent	Gentamycin
h	Hours

Hic	Factor H-binding inhibitor of complement
hRBC	Human red blood cells
IgA	Immunoglobulin A
i.n.	Intranasal
i.p.	Intraperitoneal
IPD	Invasive pneumococcal disease
JCVI	J. Craig Venter Institute, San Diego and Rockville, USA
kb	Kilobase pair (s)
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	Luria Bertani broth
LD_{50}	50% lethal dose
LPS	Lipopolysaccharide
LPxTG	Leucine-proline-x-threonine-glycine binding motif
LytA	Autolysin A
min	Minute(s)
MLST	Multilocus sequence type
MSHR	Menzies School of Health Research, Darwin, Australia
Nan(A,B,C)	Neuraminidase (A, B, C)
NCBI	National Center for Biotechnology Information at the U.S. National Library of Medicine, Bethesda, Maryland, USA
nt	Nucleotide(s)
N-terminus	Amino terminus
OM	Otitis media
OME	Otitis media with effusion
O/N	Overnight
ORF	Open reading frame
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
Ply	Pneumolysin
PP1	Pneumococcal pathogenicity island 1
PspA	Pneumococcal surface protein A
PspC	Pneumococcal surface protein C
PTS	Phosphotransferase system
PTS 1	The putative cellobiose PTS identified in Chapter 4 of this study

PTS 2	The putative cellobiose PTS described by McKessar and Hakenbeck (2007)
RD	Region of diversity
ROK	Repressor, ORF of unknown function, kinase
RT	Room temperature
RT-PCR	Reverse transcriptase polymerase chain reaction
SB	Serum broth
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
Sec	Second (s)
ST	Sequence type
TEMED	N,N,N',N'-tetramethyl-ethylene-diamine
TH	Todd-Hewitt broth
THY	Todd-Hewitt broth supplemented with yeast extract
WCH	Women's and Children's Hospital, Adelaide, Australia
WHO	World Health Organization
X-gal	5-bromo-4-chloro-3-indolyl- β-D-galactopyranoside

Chapter 1

General Introduction

1.1 Historical Background

The pathogenic gram positive bacterium *Streptococcus pneumoniae* (the pneumococcus) was most probably first visualised in 1875 by Edwin Klebs. It was subsequently independently isolated by George Sternberg (Sternberg, 1885) in the United States and Louis Pasteur (Pasteur and Chamberland, 1881) in France in the latter half of 1880 (Austrian, 1981). The bacterium was identified by both men using saliva injected into rabbits; Sternberg with his own and Pasteur from a child with rabies, culminating in the isolation of diplococci bacteria from the blood. Pasteur named the bacterium *Microbe septicemique du salive*, while Sternberg, who isolated the bacterium first but published after Pasteur, named it *Micrococcus pasteuri*. In the 1920s, the organism was assigned the name *Diplococcus pneumoniae*, due to its propensity to form pairs. Finally, in the 1970s, it was renamed *Streptococcus pneumoniae* because of its characteristic chain-like morphology in liquid media (Austrian, 1981; Watson *et al.*, 1993). In the years immediately following its initial isolation, the role of *S. pneumoniae* in human lobar pneumonia was clearly established and by the end of that decade, its involvement in other diseases, such as meningitis and otitis media (OM), was ascertained (Watson *et al.*, 1993).

Since its discovery, research on the pneumococcus has led to many important contributions to science and medicine. For example, the Gram stain (Gram, 1884) was an accidental by-product of work aimed at establishing the cause of pneumonia. Christian

Gram observed that some bacteria retained the stain aniline-gentian violet, while others decolourised, leading to bacteria, including the pneumococcus, being classified as either Gram positive or Gram negative. Although developed over 100 years ago, this classification technique is still used in research and diagnostics today. Another useful diagnostic technique, the Quellung reaction, which is used to determine the capsular serotype of human isolates, was developed due to an observation that the polysaccharide capsule that surrounds the pneumococcus swells when exposed to specific antisera (Neufeld, 1902; Armstrong, 1931). The Quellung reaction can also be used for other encapsulated bacteria such as *Klebsiella*, *Neisseria* and *Haemophilus* spp.

In the field of immunology, research on *S. pneumoniae* has been important for understanding humoral immunity and phagocytosis. This includes ending the widely held belief in immunology that only proteins can act as antigens, when the capsular polysaccharide (CPS) was demonstrated to be responsible for pneumococcal serological activity (Heidelberger and Avery, 1923) and to elicit strain specific protection in mice in the 1920s (Heidelberger, 1927).

S. pneumoniae is central to the most critical discovery in molecular biology: the establishment of DNA as the genetic material. Frederick Griffith noticed there were two types of pneumococci on blood agar (BA) plates: smooth (S) colonies, which had a polysaccharide capsule and caused fatal infections in laboratory animals, and rough (R) colonies, which neither possessed a capsule nor caused fatal infections. Griffith showed that, R strains, which do not naturally revert to the S form *in vivo*, could acquire the S phenotype by injecting the R strain into mice along with heat-killed S strain. The isolated colonies not only produced a polysaccharide capsule but the polysaccharide was exactly the same as the polysaccharide of the heat-killed S strain (Griffith, 1928). This ability to take up genetic material from the environment is known as natural transformation. Carrying on from Griffith's work, James Alloway showed that transformation could be

performed *in vitro* (Alloway, 1932). Later, Oswald Avery and his colleagues proved that DNA was the transforming principle by showing that treatment with DNase prevented transformation occurring *in vitro* (Avery, 1944).

1.2 Pneumococcal Disease

The pneumococcus is a niche bacterium of the human nasopharynx with over 90 different capsular serotypes identified. Carriage rates vary amongst studies, ranging from less than 10% to up to 100% of a population, with higher levels of carriage seen in young children and in low-income communities (Crook *et al.*, 2004). Although it is normally carried asymptomatically, the organism is able to translocate from this niche to other host tissues, causing a spectrum of diseases referred to as invasive pneumococcal disease (IPD) including pneumonia, sepsis, and meningitis. In addition, *S. pneumoniae* also causes less serious, but highly prevalent infections such as OM and sinusitis. It is the leading cause of OM and IPD (Klein, 1994; Ostroff, 1999), and IPD is the major bacterial cause of death worldwide (Ostroff, 1999). In developed countries, the mortality rate of adult pneumococcal pneumonia is estimated to be 10-20%, escalating to over 50% in high risk groups (WHO, 1999a). Pneumococcal disease particularly affects children under the age of 5 years, adults over the age of 60 years, immunocompromised individuals, and people with chronic underlying medical conditions (Butler, 2004). The highest levels of incidence of pneumococcal disease are seen in indigenous populations and developing countries (Fedson and Scott, 1999). In fact, Australian Aboriginal children have the highest rates of IPD in the world (Forrest *et al.*, 2000). The total numbers of deaths are highest in the developing world with the World Health Organization (WHO) estimating that between 700,000 to one million children under the age of 5 years alone die from IPD each year (WHO, 2005).

Prior to the advent of antibiotics, the lethality of the pneumococcus was harrowing with the mortality rate of IPD over 50% (Ostroff, 1999). The discovery of penicillin and other antibiotics in the last century has undoubtedly revolutionised the treatment of pneumococcal infections, saving millions of lives. Unfortunately, however, the impact of antimicrobials has been dramatically reduced in recent times due to an astonishing increase in resistance and the rapid spread of multi-resistant clones. Indeed, it has been reported that up to 30% of strains show resistance to multiple drug classes, and treatment failures in meningitis, OM and pneumonia have been attributed to antimicrobial resistance (Lynch and Zhanel, 2005). Being naturally transformable, the pneumococcus can easily acquire resistance genes from other pneumococci, as well as other streptococcal species, in its environment (Reichmann *et al.*, 1997). Antibiotic resistance is an even bigger problem in developing countries due to predisposing factors, such as high incidences of carriage and disease, inappropriate use of drugs, poor quality generic drugs, and the quick spread of antibiotic resistant strains as a result of overcrowding and poor sanitation (Kunin, 1993).

Due to the aforementioned shortcomings of antibiotics, global efforts are now largely focused on vaccination against pneumococcal diseases. Currently, there are two types of vaccines licensed for human use. One of these is a 23-valent vaccine consisting of purified capsular polysaccharide (CPS) preparations from the 23 most common invasive serotypes. However, CPS vaccines are poorly efficacious in the groups most at risk from pneumococcal disease, such as young children and the immunocompromised. Furthermore, CPS is a T-cell independent antigen and therefore, does not elicit immunological memory (WHO, 1999b). In order to increase the immunogenicity of the polysaccharide, it has been coupled to protein to create conjugate vaccines. The first licensed conjugate vaccine, Prevenar™, covered seven of the most common disease associated serotypes in the developed world; namely types 4, 6B, 9V, 14, 18C, 19F, and 23F (Paton, 2004). However, recently a 10-valent vaccine, Synflorix™, has been licensed,

which covers serotypes 1, 5 and 7F in addition to the aforementioned serotypes (Prymula and Schuerman, 2009). There are however, a number of shortcomings for conjugate vaccines. Firstly, they are very expensive. For example, Prevenar™ cost AU$90 per dose, and requires three doses, making it unsuitable for developing countries (Butler, 2004). Moreover, due to their low serotype coverage and temporal and geographical differences in serotype prevalence, conjugate vaccines may not provide good coverage of common serotypes in some populations. For example, in indigenous Australian communities, Prevenar™ only covers 58% of the most common disease-causing serotypes (Krause *et al.*, 2000). Even in populations with higher coverage, the phenomenon of serotype replacement, where carriage and disease by vaccine strains is replaced by non-vaccine strains, is a concern. Studies worldwide have indicated a decrease in disease by vaccine serotypes but unfortunately, also the expansion of non-vaccine serotypes causing disease in children vaccinated with Prevenar™ (McEllistrem *et al.*, 2003; Bogaert *et al.*, 2005; Beall *et al.*, 2006, Singleton *et al.*, 2007). Furthermore, "vaccine escape" has been reported, whereby a strain undergoes capsular switching to change from a vaccine serotype to a non-vaccine serotype, (McEllistrem *et al.*, 2003, Brueggemann *et al.*, 2007). In addition to serotype replacement, the impact of vaccination on antibiotic resistance is another area that has been of interest and compared to serotype replacement, the reports have been more positive due to the prevalence of antibiotic resistance in the vaccine serotypes. Spain has experienced a decrease in antibiotic resistance following the introduction of Prevenar™, including a decrease in multidrug resistant isolates, which has been attributed to the replacement of antibiotic resistant strains of vaccine serotypes with antibiotic susceptible strains of non-vaccine serotypes (Ardanuy *et al.*, 2009; Perez-Trallero *et al.*, 2009). Canada is another example, where a reduction in both erythromycin and penicillin resistance has been observed (Tyrrell *et al.*, 2009). However, the expansion of antibiotic resistant pneumococci of non-vaccine serotypes has also been reported (Huang *et al.*,

2009), including the switch of one strain from a vaccine serotype to a non-vaccine serotype together with the gain of penicillin resistance, which is an obvious concern for the future (Brueggemann *et al.*, 2007).

Due to the limitations of CPS-based vaccines, attention has focused on the development of a completely protein-based vaccine. A protein-based vaccine would be cheaper than a conjugate vaccine while retaining the advantage of being a T-cell dependent antigen and could be formulated to cover all serotypes thus solving the dilemma of serotype replacement (Paton, 2004). Without the development of new strategies to prevent pneumococcal disease it is likely that the already enormous rates of mortality and morbidity will only worsen. The most adversely affected will be the most vulnerable: young children, the elderly, patients with underlying medical conditions, as well as the socio-economically disadvantaged.

1.3 Pneumococcal Virulence

In order to solve the aforementioned problems, it is critical to continue to investigate how the pneumococcus causes the various ailments associated with it. Although much research has already been undertaken on *S. pneumoniae*, there is still a large amount that remains to be elucidated since the pneumococcus is a complex organism with a variety of factors that come into play during disease. The capsule is the most well-established factor (Figure 1.1), and is a *sine qua non* of virulence with certain serotypes believed to have a higher association with disease or carriage. However, large scale epidemiological studies are indicating that genotype (that is non-capsular factors) plays a significant role as well. Indeed, a growing number of proteins have been demonstrated to play an integral part in virulence and more are expected to be identified as the number of sequenced pneumococcal genomes increases. Although there are some conserved

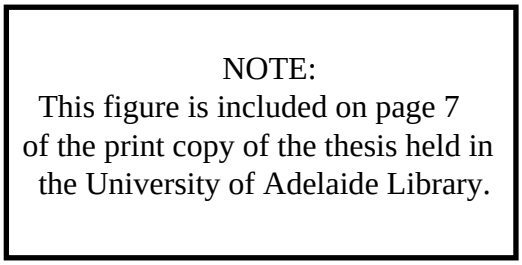


Figure 1.1. Schematic diagram of the pneumococcal surface.

Major surface components and some important virulence factors are shown. Autolysin (LytA) and dimers of PspA and PspC are shown associated with phosphorylcholine residues of teichoic and lipoteichoic acids. Other important virulence factors include capsular polysaccharide, neuraminidase A (NanA) and pneumolysin (Ply). The pneumococcal capsule and virulence proteins are discussed in Sections 1.3.1 and 1.3.3, respectively.

Adapted from Briles *et al.* (1998).

virulence proteins that have been widely characterised (Figure 1.1), an understanding of the role that genes with a variable distribution have in virulence is just beginning to emerge. Finally, morphological characteristics of a strain can influence virulence throughthe phenomenon of phase variation, and there are still outstanding questions as to the role of planktonic and biofilm growth in specific disease states.

1.3.1 The pneumococcal capsule

There are over ninety pneumococcal serotypes which are distinguished by the structure of the capsule. The capsule was the first virulence factor identified in the pneumococcus and is the target for the current vaccines. The capsule is strongly antiphagocytic and essential for colonisation of the nasopharynx and full virulence in IPD, although there have been reports of unencapsulated strains being isolated from superficial infections, such as conjunctivitis and OM (McEllistrem *et al.*, 2003; Martin *et al.*, 2003; Hanage *et al.*, 2004; Hanage *et al.*, 2006).

The enzymes for biosynthesis of the capsule are essentially encoded by the capsule locus, which can vary from only a few genes such as in serotype 3 to numerous open reading frames as found in serotype 2. The majority of serotypes use a method similar to the production of lipopolysaccharide (LPS) O-antigen with the following steps: initiation on a lipid acceptor, assembly of repeating sugar units on the cytoplasmic side of the membrane using activated nucleotide sugar precursors, translocation of the repeating unit across the membrane followed by polymerization of the nascent capsular polysaccharide chain and finally attachment of the capsular polysaccharide chain to the peptidoglycan of the cell wall (Yother, 2004). Two homologues from LPS O-antigen synthesis, which are the translocator Wzx and the polymerase Wzy, are required (Paton and Morona, 2006). On the other hand, serotypes 3 and 37 use an alternative synthesis method, which is

comparable to the synthesis of the hyaluronic acid capsule of *S. pyogenes* (Paton and Morona, 2006). This method is often referred to as the Wzy-independent method because instead of the Wzy polymerase, it utilises a synthase that is homologous to β-glycosyltransferases involved in the synthesis of a variety of polysaccharides, including hyaluronic acid, chitin and cellulose, in other bacteria and plant species (Dillard *et al.*, 1995; Llull *et al.*, 1999). Serotype 37 is unusual because the synthase gene is located outside of the capsule locus (Llull *et al.*, 1999). The synthase catalyses the formation of glycosidic bonds when activated nucleotide sugar precursors are added to the nascent capsular polysaccharide chain. Both serotypes 3 and 37 produce large amounts of capsule (Kauffmann *et al.*, 1940), which is characterized by large, mucoid colonies on blood agar (Knecht *et al.*, 1970). Serotype 3 capsular polysaccharide has been demonstrated to be noncovalently attached to the cell wall (Sorensen *et al.*, 1990). Studies into serotype 3 capsule synthesis have suggested that the capsular polysaccharide stays associated with the cell by interacting with either the synthase or the lipid acceptor on which synthesis of the capsular polysaccharide chain is initiated (Cartee *et al.*, 2005; Forsee *et al.*, 2000).

1.3.2 Serotype, sequence type and genome

In recent years, there has been much interest in the relationship between serotype, as well as genotype, and disease or carriage. Research exploring the role of genotype in virulence and carriage has been aided by the development of a pneumococcal multilocus sequence typing (MLST) scheme and corresponding internet database, which has provided researchers with a tool to assign isolates a standard sequence type (ST) and confidently compare results from different studies (Enright and Spratt, 1998). Although this database is available, care must be taken when assessing the relationship between serotype/genotype and disease since factors such as isolate sample size or underlying medical conditions of

patients can skew results. Indeed, the study by Alanee *et al.* (2007) suggested that host factors, such as age and immunosuppresion, are better predictors of morbidity and mortality than bacterial factors. However, it should be pointed out that this study looked only at the serotype of strains and did not take genotype into account (Henriques-Normark *et al.*, 2008). Another study by Chen *et al.* (2009) suggested that host factors and antibiotic resistance of bacteria were better indicators of mortality than serotype or genotype, although they felt their sample size was too small to make a conclusive judgment. Overall the studies to date suggest that there is a complex interaction between serotype, genotype and host environment, which influences disease outcome.

The use of mathematical modeling which compares the number of invasive and carriage isolates of a serotype or genotype to a reference has been used by Brueggemann *et al.* (2003) to determine the invasiveness of specific serotypes and genotypes. Through these calculations, certain serotypes were shown to have a high association with invasive disease and these serotypes usually had a statistically significant low association with carriage. In contrast, serotypes with a high association with carriage tended to exhibit low invasiveness. However, when the Prevenar™ serotypes, which are all predominant in IPD, were analysed, their association with invasiveness was found to vary. For example, serotype 1 had a high association with invasiveness, but serotype 23F had a low association and its prevalence in IPD cases was attributed to its high rates of carriage.

Certain STs have also been suggested to have a greater association with disease or carriage than others. As examples, STs 156, 482, 191, 124 and 138 have been found to have a significant association with IPD, while the STs 485 and 62 have been found to have high associations with carriage (Hanage *et al.*, 2005). However, STs often associate with the one serotype and in this study ST156 was the only ST to be found in more than one serotype. The invasive associated STs belonged to serotypes that had significant associations with IPD, apart from one ST156 isolate expressing a serotype 19F capsule.

Moreover, the serotypes of the carriage associated STs demonstrated a lower association with invasive disease. Therefore, it can be difficult to distinguish between the influence of genotype and serotype.

An ST or clonal complex can dominate within a serotype, for example, ST180 was seen in 18/20 and 21/33 serotype 3 isolates from two European studies, while ST62 was observed in 7/9 and 12/12 serotype 11A isolates in the same studies and 12/19 isolates in a third European study (Brueggemann *et al.*, 2003; Sjostrom *et al.*, 2006; Hanage *et al.*, 2005). Furthermore, ST distribution can depend on geographic location. For instance, in a large scale genotype analysis of serotype 1 clinical isolates, three distinct genetic lineages were identified (lineages A, B and C), which had different geographic distributions. Isolates from lineage A, which includes ST306 and ST227, were all from Europe and North America, while isolates from lineage cluster B, which includes ST217, were predominantly from Africa or Israel, and most isolates of lineage C, which includes ST615, were from Chile (Brueggemann and Spratt, 2003). This adds complexity to epidemiological studies when assessing serotype and genotype as illustrated by the studies of Sjostrom *et al.* (2006) and Leimkugel *et al.* (2005). In the former study, which investigated invasive isolates in Sweden, all isolates of serotype 1 belonged to either ST306 or ST227, which belong to lineage A described by Brueggemann and Spratt (2003) and only differ by only one MLST allele, and additionally all the isolates within serotype 7F belonged to ST191. The authors made the interesting observation that both these serotypes, which are of high invasive potential, produced less severe IPD compared to other serotypes in this study with low invasiveness. Still, it is difficult to resolve whether the lack of disease severity associated with serotypes 1 and 7F is due to serotype alone due to the relatedness of isolates within the serotypes. In contrast, the study by Leimkugel *et al.* (2005) was an analysis of pneumococcal isolates from a pneumococcal meningitis outbreak in Ghana with a high case-fatality rate of 44.4%. 76% of pneumococcal isolates

belonged to serotype 1 and when these serotype 1 isolates were analysed by MLST, they were found to be clonally related and from lineage B described by Brueggemann and Spratt (2003). Therefore, although Sjostrom *et al.* (2006) found that their serotype 1 isolates had a low case-fatality rate, the serotype 1 isolates of Leimkugel *et al.* (2005) from a different genetic lineage were classified as hypervirulent. However, the invasive phenotypes of the serotype 1 isolates from the two studies are difficult to compare directly because one study was conducted in a developed European country and the other in a developing country of sub-Sahara Africa and therefore, there are additional factors, such as differences in host genetic background and exposure to risk factors associated with IPD, that need to be considered.

Comparisons of capsule switch variants may help resolve the role of serotype and genotype. For example, clinical isolates of ST199 have been demonstrated to have a greater association with invasive disease when expressing a serotype 19A capsule compared to a serotype 15B/C capsule, and the serotype 9V variant of ST162 also appeared to be more invasive than the serotype 19F variant, although the difference was not statistically significant (Brueggemann *et al.*, 2003). The effects of capsule switching conducted *in vitro* have also been investigated. In the study by Kelly *et al.* (1994), strains of serotype 5, 2 and 6B were switched to serotype 3 using the capsule from a moderately virulent donor. The moderately virulent serotype 2 strain retained its virulence, but the relatively avirulent serotype 6B strain increased in virulence, although not to the level exhibited by the donor serotype 3 strain. Surprisingly, the highly virulent serotype 5 strain became avirulent. There are several explanations for these observations. One suggested by Kelly *et al.* (1994) is that expression of the serotype 5 capsule itself is almost sufficient for virulence and as such serotype 5 strains only require a few virulence factors, while the serotype 3 capsule is less virulent, and therefore, serotype 3 strains require more additional virulence factors, which are absent in the serotype 5 background. This also helps explain

why the serotype 2 strain of similar virulence to the type 3 donor strain did not change in virulence after the capsule switch (Kelly *et al.*, 1994). The relatively avirulent serotype 6B strain converted to the serotype 3 must be lacking some virulence factors found in the donor serotype 3 strain since it did not reach the same level of virulence demonstrated by the donor (Kelly *et al.*, 1994). Another explanation suggested by (Obert *et al.*, 2006) is that serotypes may have genetic requirements based on the unique physiological properties of each capsule type. For example, expression of a capsule highly resistant to phagocytosis may necessitate extra adhesins, whereas low resistance to phagocytosis would result in a greater requirement to counteract the host defence (Obert *et al.*, 2006). Whatever the reason for this phenomenon, it appears that although capsule switching may well occur in the natural environment of the pneumococcus, the impact this has on virulence can be influenced by genotype. This highlights the fact that although serotype is a useful tool to group *S. pneumoniae*, it certainly is not the only factor which determines virulence or ability to cause disease.

Although serotyping has its limitations, genotyping is also limited since MLST is restricted to the sequences of seven housekeeping genes, and as such there may be genetic differences between clones that cannot be detected through MLST alone. Some STs exhibit greater genetic differences between isolates than others. For example, in one study, there were no genetic differences detected between isolates of ST306 (serotype 1), but thirty nine genetic differences were detected between isolates of ST176 (serotype 6B) by DNA microarray analysis (Dagerhamn *et al.*, 2008). The genetic differences between strains are believed to be due to recombination rather than mutation (Dagerhamn *et al.*, 2008). Evolution through recombination has been suggested to be restricted in some strains due to non-transformability (Henriques-Normark *et al.*, 2008) and therefore, the transformability of a clonal cluster may reflect how much genetic variation is seen between clones. ST306 of serotype 1, in which no genetic differences were detected between

clones through microarray (Dagerhamn *et al.*, 2008), was found not to be able to be transformed using either CSP1 or CSP2 (Henriques-Normark *et al.*, 2008). In addition, isolates from the ST180 lineage of serotype 3, which only exhibited four genetic differences that were scattered across the genome in the same microarray study (Dagerhamn *et al.*, 2008), exhibit low transformability (Henriques-Normark *et al.*, 2008). There is a high degree of variation seen between genomes of clonally unrelated pneumococci, even within the one serotype (Dagerhamn *et al.*, 2008; Hiller *et al.*, 2007). After the coding sequences of 17 sequenced pneumococcal strains were grouped into orthologous gene clusters, only 46% were conserved across all strains (Hiller *et al.*, 2007). Approximately 20-30% of the genome of each strain was attributed to accessory genes not part of the core genome. Usually these genes are found within regions known as accessory regions or regions of diversity, whose distribution varies across strains with some present in multiple strains but others in only a few (Dagerhamn *et al.*, 2008; Hiller *et al.*, 2007). In addition, accessory genes identified within a clonal complex may not be present in every clone (Pettigrew *et al.*, 2006; Hiller *et al.*, 2007; Dagerhamn *et al.*, 2008; Blomberg *et al.*, 2009). When 25 accessory genes were analysed in clonally related isolates through DNA microarray analysis, it was found that only two genes at the most, differed in clonally related isolates (Dagerhamn *et al.*, 2008). There were at least 4 differences in the carriage of accessory genes between isolates that were not clonally related.

The genes encoded by accessory regions have a range of functions. A few of these encode putative virulence factors, such as adhesins, but the majority appear to encode for proteins involved in metabolism (Dagerhamn *et al.*, 2008). Another group of genes are due to phage insertion and are often found to be unique to one strain studied (Hiller *et al.*, 2007). A few accessory regions have been identified that are distributed differently between IPD and noninvasive isolates and have been suggested to be linked with IPD, as well as carriage, (Obert *et al.*, 2006; Blomberg *et al.*, 2009). However, it is difficult to

identify an accessory region that is consistently present in all IPD associated isolates and absent in every carriage associated isolate, and many of these regions may be redundant (Blomberg *et al.*, 2009). Despite the inconsistency in the conservation of specific accessory regions in invasive isolates, the five most invasive clonal complexes were seen to possess from four to seven out of seven sugar phosphotransferase systems (PTSs) of known accessory regions, while in comparison, four isolates of medium to low invasiveness only possessed one or two of these accessory regions, which could give the highly invasive isolates an important advantage in the bloodstream where carbohydrate metabolism is likely to be of importance (Blomberg *et al.*, 2009).

The number of accessory genes identified is increasing as more genomes are sequenced. At the start of this project only three complete pneumococcal genomes were publicly available but since then, the number has expanded considerably to over 10. Based on the level of variation between genomes sequenced to date, it has been proposed that the genetic pool of *S. pneumoniae* contains over 5,000 orthologous clusters (Hiller *et al.*, 2007). It is estimated that over 140 pneumococci need to be sequenced to cover 90% of the genetic pool in order to include genes with a low population frequency (Hiller *et al.*, 2007). To date, the publicly available genomes represent a wide variety of serotypes and STs, although there are a few examples of sequencing projects on related strains, such as D39 and R6, as well as OXC141 and SP3-BS71, which are both serotype 3 strains of ST180. Therefore, it is important that a wide range of strains are sequenced in order to cover the genetic pool, and for this depth is needed in addition to breadth. Comparisons of genomes within both STs and serotypes are integral to gaining a complete understanding of pneumococcal carriage and pathogenesis, and with recent advances in technology, such as the development of next generation sequencing, this has become easier to achieve.

1.3.3 Virulence proteins

Due to the various pneumococcal genome sequence projects occurring, a large number of potential virulence genes are being identified. However, there are a number of virulence proteins, which were identified long before the first pneumococcal genomes were sequenced and have been studied in-depth over a number of years. These proteins have been found to be carried by virtually all pneumococci to date and have a range of functions. They include the pneumococcal surface protein A (PspA), pneumococcal surface protein C (PspC), neuraminidase A (NanA), pneumolyisn (Ply) and autolysin (LytA).

1.3.3.1 Pneumococcal surface protein A (PspA)

PspA protects the pneumococcus from clearance by the host complement system (Tu *et al.*, 1999). Moreover, PspA provides protection against apolactoferrin, which is bactericidal against the pneumococcus, by binding to the iron-bound form lactoferrin (Shaper *et al.*, 2004). PspA belongs to a family of proteins known as choline binding proteins (CBP), which attach to the cell wall by binding to phosphorylcholine on teichoic acid or lipoteichoic acid (Yother and White, 1994; Swiatlo *et al.*, 2004). The *pspA* gene is highly polymorphic and can be grouped into three different families with each family divided into two clades (Brandileone *et al.*, 2004). An association between clade and ST has been proposed, but there is no apparent association between clade and serotype (Rolo *et al.*, 2009).

1.3.3.2 Pneumococcal surface protein C (PspC)

PspC is also known as choline binding protein A (CbpA), *S. pneumoniae* secretory

IgA binding protein (SpsA), C3-binding protein A (PbcA) or factor H-binding inhibitor of complement (Hic), which are all variants of PspC that have been isolated from different pneumococcal strains (Rosenow *et al.*, 1997; Hammerschmidt *et al.*, 1997; Cheng *et al.*, 2000; Janulczyk *et al.*, 2000; Iannelli *et al.*, 2002). Like PspA, it has been shown to be highly polymorphic with 11 major groups identified by Iannelli *et al.* (2002). The majority of PspC molecules attach to the cell wall via a choline binding domain but there are some which attach via an LPxTG motif, such as Hic in *S. pneumoniae* A66 (Janulczyk *et al.*, 2000; Iannelli *et al.*, 2002). PspC has been shown to be able to bind to Factor H, which protects host cells from the alternative complement pathway, and therefore, PspC may prevent complement mediated opsonisation (Dave *et al.*, 2004; Janulczyk *et al.*, 2000; Quin *et al.*, 2005). However, a broader role in virulence has been established as PspC has been shown to play a role in adherence and invasion of host cells (Rosenow *et al.*, 1997; Zhang *et al.*, 2000). Through its capacity to bind to the secretory component of immunoglobulin A (IgA), PspC binds to the polymeric immunoglobulin receptor that transports IgA and this may be important for transport across respiratory epithelial cells (Hammerschmidt *et al.*, 2000; Zhang *et al.*, 2000).

1.3.3.3 Neuraminidase A (NanA)

NanA cleaves terminal sialic acid residues from glycolipids, glycoproteins and oligosaccharides on host cell surfaces as well as host body fluid (Camara *et al.*, 1994). This is believed to expose host cell receptors and therefore is important in colonization. Moreover, the breakdown of host glycoconjugates provides a carbon source for the pneumococcus (Burnaugh *et al.*, 2008). NanA may also give the pneumococcus a competitive advantage over *Neisseria meningitidis* and *Haemophilus influenzae*, which are also present in the same host niches, because NanA has been shown to desialyate the

lipopolysaccharide of these bacteria, which may make them more susceptible to complement-mediated clearance (Shakhnovich *et al.*, 2002). As with *pspA* and *pspC*, there are a variety of *nanA* alleles (King *et al.*, 2005). In addition, a further two neuraminidases have been isolated in the pneumococcus, NanB and NanC, which have been shown to be present in 96% and 51% of pneumococci, respectively (Berry *et al.*, 1996; Tettelin and Hollingshead, 2004; Pettigrew *et al.*, 2006).

1.3.3.4 Pneumolysin (Ply)

Ply is a pore-forming toxin, which is lytic to all animal cells that contain cholesterol in their membrane and at lower concentrations, is also involved in pathogenesis by inducing apoptosis (Dockrell *et al.*, 2001; Braun *et al.*, 2002), activating host complement, which diminishes serum complement levels thus significantly reducing serum opsonic activity (Paton *et al.*, 1984; Mitchell *et al.*, 1991) and inducing proinflammatory responses in immune cells (Malley *et al.*, 2003; Cockeran *et al.*, 2001a; Cockeran *et al.*, 2001b). There are at least 15 different alleles of *ply* and several of these are non-haemolytic (Jefferies *et al.*, 2007). Surprisingly, strains that encode a non-haemolytic Ply are still capable of causing IPD (Kirkham *et al.*, 2006). Furthermore, Kirkham *et al.* (2006) found that serotype 1 strains of ST306 isolated in Scotland produced non-haemolytic Ply but the Ply from serotype 1 strains from the closely related ST227 displayed normal levels of haemolysis. These two STs differ by only one MLST allele and were reported in the study by Sjostrom *et al.* (2006) described in Section 1.3.2.

1.3.3.4 Autolysin (LytA)

Autolysins are enzymes that degrade the peptidoglycan backbones of bacteria resulting in cell lysis (Lopez *et al.*, 1997). LytA is another choline binding protein (Holtje

and Tomasz, 1975; Briese and Hakenbeck, 1985), and is believed to play a role in pneumococcal pathogenesis by allowing the release of cell wall components that are highly inflammatory for the host (Tuomanen *et al.*, 1985a; Tuomanen *et al.*, 1985b; Majcherczyk *et al.*, 1999), as well as releasing virulence factors that contain no recognizable export sequence, such as Ply (Paton *et al.*, 1993). However, autolysin independent release of Ply has also been demonstrated (Balachandran *et al.*, 2001). The *lytA* gene is upregulated during competence (Mortier-Barriere *et al.*, 1998) and it is believed that the release of DNA through autolysis in a small subset of the pneumococcal population plays an important role in transformation (Steinmoen *et al.*, 2002). Two additional cell wall hydrolases have been described in the pneumococcus, LytB and LytC (Garcia *et al.*, 1999b; Garcia *et al.*, 1999a).

1.3.4 Phase variation

The phenomenon of phase variation, whereby bacteria exist in two colony morphology phenotypes (opaque and transparent) has been observed for pneumococci when viewed under oblique, transmitted light following growth on transparent agar media (Weiser *et al.*, 1994) [Figure 1.2]. An intermediate phase has also been reported, which has been poorly characterized compared to the other phases and is hypothesized to be an unstable form of the opaque variant resulting from the absence of selective pressure *in vitro* (Serrano *et al.*, 2006; Weiser *et al.*, 1994). Switching between these phases *in vitro* occurs at a frequency of 10^{-3} to 10^{-6} per generation depending on the strain (Weiser *et al.*, 1994).

The existence of phase variants has been linked to virulence potential, the opaque variant being more virulent than the transparent phenotype (Weiser *et al.*, 1994). This presumably is because opaque colonies produce more capsular polysaccharide (Kim and

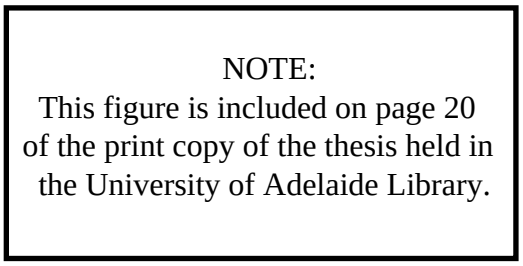


Figure 1.2. Pneumococcal colony opacity.

Single colonies of *S. pneumoniae* P62 (serotype 9v) are shown from Weiser *et al.* (1994) at 180× magnification following growth on THY/catalase agar. The colony on the left demonstrates the transparent phenotype, while the colony on the right demonstrates the opaque phenotype.

Weiser, 1998) and differ in their protein expression profile (Overweg *et al.*, 2000). Invasive isolates are found in the opaque phase more often than transparent phase with a ratio of 2:1, although this can depend on serotype (Serrano *et al.*, 2006). For example, invasive isolates of serotypes 1 and 23F were associated with the opaque phenotype more often than expected with the ratio of opaque to transparent being approximately 15:1 and 8:1 respectively (Serrano *et al.*, 2006). In contrast, invasive isolates of serotypes 3 and 14 had a lower association with the opaque phenotype because both had ratios of approximately 1:2 (Serrano *et al.*, 2006). Phase variation has also been demonstrated to have an impact on colonization in addition to virulence. Transparent bacteria have been shown to be more efficient than opaque bacteria at colonization in an infant rat nasopharyngeal model and furthermore, are more efficient at adhering to nasopharyngeal epithelial, lung epithelial, brain microvascular endothelial and middle ear epithelial cell lines *in vitro* (Kim and Weiser, 1998; Cundell *et al.*, 1995; Ring *et al.*, 1998; Li-Korotky *et al.*, 2008).

1.3.5 Physiological states

In addition to existing in two colony morphologies, pneumococci have recently been found to also exist in two physiological states, depending on growth conditions. *In vitro*, pneumococcal growth is either planktonic (in liquid media) or sessile (on agar or in

submerged biofilm) which is also characterized by differences in gene expression. *In vivo*, it is believed that pneumococci resemble planktonic growth during sepsis, but they are in a biofilm-like state during tissue infections, such as pneumonia and meningitis (Oggioni *et al.*, 2006). Furthermore, *S. pneumoniae* has been detected in the biofilms of patients with OM and chronic rhinosinusitis (Hall-Stoodley *et al.*, 2006; Sanderson *et al.*, 2006).

The work of Allegrucci *et al.* (2006) indicates that pneumococcal biofilms develop in distinct stages with protein expression varying over the course of development. The three main stages are (i) surface attachment of planktonic cells, (ii) formation of cellular aggregates, which is presumed to be due to the replication of the initial attached cells, and (iii) formation of the mature biofilm. The biomass and structure of the mature biofilms varies between strains and Allegrucci *et al.* (2006) hypothesized that this is may have an impact on virulence. For example, the strains of serotypes 3 and 23, which are serotypes associated with high mortality in IPD (Sjostrom *et al.*, 2006), were observed to have very structured biofilms with large, distinct microcolonies and pronounced water channels, but three strains of serotypes with a lower mortality rate had comparably less structured biofilms.

There is limited understanding of the genetic requirements of biofilm development. However, a range of choline binding proteins, including PspA, PspC and LytA, has been demonstrated to be essential for biofilm formation (Moscoso *et al.*, 2006). In addition, neuraminidase has been shown to be required for biofilm formation in *Pseudomonas aeruginosa* (Soong *et al.*, 2006), whose biofilms have been well-characterised, and a growing body of evidence is indicating the importance of NanA and NanB in pneumococcal biofilms. Initially, both *nanA* and *nanB* were shown to be upregulated in biofilms compared to planktonic growth (Oggioni *et al.*, 2006). More recently, a NanA inhibitor has been demonstrated to prevent biofilm formation (Parker *et al.*, 2009), while supplementation with sialic acid enhanced formation (Trappetti *et al.*, 2009). Furthermore, an interrupted *nanB* gene was identified in a screen of TIGR4 transposon mutants for inability to form biofilms (Munoz-Elias *et al.*, 2008). This study also identified 48 other genes that appear to be important for biofilm formation, including *pspC* and two sensor histidine kinases.

Capsule production is another factor important in biofilm formation. In fact, the

transposon mutagenesis study whereby *nanB* was identified was conducted using acapsular TIGR4 variants (Munoz-Elias *et al.*, 2008). Originally, transposon mutants of encapsulated TIGR4 were screened but only two genes could be identified, one which was a biofilm overproducing *cps4E* mutant in which capsule production had become disrupted. In addition, acapsular variants have been isolated from *in vitro* biofilms, which may explain why unencapsulated pneumococci are isolated from patients even though the capsule is an essential virulence factor (Waite *et al.*, 2001; Waite *et al.*, 2003; Allegrucci and Sauer, 2007; Allegrucci and Sauer, 2008).

It is important that we gain an understanding of biofilm formation because numerous studies of a wide variety of antibiotics and bacterial species, including *Streptococcus* spp., have indicated that bacteria in biofilms are more resistant to antibiotics compared to planktonic bacteria and thus present a problem for treatment of bacterial infections (Slinger *et al.*, 2006; Nickel *et al.*, 1985; Gristina *et al.*, 1987; Evans and Holmes, 1987; Larsen and Fiehn, 1996; Hall-Stoodley *et al.*, 2008). There have been a variety of mechanisms suggested for this increased resistance, including low metabolic activity or differential expression of resistance genes of bacteria within biofilms compared to planktonic growth, as well the rate of antibiotic penetration being reduced within biofilms, thus allowing the bacteria to be gradually exposed to the antibiotic and adjust by undergoing stress-induced metabolic or transcriptional changes (Walters *et al.*, 2003; Gillis *et al.*, 2005; Jefferson *et al.*, 2005). Moreover, it is believed that biofilms protect bacteria from host defences and are therefore important for survival in a hostile environment (Costerton, 1999). For example, *Staphylococcus* spp. biofilms have been reported to be resistant to phagocytosis, although penetrable by human leukocytes and antibodies (Leid *et al.*, 2002; Cerca *et al.*, 2006). Of course, the precise mechanism of host immune evasion is likely to depend on the bacterial species.

1.4 Otitis Media

Much of the above knowledge concerning pneumococcal virulence has been gathered from research focused on IPD, but pneumococcal otitis media (OM) is of major public health concern warranting as much research attention. The focus on IPD is understandable given the association with high mortality compared to OM. However, the very high morbidity associated with pneumococcal OM and the large impact it has in every society, but especially in socio-economically disadvantaged groups, such as remote Australian Aboriginal communities should not be underestimated. Unless action is taken, this impact will only grow in the future due to the ineffectiveness of current vaccine strategies and increasing antibiotic resistance. Thus, overlooking research on pneumococcal OM will have even larger associated public health costs.

OM occurs predominantly in infants and young children, and is believed to be due to an immature immune system and the anatomy of the Eustachian tube (Cripps *et al.*, 2005). It is one of the commonest childhood ailments and nearly all children have at least one episode of OM during his or her early childhood (Gates, 1998). Due to this high morbidity in children, OM is the number one medical reason for a child to visit a physician or receive antibiotics (Freid *et al.*, 1998; Gonzales *et al.*, 2001). Consequently, the health care costs associated with OM are high worldwide. As far back as 1996 it was suggested that OM costs the US $5 billion each year (Gates, 1996). The most recent study conservatively estimated the cost of acute otitis media (AOM) in children under 2 years of age alone to be $3.8 billion in the US annually, assuming no vaccination takes place, while the projected savings of Prevenar™ were $286 million per year (O'Brien *et al.*, 2009).

In addition to the significant economic cost, the social cost of OM is high. OM is the commonest cause of deafness in children (Kubba *et al.*, 2000) and has also been linked with long-term behavioural and cognitive problems (Bennett *et al.*, 2001). Pressure on the

tympanic membrane during OM can cause considerable pain for a child. The child may suffer from afflictions such as vertigo, tinnitus, headaches and vomiting as well. Children suffering from OM may misbehave at school, and their ability to learn may be affected due to pain or diminished hearing. Moreover, OM is stressful for parents and guardians, who have to attend a distressed or misbehaving child, and juggle work and other commitments with care. Further frustration occurs if the child suffers from recurrent bouts of OM (Klein, 2000). Therefore, while there are direct costs associated with OM that can be easily calculated (such as antibiotic treatment/ hearing aids/ surgery) there are many that cannot, such as family stress or people not reaching their full potential because their schooling was adversely affected by recurrent OM.

1.4.1 OM pathogenesis

The impact of OM will depend on the severity and recurrence of disease. OM classically falls into 3 categories depending on the symptoms and severity of disease (Coates *et al.*, 2002). Otitis media with effusion (OME) describes the presence of effusion in the middle ear in the absence of signs of suppurative infection. Effusion in the presence of signs of an infection is categorised as acute otitis media (AOM), of which there are 2 types: AOM without perforation, which is diagnosed by bulging of the tympanic membrane, and AOM with perforation, which is characterised by perforation of the tympanic membrane. When effusion discharges through a tympanic membrane perforation persistently for 6 weeks or more, the infection becomes known as chronic suppurative otitis media (CSOM).

A plethora of risk factors are associated with OM, including exposure to tobacco smoke, older siblings, genetic predisposition, attendance at group daycare and season. There is also a direct link between OM and poverty due to associated problems, such as

crowded living conditions, poor sanitation and inadequate medical care (Klein, 2000). Therefore, the issue of OM is compounded in developing countries. Furthermore, there are higher rates of complications from OM, such as meningitis, and one third of people in developing countries are believed to be affected by hearing impairment resulting from long-term OM infections (Berman, 1995). In cases of illiteracy, the effects of hearing loss are further amplified (Klein, 2000).

1.4.3 OM in Australian Aboriginal children

In developed countries, indigenous and low socioeconomic groups are more vulnerable to OM and its complications than other communities. In Australia, OM is seen at unacceptably high levels in the Australian Aboriginal population, particularly in remote communities. While there has been a substantial decrease in OM media in Maori and native North American populations over the last three decades, a similar improvement has not been reported in the Aboriginal population (Morris *et al.*, 2005). CSOM, which is normally rare in developed countries, affects 40% of children living in remote Aboriginal communities, which is ten times the rate considered to be a major public health problem by WHO (Coates *et al.*, 2002) [Figure 1.3]. The rates of AOM with perforation are also the highest published in the world (Leach and Morris, 2007). In one study conducted in Northern and Central Australia, it was estimated that 80% of children had some form of hearing impairment due to the prevalence of chronic OM infections (Morris *et al.*, 2005).

There are a variety of reasons why children in remote Aboriginal communities suffer from high levels of OM. They have significant exposure to many of the risk factors associated with OM such as poverty, parental smoking and malnutrition (Lehmann *et al.*, 2008). Issues such as overcrowding and poor washing facilities, as well as the high levels of persistent nasal discharge seen in children, contribute to the exposure to pathogens

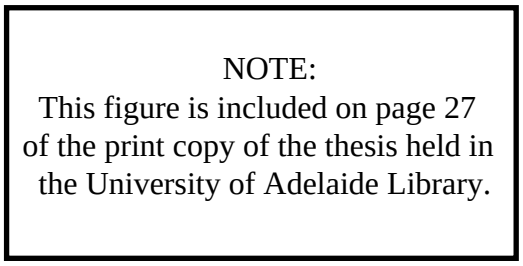


Figure 1.3. An Aboriginal child with CSOM.

Effusion, which has discharged through the perforated tympanic membrane, can be seen.

Photo provided by A.J. Leach.

(Leach and Morris, 2007). Hand contamination can facilitate the spread of pneumococci and *S. pneumoniae* is detectable on the hands of almost 40% of children living in remote Aboriginal communities, compared with 4% of children in urban childcare centres (Stubbs *et al.*, 2005). Aboriginal children have higher levels of pneumococcal carriage than non-indigenous children, even when children living in the same area are compared. For example, a study conducted on infants in the Kalgoorlie-Boulder region of Western Australia found that 49% of Aboriginal children carried *S. pneumoniae*, while only 25% of non-indigenous children were carriers (Watson *et al.*, 2006).

Babies as young as 3 weeks have been diagnosed with OM, including OME and AOM with perforation, and virtually all children in remote location are diagnosed with some form of OM within the first 3 months of life (Smith-Vaughan *et al.*, 2008). Early onset of OM is linked with the early acquisition of elevated loads of respiratory bacteria of which the likely source is heavily colonised older siblings and mothers (Smith-Vaughan *et al.*, 2008). It is thought that exposure of Aboriginal infants to high levels of bacteria or frequent exposure to new strains, results in constant inflammation that not only fails to clear the infection, but also leads to tissue damage. Non-Aboriginal children are usually exposed to low levels of bacteria, which can be cleared by the immune system so that the inflammatory response is down-regulated before high levels of tissue damage takes place (Coates *et al.*, 2002).

1.4.4 Control of OM: current and future strategies

The problem of OM is one that needs to be immediately addressed in Australian Aboriginal communities, but its treatment and prevention is complicated by the high levels of bacterial colonisation and early onset of disease. Furthermore, control of pneumococcal OM is an important issue to varying degrees in all communities, both in developing

countries, where the impacts of disease can be felt hardest, but also in developed countries, where the levels of OM are increasing. As with IPD, management of OM is impacted by antibiotic resistance and the problems encountered with the current vaccine strategies outlined in Section 1.2. However, there is also concern that OM is directly contributing to the problem of resistance through over-prescription of antibiotics to OM patients, which has lead to much discussion over the last few years in regards to the necessity of antibiotic treatment in OM and the benefits of a wait-and-see approach (Damoiseaux, 2005).

There is currently no vaccine that specifically targets OM. Whereas Prevenar™ can be used to help prevent infection by the pneumococcal serotypes included in the vaccine, it is less effective in preventing OM compared to IPD. The efficacy of Prevenar™ against IPD caused by vaccine serotypes is over 90% (Black *et al.*, 2000), but only about 67% against OM (Black *et al.*, 2000; Eskola *et al.*, 2001). Furthermore, since Prevenar™ was formulated with the aim of reducing the incidence of IPD, the serotypes chosen for the vaccine are common for IPD and therefore, serotypes, such as type 3, which are commonly associated with OM, were not included. This problem is being addressed by expanding the valency of conjugate vaccines, including the development of a 13-valent vaccine, Prevenar 13™, which includes serotype 3 (Scott *et al.*, 2007). However, it is essential for conjugate vaccines to not only include important OM serotypes but to be efficacious in preventing OM episodes (Prymula, 2009). Vaccine efficacy against AOM varies for individual serotypes. For example, Prevenar™ is 59% efficacious in preventing AOM caused by serotype 23F but only 25% efficacious against serotype 19F (Eskola *et al.*, 2001). Moreover, an 11-valent pneumococcal conjugate vaccine that included serotype 3 conferred absolutely no efficacy in preventing AOM episodes by this common OM serotype (Prymula *et al.*, 2006; Poolman *et al.*, 2009).

The introduction of Prevenar™ has resulted in only a 6-8.9% reduction in OM overall (Black *et al.*, 2000; Eskola *et al.*, 2001) with a 34% decrease in culture-confirmed

pneumococcal OM and a 33% increase in OM by non-vaccine serotypes (Eskola *et al.*, 2001). Nevertheless, there have been some positive results, such as a substantial decrease of up to 50% in recurrent OM and tympanostomy tube procedures following the introduction of Prevenar™ (Black *et al.*, 2000; Eskola *et al.*, 2001; Fireman *et al.*, 2003; Palmu *et al.*, 2004; Fletcher and Fritzell, 2007). Moreover the projected cost savings from reducing OM cases are greater than vaccine attributable savings from all other pneumococcal diseases combined (O'Brien *et al.*, 2009).

Although Prevenar™ has been shown to help prevent recurrent AOM in developed regions, in the case of Australian Aboriginal children, Prevenar™ has had no effect on OM (Mackenzie *et al.*, 2009). Not only are Australian Aboriginals burdened with OM caused by serotypes that are absent in Prevenar™, but their serotype distribution differs from the general population. For example, serotype 16F, which is not commonly found in OM in the general population (Hausdorff *et al.*, 2000; Watson *et al.*, 2007), is not only common in both carriage and OM in Aboriginal communities but has expanded in these communities since the introduction of Prevenar™ (Marsh *et al.*, 2007). Thus even though a 13-valent conjugate vaccine has been formulated, it is expected to be ineffective in Australian Aboriginal communities because their serotype distribution has not been specifically considered. However, the biggest complication to overcome with a vaccination strategy in Aboriginal communities is the problem of babies as young as 3 weeks contracting OM well before full immunization is possible. Immunisation of pregnant women to confer passive immunity in their babies is one method that has been suggested to help overcome this problem (Dunbar *et al.*, 2007). As the source of bacteria is believed to be from family members, methods of preventing transmission within families, such as access to improved hygiene facilities, are of vital importance as well (Smith-Vaughan *et al.*, 2008).

1.4.5 OM and the pneumococcus

The current limitations of pneumococcal vaccination strategies, such as serotype replacement, high costs, as well as multidrug class antibiotic resistance, have lead researchers to focus on the development of a protein based vaccine. It is imperative that future pneumococcal vaccines protect against both IPD and OM. In order to achieve this, a comprehensive understanding of the pathogenesis of pneumococcal OM is required. Unfortunately, our knowledge of how clonal properties and pathogenic factors of *S. pneumoniae* which contribute to OM is lacking compared to IPD. Although, there have been numerous epidemiological studies in the Australian Aboriginal communities examining carriage and severity of disease, there has been no characterisation of the causative strains beyond serotyping and some limited MLST (Morris *et al.*, 2005; Stubbs *et al.*, 2005; Watson *et al.*, 2006; Lehmann *et al.*, 2008).

1.4.5.1 Serotype and genotype

As with IPD, there are some serotypes that are isolated from infections more often than others, namely serotype 3 and serogroups 23 and 19 (Hausdorff *et al.*, 2000; Rodgers *et al.*, 2009). However, the exact serotype distribution can depend on a given population, such as the prevalence of serotype 16F seen in the Australian Aboriginal population (Marsh *et al.*, 2007). Nevertheless, the importance of serotype, as well as genotype, in OM is still not fully understood. It has been suggested that strains commonly isolated in OM cases may be due to their high levels of carriage rather than any special ability to cause disease (Hanage *et al.*, 2004). Indeed a study investigating the genetic relatedness of pneumococci in different sites of children suffering from AOM or OME conducted in the Netherlands found that the pneumococci isolated from the middle ear and nasopharynx in AOM patients were clonally related in 90% of children, while pneumococci from the

throat, adenoid tissue and middle ear were clonally related in 60% of OME patients (Tonnaer *et al.*, 2005). In another study, the serotypes and genotypes from the middle ears and nasopharynges of Finnish children were analysed mathematically for their association with AOM to investigate the aforementioned theory. The association between specific serotypes or genotypes with disease was slight compared to what is seen with IPD, but serotypes 19F and 23F, particularly ST37 strains of serotype 23F and ST309 strains of serotype 19F, as well as ST488 strains of serotype 6A, were suggested to have a high association with OM, while ST138 strains of serotype 6B was suggested to have a low association (Hanage *et al.*, 2004). More studies of this nature need to be conducted, including into CSOM, although this condition is more difficult to assess than AOM due to its chronic nature resulting in a considerable amount of time between onset and isolation of bacteria.

1.4.5.2 OM-associated proteins

In addition to increasing our understanding of serotype and genotype association with OM, more studies need to be conducted into understanding which proteins contribute to OM, as well as which are protective against OM when used as immunogens. The first protein shown to be an efficacious immunogen was PspA, when it was demonstrated to protect rats in an AOM model (White *et al.*, 1999). Later, immunisation of chinchillas with NanA was shown to result in a significant decrease in OME (Long *et al.*, 2004). Mutagenesis studies have also shown that NanA is required for colonisation and persistence in the nasopharynx and middle ear (Tong *et al.*, 2000). Antibodies against Ply have been observed to be at higher titres in children with AOM compared to healthy children (Rapola *et al.*, 2000), but any protective capacity in an OM model has yet to be established. Ply is believed to only play a modest role in middle ear inflammation with

cell wall components being the primary triggers of inflammation, suggesting an important role for LytA (Sato *et al.*, 1996). Nevertheless, Ply, as well as PspA, have been directly implicated in OM-associated hearing loss. Chinchillas infected with mutants of *ply* and *pspA* did not show signs of sensorineural hearing loss 28 days following inoculation of the middle ear, in contrast to animals infected with wild-type bacteria which demonstrated significant loss (Tsuprun *et al.*, 2008).

Until now, only two large scale studies have been conducted with the aim of identifying OM-associated genes (Pettigrew and Fennie, 2005; Chen *et al.*, 2007). In the first study, PCR-based subtractive hybridisation was performed using a serogroup 19 middle ear isolate, identifying only 2 genes that appeared to predominate in OM-causing strains (Pettigrew and Fennie, 2005). Both genes encoded hypothetical proteins, which were in less than 50% of middle ear isolates. In the more recent study, a genome scale screen was conducted by signature tagged mutagenesis to identify genes required for infection in a chinchilla OM model by a serotype 19F middle ear isolate of ST556 (Chen *et al.*, 2007). 169 genes were identified of which only 52 were required for nasopharyngeal colonisation in a murine model, which is suggestive of specific genetic requirements for OM. The genes identified encoded proteins with a variety of functions, including transporters, transcriptional regulators and proteases, as well as PspA, PspC and NanA. Interestingly, the two hypothetical proteins from the study of Pettigrew and Fennie (2005) were also identified by Chen *et al.* (2007).

1.4.5.3 Phase variation in OM

In addition to the aforementioned studies, the role of phase variation in OM also requires further investigation. In a chinchilla OM model, no difference in virulence between the two phase variants was observed, except when the animals were challenged

with the influenza A virus prior to inoculation, in which case opaque bacteria persisted in the middle ear for longer than the transparent variant of the same strain (Tong *et al.*, 2001). Moreover, in a rat OM model, opaque pneumococci were isolated from the middle ear at a 10 fold higher concentration than the transparent variant, with the transparent variant a more potent inducer of inflammation (Long *et al.*, 2003). Despite these studies, when transparent and opaque pneumococci were analysed *in vitro* using a human middle ear cell line under various conditions, including mimicking Eustachian tube obstruction, the transparent variant showed much greater growth, as well as adherence and killing of cells, than the opaque variant (Li-Korotky *et al.*, 2008). Subsequently, expression of the virulence genes *ply*, *lytA*, *spxB* (pyruvate oxidase) and *psaA* (manganese ABC transporter component) were compared between the two phase variants under the same conditions, and although basal expression in the inocula was comparable between the two phase variants, significant upregulation of the aforementioned genes was observed in the transparent variant compared to the opaque variant, particularly when Eustachian tube obstruction was simulated (Li-Korotky *et al.*, 2009). Therefore, the *in vitro* cell line work indicates the importance of the transparent phenotype in OM.

1.4.5.4 Biofilms in OM

Finally, the new research area of pneumococcal biofilms is expected to provide valuable information about OM as it is now becoming evident that biofilms play an important role in OM. Indeed, pneumococcal biofilms have been isolated from children with chronic OM (Hall-Stoodley *et al.*, 2006), as well as the chinchilla OM model (Ehrlich *et al.*, 2002), and pneumococci have been found to be able to grow as biofilms on tympanostomy tubes (Mehta *et al.*, 2006). Understanding the genetic requirements of biofilms may help us understand the genetic requirements of OM, although it is too

simplistic to expect the study of biofilms *in vitro* to give a complete understanding. This can be demonstrated by comparing the studies by Chen *et al.* (2007) and Munoz-Elias *et al.* (2008), who screened transposon mutants for ability to colonise the chinchilla OM model and form biofilms respectively. Chen *et al.* (2007) used a serotype 19F clinical isolate from the middle ear and Munoz-Elias *et al.* (2008) used TIGR4. In the former study, 169 genes were identified in the middle ear isolate of which 110 were annotated in TIGR4, but only 49 genes were identified in the latter study with just 4 genes common to both studies. Therefore, it appears that there are a greater number of genes required for OM compared to biofilm formation. In addition, Munoz-Elias *et al.* (2008) had issues with their biofilm model requiring them to use unencapsulated bacteria to achieve their results, while Chen *et al.* (2008) reported no problems using encapsulated bacteria. Therefore, the development of good *in vitro* biofilm models is another important step in elucidating the mechanism whereby *S. pneumoniae* causes OM.

1.5 Aims of the Project

To date there has been little research on pneumococcal OM isolates compared to IPD isolates. However, due to the prevalence of OM, which appears to be on the increase, there is a need to understand the pathogenesis of this disease, and in particular, to investigate whether there are pneumococcal proteins which have an association with OM. In addition, while there have been many epidemiological studies examining the problem of OM in remote Australian Aboriginal communities, there has been almost no molecular characterisation of middle ear isolates collected from these communities. This project aims to bridge this gap in our knowledge by genetically and phenotypically characterising a range of middle ear isolates from geographically and temporally diverse Aboriginal communities.

Accordingly, the specific aims of this project are:

1. To gain a greater molecular understanding of OM-associated strains from remote Australian Aboriginal communities at the genetic and phenotypic level.
2. To identify pneumococcal proteins that play a role in OM.
3. To investigate the distribution of these proteins in the pneumococcus.
4. To investigate the role of these proteins in pathogenesis.

Although Pettigrew and Fennie (2005) and Chen *et al.* (2007) have conducted studies with the aim of identifying OM-associated genes, this study is the first one using both multiple strains and techniques, including DNA microarray analysis, PCR-based subtractive hybridization and whole genome sequencing, to increase the likelihood of identifying genes that are positively associated with OM. It is also the first time that molecular biology techniques have been used to characterize OM isolates in-depth from remote Australian Aboriginal communities. This project should advance our understanding of the genetic basis of OM in the pneumococcus and identify potential vaccine targets to cover not only the general population but the Australian Aboriginal communities as well.

Chapter 2

Materials and Methods

2.1 General Chemicals and Reagents

The majority of chemicals used were purchased from Ajax Chemicals and were AnalR grade. Sodium deoxycholate (DOC) was purchased from BDH Biochemicals (Poole, England). Tris was purchased from Progen Industries (Queensland, Australia). Acrylamide and N,N,N',N'-tetramethyl-ethylene-diamine (TEMED) were purchased from BioRad Laboratories (California, USA). 5-bromo-4-chloro-3-indoyl-β-D-galacto-pyranoside (X-gal) was purchased from Astral Scientific (New South Wales, Australia). All enzymes unless stated otherwise, as well as deoxyribonucleoside triphosphates (dNTPs), were purchased from Roche Diagnostics (Mannheim, Germany).

2.1.1 Antibiotics

Erythromycin (Ery) was purchased from Roche. Ampicillin (Amp) was purchased from CSL (Victoria, Australia). Gentamicin (Gent) and optochin were purchased from Sigma-Aldrich (Missouri, USA). Diagnostic optochin disks were purchased from Oxoid (Hampshire, England).

2.2 Bacterial Strains and Growth

2.2.1 *S.pneumoniae* strains

Strains used in this study are listed in Table 2.1.

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

Strain	Description	Source (Reference)
D39	Capsular serotype 2	(Avery *et al.*, 1944)
D39-ABΔe5	D39 Δ*cpsAB ery*^{r}	(Morona *et al.*, 2006)
TIGR4	Capsular serotype 4	Tim Mitchell, University of Glasgow, Scotland (Aaberge *et al.*, 1995)
WU2	Capsular serotype 3	David Briles, University of Alabama, USA (Briles *et al.*, 1981)
WCH206	Capsular serotype 3 OM isolate	Andrew Lawrence, WCH (Women's & Children's Hospital, Adelaide, South Australia)
WCH207	Capsular serotype 3 OM isolate,	WCH
WCH208	Capsular serotype 3 OM isolate,	WCH
WCH209	Capsular serogroup 11 sinusitis isolate	WCH
WCH210	Capsular serogroup 11 sinusitis isolate	WCH
WCH211	Capsular serogroup 11 sinusitis isolate	WCH
WCH212	Capsular serogroup 11 sinusitis isolate	WCH

WCH213	Capsular serogroup 11 OM isolate	WCH
MSHR1	Capsular serotype 11A nasopharyngeal isolate	Amanda Leach, MSHR (Menzies School of Health Research, Darwin, Northern Territory, Australia)
MSHR2	Capsular serotype 11A nasopharyngeal isolate	MSHR
MSHR3	Capsular serotype 11A nasopharyngeal isolate	MSHR
MSHR4	Capsular serotype 11A nasopharyngeal isolate	MSHR
MSHR5	Capsular serotype 11A nasopharyngeal isolate	MSHR
MSHR6	Capsular serotype 11A nasopharyngeal isolate	MSHR
MSHR7	Capsular serotype 11A nasopharyngeal isolate	MSHR
MSHR8	Capsular serotype 11A nasopharyngeal isolate	MSHR
MSHR9	Capsular serotype 11A nasopharyngeal isolate	MSHR
MSHR10	Capsular serotype 11A nasopharyngeal isolate	MSHR
MSHR11	Capsular serotype 3 nasopharyngeal isolate	MSHR
MSHR12	Capsular serotype 3 OM isolate	MSHR
MSHR13	Capsular serotype 3 nasopharyngeal isolate	MSHR
MSHR14	Capsular serotype 3 OM isolate	MSHR
MSHR15	Capsular serotype 3 nasopharyngeal isolate	MSHR
MSHR16	Capsular serotype 3 nasopharyngeal isolate	MSHR

MSHR17	Capsular serotype 3 OM isolate	MSHR
WCH43	Capsular serotype 4 (ST205) blood isolate, virulent in mouse i.p. model	WCH
2663	Capsular serotype 11A (ST3019) lung isolate	Michael Watson, Path West, Perth, Western Australia Laboratory collection
3518	Capsular serotype 11A (ST62)	Path West
1	Capsular serotype 1 (ST304) nasopharyngeal isolate	MSHR
4	Capsular serotype 1 (ST227) nasopharyngeal isolate	MSHR
1861	Capsular serotype 1 (ST3079) blood isolate	MSHR
4496	Capsular serotype 1 (ST3018) blood isolate	WCH
3773	Capsular serotype 15B (ST199) tracheae isolate, antibiotic susceptible	Path West
4104	Capsular serotype 19A (ST199)	Path West
WCH16	Capsular serotype 6A, virulent in i.p. mouse model	WCH
WCH18	Capsular serotype 6B, Amp^s, non-virulent in mouse i.p. model at low challenge dose	WCH
WCH50	Capsular serotype 6, non-virulent in mouse i.p. model	WCH
L82016	Capsular serotype 6B	University of Alabama
WCH33	Capsular serotype 8, virulent in mouse i.p. model	WCH
WCH36	Capsular serotype 14	WCH
WCH39	Capsular serotype 14	WCH
WCH60	Capsular serotype 14, not virulent in mouse i.p. model	WCH

WCH61	Capsular serotype 14	WCH
WCH38	Capsular serogroup 19	WCH
WCH64	Capsular serogroup 19	WCH
WCH65	Capsular serogroup 19	WCH
EF3030	Capsular serotype 19F	University of Alabama
WCH67	Capsular serogroup 23, non-virulent in mouse i.p. model	WCH
WCH68	Capsular serogroup 23, non-virulent in mouse i.p. model	WCH
WCH69	Capsular serogroup 23, non-virulent in mouse i.p. model	WCH
WCH101	Capsular serogroup 23, Ery^s	WCH

2.2.2 Pneumococcal growth and storage

Pneumococci were routinely grown in Todd-Hewitt broth with yeast (THY, Todd-Hewitt broth [Oxoid] supplemented with 1% Bacto™ yeast extract [Beckton, Dickinson and Company, New Jersey, USA]), serum broth (SB, 10% (v/v) donor horse serum in nutrient broth [10 g/l peptone, 10 g/l Lab Lemco (Oxoid) and 5 g/l NaCl]), brain heart infusion (BHI) broth (Oxoid), C+Y (Lacks and Hotchkiss, 1960), or on blood agar (BA) plates [39 g/l Columbia base agar (Oxoid), 5% (v/v) defibrinated horse blood] at 37 °C in 95% air and 5% CO_2. Ery, Gent and optochin were added at concentrations of 0.2 µg/ml, 5 µg/ml and 5 µg/ml, respectively, where appropriate. For storage at –80 °C, *S. pneumoniae* were grown in SB and subsequently glycerol was added to 30%.

2.2.3 Optochin sensitvity

Strains were tested for optochin sensitivity to confirm that they were *S. pneumoniae* by plating on BA in the presence of a 5 µg optochin disk (Oxoid) and incubating for

approximately 16 h at 37 °C in 95% air/ 5% CO_2. Optochin sensitivity is an intrinsic characteristic of *S. pneumoniae*.

2.2.4 Quellung reaction

Pneumococcal serotype was assessed by the Quellung reaction. Bacteria were grown on BA O/N and inoculated into SB, then cultured to mid-exponential phase. 5 µl of culture was mixed with 4 µl of appropriate diagnostic pneumococcal typing serum (Statens Seruminstitut, Copenhagen, Denmark) and then examined microscopically for capsular swelling to confirm the production of type-specific capsule.

2.2.5 Determination of colony opacity phenotype

Pneumococcal colony opacity phenotype was determined by examining colonies, which had been grown on TH plates (Todd-Hewitt broth with 1.5% [w/v] Bacto agar) supplemented with 6300 Units (U) of catalase (Roche Diagnostics, Germany) per plate for approximately 40 h at 37 °C in 95% air/5% CO_2, under oblique, transmitted light, and designated “opaque” or “transparent” as described by Weiser *et al.* (1994).

2.2.6 Uronic acid assay

Overnight (O/N) BA plates were washed with PBS to create a bacterial suspension of A_{600} = 0.5. Ten ml of suspension was centrifuged at 3,175 × *g* for 30 min and about 9.5 ml of supernatant removed. The cells were resuspended in the remaining supernatant and centrifuged in a bench top microcentrifuge at maximum speed for 9 min to allow the removal of the remaining supernatant. Cells were resuspended in Tris-HCl pH 7.0 in a

final volume of 500 µl with Mg^{++} added to a final concentration of 1 mM. Cells were lysed by incubation at 37 °C for 30 min with 0.1% DOC. RNase A and DNase I (50 µg each) and 100 U mutanolysin were added and the lysates were incubated overnight at 37 °C. Samples were then treated with 50 µg of proteinase K and incubated at 56 °C for 4 h.

Fifty µl of CPS sample was added to 50 µl of MQ H_2O in duplicate. A duplicate blank (100 µl MQ H_2O) was used as a negative control. Samples were cooled on ice and 600 µl of sulphuric acid/ sodium tetraborate solution (12.5 mM $Na_2B_40_4$ in concentrated H_2SO_4) was added while the tubes were on ice. Samples were then vortexed and boiled for 5 min. Samples were cooled on ice before 10 µl of 0.5% NaOH (negative control) was added to one tube and 10 µl of 3-phenylphenol solution (0.15% phenylphenol in 0.5% NaOH) to the duplicate. Tubes were shaken immediately and 200 µl was transferred to a microtitre tray and the Absorption at 520 nm (A_{520}) was read using a Spectramax M2 spectrometer (Molecular Devices, California, USA).

2.2.7 Determination of growth curves

Glycerol stocks for use in growth curves were made by growing bacteria O/N on BA and inoculating into the same medium as to be used in the growth curve experiment. Bacteria were grown at 37 °C until the end of exponential phase, concentrated 10× in media supplemented with 30% glycerol. For the growth curve, cells were thawed and a sample diluted into the appropriate medium. An initial A_{600} reading was taken. Cells were incubated at 37 °C for 6-8 h with A_{600} readings taken at predetermined time points. Cultures were streaked onto BA and grown O/N to check for contamination.

2.2.8 Purple broth assay

Purple broth was made using BBL™ Purple Broth Base (Beckton, Dickinson and Company) according to the manufacturer's instructions. Bacteria were grown to midlog phase in THY and 2-3 ml pelleted by centrifugation at 3,175 × *g* for 20 min. Pellets were resuspended in 2-3 ml of purple broth supplemented with 1% (w/v) of the appropriate sugar. The negative control culture consisted of unsupplemented purple broth. The purple broth cultures were incubated at 37 °C in 95% air/ 5% CO_2. Samples (200 µl) were taken at regular intervals and centrifuged at 16,000 × *g* for 5 min. 100 µl of supernatant was used and read at A_{570} using a Spectramax M2 spectrometer.

2.2.9 Biofilm assay

Frozen stocks to seed the biofilms were prepared by growing transparent pneumococci in THY to A_{600} = 0.2. Glycerol was added to a final concentration of 30% (w/v) and stocks were stored at –80 °C.

To conduct the biofilm assay, frozen THY stocks were thawed and diluted 10-fold in 0.5× BHI in a Falcon® Microtest™ 96 well flat bottom tissue culture plate with each well having a final volume of 200 µl. One tissue culture plate was used for each time point with triplicate samples for each strain. Bacteria were grown at 37 °C in 95% air/ 5% CO_2 for up to 7 days with the medium replaced with fresh media every 24 h. At the appropriate time point, the medium was removed to a fresh microtiter plate and six 10-fold dilutions were performed in BHI/20% glycerol in order to assess planktonic growth. The biofilm plate was washed three times with 0.5× BHI and 100 µl of BHI/20% glycerol was subsequently added to each well. The plate was sealed with parafilm and adherent bacteria were resuspended by sonication in a sonicating waterbath (FXP20D, Unisonics, Sydney,

Australia) for 2 sec. The biofilm samples were serially diluted as described for the planktonic samples, and 10 µl of each dilution of the plantonic and biofilm samples were spotted onto BA and grown O/N at 37 °C in 95% air/ 5% CO_2.

2.2.10 Passaging of pneumococcal strains

Pneumococci from O/N BA plates were grown in SB until mid-exponential phase (A_{600} = 0.25) . The culture was diluted to approximately 10^5 colony forming units (CFU)/ml and 100 µl was administered intraperitoneally (i.p.) to CD-1 mice. At 24 h post-challenge, mice were sacrificed and approximately 50 µl of blood was recovered, which was serially diluted, plated onto BA plates (supplemented with antibiotics where appropriate) and incubated overnight at 37 °C in an atmosphere of 95% air and 5% CO_2. SB/glycerol stocks of the passaged bacteria were prepared and stored as described in Section 2.2.2.

2.2.11 *E. coli* strains, growth and storage

One Shot® TOP10 chemically competent *E. coli* cells (Invitrogen, Victoria, Australia) were used for cloning (Section 2.5.7). *E. coli* cells were grown in Luria-Bertani broth (LB) (10 g/l tryptone, 5 g/l yeast extract, 5 g/l NaCl), SOC medium (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20mM glucose) or on LB agar plates (LB with 15g agar per litre). Amp was added to the growth medium at 50 µg/ml where appropriate. For storage at –80 °C, bacteria were grown O/N on appropriate LB plates and resuspended in a solution of 1% peptone (Oxoid)/30% glycerol.

2.3 Oligonucleotide Primers

All oligonucleotide primers used in this study were purchased from Sigma-Genosys (New South Wales, Australia) and are listed in Table 2.2 along with their nucleotide (nt) sequences.The location of most primers are given in reference to the fully sequenced genomes of D39 (spd, GenBank accession number CP000410.1), TIGR4 (spn, GenBank accession number AE005672.3), G54 (spx, GenBank accession number CP001015.1), Hungary19A-6 (spv, GenBank accession number CP000936.1), JJA (sjj, GenBank accession number CP000919.1) or 70585 (snm, GenBank accession number CP000918.1). The GenBank accession numbers of G100 (complete coding sequences of *pspC* 4.1 and *pspC* 10.1 in strain G100) and the types 2 and 3 cps (capsular polysaccharide) locus are AF154033.1. AF026471.3 and U15171.1, respectively.

Table 2.2 Sequence and location or reference of oligonucleotide primers used in this study. AO, JM and RH denote primers that were kindly provided by A.D. Ogunniyi, J.K. Morona and R.M. Harvey, respectively.

Name	Sequence (5'-3')	Location (Reference)
NanA F	CTGATTGGAGAAAGGAGAGGGG	spd nt 1525713-1525692
NanA R	CAGGCTTGAAAACCTGCTATCTC	spd nt 1522435-1522413
NanA R2	GATAGGGTATAGGCATTTTGTCCT	spd nt 1523244-1523267
NanA R3	GTCTTCCCGTCGTCATCACTGT	spd nt 1523943-1523964
NanA F2	GTCATTTAAATGAACTGTAGTAAAAG	spd nt 1525649-1525624
NanA Insertion F	GGGACTTTTTGAGCAAAGATCTG	spd nt 1523226-1523204
NanA Insertion R	GCTGTTTTACCATTTGCCAATTGA	spd nt 1523076-1523099
NanA Deletion R	CTCTGTACCATTAACTCCCCCT	spd nt 1522679-1522700
PspA F	GTCCGCAGGCTTAAGCTTGCG	spd nt 128071-128091
PspA R	GGTTAGAGAAGACAAGGAAGAGC	spd nt 130521-130499
PspA F2	TAAAAATGACTATCAAAAAAGAGGTAA	spn nt 118390-118416

PspA R2	TTATCTTTGTTTCAAAATTAAAATGTCA	spd nt 130274-130247
PspA F3	CTTGGAAAACTTCTTGACAGCCT	WCH206 *pspA* nt 814-843
PspC F	GGACAGTGAAGCTTATGCTTGTC	spd nt 1997475-1997453
PspC R	GTGGAACATCTGGCGCTGCC	spd nt 130274-130247
PspC F2	CAGGGACTTATTACAACAAGTCG	MSHR17 *pspC* nt 373-395
PspC R2	GTTGCTGGACCTTGCTTCTTGT	MSHR17 *pspC* nt 899-878
PspC F3	CTAAACCAGAAACTCCGCAGCC	MSHR17 *pspC* nt 467-488
PspC R3	GCTTGTGGCTTGCTATTATCTGG	MSHR17 *pspC* nt 800-778
SKH63	TTTCTGGCTCATC(X/T)AACTGCTTTC	(Hollingshead *et al.*, 2006)
SKH52	TGGGGGTGGAGTTTCTTCTTCATCT	(Hollingshead *et al.*, 2006)
LSM12	CCGGATCCAGCGTCGCTATCTTAGGGG CTGGTT	(Swiatlo *et al.*, 1997)
PspC F2a	GCAGCTTTTGAGCAGTTTAAAAAAG	spv nt 2198762-2198738
PspC 3Fa	TGCGAAGTCTTCAGATTCTAGCG	MSHR5 nt 822-844
PspC 2Ra	GCTTGGTTAGAAGGTTGCTTGTC	spv nt 2193504-2193526
PspC R4	CACAAGGAAATCATCGCCAAACT	G100 nt 4566-4544
PspC 3Fb	CACCCCAACCAGGAGGACAAG	G100 nt 1658-1678
PspC F4a	GCAGACATGGATACGGTTGTTAC	G100 nt 2320-2342
PspC R3a	GCTCATCAAGCTCCTTAATAAGAG	spv nt 2059393-2059416
NanB F	GATTGCAAGTGGATTATCAGAAGG	spd nt 1517892-1517869
NanB R	CAATTCAGCTCCAAAATACCCTG	spd nt 1515686-1515708
NanC F	CAGGGAATATGTTGGTGACTTTTG	spn nt 1251742-1251719
NanC R	CCCAACTCTTGAGCAACTTCTTC	spn nt 1249236-1249258
aroE-up	GCCTTTGAGGCGACAGC	(Enright and Spratt, 1998)
aroE-dn	TGCAGTTCA(G/A)AAACAT(A/T)TTCTA	(Enright and Spratt, 1998)
ddl-up	TGC(C/T)CAAGTTCCTTATGTGG	(Enright and Spratt, 1998)
ddl-dn	CACTGGGT(G/A)AAACC(A/T)GGCAT	(Enright and Spratt, 1998)
gdh-up	ATGGACAAACCAGC(G/A/T/C)AG(C/T)TT	(Enright and Spratt, 1998)
gdh-dn	GCTTGAGGTCCCAT(G/A)CT(G/A/T/C)CC	(Enright and Spratt, 1998)
gki-up	GGCATTGGAATGGGATCACC	(Enright and Spratt, 1998)
gki-dn	TCTCCCGCAGCTGACAC	(Enright and Spratt, 1998)
recP-up	GCCAACTCAGGTCATCCAGG	(Enright and Spratt, 1998)
recP-dn	TGCAACCGTAGCATTGTAAC	(Enright and Spratt, 1998)

spi-up	TTATTCCTCCTGATTCTGTC	(Enright and Spratt, 1998)
spi-dn	GTGATTGGCCAGAAGCGGAA	(Enright and Spratt, 1998)
xpt-up	TTATTAGAAGAGCGCATCCT	(Enright and Spratt, 1998)
xpt-dn	AGATCTGCCTCCTTAAATAC	(Enright and Spratt, 1998)
Cps locus F	AGTAGAGCTTGCCGTCAAATTG	spd nt 310470-310491
Cps locus R	CCCAGTGATGGAAGGTAGAGTC	spd nt 337352-337331
Adaptor 1/*groEL* F	CCGGGCAG**GTAC**ATCATCGTAG	Adaptor 1 nt 35-44, ***Rsa*I site**, spd nt 1711620-1711607
Adaptor 2R/*groEL* F	CGGCCGAG**GTAC**ATCATCGTAG	Adaptor 2R nt 33-42, ***Rsa*I site**, spd nt 1711620-1711607
groEL F[AO]	CAGATGCCCGTTCAGCCATGGT	spd nt 1711531-1711510
groEL R[AO]	CAATCCCACGACGAATACCGATTG	spd nt 1711199-1711222
Primer 1	CTAATACGACTCACTATAGGGC	Clontech PCR-Select™ Bacterial Genome Subtraction Kit (California)
Nested Primer 1	TCGAGCGGCCGCCCGGGCAGGT	Clontech PCR-Select™ Bacterial Genome Subtraction Kit
Nested Primer 2	AGCGTGGTCGCGGCCGAGGT	Clontech PCR-Select™ Bacterial Genome Subtraction Kit
comE F[AO]	CAAGTGAGACTAGAGAGAATATTGG	spd nt 2041458-2041434
comE R[AO]	CTCTTCTTAAACATCTCCATCATTGATA	spd nt 2041136-2041162
Cps3 F	CCCTGAGCGTAATCTCGTGACAA	type 3 cps locus, nt 2766-2788
Cps3 R	CTGAGAACCTTCTGCCCACCTTA	type 3 cps locus, nt 3102-3080
M13F	GTAAAACGACGGCCAG	pCR2.1-TOPO, nt 391-406 (Invitrogen)
M13 R	CAGGAAACAGCTATGAC	pCR2.1-TOPO nt 205-221 (Invitrogen)
AD20[AO]	GGAACTTATTAGGATCCAGAAGATGGC	(Ogunniyi *et al.*, 2008) *Bam*HI site

AD21[AO]	TTGTCGCGAGCTCTCTCCTCTCCTA	(Ogunniyi *et al.*, 2008) *Sac*I site
G54 spn02106 F	CCGTAGGGATGTTTCCAGCAAG	spx nt 1632679-1632700
G54 spn02113 R	CTTCCAGTAGCGGCTAGATTGG	spx nt 1636090-1636069
G54 spn02109 F	CCCTGTGCCTCTCTTGTCAACA	spx nt 1633410-1633431
G54 spn02110 R	GTACAAGTCTGCTAGTGACAAAGA	spx nt 1633897-1633874
TIGR4 sp0303 F	AGGTGGCAAGGGAATTAGTGTTG	spn nt 276912-276934
TIGR4 sp0310 R	CCATGGAACCTGTGCGATAATTG	spn nt 283240-283218
TIGR4 sp0303/0304 F	GGCTGTAGTTGAAGATGGTGTTG	spn nt 278061-278083
TIGR4 sp0306 R	TGATGAAGACATTCCTGCTGCAC	spn nt 278739-278717
sph1924a R	GGGGATTTAATGGAGCATGTGGT	spv nt 1787860-1787882
sph1923 R	GCAGCTACTCGGTATCGATTAC	spv nt 1786264-1786285
sph1922 R	CCCCTCCGCATATAGAAGCAAG	spv nt 1785531-1785552
sph1919 R	AGTTGCCGAAGACGCTAAAGACG	spv nt 1784204-1784226
sph1932 F	CCCTATATCATGGCTGGTGACC	spv nt 1796581-1796560
sph1929 F	TGGCACCTATAACTGTTATACCTG	spv nt 1792104-1792081
sph1928 F	GCGGATGTGATACTTGTTCCAG	spv nt 1791188-1791167
sph1914/15 R	CGTCTATTGAGGGCGTGAATAC	spv nt 1782117-1782138
G54 spg1706 F	CAAGTTCGTACCATTTCGAAGAAG	spx nt 1640314-1640291
mRNA spd0016 16s F	GAGCTTGCTTCTCTGGATGAGTTG	multiple locations in pneumococcal genome
mRNA spd0016 16s R	GTGATGCAAGTGCACCTTTTAAGC	multiple locations in pneumococcal genome
mRNA sph1924 sulfatase F	CAATAAAGCTGCCCTAGCAGGAAC	spv nt 1788329-1788352
mRNA sph1924 sulfatase R	GATCAAATGAGAGCAGACGCCTTA	spv nt 1788478-1788455
mRNA sph1929 PTSIIC F	ACCTCCCAAAGCTACTTGAGCAAT	spv nt 1792409-1792432
mRNA sph1929 PTSIIC R	GATGGTGTTCCACCAGCAGTATCT	spv nt 1792555-1792532

mRNA sph1925 PTSIIc R	GGCATAGTCCCCACAAATGAATC	spv nt 1790089-1790111
comC F	TGACAGTTGAGAGAATCTT	(Whatmore *et al.*, 1999)
comC R	CTTTTCTATTTATTTGACCT	(Whatmore *et al.*, 1999)
Pcps F	GGTTCGCGGGAAGTCTACTAAG	cps2 nt 1347-1370
J214^{JM}	GAAGGAGTGATTACATGAACAA	pVA891 nt 5125-5103
J215 ^{JM}	CTCATAGAATTATTTCCTCCCG	pVA891 nt 4363-4384
J257 ^{JM}	TTAAATGCCCTTTACCTGTTCC	pVA891 nt 4980-5002
J258 ^{JM}	GAAGCTATATACGTACTTTGTTTC	pVA891 nt 4549-4525
Pcps/sph1930/31 R	CTTAGTAGACTTCCCGCGAACCTATCTC CCCTTTTTCACAATGTATAG	*cps2* nt 1370-1347, spv nt 1794617-1794642
J215/sph1920/21	CGGGAGGAAATAATTCTATGAGGAAAGC AAACAGCCTTGAAATCAAT	pVA891 nt 4384-4363, spv nt 1784704-1784680
J214/sph1930 R	TTGTTCATGTAATCACTCCTTCTAGTTGT CAGATTCTTAAAATCCTAT	pVA891 nt 5103-5125, spv nt 1793301-1793275
J215/sph1930 F	CGGGAGGAAATAATTCTATGAGACATTT AATAAATAGGCTAAAAAGAGG	pVA891 nt 4384-4363, spv1794503to1794529
J214/sph1929 R	TTGTTCATGTAATCACTCCTTCAAATATA TTTTATTATCAAAAGTTATCAATTA	pVA891 nt 5103-5125, spv nt 1793050-1793081
J215/SPH1925 F	CGGGAGGAAATAATTCTATGAGTTGC GTTTTTTGTGATAAAATAGAAATAGA	pVA891 nt 4384-4363, spv nt 1788833-178880
J214/sph1924 F	TTGTTCATGTAATCACTCCTTCAATAT GGCAAAAGAGGCACAAGAA	pVA891 nt 5103-5125, spv nt 1788535-1788558
J215/sph1923 R	CGGGAGGAAATAATTCTATGAGAAAA AGTGGTTGAGATTATTTATTTT	pVA891 nt 4384-4363, spv nt 1786190-1786165
17 File198 up	GCAGTTAAGCCAGAAAGTCAACC	spn nt 635561- 635583
17 File198 down	GGCTGGGTAAAGGTATATCTCTC	spn nt 640623- 640601
17 File340 up	GAGTACTATCCTAGAGGAAACAAC	snm nt 1038333-1038356
17 File340 down	GTTCTTTCAAGTTAACACACTTTCTA	snm nt 1039397-1039372
17 File555 up	ATCCGTGGGTCACAGGTTCGAT	spd nt 1598107-1598128
17 File555 down	GGTGTTGGATTATGATTTAAACTAGT	spd nt 1598245-1598220
17 File556 up	CTTCAGGCTTATATTGAGTATTAACG	spd nt 1599584-1599609

17 File556 down	GAAGCAGGAGTTAGTGTTGGAAC	spd nt 1603913-1602891
17 File663 up	GGACGAATACCAATACGTGGATG	spd nt 1960867-1960889
17 File663 down	CTGATAGAACTGGTAAAGAGGTTC	spd nt 1970285-1970262
WU2 File229 up	CCTATTTCAACCCCATTATCTATAG	spn nt 485911-485935
WU2 File229 down	GCGGAGATAATCTGACGCGATG	spn nt 494283-494262
WU2 File536 up	GTGCTTATCAATCTGAAGAGTGAAG	snm nt 1036666-1036690
WU2 File536 down	GCAGGAAGTCTATTATTCCTTTCG	snm nt 1040607-1040584
17 File568 up2	CCAACTTCCGCCAGCATTTTTC	spd nt 1650174-1650195
17 File568 down2	GGGGCTCCACTCATGAAAGAAG	spd nt 1656701-1656680
17 File568 internal up	GTTGCCCAAAACATTGCATTTGGT	unique sequence to MSHR17 (Primer starts 482 bp downstream of the stop codon of the galactose repressor in MSHR17)
17 File568 internal dn	CTCCCCTTCTGATTAAATATGCCA	unique sequence to MSHR17
17 File568 internal up2	TATTGGCCACTATTTAAACCTAGCA	unique sequence to MSHR17 (Primer starts 597 bp downstream of the stop codon of the galactose repressor in MSHR17)
17 File568 internal up3	CTGAGCGACCTGCTAACAATGG	unique sequence to MSHR17
17 File568 internal up4	CAAGATATCCTGTTAAGGCTTTAC	unique sequence to MSHR17
17 File568 internal up5	GATACCGTCATCTTCAAAGCGTG	unique sequence to MSHR17
17 file568 internal up6	GCGTTTTCATTTAAAAAGTCAATGC	unique sequence to MSHR17
17 File568 internal dn2	GTACTTATGGCCAATAACCTCCT	unique sequence to MSHR17

17 File568 internal dn2a	GCAAGATTGATGTGAAACCTGTC	unique sequence to MSHR17
17 File568 internal dn3	GTCAGGAAGGTGTATTGCATCG	unique sequence to MSHR17
17 File568 internal dn4	GGTTTCTGGATCTAATTTGTCATC	unique sequence to MSHR17
WU2 File888 up2	CCAAGATAGCTAAAGAAGTAAAAGC	spd nt 1981576-1981597
WU2 File888 down2	GGGAACTTTGATTGTCGGTATCG	spd nt 1987281-1987259
x4[RH]	GGTGTTATGGAGCTCCGGGC	sjj nt 948165-948184
x5 [RH]	CCGTTGCCTTTGCCCACCTG	sjj nt 950402-950383
x6 [RH]	GAAATGCGACAACCTCACCC	sjj nt 949806-949787
y3 [RH]	GAAGGTCCCAAAACTAATCCTATC	sjj nt 951145-951122
y4 [RH]	CCGTTACCATTAATAGCATTCTGG	sjj nt 948786-948809
y5 [RH]	CTAACCAAGTTGTTGGGGGAG	sjj nt 949363-949383
O972 [RH]	GGTCAAATCGTGATCTTCCCTC	sjj nt 953659-953638
spd1950 F	GTTCCTTCTGGACGGAGGGTA	spd nt 1920245-1920265
spd1951 R	CTGTATGTGCTGGTAAAGTTCA	spd nt 1920812-1920790
snm2232 F	GTGCTTTGACGCAACTCAGCCAAC	snm nt 2040990-2041013
snm2235 F	CTCCAAAGACATCTGCTGCCAA	snm nt 2045588-2045609
snm2240 F	CACATCCAACTCCAAATGCTTCG	snm nt 2050085-2050107
snm2237 R	CTGGTCAAGAAGCTGTTGAACG	snm nt 2046998-2046977
snm2241 R	GTCGAATTATTTGATGGTGAGAGC	snm nt 2051608-2051585
snm2245 R	CTGGTGACTATAATTGCAGTAGC	snm nt 2056116-2056094

2.3.2 Oligonucleotide adaptors

The two oligonucleotide adaptors, Adaptor 1 and Adaptor 2R, which were used for PCR-based subtractive hybridisation, were provided in the Clontech PCR-Select™ Bacterial Genome Subtraction Kit (Clontech, Mountain View, California USA). The adaptors are illustrated in Figure 2.1 along with their primers. PCR-based subtractive is described in Section 2.8.

NOTE:
This figure is included on page 53 of the print copy of the thesis held in the University of Adelaide Library.

Figure 2.1. Sequences of the two adaptors that were used for PCR-based subtractive hybridisation, including restriction sites and corresponding primers.

PCR based subtractive hybridisation is described in Section 2.8 with primers listed in Table 2.2. A full *Rsa*I site is obtained when the adaptors are ligated to *Rsa*I restricted DNA. The T7 promoter sequence is common to both adaptors and therefore, only Primer 1 is required in the primary PCR following the secondary subtraction. Nested Primers 1 and 2R were designed based on the unique sequence of each adaptor and are used for secondary PCR. The adaptors and primers were provided in the Clontech PCR-Select™ Bacterial Genome Subtraction Kit and this figure has been adapted from Figure 10 of the user manual of this kit.

2.4 Protein Analysis

2.4.1 Preparation of whole cell lysates

Bacteria were grown in THY to mid-exponential phase (A_{600} = 0.25) and 3 ml of culture pelleted by centrifugation at 3,175 × *g* for 15 min. Pellets were resuspended in 300 µl of LUG (5% [v/v] β-mercaptoethanol, 62.5 mM Tris, 2% [w/v] SDS, 10% [v/v] glycerol, 0.05% [w/v] bromophenol blue) and boiled for 5 min prior to analysis. Samples were stored at –20 °C.

2.4.2 SDS-PAGE

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was performed using the method described by Laemmli (1970). Separation of proteins was achieved by electrophoresis at 170 V through a stacking gel (6% acrylamide) and a separating gel (12 to 20% acrylamide). Protein staining was performed using 0.06% (w/v) Coomasie Brilliant Blue R250, dissolved in 10% (v/v) glacial acetic acid and 25% (v/v) methanol, with gentle agitation of the gel for 1 h at 65°C. This was followed by destaining by washing with several changes of 10% (v/v) acetic acid and 10% (v/v) isopropanol. Protein size was analysed by comparison with Broad Range Marker (Bio-Rad Laboratories, California, USA), which has fragment sizes of 200, 112.5, 97.4, 66.2, 45, 31, 21.5, 14.4 and 6.5 kDa.

2.4.3 Western blot analysis

Proteins were separated by SDS-PAGE, as described in Section 2.4.2 and electroblotted onto nitrocellulose (Pall Life Sciences, Michigan., USA) at 300 mA for 1 h,

as described by Towbin *et al.* (1979). Following transfer, the membrane was blocked by gentle agitation in 5% (w/v) skim milk powder (Diploma) dissolved in TTBS (20 mM Tris-HCl, 154 mM NaCl, 0.5% [v/v] Tween-20, pH 7.4) for 20 min. The blocking solution was removed and the membrane was probed with specific polyclonal antisera at a dilution of 1/3,000 in TTBS overnight at RT with gentle agitation. Three 10 min washes with TTBS were performed, followed by the addition of blotting grade goat anti-mouse or goat anti-rabbit IgG alkaline phosphatase (AP) conjugate (BioRad Laboratories) at a dilution of 1/15,000 in TTBS. The membrane was incubated for 1 h at RT with gentle agitation and then washed 4 times for 5 min with TTBS. After washing, the membrane was equilibrated by rinsing with 15 ml DIG buffer 3 (100 mM Tris, 100 mM NaCl, 50 mM $MgCl_2$, pH 9.5) for 2 mins prior to the addition of 45 µl DIG 4 (75 mg/ml nitroblue tetrazolium salt in 70% [v/v] dimethylformamide), and 35 µl DIG 5 (50 mg/ml 5-bromo-4-chloro-3-indolylphosphate toluidinium salt in dimethylformamide) in 10 ml DIG buffer 3. When the desired colour reaction had been achieved, the membrane was rinsed with 1× TE, followed by water and then dried.

2.4.4 Haemolysis assay

S. pneumoniae were cultured in THY broth until A_{600} = 0.25 (mid-log phase). Cultures were centrifuged at 3,175 × *g* for 10 min and the pellet concentrated 10-fold in PBS. The cells were lysed by incubation at 37 °C in the presence of 0.1% DOC. Supernatant protein concentration was determined by using the Bradford protein kit (BioRad Laboratories) with BSA as a standard, and lysates were then adjusted to the same concentration using PBS. 3% human red blood cell (hRBC) solution was prepared by washing 2 ml hRBC with PBS to remove lysed cells and resuspending the cells in 24 ml of PBS with 50 µl β-mercaptoethanol.

Haemolytic assays were performed as follows. 100 µl of PBS was added to each well of a 96 well round-bottomed microtiter plate (Sarstedt, South Australia). No lysate was added to the first row of 8 wells. In the other rows, pneumococcal lysates were serially diluted 2-fold in duplicate, except for the final row where pure Ply was used. To ensure each well contained a volume of 100 µl, 100 µl was removed from the last well following dilution. 100 µl of 3% hRBC solution was added to each well and the plate then incubated at 37 °C for 30 min. The plate was centrifuged at 2,500 × *g* for 10 min following incubation. 100ul of supernatant was transferred from each well into a new microtiter plate and the A_{540} was read using a Spectramax M2 spectrometer (Molecular Devices).

2.5 DNA Isolation and Manipulation

2.5.1 Agarose gel electrophoresis

DNA was electrophorised through horizontal agarose gels (0.8–2.5% (w/v) dissolved in TBE buffer [44.5 mM Tris, 44.5 mM boric acid, 1.25 mM EDTA, pH 8.4]) and immersed in TBE buffer at 150-180 V. A one tenth volume of tracker dye (15% (w/v) Ficoll, 0.1% (w/v) bromophenol blue, 100 ng/ml Rnase A) was added to DNA samples prior to loading. Gels were poststained in either ethidium bromide (0.5 µg/ml in TBE buffer) or GelRed™ Nucleic Acid Gel Stain (Biotium, California, USA) according to the manufacturer's instructions. DNA bands were visualised by short wave UV transillumination and results captured using a Biorad Gel Doc™ XR with Quantity One® software and a Sony thermal printer. The size of visualised bands was estimated by comparison with the following defined DNA size markers: 1 Kb Plus DNA Ladder (Invitrgoen; fragment sizes: 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1.65, 1, 0.85, 0.65, 0.5, 0.4, 0.3, 0.2, 0.1 kilobase pairs [kb]), *Eco*RI-digested *Bacillus subtilis* bacteriophage SPP1 DNA (Geneworks, Adelaide, SA, Australia; fragment sizes: 8.56, 7.43, 6.11, 4.90, 3.64, 2.80,

1.95, 1.88, 1.52, 1.41, 1.16, 0.99, 0.71, 0.49, 0.36, 0.08 kb), *Hpa*II-digested pUC19 DNA (Geneworks; fragment sizes: 501, 404, 331, 242, 190, 147, 111, 110, 67, 34, 26 base pairs [bp]) or *Hae*III-digested ΦX174 DNA (New England Biolabs [NEB] Massachusetts, USA; fragment sizes: 1353, 1078, 872, 603, 310, 281, 271, 234, 194, 118, 72 bp).

2.5.2 Pneumococcal chromosomal DNA isolation

Pneumococcal chromosomal DNA was isolated using the Wizard Genomic DNA purification kit (Promega Life Sciences, Wisconsin, USA) according to the manufacturer's instructions with the following exceptions: 0.1% (w/v) DOC was added to the pellet during the first lysis step and DNA was precipitated using ice cold ethanol. Strands of DNA were collected on the end of a sealed pasteur pipette and washed in 70% ethanol, followed by dissolution in 1× TE.

2.5.3 Pneumococcal chromosomal DNA isolation using phenol

Pneumococci grown O/N on a BA plate were resuspended with 2 ml of THY and added to 100ml of THY supplemented with choline chloride (0.5% w/v). Cultures were grown O/N at 37 °C/ 95% air/ 5% CO_2. Loopfulls of culture were spread onto BA in duplicate and grown O/N under the same conditions to check for contamination. The cultures were centrifuged at 25,000 × *g* for 15 min at 4 °C and the supernatant removed. Pellets were stored at –80 °C for 24 h. Pellets were thawed and resuspended in 10 ml of 0.14 M NaCl, 0.1 M trisodium citrate, pH 7.4, and DOC was added to 0.1% (w/v). The resuspension was left at room temperature (RT) for 10 min. An equal volume of phenol (Tris HCl equilibrated, pH 7.5) was added and the bottles were gently rotated in a circular motion for 10-15 min and then centrifuged at 1,620 × *g* for 5 min. Two to three phenol

extractions were performed in total, followed by two chloroform extractions. DNA was precipitated using 2.5 volumes of ice cold 100% ethanol with 0.1 volumes of 3 M sodium acetate. The solution was incubated at –20 °C for 20 min prior to pelleting the DNA by centrifuging at 25,000 × *g* for 10 min at 4 °C. The DNA was resuspended in approximately 2-3 ml 1× TE and incubated at 4 °C O/N. The following day RNase treatment was performed by incubation at 37 °C for 1 h in the presence of RNase A at a concentration of 400 µg/ml. Some preparations were also treated with proteinase K at a concentration of 100 µg/ml with incubation O/N at 56 °C. One more phenol extraction, followed by two chlorform extractions were carried out. Finally, the chromosomal preparations were dialysed O/N at 4 °C in 1× TE.

2.5.4 Restriction endonuclease digestion of DNA

Digestions of chromosomal DNA were performed in a final volume of 20 µl. Each reaction contained approximately 1 µg DNA, 1-2 units of the restriction enzyme and the buffer recommended and supplied by the manufacturer (NEB). Reactions were carried out at 37 °C for 1 h.

2.5.5 Polymerase chain reaction (PCR)

PCR reactions were carried out in an Eppendorf Mastercycler (Eppendorf, Hamburg, Germany) or a Hybaid PCR Sprint (Hybaid, Middlesex, England) using a final volume of either 25 or 50 µl. Standard reactions were performed using *Taq* DNA polymerase (Roche Diagnostics) according to the manufactuer's instructions with the exception that the buffer was used at 1.6× concentration. The Expand™ Long Template or High Fidelity PCR Systems (Roche Diagnostics) were used when high fidelity PCR was

required. Reaction conditions comprised 25-30 cycles as follows: denaturation at 94 °C for 3 min, annealing at 55-60 °C for 30 sec, and extension at 68 °C (72°C for *Taq* polymerase) for various times, depending on the length of the product (approximately 1 min for each kb of product).

Overlap extension PCR was performed using the Expand™ Long Template PCR system with the aforementioned reaction conditions. The products of the primary PCRs and amplification of the antibiotic resistance cassette were purified as described in the following section. The secondary reactions were set up so that the antibiotic resistance cassette was at a higher concentration with an approximate ratio of 5:7:5 for primary PCR product 1, antibiotic resistance cassette and primary PCR product 2.

2.5.6 Purification of PCR products and restriction digests

PCR products and restriction digests were purified using either the Mo Bio UltraClean PCR Clean-Up DNA Purification Kit (Mo Bio Laboratories, California, USA) or Qiagen MinElute® PCR Purification Kit (Qiagen, Hilden, Germany) according to the manufacturer's respective instructions. For DNA microarray analysis and PCR-based subtractive hybridisation, the Qiagen MinElute® PCR Purification Kit was used with an additional wash step using 250 µl of 80% ethanol performed following the PE wash. Tubes were centrifuged for 1 minute, the waste removed and then centrifuged for an additional 5 min with the lid open.

2.5.7 Cloning of PCR product

The TOPO TA cloning® kit (Invitrogen) with pCR®2.1-TOPO® and One Shot® TOP10 chemically competent cells was used according to the manufacturer's instructions.

PCR products were generated essentially as described in 2.5.5 using *Taq* polymerase with a 10 min incubation at 72 °C at the end of cycling. LB-Amp plates spread with 1.6 mg of X-gal were used for selection.

2.5.8 Plasmid miniprep

E.coli were cultured overnight in LB and 5 ml pelleted by centrifugation at 2880 × *g* for 10 min in a bench top centrifuge (Sigma). The UltraClean Mini Plasmid Prep kit (Mo Bio Laboratories, California, USA) was used to isolate the plasmid DNA, following the manufacturer's instructions.

2.5.9 DNA sequencing

DNA sequencing reactions were carried out using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, California, USA). Each reaction contained PCR template (10-50 ng, purified as described in Section 2.4.6), 1 µl BigDye® terminator mix, 3 µl sequencing buffer and 3.2 pmol of a single primer, made up to a final reaction volume of 20 µl with sterile MQ water. Sequencing reactions were performed using a thermal cycler (Hybaid or Eppendorf) with 25 cycles of the following conditions: heat denaturation for 30 sec at 95 °C, primer annealing and extension for 4 min at 60 °C. Following completion of the sequencing reaction, the dye terminator labelled-DNA was precipitated by the addition of 80 µl of 75% (v/v) isopropanol and 20 µg of glycogen (Roche), followed by incubation at RT for 30 min. The precipitate was pelleted by centrifugation at 16,000 × *g* at 4 °C for 30 min. The pellet was subsequently washed with 250 µl of 75% isopropanol and centrifuged at 16,000 × *g* for a further 20 min. DNA pellets were dried at 65°C for 10 min.

Sequencing was performed by the Molecular Pathology Unit at the Institute of Medical and Veterinary Science, Adelaide, SA, Australia using an Applied Biosystems 3730 DNA sequencing analyser.

2.5.10 Next generation sequencing

Sequencing and genome assembly were performed by Geneworks using chromosomal DNA prepared as described in Section 2.5.3, using an Illumina Genome Analyser *II* (California, USA) and Lasergene® 8 software (DNASTAR Inc, Wisconsin, USA). The 36 bp sequence fragments had to be 80% identical to the reference sequence to be aligned at a particular location.

2.5.11 Bioinformatics

Sequencing from the Molecular Pathology Unit was analysed using the DNAMAN program (Lynnon Biosoft®, Quebec, Canada). Next generation sequencing from GeneWorks was analysed using Lasergene® 8 SeqMan and SeqBuilder.

The BLAST programs available through the websites of the National Centre for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/Blast.cgi), the J. Craig Venter's Institute's Comprehensive Microbial Resource (JCVI-CMR, http://blast.jcvi.org/cmr-blast) and the Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/) were used to analyse sequencing data. The Clustal-W program available through the website of the European Bioinformatics Institute (EBI, http://ebi.ac.uk/Tools) was used to align multiple sequences.

2.6 Multi Locus Sequence Typing

Pneumococcal isolates were subjected to multilocus sequence typing (MLST) using the method of Enright and Spratt (1998). PCR amplification of the target gene fragments (*aroE*, *ddl*, *gdh*, *gki*, *recP*, *spi* and *xpt*) was performed as described in Section 2.5.5 using primer pairs *aroE*-up/dn to *ddl*-up/dn in Table 2.2. PCR products were purified as described in 2.5.6 and sequenced using the aforementioned primers as described in 2.5.9. The sequence type (ST) was determined by comparing the DNA sequence for the seven housekeeping gene PCR products to the pneumococcal MLST database (http://spneumoniae.mlst.net/), which is located at the Imperial College London and funded by the Wellcome Trust.

2.7 DNA Microarray

Microarray slides containing PCR products for every identified open reading frame (ORF) of the *S. pneumoniae* TIGR4 genome, supplemented with additional PCR products of ORFs identified in the sequence of *S. pneumoniae* R6, were obtained from the Bacterial Microarray Group at St George's Hospital, University of London. Slides were blocked with a solution of 1% (w/v) BSA, 0.1% (v/v) SDS and 3.5× SSC (20× SSC contains 0.15 M NaCl, 0.15 M sodium citrate) at 65 °C for 30 min. Slides were subsequently washed twice with MQ H_2O, followed by two washes with isopropanol, and dried using nitrogen gas.

S. pneumoniae of DNA (10.5 µg in 100 µl) was digested with *Sau*3a or *Rsa*I (NEB), as described in Section 2.5.4. Digests were purified using a Qiagen MinElute® PCR Purification Kit as described in Section 2.5.6.

The Genisphere Array 900 DNA™ DNA Labelling Kit for Microarrays (Genisphere, Pennsylvania, USA) was used according to the manufacturer's instructions.

20 µl of purified digest was labelled for each dye (Alexa Fluor 555 and Alexa Fluor 647) per slide. Hot tagged DNA hybridisation mix was dispensed onto pre-warmed slides and cover slips carefully applied. Slides were incubated overnight in a dark humidified chamber at 65 °C. Slides were washed the following day in a 3 step process (15 min with 2× SSC, 0.03% (v/v) SDS, at 65 °C; 15 min with 1× SSC, RT; 15 min with 0.2× SSC, at RT) and dried using nitrogen gas. A GenePix® 4000B scanner (Molecular Devices) was used to scan slides and images were acquired using Axon GenePix® Pro 6.0 (Molecular Devices).

2.8 PCR-Based Subtraction Hybridisation

Subtraction Hybridisation was performed using the Clontech PCR-Select™ Bacterial Genome Subtraction Kit (California, USA). DNA was prepared as described in Section 2.5.3 and digested for 5 h using *Rsa*I, as described by the kit manufacturer. Ligations were performed following the subtraction kit's instructions using T4 DNA Ligase and buffer (NEB). The positive control was used as described in the manual, however digested pneumococcal driver DNA instead of the supplied *E.coli* DNA and therefore, twice the concentration of *Hae*III-digested bacteriophage ϕX174 DNA (NEB) was used. Subtractive hybridisation was performed according to the manufacturer's instructions with the exception that the primary incubation was increased to 8 h. PCR amplifications were performed with a hot start, using *Taq* polymerase and 1.6× reaction buffer as outlined by the subtraction kit manual. The primary PCR consisted of 28 cycles, while the secondary PCR had 15 cycles. Secondary PCR products were used in cloning as described in Section 2.5.7.

2.9 Southern Hybridisation

2.9.1 Prepartion of membrane

DNA was digested using *Eco*RI and resolved on a 1% (w/v) agarose TBE gel (5µg/well), as previously described. The DNA was depurinated by gently agitating the gel in 0.25 M HCl for 15 min at RT. This was followed by sequential 30 min agitations in denaturing solution (0.4 M NaOH, 0.6M NaCl) and neutralising solution (1.5 M NaCl, 0.5 M Tris-HCl [pH 7.5]) at RT. DNA was transferred to a Hybond-N^+ Nylon membrane (Amersham, Buckinghamshire, UK) O/N at RT by capillary action with 10× SSC acting as the transfer buffer (Southern, 1975). The membrane was then placed DNA side up on Whatman paper, which had been soaked in 0.4 M NaOH for 20 min in order to fix the DNA, followed by brief washing in 2× SSC.

2.9.2 Hybridisation

The membrane was prehybridised by incubation in prehybridisation solution (5× SSC, 0.2% [v/v] SDS, 50% [v/v] formamide, 2% [v/v] blocking solution [10% (w/v) Blocking reagent] dissolved in DIG buffer 1 [100 mM maleic acid, 150 mM NaCl, pH 7.5]) for 4 h at 42 °C with gentle agitation. Following the removal of the prehybridisation solution, the membrane was incubated at 42 °C O/N with hybridisation solution containing the probe of interest. The following day, the membrane was washed twice with 2× SSC (5 min, 65 °C) and twice with 0.2× SSC, 0.1% (v/v) SDS (5 min, 65 °C).

2.9.3 Generation of probes

Digoxigenin-11-deoxy UTP (DIG-dUTP)-labelled probes were generated via PCR using an Eppendorf thermal cycler. Reactions were performed in 25 µl volumes, containing 0.5 µl plasmid preparation (Section 2.5.8), 10 µM oligonucleotide primers, 7 µl 1× DIG dNTP mix (Roche), 0.5 U *Taq* polymerase and *Taq* buffer to a final concentration of 1.6×.

2.9.4 Development of the membrane

The membrane was washed for 1 min in DIG buffer 1 prior to gentle agitation in DIG 2 buffer (1% w/v) Blocking reagent in DIG buffer 1) for 30 min at RT. Anti-DIG AP conjugate (Roche) was added at a dilution of 1/5000 and the membrane was gently agitated for a further 30 min at RT. The membrane was then washed twice in DIG buffer 1 (15 min) and equilibrated by incubating in DIG buffer 3 (see Section 2.4.3) for 2 min. Substrate solution (10 ml of DIG buffer 3, 45 µl DIG 4 [see Section 2.4.3] and 35 µl DIG 5 [see Section 2.4.3]) was added. The membrane was placed in the dark until the desired colour reaction had taken place. The reaction was stopped by washing the membrane in 1× TE. The membrane was then rinsed with water and dried for storage.

2.10 RNA Isolation and Manipulation

2.10.1 Hot-phenol RNA extraction

Frozen stocks of bacteria in the opaque phase were made by growing the bacteria in C + Y and then freezing with 20% glycerol. Aliquots were thawed and inoculated into appropriate media. Bacteria were grown to midlog phase and pelleted by centrifugation at 3,175 × *g* and the supernatant removed. RNA was isolated from the bacterial pellets with

acid-phenol:chloroform:isoamyl alcohol (25:24:1; pH 4.5, Ambion, Austin, Texas., USA) essentially as described by Ogunniyi *et al.* (2002). Cells were resuspended in 400 µl hot (65 °C) acid-phenol:chloroform:isoamyl alcohol. Samples were incubated at 65 °C for 5 min to lyse the cells and then 400 µl hot (65 °C) NAES solution (50 mM sodium acetate, 10 mM EDTA, 1% [w/v] SDS, pH 5.1, treated with diethyl pyrocarbonate [DEPC]), was added and mixed by vortexing before incubation for an additional 5 min at 65 °C. The lysate was then placed on ice for 2 min, followed by centrifugation at 16,000 × *g* at 4 °C for 2 min. The aqueous phase was carefully removed and the phenol extraction was repeated at least twice. Following the final phenol extraction, 30 µl of 0.05% (w/v) DEPC-treated 3 M sodium acetate and 750 µl 100% ethanol was added to the aqueous phase, and mixed by vortexing. The extract was precipitated at –80 °C for over 2 h. Nucleic acids were pelleted by centrifugation at 16,000 × *g* for 30 min at 4 °C. Pellets were washed twice using 1 ml 70% ethanol (0.7 volume 100% ethanol and 0.3 volume DEPC-treated water) and centrifuging for 5 min at 16,000 × *g* to re-pellet the nucleic acids. The supernatant was aspirated by pipetting, residual ethanol was removed by vacuum suction and pellets were suspended in 50 µl RNase-free, DNase-free water (Roche).

For DNase treatment, 30 µl of the resuspended RNA was transferred to a fresh RNase-free tube. The RNA preparation was heated at 95 °C for 5 min and then placed on ice. 10 units of RNase-free DNase I (Roche), 1× DNase reaction buffer, 1 U/µl recombinant RNasin ribonuclease inhibitor (Promega Life Sciences, Wisconsin, USA) and nuclease free water were added to give a final volume of 40 µl. DNase reactions were incubated at 37 °C for 30 min and then at 65 °C for 10 min to stop the reaction. One-step reverse transcription polymerase chain reaction (RT-PCR, Section 2.10.2) using 16s rRNA-specific primers (Table 2.2) was performed to confirm the purity of RNA preparations and ensure removal of residual DNA. For each preparation, one reaction was set up with reverse transcriptase and one reaction without reverse transcriptase. Products

were visualised after electrophoresis on a 2.5% agarose gel followed by post-staining with Gel Red (Section 2.5.1).

2.10.2 Reverse transcription PCR (RT-PCR)

The One-step Access RT-PCR system (Promega) was used according to the manufacturer's instructions to perform RT-PCR. A standard 12.5 µl reaction contained 1× AMV/*Tfl* reaction buffer, 0.2 nM dNTPs, 200 nM of each oligonucleotide, 1 mM $MgSO_4$, 1.25 U AMV reverse transcriptase, 1.25 U *Tfl* DNA polymerase and 1 ng RNA. Reactions were performed in an Eppendorf Mastercycler® Thermal Cycler and the standard program used was 45 min of reverse transcription at 48 °C, followed by 2 min denaturation at 94 °C and then amplification for 25 cycles of denaturation at 94 °C for 30 sec, annealing at 55 °C for 30 sec and extension at 72 °C for 20 sec.

2.10.3 Real-time RT-PCR

Real-time RT-PCR was performed using the Invitrogen SuperScript™ III Platinum® SYBR® Green One-Step qRT-PCR kit according to the manufacturer's instructions, and using the oligonucleotides in Table 2.2. Each reaction was performed in a final volume of 20 µl and contained 18.2 µl of reaction mix, 200 nM of each primer and 150 ng of RNA, with the exception of reactions using 16s RNA primers, which used 75 pg of RNA. Amplification was performed using a Rotorgene RG-2000 cycler (Corbett Research, Mortlake, N.S.W., Australia) with the following program: reverse transcription at 50 °C for 5 min, followed by denaturation at 95 °C for 5 min and 40 amplification cycles (denaturation at 95 °C for 15 sec, annealing and amplification at 60 °C for 15 sec). Five µl

of product from selected tubes was electrophoresed on a 2.5% agarose gel and visualised following post-staining with Gel Red.

Data were analysed using the comparative cycle threshold ($2^{-\Delta\Delta C}{}_T$) method (Livak and Schmittgen, 2001) in which the amount of target mRNA under one condition is compared to the amount under another condition normalised to an internal control (16s rRNA in this study).

2.11 Pneumococcal Transformation

2.11.1 Preparation of competent cells (CTM method)

Pneumococcal cells were grown in c-CAT medium (10 g/l Bacto Casamino acids, 5 g/l Bacto Tryptone, 5 g/l NaCl, 10 g/l Bacto yeast extract, 4% [v/v] 0.4 M K_2HPO_4, 0.002% [w/v] glucose, 150 mg/l glutamine) to A_{600} = 0.25-0.3, then diluted to A_{600} = 0.01 in 10 ml CTM medium (c-CAT supplemented with 0.2% [w/v] BSA, 1% [v/v] 0.1 M $CaCl_2$) and grown to A_{600} = 0.1. The cells (10 ml) were pelleted at 3,175 × *g* for 10 min, and resuspended in 812 µl CTM, which had been adjusted to pH 7.8 with NaOH, and 188 µl 80% glycerol. 100 µl aliquots were stored at –80 °C.

2.11.2 Transformation using CTM

An aliquot of competent cells was thawed and 500 µl of CTM-pH 7.8 and 2.5 µl of 10 µg/ml competence stimulating peptide-1 (CSP-1) (amino acid sequence: MRLSKFFRD FILQRKK [Chirontech, Victoria, Australia]) or competence stimulating peptide-2 (CSP-2) (amino acid sequence EMRISRIILDFLFLRKK [Mimotopes (Victoria, Australia) (Havarstein *et al.*, 1995)]). The cells were then incubated at 37 °C for 10-15 min before addition of approximately 1 ng of DNA, and then incubated at 32 °C for 30 min, followed

by incubation at 37 °C for 2-4 h. Following incubation, cells were plated onto appropriate BA + antibiotic plates, and incubated for approximately 16 h at 37 °C in 95% air/ 5% CO_2.

2.11.3 Transformation using THY

Pneumococci from O/N BA plates were inoculated into THY. Cultures were incubated at 37°C until A_{600} = 0.3 was obtained. The cultures were then diluted 10-fold in THY transformation media (THY/0.2% BSA/0.2% glucose/0.02% $CaCl_2$) to a final volume of 1 ml with the same concentration of CSP-1 and CSP-2 as described in Section 2.11.2. The cells were incubated for 10-15 min, which was followed by the addition of approximately 1 ng of DNA and then incubation at 37 °C for 2-4 h. Cells were plated onto appropriate BA + antibiotic plates, and incubated for approximately 16 h at 37 °C in 95% air/ 5% CO_2.

2.11.4 Transformation using THY supplemented with HCl and glycine

This method was adapted from Bricker and Camilli (1999). Pneumococci were grown in THY until mid-exponential phase and then supplemented with 30% glycerol for storage at –80 °C. Bacteria were thawed and diluted into THY supplemented with 11 mM HCl and 0.5% glycine. Bacteria were grown at 37 °C to A_{600} = 0.1. The culture was subsequently supplemented with 10 mM NaOH, 0.2% BSA and 1 mM $CaCl_2$. Six 100 µl aliquots were taken and CSP-1 was added to three and CSP-2 to the remaining three aliquots at the same concentration as described previously. The aliquots were incubated at 37 °C and DNA was added to one CSP-1 and one CSP-2 tube at 10, 15 and 20 min. The aliquots were incubated at 37 °C for 1 h, before the addition of 300 µl prewarmed THY

and incubation for a further 2 h. Cells were plated onto appropriate BA + antibiotic plates, and incubated for approximately 16 h at 37 °C in 95% air/ 5% CO_2.

For time point investigations, the above method was used except 100 µl aliquots were taken every 10 min for 2 h after the initial culture had reached A_{600} = 0.06. Samples were supplemented with the aforementioned concentrations of NaOH, BSA, $CaCl_2$ and CSP-1, and then incubated for 15 min at 37 °C before the addition of DNA.

2.11.5 Transformation using THY supplemented with antibiotics

These methods were created using Prudhomme *et al.* (2006) as a guide. Prudhomme *et al.* (2006) used C+Y but this work was performed using THY. Frozen stock for transformations were prepared as described in Section 2.11.4.

2.11.5.1 Supplementation with streptomycin

Transformation was performed in duplicate with one culture inoculated in THY and the other in THY supplemented with 11 mM HCl. Bacteria were incubated at 37 °C for 70 min to ensure they had left stationary phase, followed by the addition of streptomycin (6.25 µg/ml). NaOH (10 mM), BSA (0.2%) and $CaCl_2$ (1 mM) were also added to the culture grown in THY supplemented with HCl. Cultures were then incubated again at 37 °C and 100 µl aliquots were taken after 105 min, 135 min and 165 min to which DNA was added. Aliquots were incubated for a further 1 h at 37 °C before the addition of fresh THY (300 µl) and incubation for up to 3 h before plating as previously described.

2.11.5.2 Supplementation with mitomycin C

Transformations were performed in duplicate in THY (pH 7.3) and THY/0.2% BSA/1mM $CaCl_2$ (pH 7.3). Mitomcyin C (40 ng/ml) was added 75 min after inoculation.

100 µl aliquots were taken every 20 min commencing 60 min after the addition of mitomycin C. Sampling was increased to every 10 min from 110 min to 160 min post-supplementation. DNA was added to the samples, which were then incubated for 1 h before the addition of fresh THY and further incubation prior to plating, as previously described.

2.12 In Vivo Pathogenesis Studies

All *in vivo* pathogenesis studies were performed using four to five week old female CD1 mice obtained from either the Waite Institute, University of Adelaide, South Australia or the Animal Resource Centre (ARC), Western Australia. Animal experimentation was conducted in accordance to the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes (7th edition, 2004).

2.12.1 Intraperitoneal and pneumonia/sepsis studies

Opaque variants were cultured in SB to A_{600} = 0.08 (equivalent to mid logarithmic phase) and diluted in sterile SB so that 100 µl (intraperitoneal [i.p.] model) or 40 µl (pneumonia/sepsis model) contained the appropriate challenge dose. The actual challenge doses were determined in retrospect by plating out serially diluted challenge inocula after administration.

Wu *et al.* (1997) previously reported that mouse-to-mouse transmission of pneumococci cannot be detected in co-housed animals and therefore, mice were kept in cages of 10-15 animals in a pathogen-free environment for the duration of experiments. Mice in the i.p. study were inoculated by injection into the intraperitoneal cavity. Mice in the pneumonia/sepsis study were anaesthetised with Nembutal (pentobarbitone sodium;

Merial Australia, Sydney, Australia), which was administered by intraperitoneal injection at a dose of 66 µg per 1 g of bodyweight. The challenge strain was then pipetted into the nares and involuntarily inhaled. Mice were closely monitored until they had regained consciousness.

Following challenge, mice in both studies were monitored closely over a 3-4 week period for signs of disease and the survival time of each mouse was recorded. Blood was taken from the hearts of selected mice post-mortem, and plated on BA containing an optochin disk to confirm the presence of *S. pneumoniae*. At the end of each study, selected surviving mice were euthanased by CO_2 asphyxiation and heart blood was cultured to confirm that there were no bacteria in the blood.

2.12.2 Intraperitoneal competition

For *in vivo* competition assays, mice were inoculated as described for i.p. (Section 2.12.1) except that appropriate dilutions of the two competing *S. pneumoniae* strains were combined prior to inoculation and administered together. In this study, one strain was Ery^R whereas and the other Ery^S. The precise challenge dose and ratio of strains in the inoculum (input ratio) was determined retrospectively by plating serially diluted challenge innoculum onto both BA and BA-Ery.

The input ratio was calculated using the following formula:

$$\frac{\text{Average number of bacteria on BA-Ery}}{\text{Average number of bacteria on BA minus average number of bacteria on BA-Ery}}$$

Mice were euthanased 24 h and 48 h following inoculation by carbon dioxide asphyxiation. Blood was taken from the hearts and serially diluted in sterile SB. Dilutions were plated onto BA and BA-Ery to determine the ratios of mutant to wild-type (output ratio). The output ratio was calculated using the same formula as the input ratio. The competitive index (CI) was then determined for each time point by dividing the output ratio by the input ratio.

2.12.3 Intranasal colonisation study

Transparent variants were grown in THY to A_{600} = 0.25 (equivalent to mid logarithmic phase) and diluted in sterile THY so that 10 µl contained the appropriate challenge dose. Mice were placed in a conical tube with a hole at the narrow end and the inoculum was slowly piptted into the nares, allowing voluntary inhalation. The actual challenge dose was determined retrospectively by plating out serially diluted inoculum following administration.

Selected mice were euthanased 24 h, 48 h, 96 h and 168 h post challenge by carbon dioxide asphyxiation. The trachea of each mouse was cut below the larynx, and the trachaea washed to remove loosely adherent bacteria with 1 ml sterile trypsin buffer (0.5% trypsin, 0.02% ethylenediaminetetracetic acid [EDTA], in PBS) by insertion of a 26-gauge needle sheathed in tubing into the tracheal end of the upper respiratory tract. Bacteria had been confirmed not to be sensitive to the concentration of trypsin used in this study prior to the experiement. Buffer was slowly dripped into the nasopharynx and then collected from the nose. Following this, the entire nasopharynx was removed and placed in 5 ml of sterile PBS. Both tympanic membranes were punctured with a pipette tip and the ear cavities washed with a combined 100µl of sterile trypsin buffer. The heart was then perfused with PBS to ensure removal of contaminating pneumococci from the blood. Lungs were

removed, rinsed twice in PBS and placed in 2 ml of sterile PBS. Whole brains were removed through an incision in the skull and placed in 2 ml of sterile PBS. The lungs, brain and excised nasopharyngeal tissue were then homogenized with a CAT X120 homogenizer (CAT Ingenieurbüro M. Zipperer GmbH, Staufen, Germany) at 30,000 rpm for approximately 10 sec. All blood, lung, ear and nasal washes, and lysate samples were then serially diluted in sterile PBS or SB and plated in duplicate on BA, which was supplemented with Gent to inhibit contaminating microflora, to determine the number of viable pneumococci. Dilutions were spotted onto BA-Optochin to assess contamination where appropriate. Differences in the levels of colonization were analyzed using Student's unpaired *t*-test (two-tailed).

2.12.4 Intranasal competition

Intranasal competition was performed as described in Section 2.12.3, except the two competing strains were grown separately and appropriate dilutions combined at the time of inoculation. The challenge dose and input ratio was determined retrospectively by plating out serially diluted challenge innocula onto both BA and BA-Ery and using the formula as described in Section 2.12.2. Lungs, brains and nasopharyngeal tissue were homogenised using a Precellys 24 homogenizer (Bertin Technologies, Pessac, France) in tubes containing 2.8 mm ceramic bead tubes and 1 ml of sterile PBS with the following program: 3× 30 sec at 5000 rpm with a 5 sec wait between cycles. Dilutions were spotted onto BA-Gent in duplicate, but also BA-Gent-Ery to determine output ratio. The output ratio and CI were calculated as described in Section 2.12.2.

2.12.5 Pneumonia/sepsis competition

Mice were inoculated as described for the pneumonia-sepsis model (Section 2.12.1) except the dilutions of the two competing strains were combined prior to inoculation and administered together. The challenge dose and ratio of strains in the inoculum (input ratio) was determined retrospectively, as described in Section 2.12.2. Groups of up to five mice were euthanased 24 h, 48 h and 72 h post-challenge. The organs were removed, homogenised and plated, and the output ratio and CI for each tissue were calculated as described in Section 2.12.4.

Chapter 3

Initial Characterisation of Serotype 3 and Serogroup 11 Clinical Isolates

3.1 Introduction

Otitis media (OM) is a common childhood ailment and a burden in both developing and developed countries. The occurrence of middle ear infections is rising in developed countries (Cripps *et al.*, 2005). In addition, bacteria are becoming increasingly less susceptible to antibiotics and there is currently no effective vaccination strategy against OM.

The level of OM varies widely in differing communities; in remote Australian Aboriginal communities OM is seen at unacceptably high levels. CSOM, which is normally extremely rare in developed countries, is seen in Australian Aboriginal communities at 10 times the level defined by the WHO to be a major public health problem (Coates *et al.*, 2002). Besides the obvious malaise, there is also the associated massive levels of hearing problems that are being reported due to the prevalence of OM (Morris *et al.*, 2005).

The Menzies School of Health Research (MSHR) in Darwin monitors OM in indigenous communities of tropical and central Australia. At the start of this project, nasopharyngeal and middle ear pneumococcal isolates from a variety of indigenous communities in temporally and geographically diverse locations were obtained through the MSHR. The strains used in this study, MSHR1 to MSHR17 are listed in Table 2.1.

Strains belonging to two serotypes were obtained: serotype 3 (MSHR11-17) and serotype 11A (MSHR1-10). Serotype 3 is commonly isolated from the ears of indigenous children suffering OM with a higher percentage isolated from the ears compared to the nose across various surveillance studies (unpublished data from MSHR). In addition, this serotype has a high association with OM worldwide (Shouval *et al.*, 2006; Hausdorff *et al.*, 2000; Hausdorff *et al.*, 2002; Rodgers *et al.*, 2009) and the level of serotype 3 OM has reportedly risen following the introduction of Prevenar™ (McEllistrem *et al.*, 2005). Interestingly, although serotype 3 has a low association with IPD, it has a high case-fatality rate (Hanage *et al.*, 2005; Sjostrom *et al.*, 2006; Henriques *et al.*, 2000). Both nasopharyngeal and middle ear serotype 3 isolates were made available for this study. The second serotype, 11A, also has a low association with IPD (Hanage *et al.*, 2005), although ST62 of serotype 11A has a high case-fatality rate in patients with underlying medical conditions (Sjostrom *et al.*, 2006). It is a common nasopharyngeal isolate in Aboriginal children and in several studies was isolated from the nose but not the ears of children (unpublished data from the MSHR). Therefore, only nasopharyngeal isolates were available for use in this study. Although serotype 11A is not as commonly associated with OM as serotype 3 is worldwide (Hausdorff *et al.*, 2000; Hausdorff *et al.*, 2002), it has been isolated from middle ears in other communities (Hausdorff *et al.*, 2000; Hanage *et al.*, 2004; Porat *et al.*, 2004), including one study from Sydney, Australia (Watson *et al.*, 2007). Host factors could influence the ability of a particular serotype to cause OM in a given community and this cannot be accounted for in the scope of this study. For example, compared to Caucasian children, children from remote Aboriginal communities are exposed to high levels of bacteria and are colonised by multiple respiratory pathogens from an extremely young age. (Coates *et al.*, 2002; Watson *et al.*, 2006; Smith-Vaughan *et al.*, 2008). Such a competitive environment may affect the ability of serotype 11A to cause OM in Aboriginal children.

Incidentally, serotype 3 has been shown to be positively associated with multiple pneumococcal serotype carriage (Vestrheim *et al.*, 2008).

In addition to the MSHR isolates, clinical isolates of serotype 3 and serogroup 11 were obtained from the Women's and Children's Hospital (WCH), Adelaide, and are listed in Table 2.1 (WCH206-WCH213). This lab does not have serotype 11A antiserum to confirm whether the isolates serogrouped as 11 by the WCH are serotype 11A. The typing of the MSHR isolates was performed at the MSHR isolates, which does have access to serotype 11A specific antiserum. The serogroup 11 isolates were from sinusitis patients, except for one patient who had OM.

WU2, a serotype 3 strain that has been used in published studies for more than 20 years, was used as a control in some experiments. WU2 is a subline of a clinical isolate from the Washington University School of Medicine that was isolated in the 1960s and then passaged through mice multiple times (Briles *et al.*, 1981).

As the strains obtained from the MSHR and WCH were all clinical isolates, there was no information available regarding general characteristics, such as virulence protein expression and virulence in mouse models. Therefore, basic characterisation experiments were conducted, which included Western blot analysis, haemolysis and capsule assays, MLST and virulence studies. This then led to an investigation on the influence of serotype on virulence in which a serotype 11A isolate was switched to serotype 3.

3.2 Initial Characterisation of Isolates

3.2.1 Western blot analysis, PCR and sequencing of selected pneumococcal virulence factors

The pneumococcus expresses a range of well-characterised virulence proteins, such as PspC, PspA, Ply, LytA and NanA, whose functions are described in Section 1.3.3. Differences in these proteins may be associated with variation in ability to cause disease. Therefore, the expression and mobility of these proteins was investigated by Western blot analysis, which was conducted as described in Section 2.4.3 using the lysates of all serotype 3 and serogroup 11 clinical isolates available. Lysate of the serotype 2 strain D39 was also included as a control, because antibodies were raised against recombinant proteins purified from *E. coli* expressing the respective D39 gene. Mouse polyclonal anti-PspC, PspA, LytA and Ply, as well as rabbit polyclonal anti-NanA, were used. Proteins shown to be of interest through Western blot analysis were subsequently investigated further through PCR sequencing as described in Sections 2.5.5 and 2.5.9.

In addition, to NanA, there are two other neuraminidases, NanB and NanC. Their carriage was investigated by PCR in selected serotype 3 and serogroup 11 strains, which were chosen based on their protein diversity demonstrated by Western blot analysis.

3.2.1.1 Western blot and sequence analysis of serotype 3 strains

Western blot analysis was initially performed using lysates of the serotype 3 strains WCH206-WCH208 and MSHR11-MSHR17 (Table 2.1). The Western blots using anti-LytA and anti-Ply polyclonal antiserum indicated all strains produced both proteins and there was no difference in electrophoretic mobility or expression (Figure 3.1).

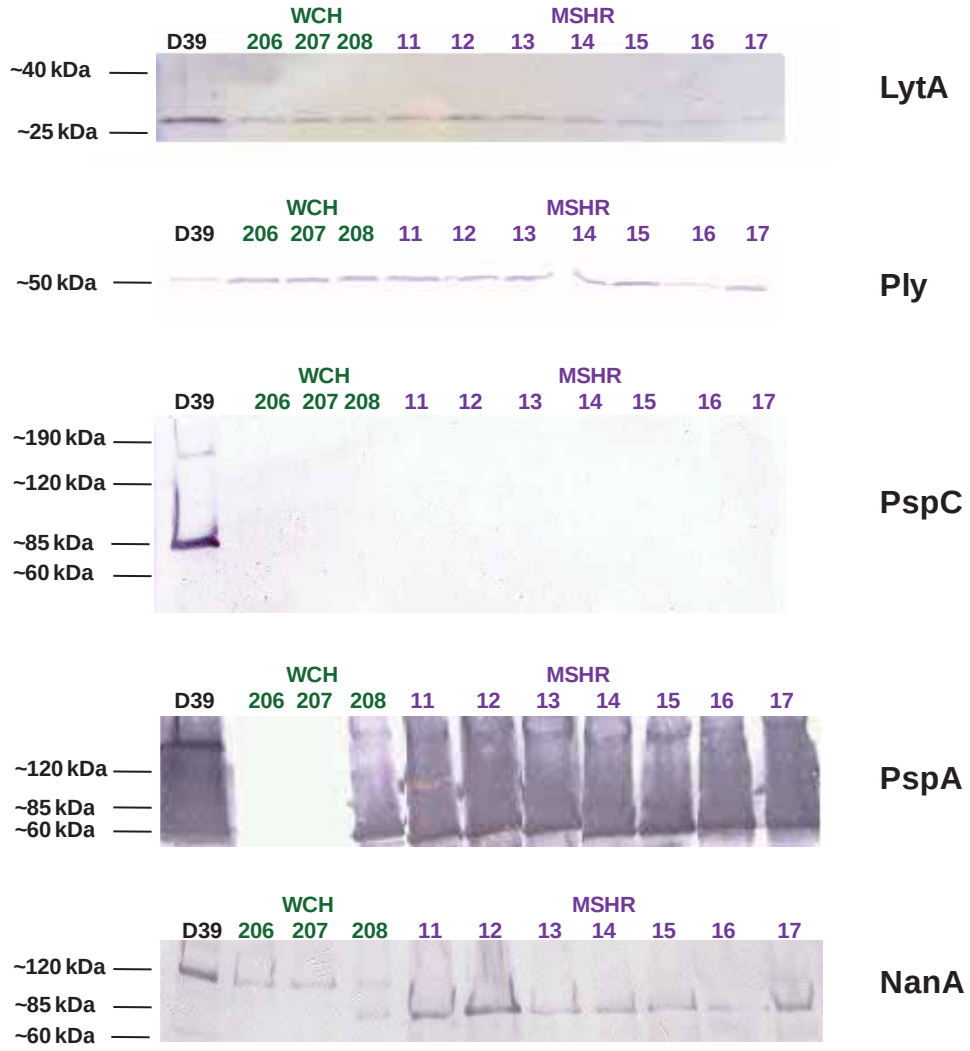


Figure 3.1. Western Blot analysis of D39, strain MSHR17 and all serotype 3 clinical isolates from the MSHR and WCH, using polyclonal LytA, PdB (Ply derivative), PspC, PspA and NanA antisera.

Lysates of the strains, which are listed in Table 2.1, were separated by SDS-PAGE, electroblotted onto nitrocellulose and probed with mouse polyclonal antisera, as described in Section 2.3. MSHR strains are written in purple and WCH strains are written in green. The mobility size marker (Invitrogen Benchmark™ Pre-stained ladder) is also indicated.

No immunoreactive bands were detected by Western blot analysis using PspC antiserum for any of the serotype 3 strains, although a band was detectable for D39 (Figure 3.1). The PspC of D39 is 79 kDa, but there was an additional immunoreactive band on the blot at about 160 kDa, which is therefore, most likely due to dimers. In addition, no immunoreactive bands for the strains WCH206 and WCH207 were detectable when probing with PspA antiserum (Figure 3.1), although all other strains had a detectable band. PspA is difficult to interpret due to the smearing. This phenomenon has been previously reported on heavily loaded gels (Talkington *et al.*, 1992). It is believed that polymers of PspA may dissociate throughout electrophoresis, continually producing monomers or lower molecular weight polymers with a higher mobility which would spread through the gel between the polymers and original monomers (Talkington *et al.*, 1992).

On the NanA Western Blot, the immunoreactive bands for the strains WCH206 and WCH207 appeared to be the same size as that for D39, but the bands for the MSHR isolates were lower revealing that the strains WCH206 and WCH207 had a different NanA to the other serotype 3 strains (Figure 3.1). On the basis of Western blot analysis, the serotype 3 clinical isolates could be divided into 2 distinct groups: one comprising the strains WCH206 and WCH207 and the second the remaining isolates.

It was decided that *pspA*, *pspC* and *nanA* were worth further investigation by PCR and sequencing, which were performed as described in Sections 2.5.5 and 2.5.9. Although no immunoreactive bands were detectable using anti-PspC, products were obtainable by PCR of selected strains when using primers flanking the 5' and 3' termini (PspCF and PspCR), which were designed based on the sequences of D39 and TIGR4 (Table 2.2). Sequencing using the additional primers listed in Table 2.2 (PspCF2-PspCR3) revealed a version of PspC in all isolates sequenced classified as group 8 according to Iannelli *et al.* (2002), which has been detected in other serotype 3 isolates but in no other serotypes, and

is different to Hic, which is the other PspC variant reported in serotype 3 (Janulczyk *et al.*, 2000; Iannelli *et al.*, 2002). Like Hic, rather than a choline binding motif, it has an LPxTG motif near its C-terminus and is attached to the cell wall in a sortase-dependent manner. It also has a large proline rich repeat region that varies in size from 104 bp in MSHR17 to over 400 bp in MSHR11. As with PspC, although no immunoreactive bands were detected using anti-PspA for WCH206 and WCH207, products were obtained in PCRs using primers flanking the PspA locus, PspA(F) and PspA(R), which were designed using the genome sequences of D39 and TIGR4 (Table 2.2). These products were sequenced using additional primers listed in Table 2.2 (pspAF2-pspAF3) and found to be *pspA* belonging to Clade 3, Family 2. The PspA of WCH206 contains two extra repeat units in the choline binding domain relative to WCH207. The PspA of D39 belongs to Family 1 (Hollingshead *et al.*, 2000) and therefore it appears that the D39 PspA antiserum cannot recognise Family 2 PspA.

The immunoreactive bands of NanA for the strains WCH206 and WCH207 suggested that their NanAs were the same size as the NanA of D39. Sequencing of the *nanA* of WCH206 using the primers NanAR2-NanAF2 (Table 2.2) revealed a *nanA* gene identical to that found in D39. The sequencing of *nanA* in MSHR17 using the same primers revealed a *nanA* gene identical to that found in *S. pneumoniae* Pn13, which is a serotype 14 strain isolated in Papua New Guinea (King *et al.*, 2005). It contained a 15 bp insertion (nt 2432 in WCH206) and 180 bp deletion (nt 2708- 2886 in WCH206) towards the 3' terminus, and a 500bp region (nt 486-1013 in WCH206) divergent to that seen in D39 and WCH206 towards the 5' terminus. There is a 1 bp insertion near the 3' terminus, which changes the reading frame so its stop codon is about 12 bp downstream of the stop codon of WCH206. The NanA of MSHR17 is predicted to be 110 kDa. However, the Western blot suggested that it was less than 85 kDa. The difference in mobility may relate

to differences in amino acid sequence or proteolytic processing. In the cell lysate of a serotype 1 strain, the breakdown of its 107 kDa NanA to an 86 kDa species, which remained enzymically active, has been reported (Lock *et al.*, 1988). Therefore, it is possible that the NanA of MSHR17 may also be proteolysed.

PCR was used to investigate the *nanA* of two other MSHR isolates, MSHR11 and MSHR14, which had the identical sized band as MSHR17 on the NanA Western blot. The primers used were NanA Insertion Forward versus NanA Deletion Reverse (Table 2.2) and the amplicon produced was the same size as that of MSHR17. Two bands were detectable in WCH208 by Western blot analysis for NanA, but when PCR analysis was performed using primers NanA Insertion Forward and NanA Deletion Reverse, the amplicon was the same size as that for MSHR17.

3.2.1.2 Western blot and sequencing analsyis of serogroup 11 strains

Lysates from the strains WCH209-215 and MSHR1-MSHR10 (Table 2.1) were used. Lysate of the serotype 3 strain MSHR17 was used for all serogroup 11 Western blots as a comparison. More variable results were obtained and unlike serotype 3, the strains could not be divided into two distinct groups based on the size and sequence of NanA, PspA or PspC. However, some strains such as MSHR1 and MSHR3 exhibited identical protein profiles. Western blotting suggested variation in PspA, PspC and NanA, but not Ply or LytA (Figures 3.2 and 3.3).

PspA was detected in the Western blot for some strains but not others. A method has been developed to type PspA as either Family 1 or Family 2 using PCR (Vela Coral *et al.*, 2001), which was used in this study to type selected strains. The primers, which are listed in Table 2.2 (LSM12-SKH52), and control chromosomal DNA of pneumococcal strains Rx1 (PspA Family 1) and EF3296 (PspA Family 2) (Hollingshead *et al.*, 2000)

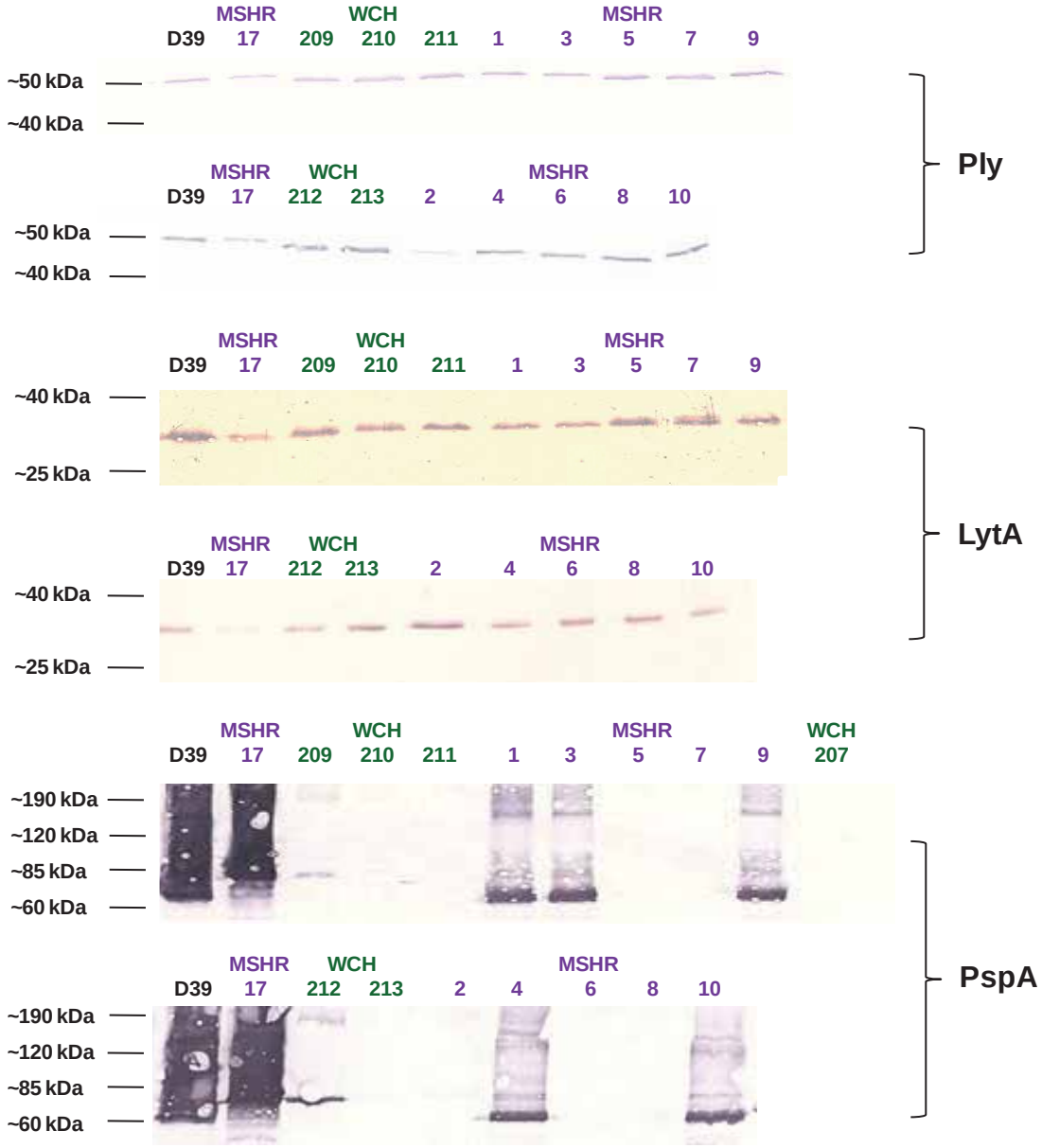


Figure 3.2. Western Blot analysis of D39, strain MSHR17 and all serogroup 11 clinical isolates from the MSHR and WCH, using polyclonal PdB (Ply derivative), LytA and PspA antisera.

Lysates of the strains, which are listed in Table 2.1, were separated by SDS-PAGE, electroblotted onto nitrocellulose and probed with mouse polyclonal antisera, as described in Section 2.3. Strain WCH207 was also included for the PspA blot because its PspA is known to be Family 2 and could not be detected by Western Blot analysis previously (Figure3.1). The PspA of D39 belongs to Family 1. MSHR strains are written in purple and WCH strains are written in green. The mobility size marker (Invitrogen Benchmark™ Pre-stained ladder) is also indicated.

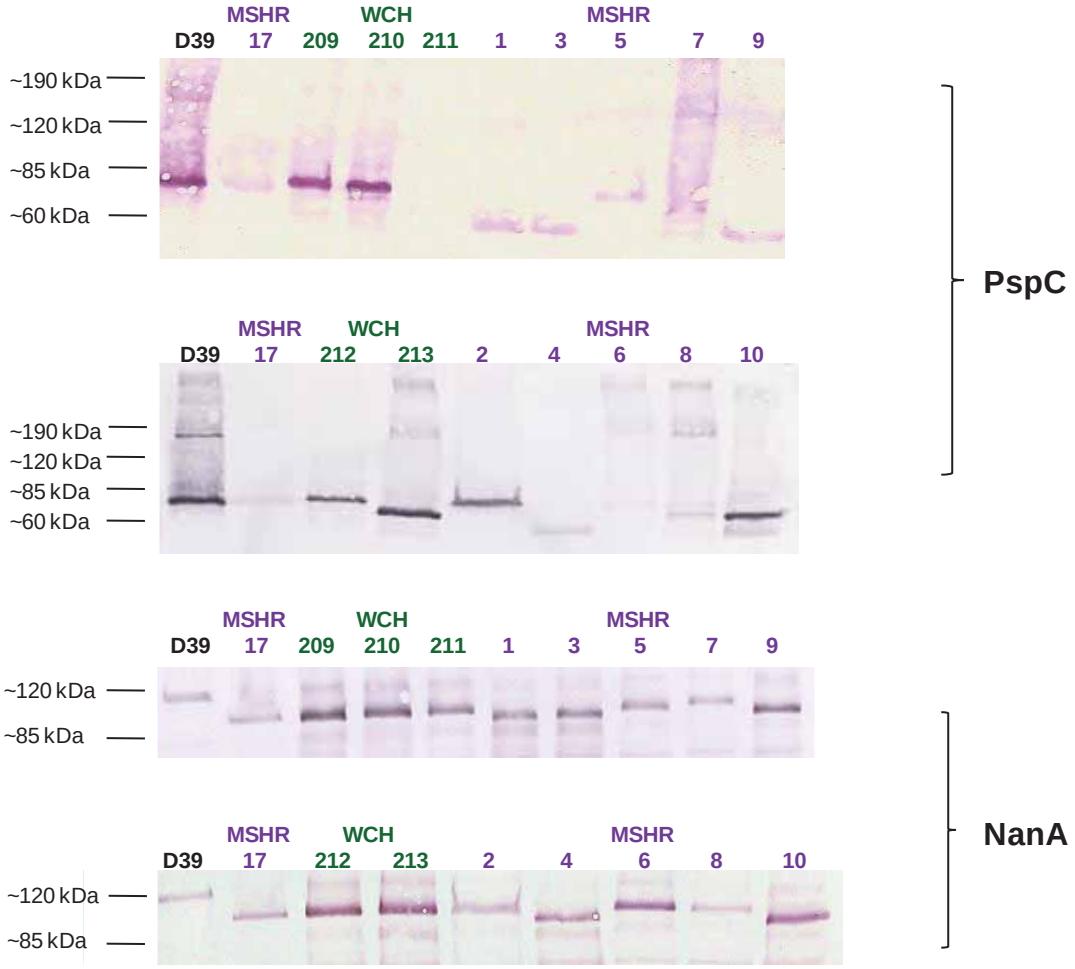


Figure 3.3. Western Blot analysis of D39, serotype 3 strain MSHR17 and all serogroup 11 clinical isolates from the MSHR and WCH using polyclonal PspC and NanA antisera.

Lysates of the strains, which are listed in Table 2.1, were separated by SDS-PAGE, electroblotted onto nitrocellulose and probed with mouse polyclonal antisera, as described in Section 2.3. MSHR strains are written in purple and WCH strains are written in green. The mobility size marker (Invitrogen Benchmark™ Pre-stained ladder) is also indicated.

were kindly supplied by D.E. Briles. PCR family typing of strains without a detectable Western blot band (MSHR5, WCH210, WCH211 and WCH213) revealed that they contained PspA belonging to Family 2. In addition, MSHR17 and MSHR1, which both produced bands on the Western blot, were verified to carry Family 1 PspA. This confirmed that the anti-serum recognised Family 1 PspA but not Family 2.

Immunoreactive bands of varying intensity were visible on the PspC Western blot (Figure 3.3). Interestingly, a very faint immunoreactive band was detectable on the serogroup 11 blot for MSHR17 that was not seen on the original blot (Figure 3.1). This is likely to be due to a different batch of polyclonal mouse anti-PspC serum being used on these blots. The position of the immunoreactive band was higher than expected given the PspC of MSHR17 is approximately half the molecular size of the PspC of D39. This faint immunoreactive band may be due to crossreactivity with another protein in MSHR17, and therefore, the faint immunoreactive bands visible for other strains may also be due to an immunoreaction with another protein and not PspC.

Sequencing was performed on *pspC* genes of selected strains and genes belonging to a variety of groups in the classification system developed by Iannelli *at al.* (2002) were detected. The *pspC* of WCH211 belonged to group 4, strain MSHR10 belonged to group 3 and strains WCH210, WCH213 and MSHR1 belonged to group 6. All attach to the cell wall via a choline binding domain. Strain MSHR5 harboured two *pspC* genes; one with a choline binding motif (group 3) and the other with an LPxTG binding motif (group 9). Surprisingly, only a faint immunoreactive band was visible for MSHR5 on the PspC Western blot. It is conceivable that the antiserum cannot recognise group 9 PspC, but the PspC of D39 also belongs to group 3 (Iannelli *et al.*, 2002) and therefore, a strong immunoreactive band was also expected for the group 3 PspC of MSHR5. Similarly, strong immunoreactive bands were visible for WCH210 and WCH213 but not MSHR1,

even though all three strains possess *pspC* alleles belonging to group 6. It is possible that there is very low or no expression of *pspC* in MSHR1 or MSHR5, or the proteins were degraded.

The NanA in MSHR17 and all the serogroup 11 isolates had a higher mobility on SDS-PAGE than the NanA of D39 (Figure 3.3). Some strains, such as WCH211, appeared to possess a NanA with the same mobility as MSHR17. Other strains, such as MSHR1, clearly have a NanA with a lower mobility than the NanA of MSHR17. However, when the *nanA* genes of WCH211 and MSHR1 were amplified using NanA Forward and NanA Reverse, identical sized product was obtained. Sequencing of the *nanA* of WCH211 revealed a gene that was identical to the *nanA* identified in MSHR17. Sequencing of the gene in MSHR1 revealed a gene that is a cross between the *nanA* of D39 and MSHR17. While the 15 bp insertion and 180 bp deletion were present, the 400bp region of diversity was not. In other words, it was identical to D39 at the 5' terminus and MSHR17 at the 3' terminus. Like the NanA of MSHR17, the NanA of MSHR1 was predicted to be 110 kDa but demonstrated higher mobility than expected, which in this case was even higher than MSHR17. Again this variation may reflect the difference in sequence or post-translational proteolysis.

A summary of sequencing results for *nanA*, *pspA* and *pspC* in serotype 3 and serogroup 11 is shown in Figure 3.4.

3.2.1.3 PCR analysis of *nanB* and *nanC* in selected serotype 3 and serogroup 11 strains

In addition to *nanA, nanB* and *nanC* were analysed by PCR using *nanB*(F) and *nanB*(R) primers, as well as *nanC*(F) and *nanC*(R) primers, designed based on the flanking regions of these genes (Table 2.2). In serotype 3, the strains WCH206, WCH207, MSHR11 and MSHR17 were analysed, as well as WU2. WCH206 and WCH207

Figure 3.4. Results of the sequencing of *nanA*, *pspA* and *pspC* in selected serotype 3 and serogroup 11 strains.

Western Blot analysis indicated diversity in NanA, PspA and PspC within both serotype 3 and serogroup 11, which was confirmed by sequencing in strains of interest and is summarised in this figure. NanA is represented by lilac. The other two NanA variants identified are identical to this NanA except for a 15 bp insertion (indicated by a purple box) and a 180 bp deletion (indicated by a black line). In addition, one variant had a 500 bp region of diversity at the N-terminus (indicated by a pink box). Family 1 of PspA is represented by magenta and Family 2 by dark purple. The PspC proteins are adapted from Iannelli *et al.* (2002). Homologous sequences are represented by boxes of the same colour. Black = signal peptide, grey = random coil, green = proline rich region, blue = anchor, other colours = α-helix.

displayed identical protein profiles in Section 3.2.1.1 but were different to MSHR11 and MSHR17, which displayed identical protein profiles to one another. In serogroup 11, the strains WCH211, WCH213, MSHR1 and MSHR5 were chosen, which all displayed different protein profiles to one another in Section 3.2.1.2.

The gene *nanB* was detected in all serotype/group 3 and 11 strains. However, although *nanC* could not be detected in any of the serotype 3 isolates, it was found to be carried in all serogroup 11 isolates except MSHR1 with the primers used in this study.

3.2.2 Haemolysis assay

The function of Ply is described in Section 1.3.3.4. Although it is lytic to all animal cells containing cholesterol in their cell membrane, erythrocytes are particularly susceptible (Paton *et al.*, 1993). However, several *ply* alleles have been associated with nonhaemolytic activity (Jefferies *et al.*, 2007). Interestingly, strains exhibiting nonhaemolytic activity are still capable of causing IPD (Kirkham *et al.*, 2006), but the effect it has on noninvasive disease has not been established yet. Nonhaemolytic Ply proteins tend to display reduced mobility on SDS-PAGE, although there are some alleles which are not associated with altered mobility (Jefferies *et al.*, 2007). Therefore, even if strains do not exhibit differences in Ply by Western blot analysis, it is still possible that there may be differences in haemolytic activity.

Haemolysis assays were performed as described in Section 2.4.4. Strains were chosen based on their genetic diversity demonstrated by Western blot analysis along with another serotype 3 strain, WU2, and D39. Each strain was grown in duplicate in THY on separate days. Furthermore, the haemolysis assay was performed in duplicate for each lysate. Therefore, there were four readings in total for each strain. Haemolytic titre was

determined as the $\log_2$ value of the reciprocal of the estimated dilution at which 50% of erythrocytes were lysed at A_{540}.

Results are displayed in Tables 3.1 and 3.2. All pneumococci were haemolytic. The haemolytic activity within serotype 3 was identical (about 20 haemolytic units per ml of lysate). The haemolytic activity for D39 and the serogroup 11 isolates ranged from about 30-40 haemolytic units per ml of lysate, and therefore, there was only about a 1.5-2 fold difference between serogroup 11 and serotype 3.

Table 3.1. Haemolytic units per ml of lysate for selected serotype 3 isolates

WU2	WCH206	WCH207	MSHR11	MSHR17
22	17	18	20.5	18.5

Table 3.2. Haemolytic units per ml for D39 and selected serogroup 11 isolates

D39	WCH211	WCH213	MSHR1	MSHR5
31	39.5	33.5	29	43.5

3.2.3 MLST typing

MLST is a portable, accurate and highly discriminating method for typing bacteria and other organisms (Enright and Spratt, 1999; Enright and Spratt, 1998). It has proven to be a useful tool in epidemiological studies of pneumococci as outlined in Section 1.3.2. Pneumococcal MLST entails the sequencing of sections of seven housekeeping genes: *aroE* (shikimate dehydrogenase), *gdh* (glucose-6-phosphate dehydrogenase), *gki* (glucose kinase), *recP* (transketolase), *spi* (signal peptidase I), *xpt* (xanthine phosphoribosyltransferase), and *ddl* (D-alanine-D-alanine ligase), which were chosen

because they are conserved and mutate slowly overtime (Enright and Spratt, 1999). The MLST genes are PCR amplified and sequenced using primers whose sequences are available on the MLST website (http://spneumoniae.mlst.net), which is hosted by the Imperial College, London and funded by the Wellcome Trust, and are listed in Table 2.2 (*aroE*-up to *xpt*-dn). The sequences are then assigned an allele number using the aforementioned website, and the allelic profile (analogous to a barcode) is then used to quickly determine ST. Information about other isolates with the same or related ST can also be obtained from the website.

Most epidemiological studies investigating ST have focused on IPD isolates. Furthermore, there is very little published work on ST distribution in Australia. Determining the STs of middle ear isolates from remote Aboriginal communities is of particular importance due to the huge burden of OM in these communities. Therefore, selected serotype 3 clinical isolates, as well as those of serogroup 11, were genotyped. Initially WCH206, WCH207, MSHR11, MSHR17, WCH211, WCH213, MSHR1 and MSHR5 were chosen for MLST typing based on the results of the Western Blot analysis. WCH206 and WCH207 were chosen because they differ in their protein profile from the Northern Territory isolates. Similarly, the serogroup 11 WCH211, WCH213, MSHR1 and MSHR5 were chosen because they all had different protein profiles to one another. Later MSHR12 and MSHR14 were also typed to gain a better understanding of STs in Aboriginal OM, as well as WCH208, which displayed an identical protein profile by Western blot analysis to the MSHR isolates, and WU2, which was used in some studies and did not have an ST listed on the MLST website. Results are listed in Table 3.3.

The strains WCH206 and WCH207, which exhibited identical protein profiles in the Western blot analysis (Section 3.2.1), had the identical ST, which was ST180. This ST

Table 3.3. STs of Selected Serotype 3 and Serogroup 11 Isolates.

Strain	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	ST
WU2	13	9	15	14	10	16	19	**378**
WCH206	7	15	2	10	6	1	22	**180**
WCH207	7	15	2	10	6	1	22	**180**
WCH208	2	32	9	47	6	21	17	**458**
MSHR11	2	32	9	47	6	21	17	**458**
MSHR12	2	32	9	47	6	21	17	**458**
MSHR14	2	32	9	47	6	21	17	**458**
MSHR17	2	32	9	47	6	21	17	**458**
WCH211	2	13	2	66	9	1	5	**3020**
WCH213	2	5	29	12	16	3	14	**62**
MSHR1	2	14	4	12	1	14	8	**3021**
MSHR5	5	35	29	12	9	39	18	**662**

Serotype 3 strains are indicated in blue, while serogroup 11 strains are indicated in red. Allele numbers are shown for each gene, along with the ST.

is also referred to as Netherlands^{3}-31 in the literature. The four serotype 3 strains typed from the MSHR, which also displayed identical protein profiles to one another in Section 3.2.1, had the same ST as well, which was ST458. WU2 was found to belong to ST378.

As predicted from the divergent protein profiles (Section 3.2.2), WCH211, WCH213, MSHR1 and MSHR5 all had different STs and shared 1 or 2 identical alleles at the most. WCH13 was ST62, while MSHR5 was ST662. The allelic profiles for WCH211 and MSHR1 were not listed on the database at the time of sequencing. Therefore, their allelic profiles were submitted to the MLST website and they were assigned ST3020 and 3021, respectively.

3.2.4 Uronic acid assay

Serotype 3 strains characteristically form large, mucoid colonies when grown on BA. The strains WCH206 and WCH207 produced large, homogenous colonies on BA. However, the other strains produced colonies that were comparatively smaller and heterogenous. It was not known whether this was because of differences in capsule production or growth on BA. The uronic acid assay was used to compare capsule production between selected strains.

The level of capsule production has been known to be an important contributor to virulence for sixty years (Macleod and Kraus, 1950). Studies with serotype 3 capsule expression mutants have shown that at least 6% of the parental level of capsule is needed to colonise with approximately 20% required to colonise mice as effectively as the parent (Magee and Yother, 2001). Mutants of serotype 3 strains expressing only 20% of the

amount of the parental capsule were still virulent in mouse intraperitoneal models (Magee and Yother, 2001).

Serotype 3 strains are surrounded by a cellubiuronic acid capsule and therefore, capsule production was compared in the serotype 3 strains WCH206, WCH207, MSHR11, MSHR17 and WU2 using the uronic acid assay as described in 2.13. Capsule production in the serogroup 11 strains could not be measured using this assay due to the absence of uronic acid in their capsular polysaccharide. Results are shown in Figure 3.5. Capsule production was compared between strains using the Mann Whitney *U*-test (two-tailed) and *P* values are shown in Table 3.4.

There was no difference in capsule production between strains of the same ST. However, the strains MSHR11 and MSHR17 produced a statistically significant higher level of capsule than WCH206 and WCH207, although WCH206 and WCH207 still produced 60% of the amount of capsule as strains MSHR11 and MSHR17. Interestingly, the large mucoid colonies produced by WCH206 and WCH207 on BA compared to MSHR11 and MSHR17 do not seem to be reflective of total capsule production measured by the uronic acid assay.

3.2.5 Animal studies

Selected serotype 3 and serogroup 11 strains were tested in intraperitoneal (i.p.), intranasal (i.n.) and pneumonia/sepsis mouse models in order to evaluate their virulence properties and their capacity to colonise various niches. The serotype 3 strains used were WU2, WCH206, WCH207, MSHR11, MSHR17, along with the serogroup 11 strains

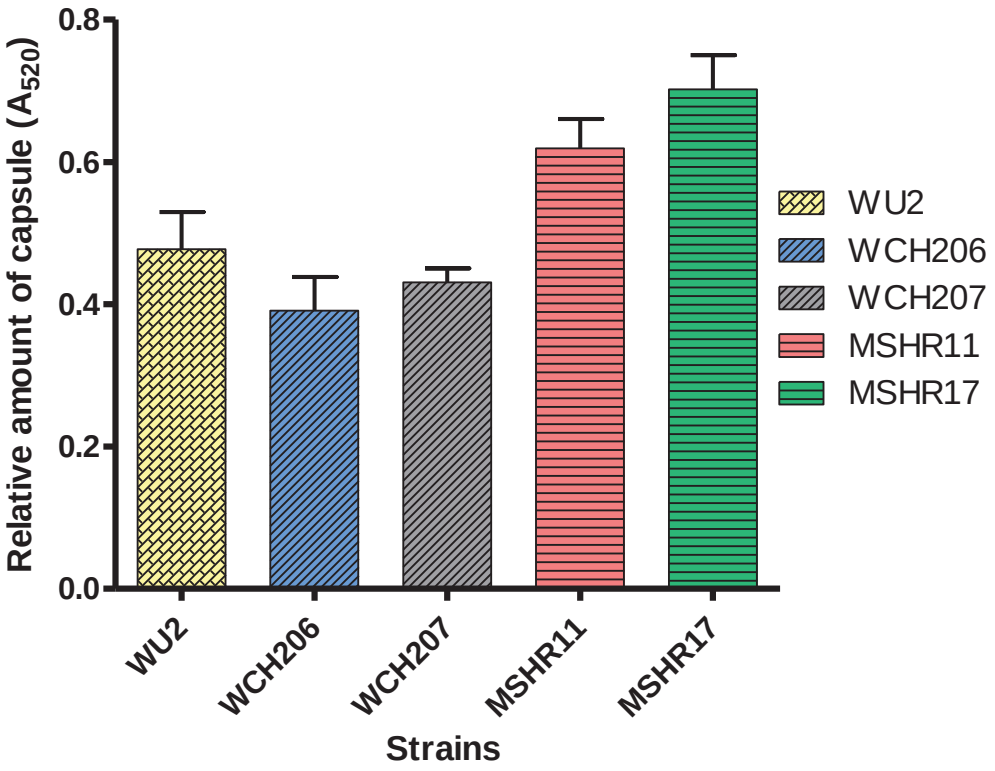


Figure 3.5. Capsule production by selected serotype 3 strains.

Pneumococci were grown on O/N BA plates and resuspended in PBS to A_{600} = 0.5. Capsule production was measured using the uronic acid assay as described in Section 2.13. Data are mean ± SD A_{520} for quintuplicate experiments. Isolates of the same ST are represented by an identical pattern.

Table 3.4. Statistical comparison of capsule.

	WU2	WCH206	WCH207	MSHR11	MSHR17
WU2	-	NS	NS	NS	0.0155
WCH206	NS	-	NS	0.0317	0.0079
WCH207	NS	NS	-	0.0079	0.0079
MSHR11	NS	0.0317	0.0079	NS	-
MSHR17	0.0159	0.0079	0.0079	NS	-

Uronic acid levels (A_{520}) were compared using the Mann Whitney *U*-test (two-tailed). *P* values are specified. NS = not significant ($P > 0.05$).

WCH211, WCH212, MSHR1 and MSHR5. These strains were chosen to ensure a genetically diverse range of isolates as possible.

3.2.5.1 Virulence models

The pneumonia/sepsis and i.p. models were used to assess the strains' abilities to cause IPD, and were performed as described in Section 2.12.1. Bacteria were inoculated in the opaque phase, because this phase has been shown to be more virulent in systemic disease in the mouse, compared to the transparent phase (Kim and Weiser, 1998).

In the pneumonia/sepsis model, ten mice were inoculated intranasally under anaesthesia with 1.5-4 × 10^7 CFU/mouse. The duration of the experiment was 4 weeks. At the end of the study, five mice from the WU2, WCH211, WCH213, MSHR1 and MSHR5 groups, as well as the remaining mice in the other groups, were euthanased and a nasal wash performed and blood taken as described for the colonisation study (Section 2.12.3).

In the i.p. model, bacteria were injected directly into the peritoneal cavity. Therefore, this model assessed virulence without the need of the bacteria to initially infect the lungs prior to translocation to the bloodstream. Two differing doses were used, 5 × 10^2 CFU for serotype 3 and 1 × 10^7 CFU for serogroup 11. The course of the experiment was 3 weeks.

Serotype 3 strains were highly virulent compared to serogroup 11 strains in both the i.p. and pneumonia/sepsis models (Figures 3.6 and 3.7). A dose of 1 x 10^7 CFU per mouse was not large enough to kill 50% of mice for serogroup 11 in the i.p. model and no deaths were recorded using the pneumonia/sepsis model. In contrast, more than 50% of mice died within 62.5 hours, when inoculated with 5 x 10^2 CFU of serotype 3 in the i.p.

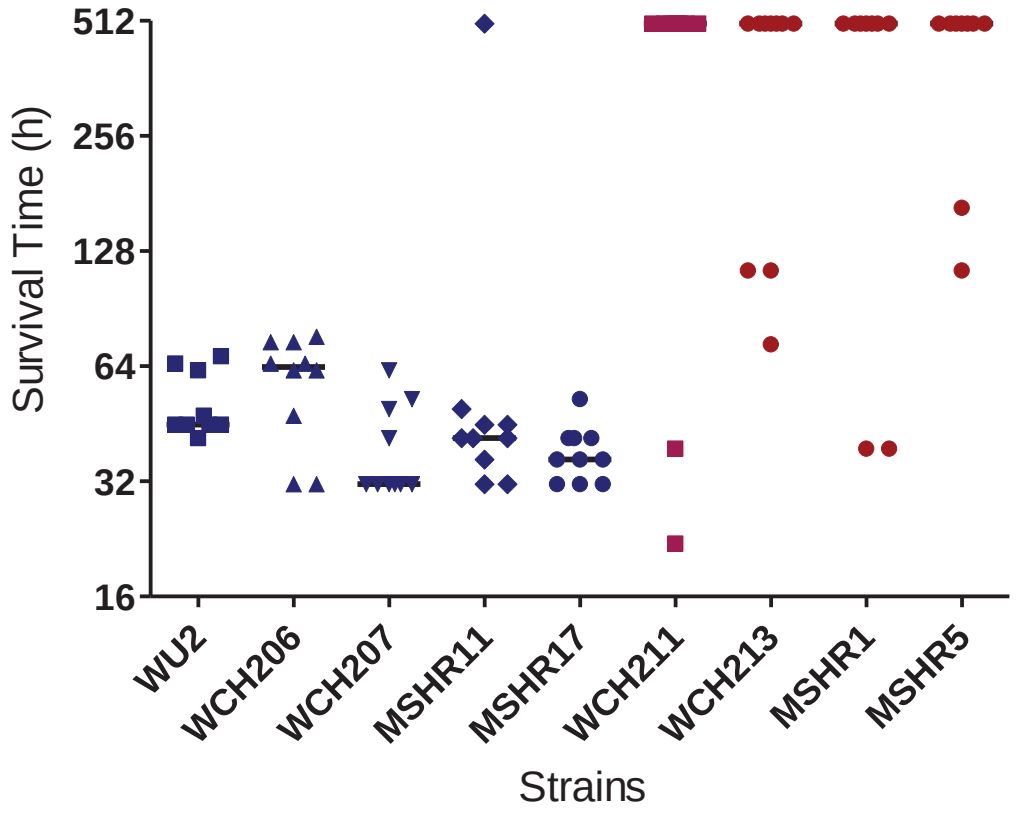


Figure 3.6. Murine intraperitoneal challenge with serotype/ group 3 and 11 strains.
Groups of 10 Swiss mice were injected with either 5 x 10^2 CFU (serotype 3 strains) or 1 x 10^7 CFU (serogroup 11 strains) of bacteria. Survival was monitored over a 3 week period. Survival time for each mouse is shown (log_2 scale) and the horizontal bar indicates the median survival time for each group.

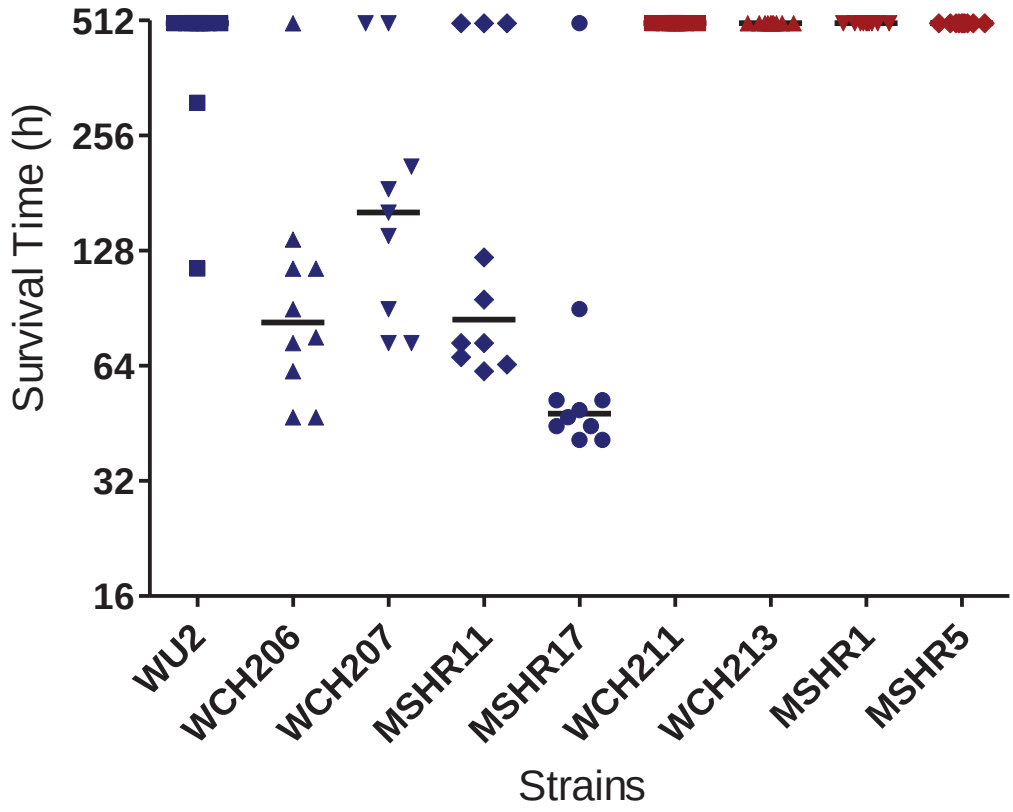


Figure 3.7. Murine pneumonia/sepsis model with serotype/ group 3 and 11 strains.
Groups of 10 Swiss mice were intranasally inoculated under anaesthesia with 1.5 – 4 x 10^7 CFU of the strains indicated. Survival was monitored over a 3 week period and the median survival time, which is represented in the graph by black bars, was calculated.

model. In the pneumonia sepsis model, more than 50% of mice had died within 57 hours, with the exception of the group inoculated with WU2 in which only two out of ten mice died.

When the median survival time was compared between serotype 3 strains using the Mann Whitney *U*-test (Tables 3.5 and 3.6), strain to strain variation was observed. In particular, WU2 was less virulent than the other strains in the pneumonia/sepsis model. It is interesting to note that there were also differences between strains with the same STs. There was significant difference between MSHR11 and MSHR17 in the pneumonia/sepsis model and between WCH206 and WCH207 in the i.p. model. In addition, there were significant differences between WU2 and WCH207, and WU2 and MSHR17, but not between WU2 and WCH206 or MSHR11.

At the end of the pneumonia/sepsis study, bacteria were not detectable in the blood, but were detected in the nasal washes of mice infected by both serotypes with the level of colonisation indicated in Table 3.7. This indicates that both serotypes can be carried for at least 4 weeks in the mouse.

3.2.5.2 Colonisation model

The i.n. model was used to investigate colonisation of serotype3 and serogroup 11 strains in the nasopharynx, lungs, blood and ears over a seven day period, as described in Section 2.12.3. Transparent phase bacteria were used because bacteria in this phase have been reported to colonise the nasopharynx better than bacteria in the opaque phase (Weiser *et al.*, 1994; Kim and Weiser, 1998). Twenty mice were intranasally inoculated with approximately 1-2 x 10^7 CFU. Mice were sacrificed 1, 2, 4 and 7 days post-inoculation and the number of bacteria on the surface of the nasopharynx (nasal wash), as well as in the nasopharyngeal tissue (nasal cut), lungs, blood, ears and brain were determined.

Table 3.5. Statistical analysis of the murine intraperitoneal study for serotype 3.

	WU2	WCH206	WCH207	MSHR11	MSHR17
WU2	-	NS	0.0355	NS	0.0015
WCH206	NS	-	0.0147	NS	0.0147
WCH207	0.0355	0.0147	-	NS	NS
MSHR11	NS	NS	NS	-	NS
MSHR17	0.0015	0.0147	NS	NS	-

Survival time data were analysed using the Mann Whitney *U*-test (two-tailed). *P* values for each pairwise comparison are indicated. NS= not significant ($P > 0.05$).

Table 3.6. Statistical analysis of the pneumonia/ sepsis study for serotype 3.

	WU2	WCH206	WCH207	MSHR11	MSHR17
WU2	-	0.0007	0.0172	0.0185	0.0003
WCH206	0.0007	-	NS	NS	0.0355
WCH207	0.0172	NS	-	NS	0.003
MSHR11	0.0185	NS	NS	-	0.0039
MSHR17	0.0003	0.0355	0.003	0.0039	-

Survival time data were analysed using the Mann Whitney *U*-test (two-tailed). *P* values for each pairwise comparison are indicated. NS= not significant ($P > 0.05$).

Table 3.7. Colonisation in the pneumonia/sepsis model 4 weeks post-challenge.

Infection Strain	CFU in 1 ml of nasal wash
WU2	80
WU2	40
WU2	200
WU2	160
WU2	160
WCH206	9000
WCH207	400
WCH207	80
MSHR11	120
MSHR11	21000
MSHR11	20
MSHR17	25000
WCH211	1080
WCH211	1040
WCH211	440
WCH211	20
WCH211	21000
WCH213	20
WCH213	200
WCH213	120
WCH213	40
WCH213	120
MSHR1	1000
MSHR1	360
MSHR1	8000
MSHR1	18000
MSHR1	20
MSHR5	5000

MSHR5	80
MSHR5	5000
MSHR5	160
MSHR5	20

Bacteria were isolated from the nasopharyngeal surface of mice that survived the 4 week duration of the pneumonia/sepsis study. The nasal wash was performed as described Section 2.12.3. The CFU in 1 ml of nasal wash is indicated. Serotype 3 strains are indicated by blue writing, while serogroup 11 strains are indicated by red writing. 20 represents the limit of detection.

Serogroup 11 was isolated from each site more frequently than serotype 3, apart from the blood, from which only serotype 3 was isolated (Figure 3.8, Table 3.8). On Days 1, 2 and 4, serogroup 11 was at statistically significantly higher levels in the nasopharynx than serotype 3, although the levels of serotype 3 isolated from the nasopharynx increased over time (Figure 3.8).

3.3 Capsule Switch Study

As previously mentioned, serotype 3 is commonly isolated from the middle ears of children in remote Aboriginal communities, while serotype 11A is not. Although serotype 11A has been reported in OM in other communities (Porat *et al.*, 2004; Hanage *et al.*, 2004; Watson *et al.*, 2007), it is serotype 3 that causes more concern (Prymula, 2009). In Section 3.2, basic characterisation experiments were conducted on clinical isolates of serogroup 11 and serotype 3. Serotype 3 strains of three unrelated genetic lineages were highly virulent in i.p and pneumonia/sepsis murine models; however, four serogroup 11 strains of unrelated genetic lineages were avirulent in these models. Nonetheless, the same serogroup 11 strains were isolated at statistically higher numbers compared to the serotype 3 isolates from the nasopharynx on Days 1 and 2 post-challenge in a murine colonisation model. Therefore, it was wondered if the difference in virulence between serotype 3 and serogroup 11 was simply due to capsule alone.

In order to investigate the influence of serotype on virulence, a serotype 11A isolate from the MSHR was converted to serotype 3. The reverse was not performed due to time constraints within this study. It is very simple to select for serotype 11A isolates that have been transformed into serotype 3 due to the large, mucoid colonies this serotype characteristically forms on BA. However, this makes selecting for a serotype 3 that has been switched to serotype 11A difficult and therefore, transforming a serotype 3 into

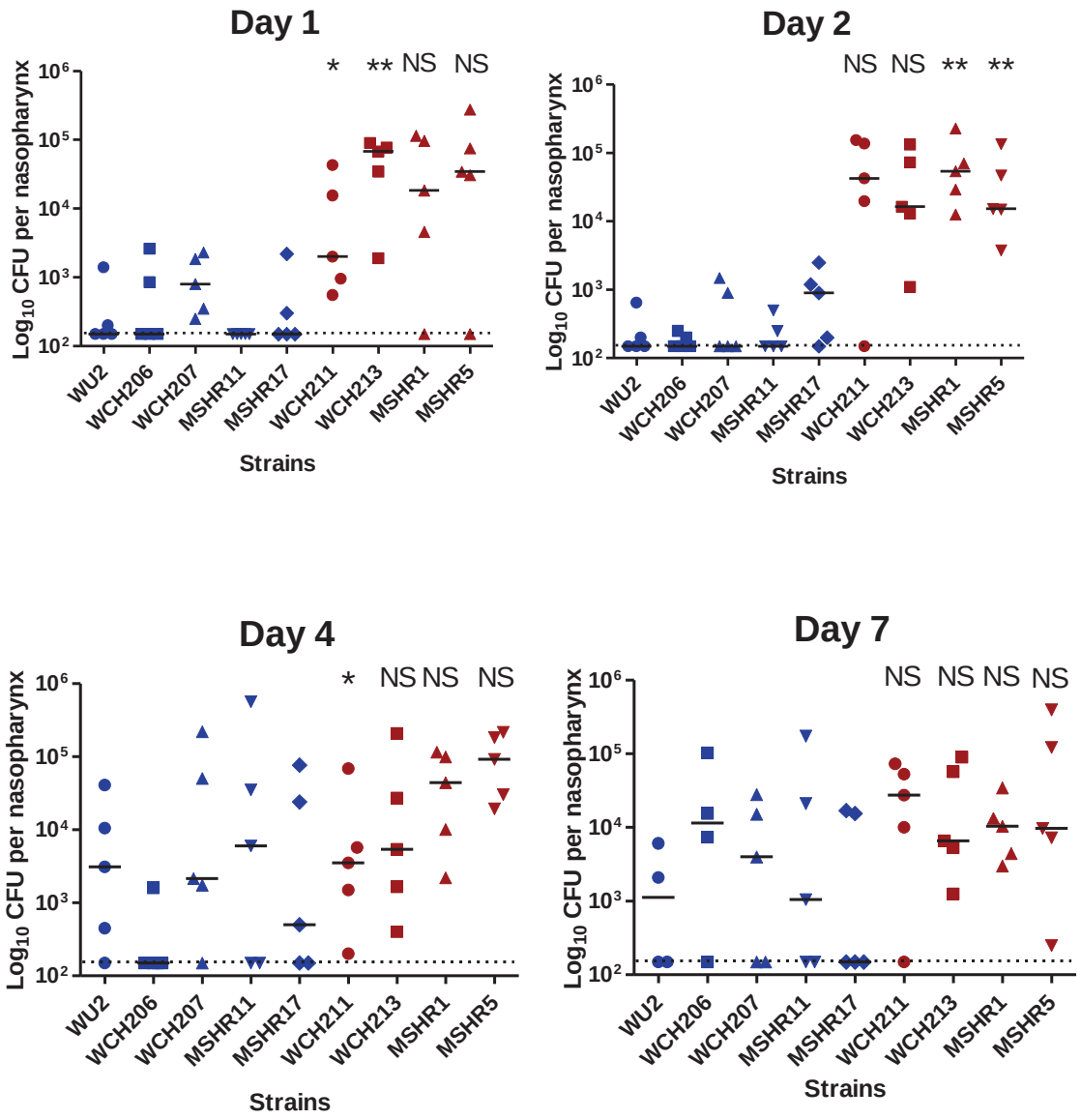


Figure 3.8. Colonisation of the Nasopharynx by Serotype/group 3 and 11 strains.

Groups of 5 mice for each strain were euthanased 1, 2, 4 and 7 days post-inoculation. The nasopharynx was washed and then removed and homogenised in PBS. The number of pneumococci in both the nasal wash and homogenate of the nasal tissue were ascertained and combined to calculate the total number of bacteria in the nasopharynx. The median number of bacteria in the nasopharynx for each strain is represented by a black bar. The broken line represents the limit of detection, which is 40 organisms. The Mann Whitney *U*-test was performed on all serogroup 11 strains versus WU2. *= $P < 0.05$, **= $P < 0.01$, NS= Not Significant.

Table 3.8. Frequency of isolation of bacteria from the blood, brain, lungs and ears over a 7 day period in the i.n. model.

Serotype 3					
	WU2	**WCH206**	**WCH207**	**MSHR11**	**MSHR17**
Blood	1/19	0/19	1/20	2/20	1/20
Brain	1/19	0/19	3/20	1/20	0/20
Lungs	1/19	0/19	1/20	1/20	1/20
Ears	0/19	3/19	4/20	1/20	2/20
Serogroup 11					
	WCH211	**WCH213**	**MSHR1**	**MSHR5**	
Blood	0/20	0/20	0/20	0/20	
Brain	3/20	3/20	5/20	6/20	
Lungs	1/20	4/20	2/20	1/20	
Ears	7/20	6/20	8/20	3/20	

Serotype 3 strains are indicated by blue, while serogroup 11 strains are indicated by red.

serotype 11A is a more complicated process, requiring the introduction of antibiotic resistance genes for selection.

3.3.1 Creation of a capsule switch mutant

The capsule locus of the serotype 3 strain WCH206 was amplified by PCR using the Expand Long Range PCR kit (Section 2.5.5) and the primers *cps* locus F and R (Table 2.2). This product was used to transform competent MSHR5 (serotype 11A, ST662) cells as described in Sections 2.11.1 and 2.11.2. A large, mucoid colony was picked from a BA plate for further analysis, MSHR5^3.

When a Quellung reaction was performed, MSHR5^3 reacted with serotype 3 antiserum like the capsule donor strain WCH206, but no longer reacted with serogroup 11 antiserum, to which the parent MSHR5 reacted (Section 2.2.4). The background of MSHR5^3 was also confirmed by sequencing its *aroE* allele as described in Section 3.2.3. The strain WCH206 carries *aroE* 9, while MSHR5 carries *aroE* 5 and was the only isolate sequenced in this study to carry this allele (Table 3.3). MSHR5^3 carried *aroE* 5 confirming that the genetic background of MSHR5^3 is still the same as the parent MSHR5.

Colonies of MSHR5^3 were not as large as the colonies of WCH206 when grown on BA and were more similar in appearance to the colonies of WU2 and MSHR17. A uronic acid assay was performed as described in Section 2.13 to compare capsule production between the capsule donor strain WCH206 and the capsule switch mutant MSHR5^3. The capsule switch variant produced about 70% of the level of capsule as the parent. The difference may be due to genomic differences in WCH206 and MSHR5^3 outside the capsule locus.

3.3.2 Growth curve of the capsule switch variant

A growth curve was constructed in BHI as described in Section 2.2.6 and is shown in Figure 3.9. The strains MSHR5^3, MSHR5 and WCH206 were used, as well as WU2 and MSHR17, which were included for further comparison with other serotype 3 strains. No difference in growth rate was detectable between the strains MSHR5^3, MSHR5, WCH206 (Figure 3.9). However, WU2 did not reach as high an OD_{600} as these strains, and MSHR17 grew slower.

3.3.3 Pneumonia sepsis model

The impact of capsule switching on virulence was then investigated using the pneumonia sepsis model, which is described in Section 2.12, and the strains MSHR5^3, MSHR5 and WCH206. Results are shown in Figure 3.10 with statistical analysis shown in Table 3.9. Retrospective plating indicated that the mice had been infected with 0.7-1.3 × 10^7 CFU/mouse. Instead of measuring survival as in Section 3.2.4.1, groups of five mice were dissected 24 h, 48 h and 72 h post-inoculation to examine colonisation of the nasopharynx, ears, lungs, blood and brain. However, the MSHR5 group contained 14 mice and therefore, only 4 mice were dissected on the third day.

Using the sepsis model it was found that mice which were challenged with WCH206 and MSHR5 were positive more often for an infection in the brain, lungs and ears compared to the colonisation model (Section 3.2.5.2, Table 3.8). For strain WCH206 this extended to the blood as well. WCH206 was detected in the blood of nine mice but MSHR5^3 like its MSHR5 parent, was not isolated from the blood of any mouse. On Day 2, the number of bacteria in the blood of WCH206 was shown to be statistically significantly higher than either MSHR5 or MSHR5^3 with a *P* value of 0.0259 for both strains.

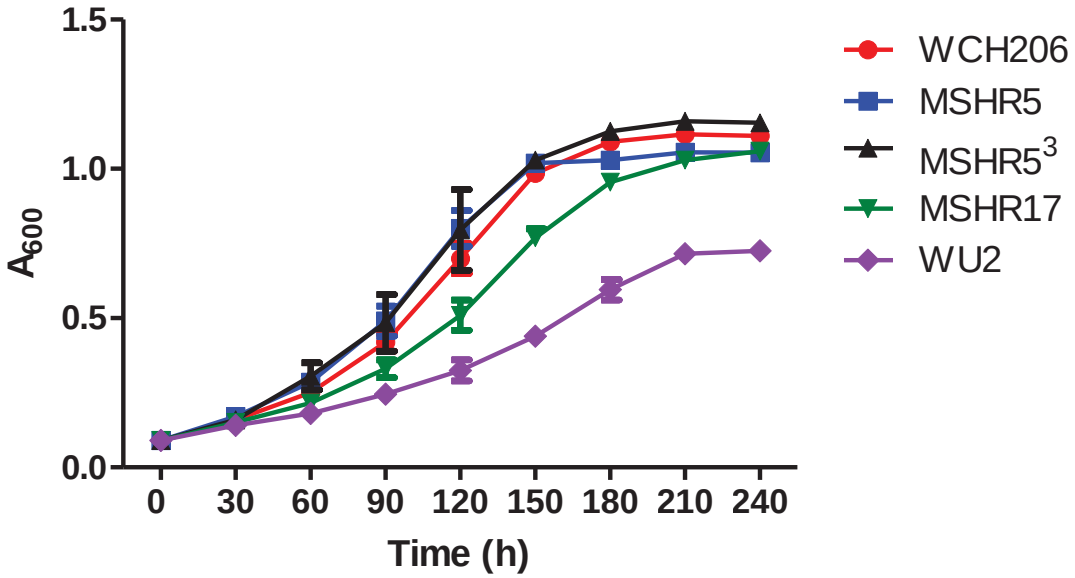


Figure 3.9. Average A_{600} readings for selected pneumococcal strains grown over a 4 h period in BHI. Measurements were taken using an OD_{600} Diluphotometer (Implen, California, USA).

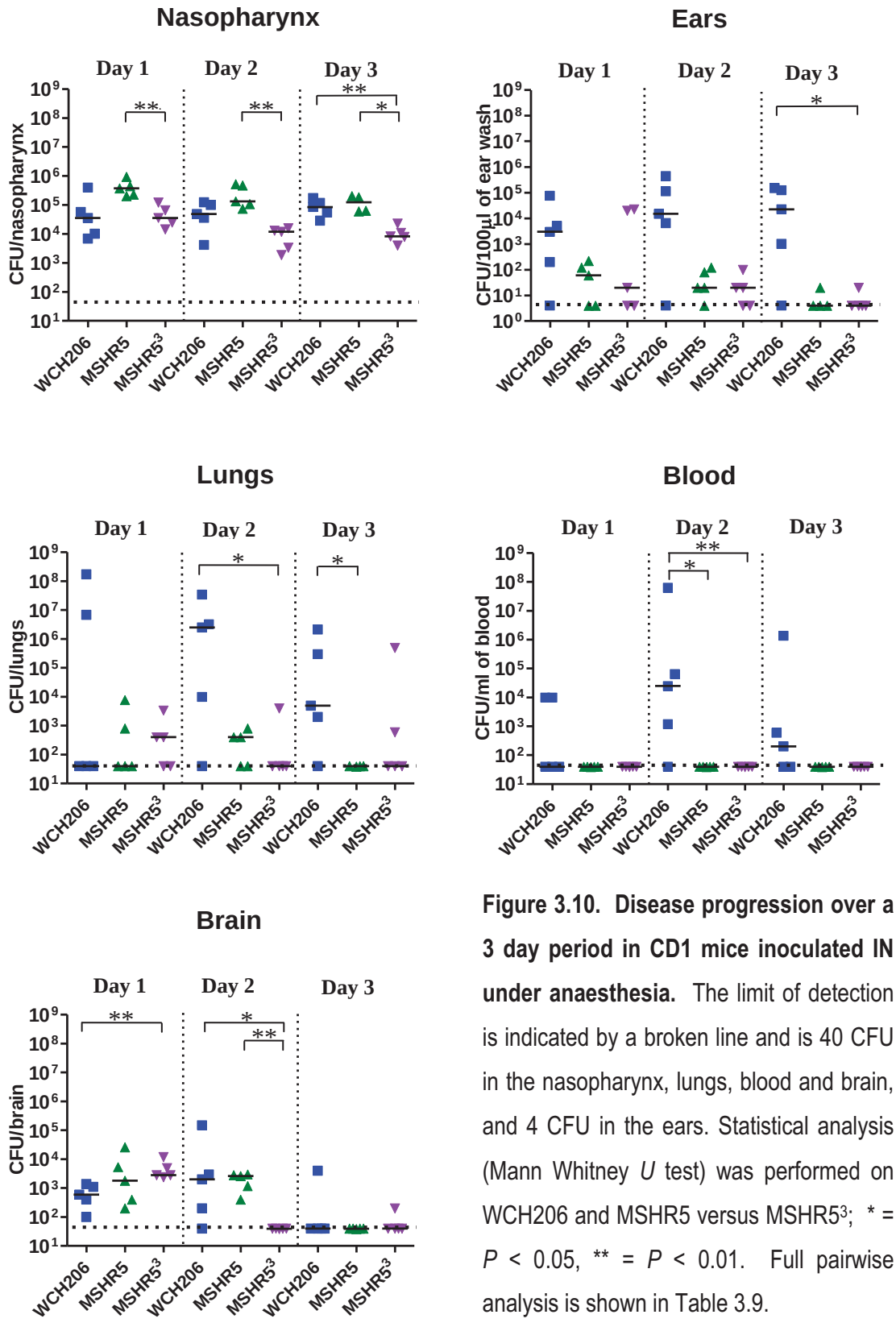


Figure 3.10. Disease progression over a 3 day period in CD1 mice inoculated IN under anaesthesia. The limit of detection is indicated by a broken line and is 40 CFU in the nasopharynx, lungs, blood and brain, and 4 CFU in the ears. Statistical analysis (Mann Whitney *U* test) was performed on WCH206 and MSHR5 versus MSHR5^3; * = $P < 0.05$, ** = $P < 0.01$. Full pairwise analysis is shown in Table 3.9.

Table 3.9. Statistical analysis for the capsule switch study.

Nasopharynx			
	Day 1	**Day 2**	**Day 3**
WCH206/MSHR5[3]	NS	NS	0.0079
MSHR5/MSHR5[3]	0.0079	0.0079	0.0452
WCH206/MSHR5	NS	NS	NS
Ears			
	Day 1	**Day 2**	**Day 3**
WCH206/MSHR5[3]	NS	NS	0.0449
MSHR5/MSHR5[3]	NS	NS	NS
WCH206/MSHR5	NS	NS	NS
Lungs			
	Day 1	**Day 2**	**Day 3**
WCH206/MSHR5[3]	NS	0.0449	NS
MSHR5/MSHR5[3]	NS	NS	NS
WCH206/MSHR5	NS	NS	0.0407
Blood			
	Day 1	**Day 2**	**Day 3**
WCH206/MSHR5[3]	NS	0.0259	NS
MSHR5/MSHR5[3]	NS	NS	NS
WCH206/MSHR5	NS	0.0079	NS
Brain			
	Day 1	**Day 2**	**Day 3**
WCH206/MSHR5[3]	0.0079	0.0259	NS
MSHR5/MSHR5[3]	NS	0.0079	NS
WCH206/MSHR5	NS	NS	NS

Median CFU of bacteria isolated from the nasopharynx, ears, lungs, blood and brain were compared using the Mann Whitney *U*-test (two-tailed) for Days 1, 2 and 7 post-challenge using the pneumonia/sepsis mouse model. *P* values are shown. NS= not significant ($P > 0.05$).

Furthermore, by Day 3, WCH206 was present at a statistically higher number in the ears and nasopharynx than MSHR5^3 (P = 0.0449, P = 0.0079). Although there was no statistically significant difference between WCH206 and MSHR5^3 in the lungs on Day 3, MSHR5^3 was present at a statistically significant lower level than WCH206 on Day 2 (P=0.0449) in this niche. In the brain, a statistically significant lower number of WCH206 was isolated from the brain compared to MSHR5^3 on Day 1, but by Day 2, colonisation of this niche in both WCH206 and MSHR5 was statistically significantly higher than MSHR5^3. However, by Day 3, infection of the brain was below the limit of detection for most mice. Pneumococci have been shown to be capable of entering the brain via axonal transport via olfactory nerves and therefore, independent of their ability to enter the blood (van Ginkel *et al*.,2003).

Overall, the statistics indicate that MSHR5^3 was attenuated in the mouse pneumonia-sepsis model compared to WCH206. Therefore, although it possessed the capsule 3 locus from WCH206, it did not consequently give it the same ability to colonise in the various niches of this model, indicating other genetic factors play a role. However, there were some differences between MSHR5 and MSHR5^3 as well, signifying changing the capsule can have an impact on the ability to colonise. In particular, MSHR5^3 did not colonise the nasopharynx as effectively and was isolated at a statistically significant lower amount at every time point (P=0.0079, P=0.0079, P=0.0452). Additionally, on Day 2, MSHR5^3 was present at a statistically significant lower number than MSHR5 in the brain.

3.4 Discussion

3.4.1 Protein analysis and MLST

All clinical isolates from the WCH and MSHR were analysed by Western blot analysis looking at various known virulence factors. Selected strains were then subjected to haemolysis assays and MLST typing. These data were used to select a representative subset for subsequent more detailed analyses. For the serotype 3 strains, WCH206 and WCH207 displayed identical protein profiles when analysed by Western blotting Blot followed by sequencing of well-known virulence factors, while the strain WCH208 and the MSHR isolates displayed protein profiles that were different to WCH206 and WCH207, but identical to one another. When WCH206 and WCH207 were MLST typed they were found to be ST180, while WCH208, MSHR11, MSHR12, MSHR14 and MSHR17 were all ST458. It is noteworthy that the isolates of the same ST displayed identical protein profiles. Sequencing work on *pspA* suggests that clades are ST-related (Rolo *et al.*, 2009) and the four ST180 isolates in that study were all PspA Clade 3, which is the same as the two ST180 isolates in this study. This may well be the case for other virulence proteins too, and is certainly suggested by the Western blot analysis in this study, although more strains would need to be investigated in order to draw a more definitive conclusion.

The relatedness of the MSHR serotype 3 isolates is also noteworthy. Despite many of the communities the MSHR surveys being remote and the vast distances between communities, as well as the large variation in climate between Central Australia, which is arid, and Northern Australia, which is tropical, the serotype 3 isolates in this study that colonise and cause OM appear to be clonally related. The result is also interesting because although the MSHR isolates typed belong to ST458, in numerous epidemiological studies conducted in the United States, Europe and northern Asia, the dominant ST seen within

serotype 3 in OM, IPD and carriage has been ST180, which is unrelated to ST458 (McEllistrem *et al.*, 2005; Isozumi *et al.*, 2008; Bozdogan *et al.*, 2004; Vestrheim *et al.*, 2008; Porat *et al.*, 2008; Brueggemann *et al.*, 2003). WCH206 and WCH207 both belong to ST180 and there are also over 80 ST180 isolates listed on the MLST database (http://spneumoniae.mlst.net). American studies have reported that ST180 has expanded in both IPD and OM post-Prevenar™ (Beall *et al.*, 2006; McEllistrem *et al.*, 2005). In Japan, which did not approve the implementation of Prevenar™ until 2009, ST180 is also prevalent and is causing concern, because a large proportion of isolates are resistant to erythromycin (Isozumi *et al.*, 2008). Macrolide resistant ST180 is of concern in other Asian countries as well (Chen *et al.*, 2006; Bozdogan *et al.*, 2004) with one report of penicillin resistance as well as erythromycin resistance in Taiwan (Chen *et al.*, 2006). Interestingly, however, ST180 has a low association with invasive disease (Brueggemann *et al.*, 2003).

ST458 has been reported in Portugal (IPD) and Israel (noninvasive infection), although at levels lower than ST180 or related STs (Porat *et al.*, 2008; Serrano *et al.*, 2005). However, in South Africa, in a study involving the typing of just over 100 serotype 3 isolates, this ST represented 35% of isolates, while ST180 only represented 14% (Mothibeli, 2007). There are only 11 ST458 isolates listed on the MLST database to date and all but one come from Africa with most submitted from Ghana. All isolates belonged to serotype 3, except one from Egypt that had been serotyped as 15C.

Since almost all ST458 isolates on the MLST database are from Africa, and the South African study finding that ST458 dominates over ST180, it is easy to speculate that there may be a link between poverty and ST. Much of the published ST distribution studies come from Europe and the United States. Thus further studies done in developing countries and high-risk communities are needed to determine whether there is a link

between low socio-economic status or geographic region and ST. In particular, more isolates from the general population of Australia need to be typed, in addition to isolates from Aboriginal communities, for comparison.

WU2, which is a subline of an American clinical isolate from the 1960's (Briles *et al*., 1981), was ST378. This ST was associated with a conjunctivitis outbreak at an American military training site in the 1980's of non-typeable *S. pneumoniae* (Shayegani *et al.*, 1982; Crum *et al.*, 2004). A clinical isolate from Poland in 2000 and a clinical isolate from the Czech Republic in 1997, which both belonged to serotype 3, were listed as this ST in the MLST database. Therefore, this ST seems to have been present for a few decades, but does not seem to be as abundant as other serotype 3-associated STs, such as ST180.

In contrast to serotype 3, Western blot analysis of virulence proteins showed that there appears to be a greater genetic diversity within the (albeit small) collection of serogroup 11 isolates used in this study. Four strains were chosen based on the diversity of their protein profiles and found to belong to four different STs, two of which were novel (ST3020 for WCH211 and ST3021 for MSHR1). ST178 is the closest ST to ST3021 and differs by one allele. There has only been one ST178 isolate submitted to the MLST database to date, which was a 19F strain isolated in Australia in 1991. The closest STs in the database to ST3020 all differed by 3 or 4 alleles and belonged to different serotypes, such as 24F, 15 and 6B. Strain MSHR5 is ST662 and as with ST180, this ST has been reported to have expanded after the introduction of Prevenar™, although in serotype 33F (McEllistrem *et al.*, 2005). On the MLST website, this ST has also been seen in a NT isolate from a pneumonia patient, as well as from a serotype 33F and a serogroup 33 carriage isolates.

WCH213 is ST62, which has been shown to have a much higher association with carriage than IPD (Hanage *et al.*, 2005) and is another ST that has increased in the United States following the introduction of Prevenar™ (Hanage *et al.*, 2007). There are almost forty ST62 isolates on the MLST website to date with all but one (type 23B) serotyped/grouped as 11A or 11. ST62 is common in Europe (Vestrheim *et al.*, 2008; Serrano *et al.*, 2005; Ardanuy *et al.*, 2006) and it is the third most frequent erythromycin-nonsusceptible pneumococcus in Spain (Ardanuy *et al.*, 2006). Clones of ST62 appear to be opportunistic: one study found that they only caused disease in people already with an underlying disease, and not unsurprisingly, they have a high mortality rate (Sjostrom *et al.*, 2006).

Although the serogroup 11 strains did not fall into two neat groups through the Western Blot analysis, the protein profiles suggest that some of the strains may be related. For example, MSHR1 and MSHR3 have identical banding patterns on all gels (Figures 3.2 and 3.3). Therefore, it is predicted that MSHR3 may be ST3021 or a very closely related ST. WCH209 and WCH210 have identical protein profiles as well. They only differ from WCH213 in that their PspC has a slower mobility on SDS-PAGE. Interestingly, sequencing revealed that the PspC of WCH213 and WCH210 both belong to family 6 although the electrophoretic mobility of their respective PspC is different. Therefore, it is predicted that all three isolates are ST62 or that WCH209 and WCH210 are closely related to ST62. In Europe and the USA, although exact numbers vary between epidemiological studies, over 75% of serotype 11A isolates typically belong to ST62 with one report of 100% of serotype 11A isolates (Gertz *et al.*, 2003; Brueggemann *et al.*, 2003; Serrano *et al.*, 2003; Sjostrom *et al.*, 2006). In Australia, the distribution is expected to be similar, but the Western blot analysis and MLST data suggest at least 60% of the serotype 11A isolates from the remote Aboriginal communities are unrelated to ST62.

Although the limited number of isolates typed in this study means that it is not possible to gain a true picture of the prevalence of different STs in Australia, it can be seen that STs that are common in other countries, namely ST180 and ST62, are present in the general population of Australia. This study also suggests that the ST profile of pneumococci in remote Aboriginal communities differs from those found in the general population, although again the limited number of isolates and serotypes means that a definitive conclusion cannot be drawn.

3.4.2 Virulence and colonisation studies

Both serotypes 3 and 11A have a low association with IPD (Hanage *et al.*, 2005; Sjostrom *et al.*, 2006; Henriques *et al.*, 2000). However, when serotype 3 does cause IPD, it is severe with a very high mortality rate (Harboe *et al.*, 2009; Sjostrom *et al.*, 2006; Henriques *et al.*, 2000). Although serotype 11A has high mortality rate, it is associated with comorbidity (Sjostrom *et al.*, 2006; Harboe *et al.*, 2009). This difference in virulence was evident in the pneumonia/sepsis and i.p. models. Serotype 3 strains were much more virulent than serogroup 11. When groups of 10 mice were challenged with 5×10^2 CFU/mouse of the serotype 3 strains, more than 50% of mice died, with median survival times between 31.5 h and 62.5 h. However, even with inocula 20,000 times that used for serotype 3, less than 50% of mice inoculated with the serogroup 11 strains died. In the pneumonia/sepsis study, WU2 (ST378) was not able to kill 50% of the mice, but for the remaining serotype 3 strains of ST180 and ST458, 50% of mice were dead by 47 h to 161 h, depending on the strain. There were no deaths in the groups of mice challenged with serogroup 11 strains. In contrast, more than 50% of mice died within 62.5 hours, when

inoculated with 5 x 10^2 CFU of serotype 3 in the i.p. model. In the pneumonia sepsis model, more than 50% of mice had died within 57 hours with the exception of WU2.

No serogroup 11 bacteria were isolated from the blood in the colonisation study (3.2.5.2). In addition, the serotype 11A strain MSHR5 was not isolated from the blood in the pneumonia-sepsis model during the capsule switch study (Section 3.3.3). However, serogroup 11 was isolated from the brain in both these models and in fact, serogroup 11 strains were isolated from the brains of more mice compared to serotype 3 in the colonisation study (Table 3.8). In the pneumonia-sepsis model, the number of bacteria in the brain had fallen below the limit of detection by day 3 in most mice for both serotype 3 and serotype 11A strains (Figure 3.10). Pneumococci have been reported to be able to enter the brain via axonal transportation through olfactory nerves (van Ginkel *et al.*, 2003). The immune system is capable of clearing these bacteria and therefore, meningitis does not automatically result (van Ginkel *et al.*, 2003). This is believed to be what has happened during this study.

WCH206 and WCH207 (ST180) showed statistically significant differences in virulence in the i.p. model and MSHR11 and MSHR17 (ST458) showed statistically significant differences in virulence in the pneumonia/sepsis model, suggesting there are differences between the clones either at the gene or protein level. Silva *et al.* (2005) also found that pneumococci of the same serotype and ST could produce different levels of bacteremia in an i.p. murine model. Although strains with the same ST are classed as clones, this does not mean they are 100% identical. MLST is limited to the sequence of seven housekeeping genes and therefore, does not give information on other genes (Obert *et al.*, 2007). Furthermore, the housekeeping genes are believed to mutate slowly overtime, but there are other regions of the genome that may change much more rapidly (Obert *et al.*, 2007). Indeed, isolates of the same or highly related STs have been shown to

differ in their carriage of accessory genes by DNA microarray analysis, although the difference is of course minor compared to unrelated isolates (Dagerhamn *et al.*, 2008). In addition, isolates with the same ST have been shown to differ in whether they carried *nanB* and *nanC*, although there were no differences in ST180 isolates, which carried *nanB* but not *nanC* like what was found for WCH206 and WCH207 in Section 3.2.1.3 (Pettigrew *et al.*, 2006). ST180 is said to have low transformability, which could affect its evolution through recombination (Henriques-Normark *et al.*, 2008). However, it is known that some ST180 isolates carry *ermB*, while others do not (Isozumi *et al.*, 2008; Rolo *et al.*, 2009) so there are real differences between ST180 isolates. In this study, it was found that the proline rich repeat region in the PspC of MSHR11 was approximately four times the length of the one in MSHR17 and the PspA of WCH206 contained two extra choline binding repeat units compared to WCH207 (Section 3.2.1). It is currently not known how these types of small differences affect virulence.

3.4.3 Influence of serotype on virulence

The strain MSHR5 was switched to serotype 3 to see what effect this would have on virulence. The capsule switch variant MSHR5^3 produced at least 70% of the amount of capsule as the capsule donor strain WCH206. Although the capsule locus was amplified from the strain WCH206, the capsule switch variant may not produce exactly the same amount of capsule as the donor strain due to differences in their genetic background. There are a variety of contributors to capsule production, including genes outside of the capsule locus, such as phosphoglucomutase, which is used for the synthesis of UDP-glucose, one of the sugar precursors (Hardy *et al.*, 2001). Differences in capsule production were also observed between selected serotype 3 clinical isolates (Section 3.2.4)

but this did not correlate with differences in virulence (Section 3.2.5.1). These findings are consistent with those of Magee and Yother (2001) who showed that a serotype 3 mutant displaying an 80% reduction in capsule production was still virulent in an i.p. mouse models and colonised the nasopharynx as effectively as the parent strain after i.n. inoculation.

The difference in virulence between WCH206 and MSHR5 in the pneumonia/sepsis and i.p. models is not due to a difference in growth, since they displayed identical growth curves (Section 3.2.6.2). Furthermore, switching the capsule (MSHR5^3) did not affect the growth rate either.

Kelly *et al.* (1994) demonstrated that converting various strains to serotype 3 did not always result in the strains possessing identical virulence properties as the donor serotype 3 strain. Furthermore, changing the capsule could both increase and decrease virulence. Kelly *et al.* (1994) used an i.p. model. However, in this study, we used the pneumonia/sepsis model to further investigate the role capsule switching can have on virulence. At all time points, the serotype 11A strain MSHR5 was isolated from the nasopharynx at higher numbers than its serotype 3 capsule switch variant MSHR5^3. However, switching to capsular serotype 3 did not make it identical in terms of colonisation to the serotype 3 strain WCH206. Unlike WCH206, MSHR5^3 was never isolated from the blood. Furthermore, by Day 3, WCH206 was at a statistically higher number in the ears and nasopharynx compared to MSHR5^3. On Day 2, WCH206 was present at a statistically higher level in the lungs compared to MSHR5^3. Thus non-capsular genes in WCH206 play a significant role in terms of capacity to cause IPD.

The observation that switching a serotype 11A isolate to serotype 3 did not consequentially endow it with the same colonisation abilities as the capsule donor strain supports the idea that there are genetic factors aside from serotype that influence virulence,

including the ability to cause OM. This is despite serotype 3 being found to have a common association with OM (Hausdorff *et al.*, 2000; Hausdorff *et al.*, 2002). The differences demonstrated in Section 3.2.5.1 between survival time in the i.p. and pneumonia-sepsis models of the different serotype 3 isolates, also supports this idea. In particular, WU2, which is a serotype 3 isolate of ST378, demonstrated a dramatic reduction in virulence compared to the ST180 and ST458 isolates chosen for this study (Figure 3.7).

Overall the results of this chapter suggested that the genomic analysis of serotype 3 isolates would give meaningful data as would the comparison of serotype 3 to serogroup 11. The Western blot analysis and MLST typing revealed that unrelated genetic lineages were represented within both serotype 3 and serogroup 11, and this would need to be considered when choosing strains for genomic analyses, which were conducted in Chapters 4 and 7. The capsule switch study was also important in identifying a model that could be used to investigate ear colonisation and the pneumonia/sepsis model was adapted in Chapter 6 for this purpose.

Chapter 4

DNA Microarray Analysis and PCR-based Subtractive Hybridisation

4.1 Introduction

In the previous chapter, a variety of characterisation experiments were performed on serotype 3 and serogroup 11 clinical isolates. Serotype 3 is an important serotype in OM worldwide, including in remote Aboriginal communities, but serogroup 11 is rarely identified in the middle ears of children from these remote communities and generally has a lower association with OM compared to serotype 3 in other communities (Prymula, 2009; Hausdorff *et al.*, 2000; Hausdorff *et al.*, 2002). When selected isolates were analysed by MLST, two STs were identified in the serotype 3 clinical isolates (ST180 and ST458), while four different STs were identified within serogroup 11 (ST62, ST662, ST3020, ST3021). Serotype 3 isolates of ST180 and ST458 were highly virulent in mouse virulence models, but the four serogroup 11 isolates tested, which were unrelated to one another, were all avirulent. However, the importance of factors other than serotype on virulence was clearly demonstrated in the capsule switch study in which a serotype 11A strain from MSHR (MSHR5, ST662) was switched to serotype 3. In a pneumonia/sepsis model, it was not isolated from the blood unlike the donor capsule strain WCH206. Furthermore, it was less fit in the middle ears compared to WCH206.

In this chapter, serotype 3 has been directly compared to serogroup 11 by pooling the genomic DNA of isolates from the same serotype/group and performing DNA

microarray analysis in order to identify any genes common to one serotype/group, but absent in the other, apart from the obvious capsule loci. Genes unique to serotype 3 would be of particular interest and would be explored further for a potential role in OM. Prior to this, isolates of the same serotype/serogroup were compared by DNA microarray analysis. This helped assess the data from the cross serotype/group comparison. Although it was difficult to identify potential OM associated genes through comparing OM isolates to one another due to the large amount of data generated, genes that were present in one isolate but not the other could be discarded as potential candidates for further study. Furthermore, accessory regions which are differentially distributed within serotype 3 or serogroup 11 could be identified.

DNA microarray is a quick method for obtaining an overall picture of which genes are present or absent in strains of interest, compared to the genomes of reference strains on a chip. However, the use of microarrays for genome comparison has one significant limitation in that only genes which are present on the array can be analysed. The array gives no information on genes which may be present in the strain to be examined but absent on the array. In order to overcome this, a PCR-based subtractive hybridisation method was adapted to detect genes present in serotype 3 isolates of both ST180 and ST458 which were not represented on the microarray slide. Using these two complementary techniques increased the probability of identifying genes to be explored for a role in OM, as well as obtaining a more complete understanding of the two genomes.

4.2 DNA Microarray Analysis

In this study, microarray slides from the Bacterial Microarray Group at St George's, University of London were used, which contain the open reading frames (ORFs) present in the genomes of R6 and TIGR4. R6 (a rough derivative of the serotype 2 strain

D39) and TIGR4 (a virulent serotype 4 clinical isolate from a Norwegian patient) are both well-characterised and have been used in numerous studies. Their genomes have also been fully sequenced and are available publicly.

A number of comparisons were performed. First, the genomes of the serotype 3 middle ear isolates WCH206 (ST180) and MSHR17 (ST458) were compared by DNA microarray analysis. Following this, the genomes of strains WCH208 and MSHR12, which are other serotype 3 isolates of ST458, were compared to one another, as were the genomes of the serotype 3 strains MSHR13 and MSHR15. MSHR13 and 15 have not been analysed by MLST, but displayed identical protein profiles to WCH208, MSHR12 and MSHR17 (Section 3.2.1). A number of serogroup 11 strains were also analysed by comparing the genomes of WCH211 (ST3020) with that of MSHR1 (ST3021), and WCH213 (ST62) with that of MSHR5 (ST662). Finally, a comparison was carried out between a number of strains from serotype 3 and serogroup 11 (WCH206, WCH207 [ST180], MSHR11 [ST458] and MSHR17 of serotype 3, and WCH211, WCH213, MSHR1 and MSHR5 of serogroup 11). Four isolates were used for each serogroup in order to increase the chance of identifying genes unique to a given serotype as opposed to unique to an individual strain.

4.2.1 Genomic comparison of representative serotype 3 isolates of ST180 and ST458

The genome of the serotype 3 strain WCH206 (ST180) was compared to the genome of the serotype 3 strain MSHR17 (ST458). In brief, 10 ng of chromosomal DNA was digested with *Sau*3A as described in Section 2.5.4. Four hybridisation reactions in total were performed. For the first two replicate reactions, WCH206 was labelled with Alexa Fluor 647, while MSHR17 was labelled with Alexa Fluor 555. The dyes were

reversed for the subsequent two replicate reaction slides. The results are displayed in Tables 4.1-4.3.

Ten ORFs were detectable in MSHR17 but absent in WCH206 (Tables 4.1 and 4.2). In Chapter 1, the concept of accessory regions, which have variable distribution in the pneumococcus, was introduced (Section 1.3.2) and several known accessory regions were identified in the microarray analysis (Figure 4.1). In particular, there appeared to be a large deletion or major sequence variation in strain WCH206 from approximately spn2159 (fucolectin-related protein) until spn2166 (L-fuculose phosphate aldolase) [spn number are based on the TIGR4 genome designation] of genes putatively involved in fucose metabolism. This accessory region is called AR (accessory region) 41 by Blomberg *et al.* (2009), but has been named C* (cluster*) 13 (Bruckner *et al.*, 2004), RD (region of diversity) 25 (Silva *et al.*, 2006) or RD13 (Obert *et al.*, 2006) in previous studies. However, for consistency the accessory regions described in this study have been labelled according to Blomberg *et al.* (2009). In addition to AR41, genes from other known accessory regions, spn0691-spn0699 (AR17, ABC transporter) and spn0473-spn0478 (AR12, lactose or cellobiose metabolism) were identified in MSHR17 but not WCH206. Only one ORF (spn1857), which is not part of a known accessory region, was detectable in WCH206 but not MSHR17 (Table 4.3). Although not all genes of the accessory regions appear in Tables 4.1 and 4.2, many of these were not detectable in either strain on the microarray slides. This indicates that these genes could be absent or vary in sequence to the ORF represented on the microarray slide in both strains, but may have also simply been due to the microarray conditions. The gene spn2165 appeared to be present in MSHR17 and absent in WCH206 on two out of four slides. Finally, it is also possible that the accessory regions which were not detectable in WCH206 are actually present but were not detected due to sequence variation relative to R6 and TIGR4.

Table 4.1. Genes that are detectable in strain MSHR17 but not in strain WCH206 by microarray analysis for four out of four slides.

TIGR4 ID (spn)	Gene
spn2159	fucolectin-related protein
spn2160	hypothetical (large secreted protein)
spn2161	PTS, IID component
spn2166	L-fuculose phosphate aldolase

Table 4.2. Genes that are detectable in MSHR17 but not WCH206 by microarray analysis for three out of four slides.

TIGR4 ID (spn)	Gene
spn0474	PTS-system, cellobiose specific component
spn0477	6-phospho-β-galactosidase
spn0695	HesA/MoeB/Thi family protein
spn0697	ABC transporter, ATP-binding protein, authentic point mutation
spn2162	PTS, IIC component
spn2163	PTS, IIB component

Table 4.3. Genes that are detectable in WCH206 but not MSHR17 by microarray analysis for three out of four slides.

TIGR4 ID (spn)	Gene
spn1857	cation efflux system protein

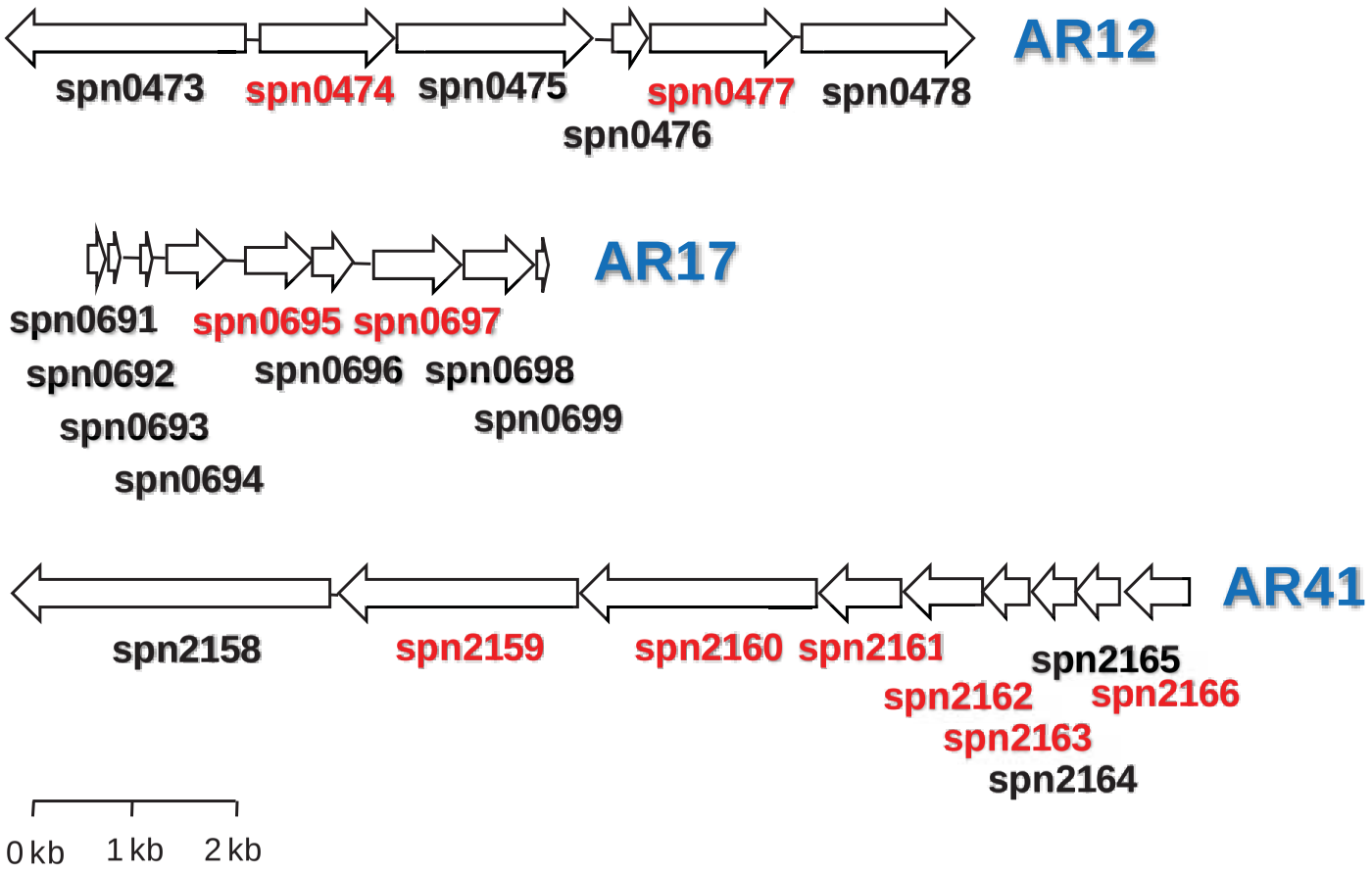


Figure 4.1. Accessory Regions identified through DNA microarray analysis as differentially distributed in the Serotype 3 Isolates. Accessory regions (AR) are numbered according to Blomberg *et al.* (2009). Genes from Tables 4.1-4.2 are shown in red. Genes are annotated according to the genome of TIGR4 (spn).

4.2.2 Genomic comparison of highly related serotype 3 isolates

In this section, serotype 3 isolates of ST458 (Section 3.2.3) or serotype 3 isolates which displayed identical protein profiles to the known ST458 isolates in the Western blot analysis (Section 3.2.1.1) have been compared by DNA microarray analysis. The genomes of ST180 isolates or highly related isolates have been compared in other studies. Dagerhamn *et al.* (2008) performed DNA microarray analysis to compare the genomes of two serotype 3 isolates of ST180 and ST1823 (single locus variant of ST180), respectively, which were found to be identical in regards to their possession of known accessory regions from TIGR4 and R6, and only differed in genetic content by four genes overall. The comparison of the whole genome sequences of two serotype 3 isolates of ST180 also found that there were only a small number of differences between the isolates (Hiller *et al.*, 2007).

The DNA from the serotype 3 strains WCH208 (ST458), MSRH12 (ST458), MSHR13 and MSHR15 was digested, labelled and the slides probed as described for strains WCH206 and MSHR17. One array slide hybridisation was performed for strains WCH208 and MSHR12 and another for strains MSHR13 and MSHR15. No differences were detectable between these strains by DNA microarray analysis. This correlates with the Western blot analysis (Section 3.2), and further demonstrates the high level of relatedness between these strains despite being isolated from temporally and geographically distinct locations.

4.2.3 Genomic comparison of serogroup 11 pneumococcal isolates

For these analyses, strain WCH211 was compared to strain MSHR1, and strain WCH213 was compared to strain MSHR5. The DNA was digested and labelled, and the slides probed essentially as described for strains WCH206 and MSHR17 above. Two array slides were used for each pair of strains. One strain was labelled with Alexa Fluor 647 and the other with Alexa Fluor 555 for one slide, with the dyes reversed for the second slide. The results of the microarray comparison are summarised in Tables 4.4 to 4.7. ORFs from two of the same accessory regions identified in Section 4.2.1 (AR12 and AR41) were also identified in this section. The ORFs spn0474 and spn0477 were detectable in strain MSHR1 but not in strain WCH211 (Table 4.4), while the ORFs spn2159 to spn2163 were detectable in strain WCH211 but not in strain MSHR1 (Table 4.5). Moreover, differences in seven additional accessory regions were detected (Tables 4.4-4.7, Figure 4.2). The region from spn1615 to spn1621 (AR31) was detectable in strain MSHR1 but not in strain WCH211 (Table 4.4). Half of these genes encode a PTS (spn1617-1620). A large number of genes from spn1315 through to spn1331 (AR27) were detectable in strain WCH211 but not in strain MSHR1 (Table 4.5) with most genes encoding a v-type sodium ATP synthase (spn1315, 1316, 1317, 1319, 1321 and spn1322). One ORF (spn1797) from AR35 (ABC transporter) was detectable in WCH213 but not MSHR5 (Table 4.6). There are three possible large deletions in strain WCH213 of genes that appear to be present in strain MSHR5 (Table 4.7): spr1618 to spr 1620 (AR36; annotated according to the R6 genome due to the absence of this accessory region in TIGR4), which encode components of an ABC transport system, whereas most of the genes from spn1759 to spn1772 (part of AR34) encode glycosyl-transferases and secretory proteins. The region from spn1949 to spn1954 (AR40) appears to encode for bacteriocin production and secretion. Furthermore,

Table 4.4. Genes that are detectable in strain MSHR1 but not in strain WCH211 by microarray analysis.

TIGR4 ID (spn)	Gene
spn0474	PTS, cellobiose-specific IIC component
spn0477	6-phospho-beta-galactosidase
spn1615	transketolase, authentic frameshift
spn1616	ribulose-phosphate 3-epimerase family protein
spn1617	PTS, IIC component
spn1618	PTS, IIB component
spn1619	PTS, IIA component
spn1620	PTS, nitrogen regulatory component IIA, putative
spn1621	transcription antiterminator BglG family protein, authentic frameshift

Genes were not detectable in WCH211 for two out of two slides

Table 4.5. Genes that are detectable in strain WCH211 but not in strain MSHR1 by microarray analysis.

TIGR4 ID (spn)	Gene
spn1315	v-type sodium ATP synthase, subunit D
spn1316	v-type sodium ATP synthase, subunit B
spn1317	v-type sodium ATP synthase, subunit A
spn1319	v-type sodium ATP synthase, subunit C
spn1321	v-type sodium ATP synthase, subunit K
spn1322	v-type sodium ATP synthase, subunit I
spn1325	oxidoreductase, Gfo/Idh/MocA family
spn1326	neuraminidase, putative
spn1327	conserved hypothetical protein
spn1769	glycosyl transferase, authentic frameshift
spn1951	conserved hypothetical protein
spn1952	hypothetical (large secreted protein)
spn2159	fucolectin-related protein
spn2160	conserved hypothetical protein
spn2161	PTS, IID component
spn2162	PTS, IIC component
spn2163	PTS, IIB component

Genes were not detectable in MSHR1 for two out of two slides

Table 4.6. Genes that are detectable in strain WCH213 but not in strain MSHR5 by microarray analysis.

TIGR4 ID (spn)	Gene
spn1797	ABC transporter, permease protein

Genes were not detectable in MSHR5 for two out of two slides

Table 4.7. Genes that are detectable in strain MSHR5 but not in strain WCH213 by microarray analysis.

R6 ID (spr)/ TIGR4 ID (spn)	Gene
spr1618	ABC transporter membrane-spanning permease - sugar transport
spr1619	ABC transporter membrane-spanning permease - sugar transporter
spr1620	ABC transporter substrate-binding protein - sugar transport
spn0391	choline binding protein F
spn0394	PTS, mannitol-specific IIBC components
spn0397	mannitol-1-phosphate 5-dehydrogenase
spn1722	PTS IIABC components
spn1759	preprotein translocase, SecA subunit
spn1761	hypothetical (accessory secretory protein Asp2)
spn1762	Hypothetical (accessory secretory protein Asp1)
spn1763	preprotein translocase SecY family protein
spn1764	glycosyl transferase, family 2
spn1766	glycosyl transferase, family 8
spn1767	glycosyl transferase, family 8
spn1769	glycosyl transferase, authentic frameshift
spn1772	cell wall surface anchor family protein
spn1949	hypothetical
spn1950	bacteriocin formation protein, putative
spn1951	conserved hypothetical protein
spn1952	hypothetical
spn1953	toxin secretion ABC transporter, ATP-binding/permease protein
spn1954	serine protease, subtilase family, authentic frameshift

Genes were not detectable in WCH213 for two out of two slides

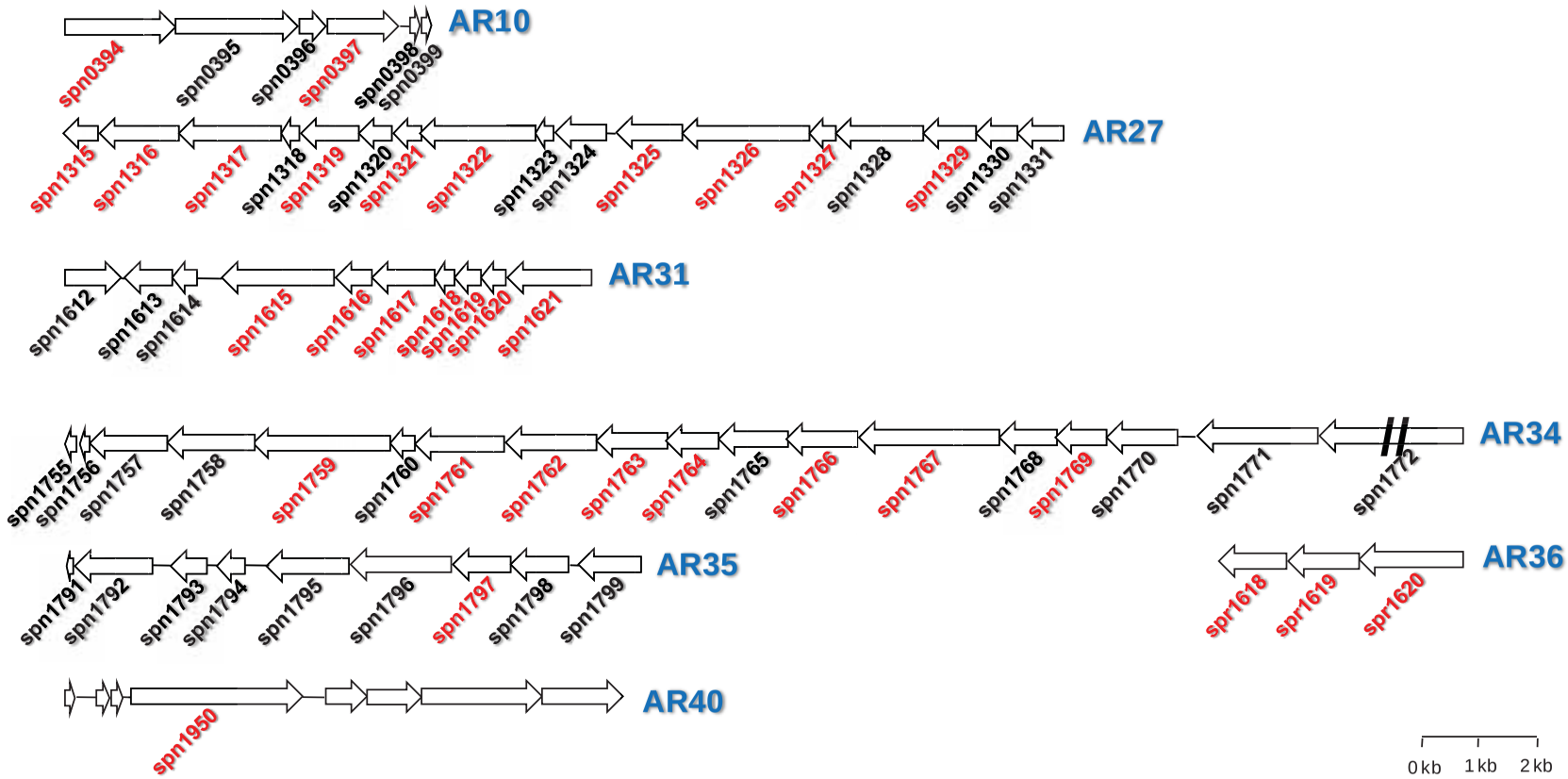


Figure 4.2. Accessory Regions identified through DNA microarray analysis as differentially distributed in the serogroup 11 isolates in addition to those also seen in serotype 3.

Accessory regions (AR) are numbered according to Blomberg *et al.* (2009). Genes from Tables 4.4-4.8 are shown in red font. Genes are annotated according to the genomes of TIGR4 (spn) or R6 (spr). The gene spn1772 of AR34 is 14.3 kb.

spn0394 and spn0397 from AR10 (mannitol metabolism) were detectable in strain MSHR5 but not in strain WCH208 (Table 4.7). Many of the genes of the accessory regions that did not appear in the above tables were detectable in neither strain or only appeared differentially distributed on one of the two slides.

4.2.4 Genomic comparison of serotype 3 and serogroup 11 isolates

In order to compare the two serotype/groups, chromosomal DNA of the serotype 3 strains WCH206, WCH207, MSHR11 and MSHR17, which were the same isolates used in the PCR-based subtractive hybridisation study in Section 4.3, were combined in equal amounts, as was the chromosomal DNA of the serogroup 11 strains WCH211, WCH212, MSHR1 and MSHR5. The strains WCH206 and WCH207 are ST180 isolates, while the strains MSHR11 and MSHR17 are both ST458 isolates. The four serogroup 11 strains represent 4 unrelated STs (Table 3.3). The pooled DNA for each serotype/group was digested using *Rsa*I, which was the same restriction endonuclease used in the PCR-based subtractive hybridisation. Microarrays were performed in duplicate. Serotype 3 DNA was labelled with Alexa Fluor 647 and serotype 11A was labelled with Alexa Fluor 555 for one array and the dyes were reversed for the second slide. Results are shown in Tables 4.8 and 4.9. The results were also compared to the previous tables and the microarray slides of Sections 4.2.1 and 4.2.3. The distribution of genes listed in Tables 4.8 and 4.9 in the strains WCH206, MSHR17, as well as the four serogroup 11 strains, indicated by the DNA microarray analysis performed in Sections 4.2.1 and 4.2.3 is displayed in Table 4.10.

Three genes (spn0695, spn0697 and spn1850) appeared to be absent in the serogroup 11 strains, but present in the serotype 3 strains in the serogroup comparison using microarray analysis (Table 4.8). However, when the genomes of the serotype 3

Table 4.8. Genes that are detectable in serotype 3 strains but not in serogroup 11 strains by microarray analysis.

TIGR4 ID (spn)	Gene
spn0697	ABC transporter, ATP-binding protein, authentic point mutation
spn0695	HesA/MoeB/ThiF family protein
spn1850	type II restriction endonuclease DpnI

Genes were not detectable in serogroup 11 for two out of two slides

Table 4.9. Genes that are detectable in the serogroup 11 strains but not the serotype 3 strains by microarray analysis.

TIGR4 ID (spn)/ R6 ID (spr)	Gene
spn1988	immunity protein
spn1067	cell division protein FtsW, putative
spn0330	sugar binding transcriptional regulator RegR
spn0506	integrase/recombinase, phage integrase family
spr0113	hypothetical protein (no conserved domains)
spn1763	preprotein translocase SecY family protein
spn1950	bacteriocin formation protein, putative
spn1953	toxin secretion ABC transporter, ATP-binding/permease protein
spn1331	phosphosugar-binding transcriptional regulator, putative
spn1328	sodium: solute symporter family protein
spn1326	neuraminidase, putative
spn1317	v-type sodium ATP synthase, subunit A

Genes were not detectable in serotype 3 for two out of two slides

Table 4.10. The distribution of genes listed in Tables 4.8 and 4.9 suggested by the microarray slides of Sections 4.2.1 and 4.2.3.

	TIGR4 ID (spn)/ R6 (spr)	Serotype 3		Serogroup 11			
		WCH206	MSHR17	WCH211	WCH213	MSHR1	MSHR5
Table 4.8	spn0697	-	+	-	-	-	-
	spn0695	-	+	-	-	-	-
	spn1850	+	+	+	-	+	-
Table 4.9	spn1988	+	+	+	+	+	+
	spn1067	+	+	+	+	+	+
	spn0330	-	-	-	+	-	+
	spn0506	+	+	+	+	+	+
	spr0113	-	-	+	+	+	+
	spn1763	-	-	+	-	-	+
	spn1950	-	-	+	-	-	+
	spn1953	-	-	+	-	-	+
	spn1331	-	-	-	+	-	+
	spn1328	-	-	+	+	-	+
	spn1326	-	-	+	+	-	+
	spn1317	-	-	+	+	-	+

The genomes of MSHR17 and WCH206 were compared in Section 4.3.1. The genomes of MSHR1 and WCH211 were compared in Section 4.3.3, along with the genome of MSHR5 to the genome of WCH213. The slides from these two sections were reanalysed to determine the distribution in these strains of genes listed in Tables 4.9 and 4.10. + indicates a gene that was detectable on the slides and – indicates a gene that was not detectable.

strains WHC206 (ST180) and MSHR17 (ST458) were compared in Section 4.2.1, spn0695 and spn0697 were only detected in MSHR17 (Table 4.2). Reanalysis of the 4 microarray slides used in Section 4.2.1 indicated that the third gene spn1850 was present in both WCH206 and MSHR17 (Table 4.10). Furthermore, reanalysis of the microarray slides of the serogroup 11 strains from Section 4.2.3 indicated that spn1850 is absent in strains MSHR5 and WCH213, but is present in MSHR1 and most likely WCH211 as well (Table 4.10).

Twelve genes were listed in Table 4.10 as being absent in serotype 3 but present in serogroup 11. However, when the data from Sections 4.2.1 and 4.2.3 was reanalysed, there was no gene that was absent in both serotype 3 strains, but carried by all four serogroup 11 strains (Table 4.10). Therefore, no genes which could distinguish serotype 3 from serogroup 11 apart from the capsule locus were identified and the capsule loci remained the only distinguishing genetic feature between serotype 3 and serogroup 11. The microarray analysis also highlighted the great genetic variability that can be found within a serotype and the importance of analysing a number of unrelated isolates when studying a specific serotype.

4.3 PCR-based Subtractive Hybridisation

In this section, PCR-based subtractive hybridisation was used to complement the DNA microarray analysis of Section 4.2.1 by identifying genes present in serotype 3 isolates of ST180 and ST458, which are not represented on the microarray slide. Subtractive hybridisation was first used in the 1960's by Bautz and Reilly to purify bacteriophage T4 mRNA (Bautz and Reilly, 1966). It is a well-established technique for identifying both genomic deletions and the up or down regulation of genes (Sagerstrom *et al.*, 1997). More recently (Akopyants *et al.*, 1998) developed a PCR-based subtractive hybridisation strategy

(Figure 4.3) to identify genes present in one bacterial strain (referred to as the tester) which are not present in another strain of the same species (referred to as the driver). The method is an adaptation from suppression subtractive hybridisation (Diatchenko *et al.*, 1996; Gurskaya *et al.*, 1996), which is used to investigate eukaryotic gene expression, and is described in Figure 4.3 using strains from this study. The method has been modified slightly for this study by increasing the number of driver and tester strains used. The two pneumococcal strains D39 and TIGR4 were chosen as the driver strains because the slides used in the microarray work contained ORFs from these strains. Furthermore, the availability of fully sequenced genomes also made them attractive strains to use since this makes it easy to detect false positives. This is particularly important, since false positives, which are sequences present in the driver strain that have been amplified, are always an issue in subtractive hybridisation and this background amplification can never be eliminated completely. Genomic DNA from strains WCH206 and WCH207 (serotype 3 isolates of ST180) were used as the first tester, and subsequently the DNA from strains MSHR11 and MSHR17 (serotype 3 isolates of ST458) were used as the second tester. In total four strains of serotype 3 were used in order to increase the chance of identifying genes unique to the isolates as a whole rather than a specific strain. The genomic DNA of strains WCH206 and WCH207 was combined because they belong to the same ST. The same reasoning was applied for strains MSHR11 and MSHR17.

The first step in the technique devised by Akopyants *et al.* (1998) is the digestion of the driver and tester strains, followed by the ligation of unique sticky ended adaptors to separate aliquots of digested tester DNA. However, in this study, one adaptor was ligated to digested DNA of one ST and the second adaptor was ligated to digested DNA of the other ST. The adaptors contain a T7 promoter at the 5' end and a unique sequence at the 3' end. As the adaptors lack a phosphate group at the 5' end, they can only be ligated to the 5' end of the digested tester DNA. Following adaptor ligation, hybridisation can be

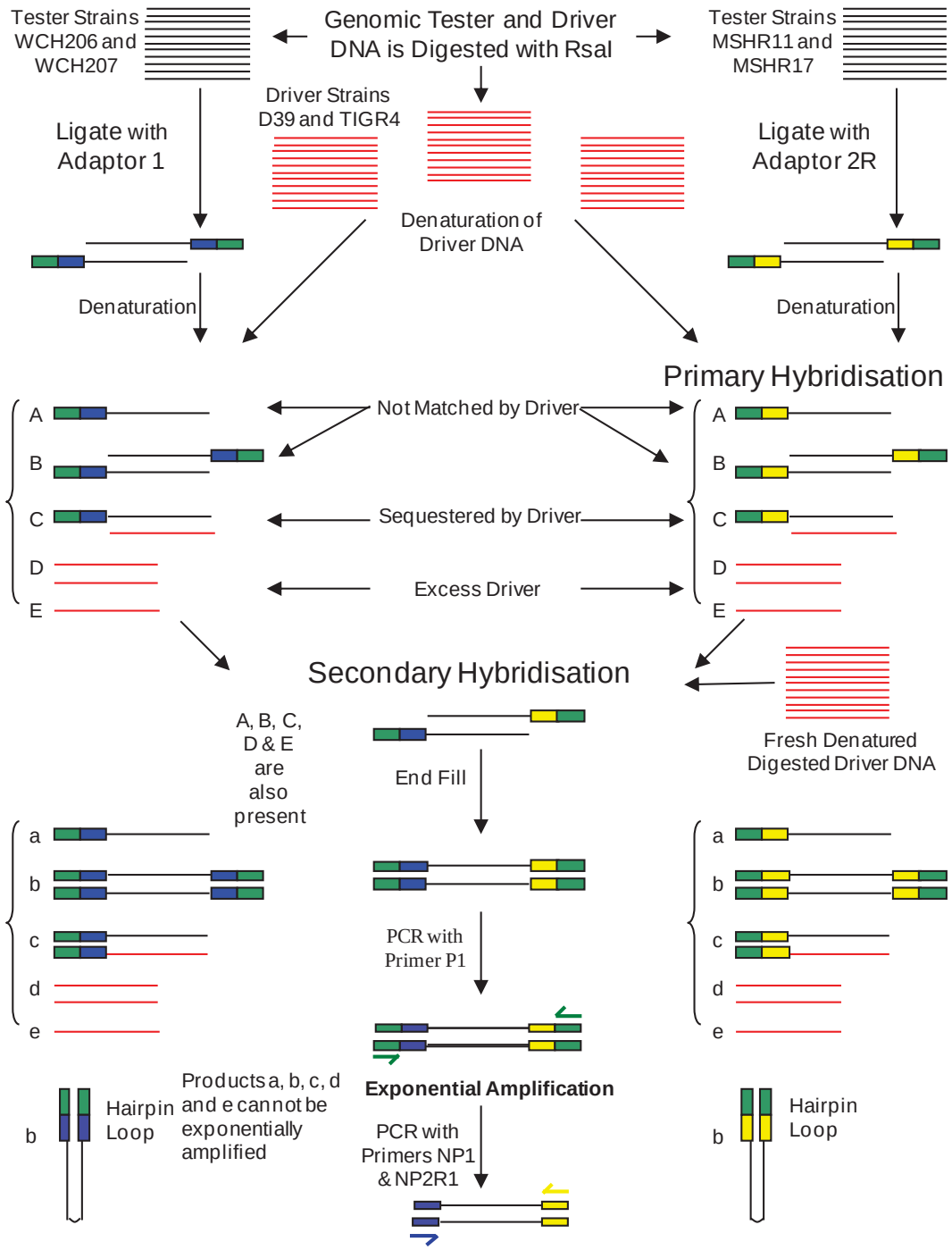


Figure 4.3. PCR-based subtractive hybridisation as performed in this study.

Solid lines represent digested tester (black) and driver (red) DNA. Boxes represent adaptors. Adaptors 1 and 2R are identical at their 5' end (green) but differ at their 3' end (blue and yellow, respectively). This diagram was adapted from Akopyants et al. (1998).

performed. In the primary hybridisation, excess digested driver DNA is added to the two different tester populations. The DNA is denatured and incubated to allow the formation of tester and driver hybrids, thus sequestering fragments of DNA that are present in both the tester and driver strains. Subsequently the two tubes are combined with fresh denatured digested driver DNA and allowed to incubate further so that tester-tester hybrids can form with the single tester molecules remaining from the first hybridisation. The ends of the hybrids are filled and nested PCR is performed. The primary PCR amplification uses a primer for the T7 promoter on both adaptors and the secondary PCR amplification uses two primers (one for each of the unique regions of the adaptors) [Figure 4.3]. If two molecules from the same tester strain anneal, identical adaptors will be present on both the 5' and 3' ends after end filling. The adaptors have been designed in such a way that if this happens, a hairpin loop will form that prevents amplification during PCR. Therefore, only when two molecules from different tester populations anneal, which results in different adaptors being present on the 5' and 3' ends, does exponential amplification occur.

Akopyants *et al.* (1998) designed their method to eliminate several disadvantages of previous bacterial genomic subtraction methods, such as the requirement of multiple rounds of hybridisation and physical removal of tester-hybrid complexes or the removal of adaptors after PCR amplification, which was followed by the ligation of a new adaptor for the subsequent round of PCR amplification. Furthermore, they were also able to reduce the primary incubation time due to the low copy number of bacterial genes compared to eukaryotic mRNA in suppression subtractive hybridisation.

In this study, PCR-based subtraction was performed using the Clontech PCR-Select™ Bacterial Genome Subtraction Kit, which is based on the method developed by Akopyants *et al.* (1998). PCRs were performed using *Taq* polymerase (Roche).

4.3.1 Genomic DNA

Chromosomal DNA for use in the PCR-based subtractive hybridisation was obtained as described in Section 2.5.3. This method was used in order to ensure that the DNA was of the highest quality. Chromosomal preparations of strains D39, TIGR4, WCH206, WCH207, MSHR11 and MSHR17 were diluted to 200 μg/ml and then combined to ensure a 1:1 ratio of genomes with a final concentration of 200 μg/ml.

4.3.2 Adaptor ligation

DNA was digested with *RsaI* as described in the instruction manual but the restriction digest was desalted as described in Section 2.5.6. Adaptor 1 was ligated to digested WCH206 and WCH207 DNA, while adaptor 2R was ligated to digested MSHR11 and MSHR17 DNA as described in Section 2.8. The adaptors used in this study are shown in Figure 2.1 of Section 2.3.2. A sample from each ligation mixture was mixed together prior to incubation as described in the instruction manual. Following incubation, this was diluted in water to act as the unsubtracted tester control for nested PCR amplification.

In order to ensure efficient subtraction, at least 25% of the tester DNA must have adaptors attached. This was confirmed by performing a PCR test described by the manufacturer but using pneumococcal specific primers (Figure 4.4). In essence, two PCR reactions were set up for each tester ligation: one using the primers *groEL* F and *groEL* R and the second Primer 1 and *groEL* R. The gene *groEL* was chosen because it is conserved in the pneumococcus and the region amplified did not contain a *RsaI* restriction site. In some PCR amplifications, Primer 1 was replaced with adaptor1/*groEL* F or adaptor2R/*groEL* F, depending on which adaptor was used in the ligation. These primers were designed for the junction where adaptor 1 and adaptor2R ligated to the *groEL*

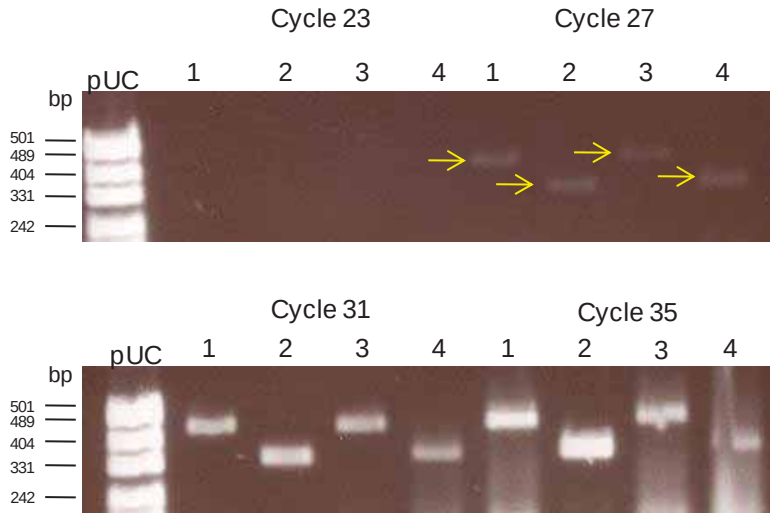


Figure 4.4. PCR for testing the efficiency of ligation.

Reactions 1 and 2 used the product from a ligation reaction with adaptor 1. Reactions 3 and 4 used the product from a ligation reaction with adaptor 2R. The following primer combinations were used:

Reaction 1- *groEL* R and adaptor1/*groEL* F

Reaction 2- *groEL* R and *groEL* F

Reaction 3- *groEL* R and adaptor2R/*groEL* F

Reaction 4- *groEL* R and *groEL* F

The number of cycles at which samples were taken is indicated on the photos above. 5 µl samples were electrophorised on a 2% agarose gel and stained using ethidium bromide. The intensity of the bands for reactions 1 and 3 were compared to the intensity of the bands for reactions 2 and 4, respectively. The intensity appeared similar at all cycles indicating that both ligation reactions had worked satisfactorily.

*Hpa*II-restricted pUC19 plasmid DNA (pUC) was used as the DNA size marker.

fragment, and contained *groEL* sequence, as well as the last 10 bp of the adaptor. The use of Primer 1 required a 2 min extension at 72 ^{o}C before the commencement of cycling in order to extend the adaptors, but the design of the primers adaptor1/*groEL* F or adaptor2R/*groEL* F bypassed this requirement. Whichever primer was used, 5 µl samples were taken from the PCR reactions every four cycles from cycle 23 and ligation efficiency was assessed by comparing band intensity of the two PCR reactions on a 2% agarose gel. Only when there was less than a four-fold difference in band intensity were ligation reactions used.

4.3.3 Hybridisations and nested PCR

Hybridisation and nested PCR reactions were initially performed as described per the instruction manual. The primary PCR reaction used Primer 1 and the secondary PCR reaction used Nested Primer 1 and Nested Primer 2R (Table 2.2). Nested PCR was performed using *Taq* polymerase with a manual hot start. The conditions *Taq* polymerase was employed under required change, which was most likely due to the high annealing temperatures specified for the nested PCR. However, the annealing temperatures could not be lowered due to an increase in background amplification. In a report on multiplex PCR, it was stated that raising KCl or buffer concentration greatly enhanced the amplification of smaller products with the optimal buffer concentration for their particular reaction being 1.2-1.6× (Henegariu *et al.*, 1997). When the concentration of the supplied *Taq* buffer was increased from 1× to 1.6×, amplification was possible for both the primary and secondary PCRs with annealing temperatures of 66 °C and 68 °C, respectively.

It is difficult to see in a scanned gel photo but generally a faint smear appeared on a 2% agarose gel for the hybridised sample following 28 cycles of the primary PCR with bands clearly visible at cycle 31 (Figure 4.5). For most PCR amplifications, product from

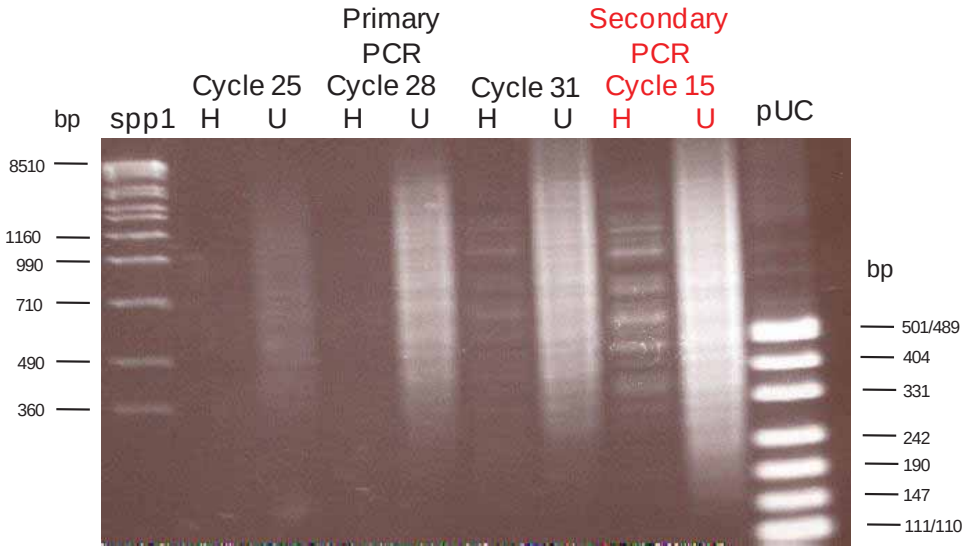


Figure 4.5. Product from the primary and secondary PCR reactions at various cycles.

7 µl of product was run on a 2% agarose gel and stained using ethidium bromide.

The primary PCR reactions used diluted hybridised (H) or unhybridised DNA (U). 31 cycles were performed following the conditions outlined in the user's manual and 7 µl of product was sampled at cycles 25, 28 and 31. Product from cycle 28 was diluted and used for the secondary PCR amplifications.

*Eco*RI-digested *Bacillus subtilis* bacteriophage SPP-1 DNA and *Hpa*II-restricted pUC19 plasmid DNA were used as size markers.

cycle 28 of the primary PCR was diluted for the secondary PCR and the product from cycle 15 was subsequently used. In an attempt to optimise the subtractive hybridisation, other cycle combinations, such as 25 primary cycles, and 15 secondary cycles, were investigated through the course of this study. Ideally, as few cycles as possible should be performed to reduce background amplification. However, altering the conditions and using different cycle combinations was not sufficient to resolve the subtraction efficiency problem described in Section 4.2.5.

4.3.4 Subtractive hybridisation efficiency

To investigate subtractive hybridisation efficiency, the user manual described a PCR test for subtraction and enrichment. This required diluted product from the secondary PCRs of the hybridised and unhybridised tester DNA, as well as primers for a conserved pneumococcal gene. In this case, *comE* F and R or *groEL* F and R (Table 2.2) were chosen to amplify short regions that did not contain *Rsa*I restriction sites. The PCR product was sampled every 3 cycles from cycle 18 onwards and electrophorised on a 2% agarose gel. When the subtractive hybridisation was performed optimally, product for the PCR using unhybridised DNA because visible before product for the PCR using hybridised DNA; an example is shown in Figure 4.6 where there is a difference of approximately 6 cycles. When it had not performed optimally, the product for both hybridised and unhybridised DNA amplifications became apparent at the same cycle. As a positive control for enrichment, a primer pair designed for a region of the serotype 3 capsule locus not present in either TIGR4 or D39 and containing no *Rsa*I restriction sites was designed (Table 2.2). The product for the unhybridised DNA was expected to appear later than the product for the hybridised DNA. However, this particular sequence of the capsule locus was never enriched and the product for both subtracted and unsubtracted template appeared at the

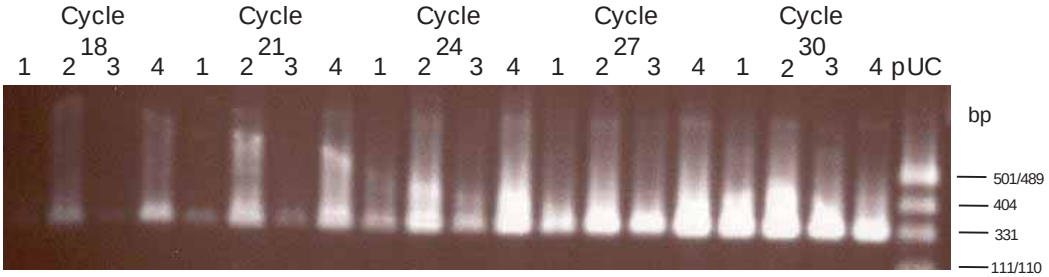


Figure 4.6. PCR for testing the efficiency of subtraction using *comE* primers.

Reactions 1 and 3 are duplicates, which used nested PCR product of hybridised DNA. Reactions 2 and 4 are duplicates, which used nested PCR product of unhybridised DNA. 5 µl of product was electrophorised on a 2% agarose gel and then stained with ethidium bromide. The amplicons of Reactions 2 and 4 are clearly visible under UV light at cycle 18, but the amplicons for Reactions 1 and 3 do not reach a similar level of intensity until cycle 24. This indicates that the gene *comE* has been subtracted in the hybridised sample.

*Hpa*II-restricted pUC19 plasmid DNA (pUC) was used as the DNA size marker.

same cycle even with the optimal conditions described in section 4.2.5.

4.3.5 Reduction of subtractive hybridisation background

Experiments carried out according to the manufacturer's instruction did not result in the expected outcomes. The PCRs of the capsule 3 locus gene, *comE* and *groEL* described in the previous section suggested that the subtractive hybridisation had not worked optimally, with high levels of background amplification which did not allow the identification of any sequences of interest. Southern Blot analysis was carried out to confirm this. The products generated from the above described subtractive hybridisation were cloned and used in Southern blot analysis against DNA from D39 and TIGR4 as described in Section 2.9. The probes hybridised to DNA from both TIGR4 and D39 at high stringency indicating that the insertion sequences were present in these genomes (data not shown). This indicated that the conditions of the PCR-based subtractive hybridisation required modification. In order to optimise the conditions, a positive control described in the instruction manual to show effective subtractive hybridisation had taken place, which used *Rsa*I digested *E. coli* DNA, was adapted by substituting pneumococcal DNA for *E. coli* DNA. Tester DNA was made by combining *Rsa*I digested D39 and TIGR4 DNA with *Hae*III-digested ΦX174 DNA so that there was one copy of the ΦX174 genome for every pneumococcal genome. Adjustments were made to the amount of ΦX174 DNA used to compensate for the difference in genome size between *E. coli* (approximately 5×10^6 bp) compared to *S. pneumoniae* (approximately 2×10^6 bp). Adaptors 1 and 2R were ligated to the control tester DNA in two separate reactions, as well as a combined reaction for the unhybridised control. Subtractive hybridisation was then performed using *Rsa*I-digested D39 and TIGR4 DNA as the driver. The D39 and TIGR4 DNA molecules with adaptors should bind to the driver DNA and as such only the phage DNA should be amplified

during PCR. After the secondary PCR, a ladder should be visible on an agarose gel identical to the ladder formed when pure *Hae*III-digested ΦX174 DNA is electrophorised.

The first modification to the protocol investigated the effect temperature had on the primary and secondary subtractions. Incubations were performed in tandem at approximately 43 ^{o}C, 53 ^{o}C and 63 ^{o}C, using an Eppendorf Gradient Thermocycler. Also, a fourth hybridisation was performed where the primary hybridisation was conducted at 63 ^{o}C, but the secondary hybridisation was commenced at 63 ^{o}C and then gradually reduced to 43 ^{o}C over time. This was achieved by setting a Eppendorf Thermocycler to reduce the incubation temperature by 1 ^{o}C every 45 min. When subtraction was performed at 43 ^{o}C, a smear rather than a ladder was present on a 2% agarose gel, indicating that this temperature was too low for successful subtractive hybridisation to take place (Figure 4.7). Although bands identical to the ΦX174 *Hae*III ladder were visible for subtractions performed at 53 ^{o}C and 63 ^{o}C, as well as the gradient hybridisation, PCR with *groEL* primers indicated that there were still problems with background. Performing additional incubations with fresh denatured driver DNA following secondary hybridisations at 53 ^{o}C and 63 ^{o}C did not appear to reduce the background.

Subsequent work concentrated on altering the length of hybridisation in order to optimise the subtractive hybridisation. In their adaptation of suppression subtractive hybridisation, Akopyants *et al.* (1998) reduced the length of primary hybridisation to 1.5 h compared to 8 h. In order to investigate if returning to the original primary hybridisation length had any effect, hybridisation was repeated at 63 ^{o}C with the length of the primary hybridisation increased from 1.5 h to 8 h. PCR analysis indicated that background had decreased. Therefore, subtractive hybridisation was repeated in duplicate using the DNA of interest with a primary hybridisation of 7-8 h. PCR analysis of products as described in Section 4.2.4 indicated that the background had been sufficiently reduced as shown in Figure 4.6, which depicts a gel photo of a PCR using primers for *comE*.

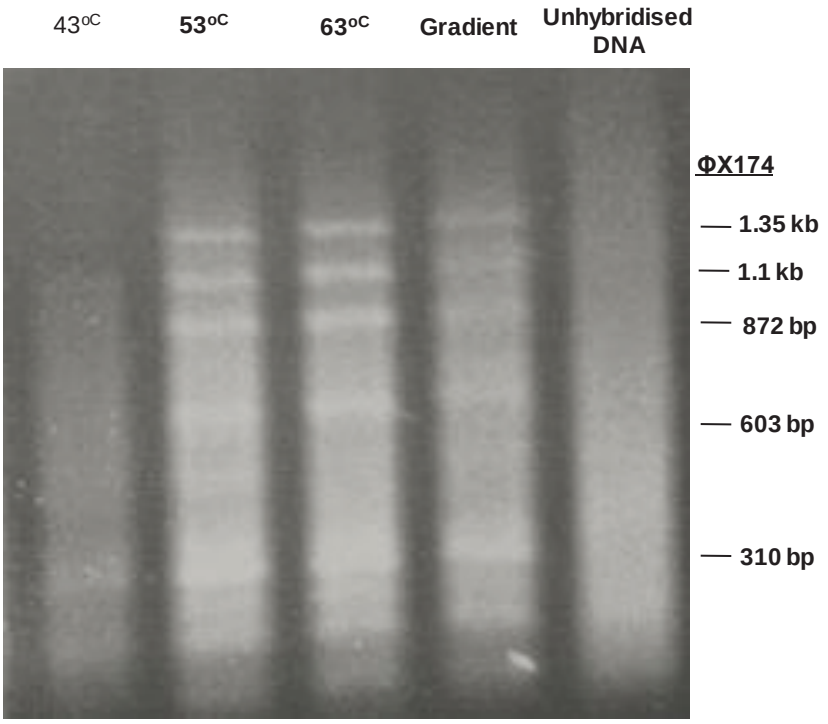


Figure 4.7. Hybridisation under various temperatures.

7 µl of nested product was electrophorised on a 2% agarose gel and stained in ethidium bromide. The first three reactions were for primary and secondary hybridisations performed at 43 °C, 53 °C and 63 °C, respectively. The fourth reaction's primary hybridisation was carried out at 63 °C, but the secondary hybridisation was carried out at 63 °C initially, with the temperature reduced overtime to 43 °C. The fifth PCR was a control using unhybridised DNA. The banding patterns and intensities of the amplicons for the hybridisations carried out at 53 °C and 63 °C appeared to be identical. The same ΦX174 *Hae*III-digest banding pattern was observed for the gradient hybridisation, but the bands were less intense. However, it was difficult to distinguish banding for the hybridisation carried out at 43 °C.

The mobility of the DNA size marker (*Hae*III-digested bacteriophage ΦX174 *am3 cs70* DNA) is shown.

4.3.6 Sequencing

Amplicons were cloned using the Invitrogen TOPO TA Cloning® system and plasmids were extracted from randomly selected white clones, using a Qiagen Miniprep kit, as described in Sections 2.5.7 and 2.5.8. Plasmids were sequenced using M13F and M13 R (Table 2.2) and sequences were identified using NCBI BLAST-N analysis (Section 2.5.10). Clones were picked and their plasmids sequenced until previously identified sequences were being re-identified. At this point it was assumed that all potential clones which could be cloned had been examined.

The percentage of false positives for two separate PCR-based subtractive hybridisations using *Rsa*I was 79.55% (35/44) and 79.48% (31/39). False positives are defined as sequences which are found in either D39 or TIGR4 and as such are not unique to the serotype 3 strains. A hypothetical protein that had the first 57 bp of sequence missing in both D39 and TIGR4, as well as a reductase that demonstrated regions of sequence variation in both D39 and TIGR4, which were both identified in the second subtractive hybridisation, were not included in the calculation of false positives. Sequencing revealed that the false positives contained both adaptors, indicating the annealing of homologous sequences from two different testers. Some sequences detected were absent in D39 but present in TIGR4, including TIGR4 capsule sequence. However, the subtractive hybridisation did allow the cloning of unique sequences including the serotype 3 capsule gene sequence, which acted as the positive control. Furthermore, the IIA and IIB components of a putative cellobiose PTS were also identified in both subtractive hybridisations. Although these components could not be identified in D39, R6 or TIGR4 using the BLAST-N program of NCBI, these components could be identified in G54 using the BLAST-N program available through the website of the J. Craig Venter Institute's (JCVI) Comprehensive Microbial Resource (CMR) [http://cmr.jcvi.org]. The

genome of G54 was not included in the database of NCBI at the time. Furthermore, the components were associated with other PTS components. Shortly after the PCR-based subtractive hybridisation work was complete, the fully sequenced genomes of the pneumococcal strains Hungary19A-6 and CGSP14 became available through the Kyoto Encyclopedia of Genes and Genomes (KEGG) [http://www.genome.jp]. The IIA and IIB components could be identified in these two strains and were also associated with other PTS components. Figure 4.8 illustrates the putative cellobiose PTS in Hungary19A-6, available through KEGG.

The putative PTS was confirmed to be present in all 4 tester isolates by PCR analysis using the primers G54 spn02109F and spn02210R (Table 2.2). The sequence of the PTS in strains of serotype 3 ST180 and ST458 will be further discussed in Chapter 5.

4.4 Discussion

In this chapter, DNA microarray and PCR-based subtractive hybridisation were employed to complement each other in order to maximise the probability of identifying genes with a possible role in OM. The microarray approach yielded no genes to be investigated further for a role in OM, although a few interesting observations were made. No differences were detectable between the strains WCH208 and MSHR12, as well as strains MSHR13 and MSHR15, which MLST analysis and/or Western blot analysis suggest are highly related. This is a similar result to another study comparing the genomes of two serotype 3 isolates from the ST180 lineage using a microarray slide also containing ORFs from R6 and TIGR4 (Dagerhamn *et al.*, 2008). Due to the limitation of the microarray slide only containing genes annotated in R6 and TIGR4, there may therefore be differences in genes that are not represented on the microarray slide.

Most of the genes detectable in MSHR17 (ST458) but not in strain WCH206

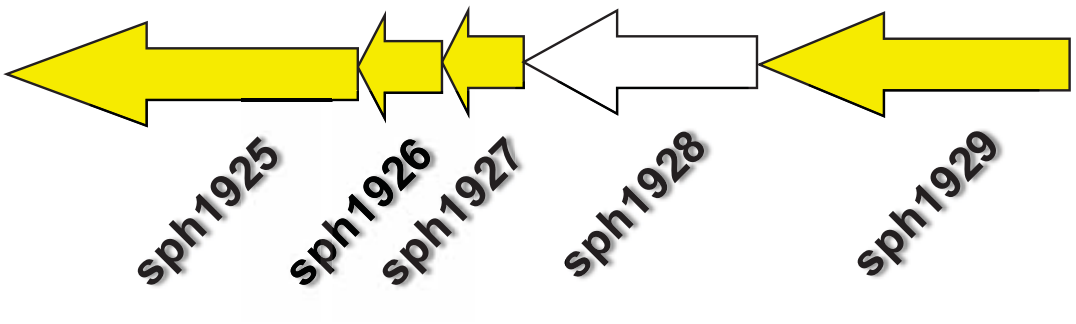


Figure 4.8. The putative cellobiose PTS in *S. pneumoniae* Hungary19A-6 (nt 1788843-1793033).

sph1925; PTS, cellobiose-specific IIC component phosphotransferase system, sph1926; lichenan-specific phosphotransferase enzyme IIA component, sph1927; lichenan-specific phosphotransferase enzyme IIB component, sph1928; hypothetical protein, sph1929; lichenan permease IIC component.

Genes were assigned colour according to their KEGG pathway colour code with yellow indicating membrane transport and white indicating an unassigned pathway.

(ST180) were from three known accessory regions (AR12, AR17 and AR41) that are differentially distributed amongst pneumococcal strains (Bruckner *et al.*, 2004; Silva *et al.*, 2006; Obert *et al.*, 2006; Blomberg *et al.*, 2009). However, it is also possible that the genes are actually present but were not detectable due to sequence variation with respect to R6 and TIGR4. One of these genes, spn0474, is annotated as the cellobiose-specific IIC component of a PTS-system, but this is not from the same PTS identified through the PCR-based subtractive hybridisation (Section 4.3). This gene is part of AR12 (Figure 4.1), which includes IIA and IIBC components of a lactose-specific PTS. It is not known if spn0474 forms part of this PTS. Furthermore, the annotation of spn0474 as cellobiose-specific is also questionable and the homologue of spn0474 in Hunagry19A-6 (sph0586) has been annotated as lactose-specific. Another region of interest is spn2158-spn2166 (AR41 [Figure 4.1]), which has been associated with virulence in TIGR4 (Embry *et al.*, 2007). This region is discussed in more detail in Chapter 7.

Genes from the same accessory regions were also found to be differentially represented in the serogroup 11 isolates. In addition, genes from seven additional accessory regions (AR10, AR27, AR31, AR34, AR35, AR36 and AR40), which were shown in Figure 4.2, were identified as being differentially carried. One gene, spn1326 (*nanC*), was shown to be absent in MSHR1 by DNA microarray analysis (Table 4.5) had previously been shown to be absent in this strain by PCR analysis (Section 3.2.1).

When serotype 3 and serogroup 11 were compared by DNA microarray analysis, it was not possible to identify an ORF that was carried in all strains of one serogroup but not the strains of the other serogroup, which indicates there is a high degree of genetic variability within the serotype/groups. This is consistent with the observation made by Dagerhamn *et al.* (2008) following DNA microarray analyses of pneumococcal strains of various serotypes. Initially, 3 genes were identified as being in serotype 3, but not in serogroup 11 (Table 4.8). However, two of these genes had previously been suggested to

be absent in WCH206 (serotype 3) but present in MSHR17 (serotype 3) when the two genomes were compared (Section 4.2.4). The third gene encoded a putative endonuclease and although it was carried in both WCH206 and MSHR17, it was also found to be present in some serogroup 11 strains (Table 4.10). In comparison to serotype 3, the list of genes suggested to be unique to serogroup 11 was longer. This may be due to the larger number of unrelated STs represented in the serogroup 11 sample used in this study. However, as for serotype 3, the genes suggested to be absent in serotype 3 but present in serogroup 11 were found in reality to be differentially distributed within serogroup 11 or carried by at least one serotype 3 strain. For example, genes from the region spn1315-spn1331 (AR27) were detected in serogroup 11 and not serotype 3 (Table 4.9) but when the serogroup 11 arrays were reanalysed, this region appears to be absent in strain MSHR1 but present in the other three strains (Table 4.10). One of the genes from this region, *nanC*, which was detectable in three of the serogroup 11 isolates used in Sections 4.2.3 and 4.2.4 by PCR (Section 3.2.1) was not detectable in the serotype 3 isolates used in Sections 4.2.1 and 4.2.4. This region, which was named RD8a by Obert *et al*. (2006), is of particular interest because it has been suggested to be associated with invasive isolates, although there has been disparity regarding this association (Blomberg *et al.*, 2009). Interestingly, this region has been detected in serogroup 11, which was demonstrated to be avirulent in mice, but not in serotype 3, which was shown to be highly virulent in mice (Section 3.2.5.1). The data presented here would suggest that the role of NanC in invasion is either overrated or is only required for the invasive capabilities of some serotypes.

Following DNA microarray analysis, PCR-based subtractive hybridisation was performed to determine if there were any genes of interest not represented on the microarray slides. Two modifications were made: a) the final concentration of *Taq* polymerase buffer was increased to 1.6×, and b) the length of the primary hybridisation was increased to 8 h. The latter change was made due to the inefficiency of the subtractive

hybridisation. The PCR-based subtractive hybridisation method described in the kit used an adapted method of investigating gene expression in eukaryotes and one of the changes from the originally published method was a reduction in the length of the primary hybridisation. This was done because of the low copy number of bacterial genes compared to eukaryotic transcripts, and results in a larger number of single tester molecules available at the start of the secondary hybridisation (Akopyants *et al.*, 1998). Although this is believed to increase the efficiency of the secondary hybridisation (Akopyants *et al.*, 1998), it was unknown what influence this would have on background. Therefore, it was decided to increase the length of the primary hybridisation to that reported for eukaryotic subtractive hybridisation (from 1.5 h to 8 h). This was found to reduce the amplification of non-tester specific sequences sufficiently. It is thought that increasing the incubation time gave the tester molecules more time to bind to the driver DNA in the primary hybridisation and thus be unavailable to bind to tester molecules with the alternative adaptor in the secondary hybridisation. However, the need to increase the primary hybridisation most likely reflects the complexity of the system used in this study due to the large number of strains, as well as their unrelatedness. PCR-based subtractive hybridisation was developed using one tester strain and one driver strain (Akopyants *et al.*, 1998) and it has been successfully performed with the Clontech PCR-Select Bacterial Genomic Subtraction Kit previously on *S. pneumoniae* using just one strain as the tester and one strain as the driver (Pettigrew and Fennie, 2005) without any disclosed changes to the methodology. In addition, it is known that there is higher background with unrelated strains due to sequence variation and differences in restriction fragment length affecting the binding of the driver DNA to homologous tester sequences (Akopyants *et al.*, 1998). In this study, even with the increased primary hybridisation time, almost 80% of cloned sequences still represented background. However, Agron et al (2002) found in parallel subtractive hybridisations using two unrelated *Helicobacter pylori* strains and a variety of restriction enzymes that

only 11-30 % of their sequences were tester specific. *H. pylori* has a GC content of 39% (Tomb *et al.*, 1997), which is similar to *S. pneumoniae* (Tettelin *et al.*, 2001).

Through PCR-based subtractive hybridisation, serotype 3 capsule genes, which acted as a positive control, and the IIA and IIB components of a putative cellobiose PTS were identified. This PTS was confirmed to be absent in R6 and TIGR4 and therefore, not represented on the microarray slide. Therefore, this chapter has successfully demonstrated that PCR-based subtractive hybridisation can be used to complement DNA microarray analysis by identifying genes not represented on a microarray slide. Furthermore, PCR-based subtractive hybridisation has been successfully adapted to identify genes unique to two unrelated testers. This improvement was significant in this study, since both STs were found in OM isolates and therefore, there was interest in genes common to both STs rather than simply one ST. This is important in light of the high level of variability seen in some serotypes and the number of potential accessory genes which may not play a role in virulence per se or at least not in OM.

Work to date has shown that a putative cellobiose PTS is upregulated in biofilms of *S. mutans* when compared to planktonic growth (Shemesh *et al.*, 2007). Furthermore, mutants in another pneumococcal putative cellobiose PTS were found to be attenuated in a chinchilla OM model during a signature tagged mutagenesis study (Chen *et al.*, 2007). Therefore, it was concluded that it is possible that the putative cellobiose PTS identified during the PCR-based subtractive hybridisation could play a role in OM and was worthy of further investigation, which was carried out in Chapters 5 and 6.

Chapter 5

Identification of a Genomic Island Carrying a Putative Cellobiose PTS

5.1 Introduction

In the previous chapter, a putative cellobiose PTS (PTS 1) was identified in serotype 3 OM isolates by PCR-based subtractive hybridisation. This PTS is present in other pneumococci, including G54 (serotype 19F) and Hungary 19A-6 (serotype 19A), but is absent in D39 and TIGR4. It has previously been reported in genome sequencing studies by Shen *et al.* (2006) and Croucher *et al.* (2009). However, a second putative cellobiose PTS (PTS 2) has also been reported in *S. pneumoniae* (McKessar and Hakenbeck, 2007), which is present in both D39 (spd0279-spd0283) and TIGR4 (spn0305-spn0310), and a signature-tagged mutagenesis study has provided evidence that it plays a role in middle ear colonisation (Chen *et al.*, 2007). This second system is also not ubiquitous and is absent in strains such as Hungary19A-6 and Spain23F-1 (McKessar and Hakenbeck, 2007). Therefore, the distribution of these two putative cellobiose PTSs was investigated in this chapter by PCR analysis of a variety of clinical isolates. The significant expansion of fully sequenced genomes available publically also allowed the distribution in other serotypes to be investigated further using bioinformatics. Genome sequence data from G54,

Hungary19A-6 and CGSP14 were also utilised in this chapter to investigate the genes surrounding PTS 1.

5.2 Results

5.2 1 PCR analysis of the distribution of PTS 1 and PTS 2

PCR was used in Chapter 4 to confirm that PTS 1 was present in all tester isolates used for PCR-based subtractive hybridisation (Section 4.3.6). PCR product was not obtainable for the two driver strains, TIGR4 and D39. The distribution in other serotype 3 and serogroup 11 strains used in this study, as well as other clinical isolates from our collection (Table 2.1), was investigated for both PTS 1 and PTS 2. Results are displayed in Table 5.1. The two systems were found in a wide range of clinical isolates of varying virulence potential. All isolates possessed at least one system with most harbouring both. It is difficult to link the number of systems present to virulence potential because strains that have exhibited avirulence in murine virulence models, such as WCH67, WCH68 and MSHR1 (Table 2.2, Section 3.2.5.1), have been shown to possess both systems. Distribution was also found to vary among isolates of the same serotype or serogroup, for example, PTS 1 is not present in the serotype 3 strain WU2 (ST378). However, distribution was the same for isolates of the same ST. This included two capsular variants of ST199 (serotype 19A and serotype 15B), which both possessed PTS 1 but not PTS 2.

Table 5.1. Identification of the two putative cellobiose PTS in multiple pneumococcal strains by PCR analysis.

Strain	Serotype/Serogroup (ST)	PTS 1	PTS 2
D39	2	-	+
TIGR4	4 (ST205)	-	+
WCH43	4 (ST205)	-	+
WCH206	3 (ST180)	+	+
WCH207	3 (ST180)	+	+
WCH208	3 (ST458)	+	+
MSHR11	3 (ST458)	+	+
MSHR12	3 (ST458)	+	+
MSHR13	3	+	+
MSHR14	3 (ST458)	+	+
MSHR15	3	+	+
MSHR17	3 (ST458)	+	+
WU2	3 (ST378)	-	+
WCH211	11 (ST3020)	+	-
WCH213	11 (ST62)	+	-
MSHR1	11A (ST3021)	+	+
MSHR5	11A (ST662)	+	-
2663	11A (ST3019)	+	-
3518	11A (ST62)	+	-
1	1 (ST304)	+	+
4	1 (ST227)	+	+
1861	1 (ST3079)	+	-
4496	1 (ST3018)	+	+
3773	15B (ST199)	+	-
4104	19A (ST199)	+	-
WCH16	6A	+	+
WCH18	6B	+	+
WCH50	6	+	+
WCH132 (L82016)	6B	+	+

WCH33	8	+	+
WCH36	14	+	+
WCH39	14	+	-
WCH60	14	+	-
WCH61	14	+	+
WCH38	19	+	+
WCH64	19	+	+
WCH65	19	+	+
EF3030	19F	+	+
WCH67	23	+	+
WCH68	23	+	+
WCH69	23	+	-
WCH101	23	+	-

Strains from the laboratory's collection were analysed. + indicates that PCR product was obtained, while – indicates no product was obtained. Mulitplex PCR was performed using the primers AD20 and AD21 (Table 2.2), which amplify *ply*, as a control. The primers spn02109F and spn02110R were used to amplify PTS 1, while the primer pair sp0303/0304F and sp0306R were used to amplify PTS 2. In some cases, primer pairs G54 spn02106F and spn02113R or sp0303F and sp0310R (Table 2.2), which targeted PTS 1 and PTS 2 respectively, were used to confirm the results

5.2.2 Bioinformatic analysis of the distribution of putative cellobiose PTSs

The number of fully sequenced genomes available online has expanded considerably since the PCR-based hybridisation work described in Chapter 4 was completed and has allowed a greater understanding of the distribution of these two PTSs using online search engines. BLAST-N and BLAST-X (Section 2.5.11) searches identified PTS 2 in thirteen pneumococcal strains (Table 5.2). Whereas, searches using the same programs with the sequence identified through PCR-based subtractive hybridisation indicated that this system is present in 17 pneumococcal strains. Two strains of the database, SP3-BS71 and SP11-BS70, are isolates of serotype 3 ST180 like WCH206 and WCH207, and serogroup 11 ST62 like WCH213, respectively (Shen *et al.*, 2006). SP3-BS71 has the same PTS distribution as WCH206 and WCH207, and SP11-BS70 has the same PTS distribution as WCH213. Another strain that also originated from the comparative genomic analysis study of Shen *et al.* (2006), SP14-BS69, does not appear to carry either PTS. In this same study, the authors identified 58 novel sequences that were not in TIGR4 or R6 and one of these, which demonstrated 63% identity to a PTS cellobiose-specific protein from *Enterococcus faecium* (Shen *et al.*, 2006), is believed to be from PTS 1. Shen *et al.* (2006) showed that this sequence was expressed in the identical strains (SP9-BS68, SP11-BS70, SP3-BS71, SP23-BS72, SP19-BS75) identified as carrying PTS 1 by BLAST-X analysis in this section (Table 5.2).

Due to the large number of partial and complete genome sequences available, other transporters in the pneumococcal genome may have been annotated as putatively cellobiose-specific. Therefore, searches in the available pneumococcal genome annotations were conducted through the website of KEGG (Section 2.5.11) using the keyword “cellobiose”. One of the genes identified (annotated as spn0474 in TIGR4) was

Table 5.2. Presence of putative cellobiose PTSs in pneumococcal genome sequences.

Strain	PTS 1	PTS 2
R6	-	+
D39	-	+
TIGR4	-	+
G54	+	-
Hungary19A-6	+	-
CGSP14	+	+
Taiwan19F-14	-	+
70585	+	+
JJA	+	-
P1031	+	-
ATCC700669 (Spain23F-1)	+	-
SP9-BS68 (type 9V, ST1269)	+	-
SP14-BS69 (type 14, ST23)	-	-
SP11-BS70 (type 11, ST62)	+	-
SP3-BS71 (type 3, ST180)	+	+
SP23-BS72 (type 23F, ST37)	+	+
SP6-BS73 (type 6A, ST460)	-	+
SP18-BS74 (type 18C, novel ST)	-	+
SP19-BS75 (type 19F, ST485)	+	+
CDC1873-00 (type 6A)	+	+

CDC1087-00	-	+
CDC3059-06	+	-
SP195	+	-
MLV-016	+	-
CCRI1974	+	-

The programs BLAST-N and BLAST-X of NCBI (Section 2.5.11) were used to analyse the distribution. + indicates the strain carries the PTS and indicates it does not. In order to identify PTS 1, BLAST searches were conducted using the sequence identified in the PCR-based subtractive hybridisation study (Section 4.3.6). In order to identify PTS 2, the sequence of spr0282 (IIC component of PTS 2) was used. Results were confirmed by searches using the sequences of the whole PTSs (nt 1788843-1791773 in Hungary19A-6 for PTS 1 and nt 276406–281092 in D39 for PTS 2).

Information in regards to serotype and ST for strains SP9-BS68 – SP19-BS74 is from Shen *et al.* (2006).

previously identified during DNA microarray analysis (Sections 4.2.1 and 4.2.3) and is part of AR12. However, other PTS components from this accessory region have been annotated as lactose-specific, as described in Section 4.4. Two additional putative cellobiose PTSs that appeared to be ubiquitous were also identified (spd0232-spd0234 and spd1831-spd1833 in D39) and it appears pneumococcal strains carry three to four putative cellobiose PTSs.

5.2.3 Isolation of a genomic island carrying PTS 1

PTS 2 is part of a genomic island controlled by two component system TCS08, which is illustrated for D39 in Figure 5.1. Whereas the streptococcal species that the components of PTS 1 display the highest identity with are *S. sanguinis* and *S. equi* subsp. *Zooepidemicus* (61-67%), PTS 2 is believed to have originated from *S. gordonii* and includes genes for two hypothetical membrane proteins, a putative transcriptional regulator, BglG and a predicted 6-phospho-β-glucosidase, CelA or BglA (McKessar and Hakenbeck, 2007). In TIGR4, the putative *bglG* sequence has been annotated as two genes; a putative transcriptional regulator and the IIA component of a PTS. This discrepancy is due to a base pair substitution in the *bglG* gene that converts CAG to TAG which results in the gene being annotated as 1482 nt in TIGR4 instead of 1980 nt like it is in D39. The sequence of the IIA component in TIGR4 represents the last 498 bp of the *bglG* gene. The putative cellobiose PTS in D39 and TIGR4 is part of a genomic island (McKessar and Hakenbeck, 2007) and therefore, it was of interest to determine if the PTS identified in Chapter 4 was also part of a larger island. These potential genes making up the larger island may be absent in D39 and TIGR4 but not identifiable through PCR-based

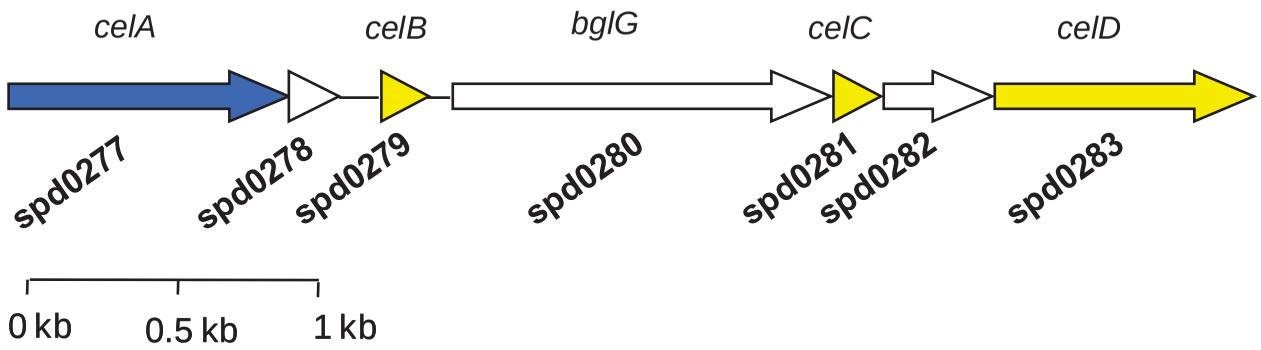


Figure 5.1. Map of the genomic island in *S. pneumoniae* D39 that carries PTS 2 (nt 274549- 281092).

spd0277, celA (6-phospho-β-glucosidase); spd0278, hypothetical; spd0279, celB, (cellobiose PTS IIB component); spd0280, putative transcriptional regulator; spd0281, celC, (cellobiose phosphotransferase system IIA component); spd0282, hypothetical; spd0283, celD (cellobiose phosphotransferase system IIC component).

Genes are coloured according to their KEGG pathway colour code with yellow indicating membrane transport, blue indicating carbohydrate metabolism, and white indicating an unassigned pathway. Black lines indicate intergenic sequence.

subtractive hybridisation for a number of reasons, such as deletions in some of the tester strains, sequence variation between tester strains or the location of restriction sites within the island. The optimal size of restriction fragments in PCR-based subtractive hybridisation is 200 bp to 1.2 kb and therefore, a gene will not be detectable by PCR-based subtractive hybridisation if by chance it is not digested into fragments of the optimal size by the restriction enzyme chosen for the experiment (Agron *et al.*, 2002). Repeating PCR-based subtractive hybridisation with restriction enzymes that recognise different sequences has been shown to increase the number of unique ORFs identified (Agron *et al.*, 2002). Therefore, the genes surrounding the putative cellobiose PTS identified through PCR-based subtractive hybridisation were analysed using KEGG genome maps of the pneumococcal strains Hungary19A-6, CGSP14 and G54 and appeared to be part of a genomic island between the genes for a transposase (spn1806 in TIGR4) and acetyl transferase, GNAT family protein (spn1807). In some strains, including D39 and G54, the transposase has not been annotated, although the sequence is present. At a similar time to when this analysis was performed, Forbes *et al.* (2008) reported that the cellobiose PTS identified by Shen *et al.* (2006), which is the same as the one found through PCR-based subtractive hybridisation, was part of a ten gene mobile genetic element, although the authors did not name the other genes.

At each end of the genomic island is the 18 nt sequence: CTATCTCCCCTTTTTCAT. This sequence is only found once in the intergenic region between the transposase and acetyl transferase in strains such as TIGR4 that do not harbour the genomic island. In Hungary19A-6, the overall GC content of the island was calculated to be 54.65% compared to 39% for the entire pneumococcal genome (Tettelin *et al.*, 2001). This suggests that the region is from a non-pneumococcal source and has been transposed into the chromosome at this site.

The islands of Hungary19A-6 and G54 are shown in Figure 5.2, respectively, as examples. The first gene encodes a small hypothetical protein (sph1921 and spg1692), although this sequence is considered intergenic in CGSP14. Immediately upstream of this gene is the gene for a hypothetical protein of about 250 amino acids (sph1922, spg1693). This is preceded by a sulfatase-modifying factor and a sulfatase (sph1923 and sph1924, respectively), in Hungary19A-6. The sulfatase carries the 12 amino acid sequence CVPARAALLTGL in the N-terminal region, which suggests it is a Cys-type sulfatase (Hanson *et al.*, 2004). However, in G54, the sulfatase is absent and the first 228 bp of the sulfatase-modifying factor has been deleted. The portion of sulfatase-modifying factor remaining has been annotated as a hypothetical protein of 208 amino acids (compared to 282 amino acids for the Hungary19A-6 ORF). Upstream of the sulfatase in Hungary19A-6 is the operon for the putative cellobiose PTS, which is also encoded on the complementary strand. The system consists of five components: IIC component (sph1925), IIA component (sph1926), IIB component (sph1927), hypothetical (sph1928) and IIC component (sph1929). However, the system is annotated as 7 ORFs in G54: IIC component (spg1695), IIC component (spg1696), IIA component (spg1697), IIB component (spg1698), hypothetical (spg1699), hypothetical (spg1700) and IIC component (spg1701). Sequence analysis of this region in G54 showed that there is a premature stop codon in the gene for the hypothetical protein that is annotated as sph1928 in Hungary19A-6 due to a single base-pair insertion. Therefore, spg1700 is only 309 nt compared to 897 nt for sph1928. The gene for the subsequent hypothetical protein (spg1699) resulted from the way the G54 genome was annotated and follows the original reading frame from 11 nt upstream of the early stop codon until the “original” stop codon. G54 also has a single nucleotide insertion in the IIC component (sph1925) closest to the sulfatase gene, which again results in a premature stop codon and the gene is therefore annotated as two ORFs.

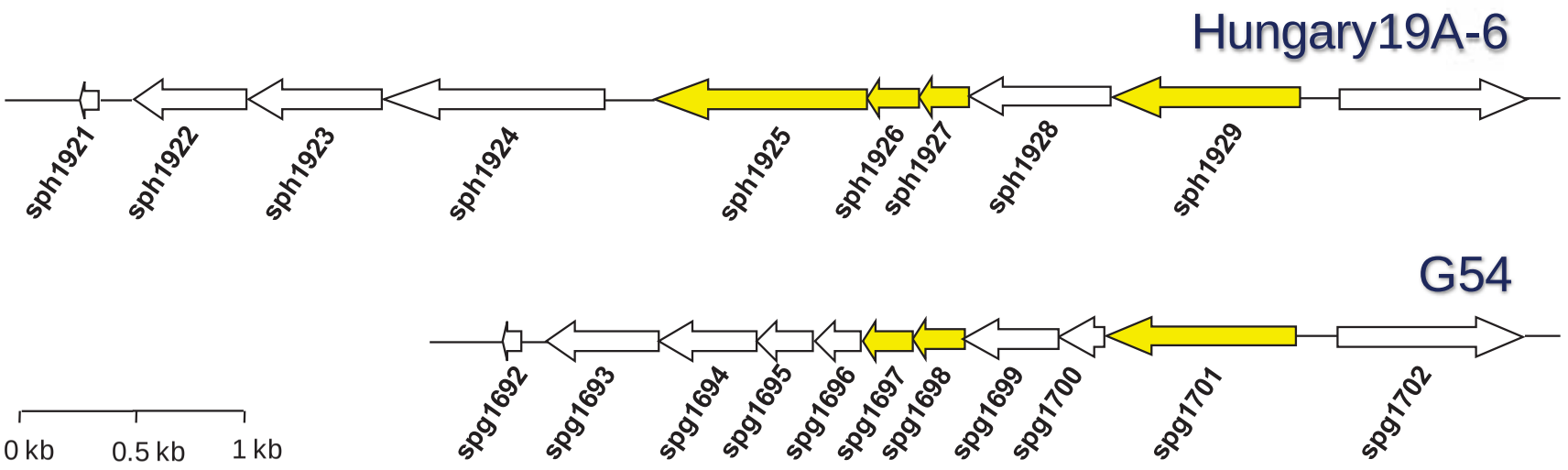


Figure 5.2. The genomic islands in *S. pneumoniae* Hungary19A-6 (nt 1784704-1794633) and G54 (nt 1630647-1637715), which includes the putative cellobiose PTS system identified in Chapter 4.

Diagram is based on the genome maps of Hungary19A-6 and G54. Genes were assigned colour according to their KEGG pathway colour code with yellow indicating membrane transport and white indicating an unassigned pathway.

sph1921/spg1693, hypothetical protein; sph1922/spg1693, hypothetical; sph1923, sulfatase-modifying factor 1, (spg1694,hypothetical); sph1924, sulfatase; sph1925, PTS, cellobiose-specific IIC component phosphotransferase system(spg1995,hypothetical; spg1696,hypothetical); sph1926/spg1697, lichenan-specific phosphotransferase enzyme IIA component; sph1927/spg1698, lichenan-specific phosphotransferase enzyme IIB component; sph1928, hypothetical protein (spg1699,hypothetical; spg1700,hypothetical); sph1929/spg1701, lichenan permease IIC component; sph1930/spg1702,ROK family protein.

In addition, the last 830 nt of the IIC component gene is deleted in G54. Therefore, there is a deletion in G54 that has removed the sulfatase and most of the second IIC componentof the putative cellobiose PTS, as well as over a quarter of the sulfatase-modifying factor.

Following the PTS in all three strains is the gene for a ROK family protein, which is 280 nt away from the first IIC component (sph1929) and is encoded on the opposite strand as the PTS. Repressors of the ROK family have a characteristic helix-turn-helix DNA-binding motif (Titgemeyer *et al.*, 1994), but sph1930 does not have this motif. The pneumococcus possesses other genes annotated as ROK family proteins, including one in AR12, the accessory region mentioned in Section 5.2.2. However, no other sulfatases or sulfatase modifying factors have been identified. Overall the island is approximately 10 kb in length in Hungary19A-6 and CGSP14, but in G54 it is only 7 kb due to the large deletion.

The premature stop codons in G54 are presumed to result in truncated products. However, there have been reports in both bacteria and viruses of translation being reinitiated due to internal start codons within close proximity to premature stop codons, and functional protein being obtained (Sarabhai and Brenner, 1967; Grodzicker and Zipser, 1968; Adhin and van Duin, 1990). Reinitiation frequencies decrease as the distance between the premature stop codon and internal start codon increase, but distances of up to 46 bp have been reported (Adhin and van Duin, 1990). In the case of the hypothetical proteins from G54's PTS 1 (spg1699 and spg1700), there are potential start codons 11 bp upstream (TTG) and 14 bp downstream (ATG) of the stop codon of spg1700 which could restore the original reading frame. The gene spg1696 (IIC component) has a potential start codon (TTG) 14 bp upstream of its stop codon, which could also restore the original reading frame. Reinitiation may also occur for BglG of TIGR4, which is associated with

PTS 2. There is a potential start codon following the nonsense stop codon, as well as another potential start codon two codons upstream.

5.2.4 Distribution of the sulfatase in the tester strains

The genomic island identified in the previous section is not present in the driver strains (D39 and TIGR4) used for PCR-based subtractive hybridisation (Section 4.3). Furthermore, in the previous section, it was found that the sulfatase is absent in G54, although it carries the island. Sequence from the sulfatase was not detected by PCR-based subtractive hybridisation and therefore, may have been missing in some of the tester strains. Therefore, the distribution of the sulfatase in these strains was explored. First, a BLAST-X search (Section 2.5.11) using the sequence of the sulfatase from Hungary19A-6 was conducted. The sulfatase was found in the ST180 strain SP3-BS71, which suggested that it would be in strains WCH206 and WCH207, which are also ST180. PCR analysis using sulfatase-specific primers (mRNA sph1924 sulfatase F and R [Table 2.2]) showed that the sulfatase gene was indeed present in WCH206 and WCH207, but could not be detected in the ST458 isolates MSHR11 and MSHR17.

The island of G54 has two insertions, which result in premature stop codons, in addition to the large deletion that includes the sulfatase gene. Therefore, it was felt that this region of the island containing the sulfatase gene and PTS genes should be analysed by sequencing in case these truncations are associated with the deletion. MSHR11 was sequenced using primers spn02106 F, spn02113 R, sph1928 F and sph1929 F (Table2.2), and compared to the sequences of G54, Hungary19A-6 and CGSP14. The strain MSHR11 did not contain the two single nt insertions that caused the premature stop codons in the two genes of G54. However, the same large deletion encompassing the end of the PTS

operon, the whole sulfatase gene and the start of the sulfatase-modifying factor, was identified. In addition, next generation sequencing data described in Chapter 7 shows that this region in MSHR17 lacks the two single-bp insertions, despite possessing the same large deletion as G54.

5.2.5 General distribution of the sulfatase

The distribution of the sulfatase gene was analysed in the strains that were identified as carrying PTS 1 in Section 5.2.1. The primers mRNA sph1924 sulfatase F and mRNA sph1924 sulfatase R, as well as mRNA sph1929 PTS IIC F and mRNA sph1929 PTS IIC R, which acted as the positive control, were used. In some cases, sph1922 R and sph1928 F with strains WCH206 and MSHR11 as size comparisons were used. All strains examined carried the sulfatase gene with the exception of 1861, 3773 and 4104.

The distribution of the sulfatase was also analysed using bioinformatics by conducting BLAST-N and BLAST-X searches (Section 2.5.11) against the nucleotide sequence of the sulfatase from Hungary19A-6 (sph1924). The sulfatase is found in strains ATCC700669 (Spain23F-1), Hungary19A-6, CGSP14, JJA, 70585, CDC1087-00, SP23-BS72, SP9-BS68, SP195, SP11-BS70, MLV-016, CCRI 1974 and SP3-BS71 but not in G54, P1031, SP19-BS75 or CDC3059-06. Therefore, most strains examined which harbour PTS 1, also carry the sulfatase.

5.2.6 *Rsa*I restriction sites of the island

*Rsa*I restriction sites were identified in the genomic island using the OXC141 genomic sequence released on the website of the Wellcome Trust Sanger Institute

(http://www.sanger.ac.uk/). The genes for the first IIC component of the putative cellobiose PTS operon (sph1929) and the ROK family protein (sph1930), as well as a large portion of the gene for the hypothetical protein (sph1928) are present on a restriction fragment of over 3.6 kb, which is well outside the optimal size for the subtractive hybridisation.

The sulfatase and much of the last IIC component of the PTS were found to be absent in MSHR11 and MSHR17 and as such, these genes could not be detected through PCR-based subtractive hybridisation. The sulfatase modifying factor (sph1923) and parts of the hypothetical protein (sph1922) and sulfatase (spg1924) are on a fragment that is 1.5 kb in length in OXC141. The 1.5 kb *Rsa*I fragment containing the genes of interest is not only outside the optimum fragment size, but the fragment varies in size and sequence in MSHR11 and MSHR17 due to the effect of the large deletion. Therefore, the portion of the sulfatase modifying factor that is present in MSHR17 and MSHR11, as well as the adjacent hypothetical protein, would be difficult to detect by PCR based subtractive hybridisation.

5.2.7 Expression of the putative cellobiose PTS and sulfatase

Real time RT-PCR was used to confirm that the cellobiose PTS was expressed in both strains WCH206 and MSHR17 and in addition, that the sulfatase was expressed in WCH206. RT-PCR was used to confirm the components of the cellobiose PTS are expressed as an operon and that the sulfatase is not included on the same transcript.

Gene expression was analysed in C+Y, SB and THY (Section2.10). Frozen stocks of opaque WCH206 and MSHR17 were made by growing bacteria in C+Y and freezing with 20% glycerol, which were then used to inoculate the three different types of media

used in this study. Cultures were grown until a concentration of approximately 10^8 CFU/ml was obtained. RNA extraction and DNase treatment were performed as described in Sections 2.10.1. RNA was confirmed to be free of contaminating DNA by RT-PCR analysis using the primers mRNA spd0016 16s F and R (Table 2.2) according to Section 2.10.2. RT-PCR products were electrophorised on a 2.5% TBE agarose gel and stained as described in Section 2.5.1. Real time RT-PCR was performed according to Section 2.10.3. The primers mRNA spd0016 16s F and R were used as an internal control with the primer pairs mRNA sph1929 PTSIIC F and R, and mRNA sph1924 sulfatase F and R, which are also listed in Table 2.2, used to analyse gene expression of the PTS and sulfatase, respectively.

Results are shown in Tables 5.3 and 5.4. Both genes were expressed in all media tested. Furthermore, growth media did not influence expression and there was no significant difference in expression of the first IIC gene of the putative cellobiose PTS between WCH206 and MSHR17.

Subsequent transcript analysis was performed using the RNA of WCH206 grown in THY. Bacterial genes are generally transcribed polycistronically (Kozak, 1983) and analysis of the sequence of the cluster of PTS genes suggested there were no internal promoters that could be used to transcribe genes indiviually. To confirm that the components of the PTS were expressed on the one transcript, the primers mRNA sph1925 PTSIIC F and mRNA sph1929 PTSIIC R (Table 2.2) were used in RT-PCR, yielding an amplicon of the expected size (2.4 kb) was generated using RT-PCR. This indicates the genes (spn1925 to spn1929) are co-transcribed and that the level of expression of the first IIC gene reflects expression of the entire operon. To determine whether the sulfatase is included on the same transcript, RT-PCR was performed using the primers sph1923R and

Table 5.3. Expression of the first IIC gene of the putative cellobiose PTS in WCH206 and MSHR17 following culture in various media.

	ΔC_T	
Medium	WCH206	MSHR17
C+Y	9.34	10.84
SB	9.64	10.35
THY	11.01	9.34

Table 5.4. Expression of the sulfatase in WCH206 following culture in various media.

Medium	ΔC_T
C+Y	3.76
SB	5.53
THY	4.63

Gene expression was analysed in C + Y, SB and THY (Section 2.2.2). Bacteria were cultured and the RNA extracted as described in Section 2.10.1. RNA was confirmed to be free of contaminating DNA by RT-PCR analysis using primers mRNA spd0016 16s F and R (Table 2.2) as described in Section 2.10.2. Real time RT-PCR was performed as described in Section 2.10.3. The primers mRNA spd0016 16s F and R were used as an internal control with the primer pairs mRNA sph1929 PTS IIC F and R, and mRNA sph1924 sulfatase F and R to measure expression of PTS 1 and sulfatase, respectively. C_T represents the average threshold cycle from three reactions. ΔC_T is C_T(PTS or sulfatase) - C_T(16s).

spn02210R (Table 2.2), and no product was obtainable under the conditions used. This is not surprising since the sulfatase gene is about 320 bp away from sph1925 (PTS IICcomponent) in the genome of Hungary19A-6 and indicates that the sulfatase is a separate transcriptional unit from the putative cellobiose PTS.

5.3 Discussion

The putative cellobiose PTS identified in Chapter 4 was confirmed to be present in all serotype 3 OM clinical isolates from this study (Table 5.1). In addition, this cellobiose PTS was also found in a wide range of isolates of various serogroups and virulence potential, including serogroups 23 and 19, which have been suggested to have an association with OM (Hausdorff *et al.*, 2000; Hanage *et al.*, 2004). However, the system is not restricted to OM isolates and it was identified in isolates of both high and low invasive potential. A second putative cellobiose PTS is also carried in a large number of the clinical isolates, although not as many as the former system (Section 4.3).

PTS 1 was found to be part of a genomic island harbouring a ROK family protein and a sulfatase with an associated modifying factor, as well as hypothetical proteins. The PTS was confirmed to be expressed as an operon in Section 5.2.7, but the surrounding genes appear to be transcribed separately. The ROK family protein is encoded on the complementary strand to the PTS operon, while the sulfatase is over 300 bp downstream of the PTS operon and was not found to be on the same transcript as the PTS operon when analysed by RT-PCR.

Another *S. pneumoniae* strain, G54, carries the PTS 1 genomic island comprising the putative cellobiose PTS and neighbouring genes, but has a deletion of about 2.5 kb, which includes the sulfatase gene, the sulfatase modifying factor, and the last gene of the

putative cellobiose PTS (Section 5.2.3). The same deletion is apparent in the serotype 3 isolates MSHR11 and MSHR17, which both belong to ST458. However, most strains carrying the PTS also have the sulfatase gene, although several additional strains that lack it were identified through either PCR or bioinformatics (Section 5.2.5). In addition to the deletion, G54 has two single-bp insertions in genes of the PTS that result in premature stop codons (Section 5.2.3). This could result in truncated product, although reinitiation may also be possible due to internal start codons nearby (Section5.2.3). MSHR11 and MSHR17 do not have these single bp insertions and therefore, they do not appear to be characteristic of strains with the large deletion (Section 5.2.4). It is not known whether the deletion in the IIC component affects the PTS because the first gene of the system also encodes a IIC component, which suggests a redundancy in the system. However, it was demonstrated in Section 5.2.7 that the cellobiose PTS is at least expressed in MSHR17 *in vitro* to a similar level as WCH206. The absence of the sulfatase gene in MSHR11 and MSHR17 explains why it was not detectable through PCR-based subtractive hybridisation. The other genes of the island, such as the ROK family protein, were not presumably detected by PCR-based subtractive hybridisation due to the large size of the *Rsa*I restriction enzyme fragment, or restriction length polymorphism.

5.3.1 Potential role of the proteins identified in the genomic island

It is tempting to speculate that the genes surrounding the putative cellobiose PTS operon may have a relationship with the operon. However, if the PTS is functional in MSHR11 and MSHR17, it suggests that it is not essential for the sulfatase to be carried with the PTS operon. Only one sulfatase has been identified in the pneumococcus to date but there are several other ROK family proteins in the pneumococcus, including one in AR12 (annotated as spn0473 in TIGR4) that is adjacent to spn0474 (cellobiose-specific

component of a PTS), which was identified during DNA microarray analysis (Section 4.2.1) and is briefly mentioned in Section 5.2.3. The ROK family comprises transcriptional repressors, ORFs of unknown function and sugar kinases, and includes a xylose repressor from *Bacillus subtilis*, a N-acetylglucosamine repressor, *nagC*, from *E. coli*., and a fructose kinase from *S. mutans* (Titgemeyer *et al.*, 1994). No DNA binding motif is detectable within the ROK family protein sequence, which suggests it is likely to function as kinase.

A link between the sulfatase and PTS has also been hypothesised (Croucher *et al.*, 2009). Sulfatases are highly conserved across eukaryotes and prokaryotes and are believed to be used by prokaryotes to hydrolyse sulphate ester groups from proteoglycans, glycosaminoglycans (GAGs) and choline sulphate (Hanson *et al.*, 2004). It is believed that sulfatases play an important role in virulence in many prokaryotic species and sulfation of host carbohydrates is used as a protective mechanism against bacterial degradation and adhesion (Hanson *et al.*, 2004). Mucin sulfatase activity has been reported in other streptococci, e.g. some *S. mutans* and *S. gordonii* isolates (Smalley *et al.*, 1994).

The number of sulfatases in a prokaryotic species can range from zero to over a hundred and in most cases they are associated with a sulfatase-modifying factor gene (Sardiello *et al.*, 2005). Prokaryotic sulfatases require activation by the post-translational modification of a cysteine (Dierks *et al.*, 1998) or serine (Miech *et al.*, 1998) residue in the active site to formylglycine, which is catalysed by the sulfatase modifying factor (Landgrebe *et al.*, 2003). In this study, it is believed to be a cysteine residue (Section 5.2.3). Apart from the sulfatase-modifying factor, no gene has been found to be associated with sulfatases across bacterial species, except for the components of an ABC-type permease, which was seen in a few genomes (Sardiello *et al.*, 2005).

The sulfatase and PTS have been hypothesised to be involved in choline acquisition in *S. pneumoniae* ATCC 700669, which is a fully sequenced strain of the Spain23F-1 lineage (Croucher *et al.*, 2009). However, the sulfatase protein together with the cellobiose system may be used to metabolise other sulphated host molecules, such as GAGs. Many GAGs contain the same β1-4 glycosidic bond as is present in cellobiose. The pneumococcus has been demonstrated to be able to use complex human glycoconjuagtes as its only carbon source, which requires hydrolysis and uptake of the complex sugar (Burnaugh *et al.*, 2008). The genomic island containing PTS 2 encodes a 6-phospho-β-glucosidase and therefore, this island may also be involved in the metabolism of complex host glycoconjugates. As previously mentioned, not all strains carrying the PTS 1 genomic island identified in this chapter possess the sulfatase. PTS 1 may transport a substrate that does not require liberation by this hydrolase or alternative hydrolases encoded elsewhere in the genome may be able to compensate. Capacity to acquire carbon sources from complex host glycoconjugates may be important for virulence, and so this possibility will be investigated in the subsequent chapter by mutagenesis analysis.

Chapter 6

Mutational Analysis of the PTS 1 Genomic Island

6.1 Introduction

The putative cellobiose PTS identified using PCR-based subtractive hybridisation in Chapter 4 (PTS 1) was found to be part of a genomic island in Chapter 5 (Figure 5.2). This island is not present in D39, R6 or TIGR4, but is carried by other pneumococcal strains, including those with fully sequenced genomes, such as Hungary19A-6 and G54. Other genes of note in the island encoded a ROK family protein and a sulfatase with its associated modifying factor. It is possible that these genes could play a role in pathogenesis, as well as PTS 1. However, this entire region is not essential for virulence in all strains, because the virulent strain G54 carries a large deletion which removes the sulfatase, as well as part of the sulfatase modifying factor and the last gene of the PTS 1 operon. When serotype 3 isolates were analysed, WCH206 (an isolate of ST180) carried the full genomic island, but MSHR17 and MSHR11 (isolates of ST458) had the same deletion that was found in G54. Still, as the genomic island contained several potential virulence factors, it was felt it was worthy of further investigation. Therefore, in this chapter, the importance of this island in IPD and OM has been investigated by mutagenesis and subsequent *in vitro* and *in vivo* analyses.

6.2 Mutational Analysis of the Whole Genomic Island

6.2.1 Construction of mutants in WCH206 and MSHR5

Initially, the whole genomic island containing PTS 1 was removed and replaced in both WCH206 and MSHR5 with an erythromycin resistance cassette attached to the promoter of the serotype 2 capsule locus. This cassette and promoter construct was amplified from the strain D39-ABΔe5 (Table 2.1) using the primers J215 and Pcps (Table 2.2). Overlap extension PCR (Horton *et al.*, 1993; Lee *et al.*, 2004), which is shown in Figure 6.1, was used to produce the construct for transformation (Section 2.5.5). The primers Pcps/sph1930/31R paired with sph1932F, as well as J215/sph1920/21R paired with sph1914F, were used for the primary PCR and the primers spg1706F and sph1914/1915R were used for the secondary PCR (Table 2.2). The resultant construct of approximately 6 kb was then used to transform WCH206 and MSHR5. Growth of competent cells and transformations were performed as described in Section 2.11.1. Sequencing with the primers J257 and J258 (Table 2.2) was performed as described in Section 2.11.2 in order to confirm the mutations. Confirmed mutants were designated WCH206 ΔIsland and MSHR5 ΔIsland, respectively.

6.2.2 Attempts to mutate MSHR17

The strain MSHR17 was not transformable in CTM under the conditions outlined in Sections 2.11.1 and 2.11.2. Additional attempts were made to transform MSHR17 using THY instead of CTM (Section 2.11.3), but no transformants were obtained. Therefore, other alternative approaches were used. The transformation efficiency of TIGR4 has been reported to be improved by lowering the pH of THY and supplementing with glycine (Bricker and Camilli, 1999). In addition, competence has been hypothesised to be a

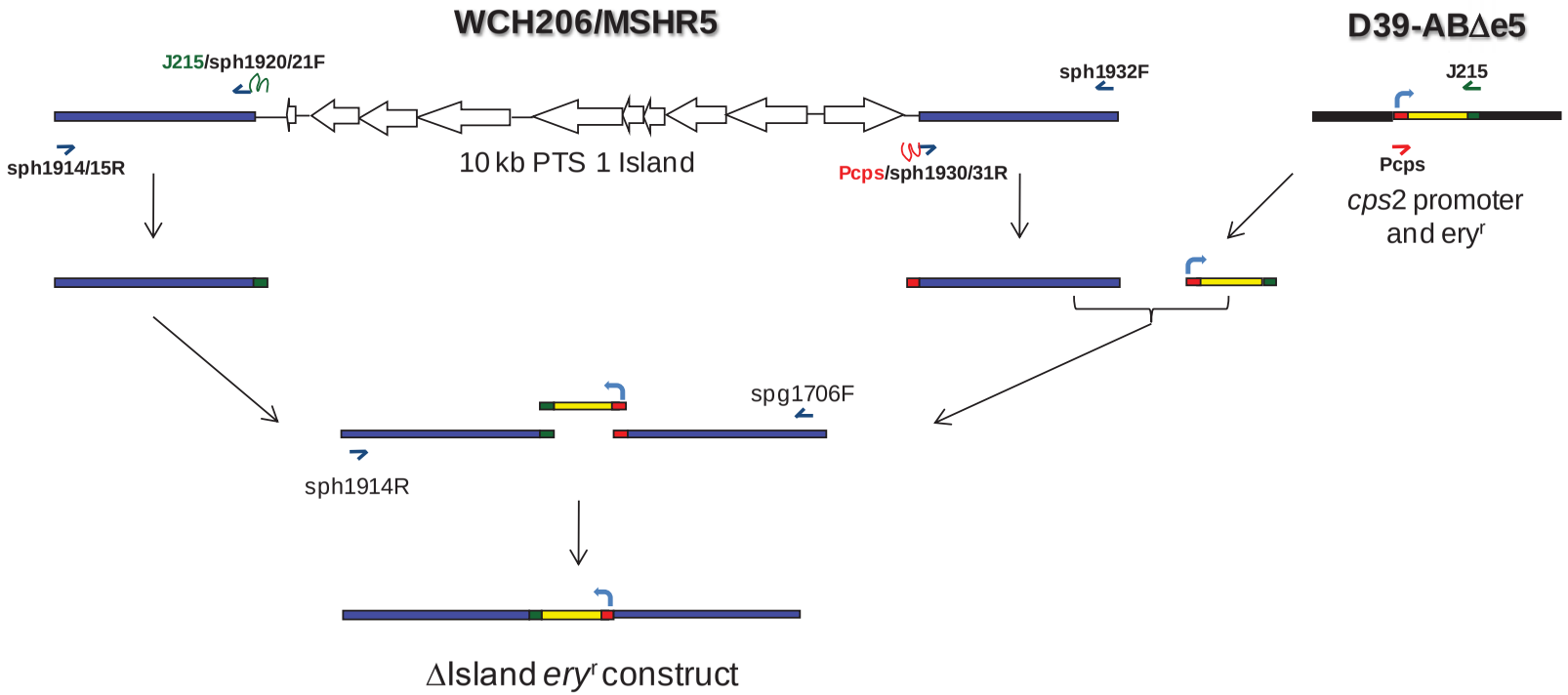


Figure 6.1. Generation of the ΔIsland *ery*^{r} construct using overlap extension PCR.

The constructs for the transformations of WCH206 and MSHR5 were obtained in a two PCR-step process. The first step involved three independent PCRs to generate the insertion fragment and two adjacent fragments. The insertion fragment in this study was a gene conferring erythromycin resistance (yellow) attached to a capsule locus promoter (pale blue arrow) in D39-ABΔe5. The adjacent fragments were the regions up and downstream of the PTS 1 island (blue). The two primers which anneal adjacent to the island had "tails" (red and green), which could not anneal to WCH206/MSHR5 DNA but were incorporated into the amplicons. The sequences of the "tails" were complimentary to the termini of the insertion fragment, which allowed the three fragments to recombine in the subsequent ligation PCR to create the construct.

general pneumococcal stress response and antibiotics have been shown to be able to induce competence without competence stimulating peptide (CSP) supplementation (Prudhomme *et al.*, 2006). A variety of antibiotics can induce competence, including the two that were investigated in this study, streptomycin and mitomycin C, which inhibit protein synthesis and damage DNA, respectively (Prudhomme *et al.*, 2006). Transformation was repeated using methods adapted from Bricker and Camilli (1999) and Prudhomme *et al.* (2006) [Sections 2.11.4 and 2.11.5]. However, despite investigating competence over a length of time or using two different antibiotics, none of these methods worked even for D39, which is transformable in CTM under the conditions described in Section 2.11.2.

CSP, which is encoded by the *comC* gene, was investigated to confirm that the issue with transformability was not due to absence of CSP or that this strain possessed a different allele. The CSP sequence of MSHR17 was therefore determined. PCR amplification using the primers *comC* F and *comC* R (Table 2.2) resulted in product, indicating that this strain does carry CSP. When this product was sequenced using the same primers, MSHR17 was found to carry CSP-1 and so the inability to transform MSHR17 is not due to use of the incorrect CSP. No further attempts were made to make MSHR17 competent for transformation.

6.2.3 Cellobiose fermentation

The purple broth fermentation assay was used to investigate fermentation of sucrose (positive control) and cellobiose by WCH206 and WCH206 ΔIsland grown in THY (Section 2.2.7). Purple broth contains bromocresol purple indicator, and therefore, when the bacteria ferment the carbohydrate supplemented in the purple broth, a colour change from purple to yellow is observed due to formation of acidic fermentation products.

This is measured as a decrease in A_{570}. Results for two independent assays are shown in Figure 6.2. Samples were taken regularly over the first 3 h period, and then a final reading was taken at 24 h. 1% chondroitin sulphate, which has a β1-4 glycosidic bond like cellobiose, was used in the first assay to see if the bacteria could ferment a more complex sugar, but no colour change was observed (data not shown). Data were analysed using unpaired *t*-tests. There was no statistically significant difference between WCH206 and WCH206 ΔIsland to ferment either sucrose or cellobiose ($P > 0.05$). Fermentation of cellobiose appeared to be less efficient for both WCH206 and WCH206 ΔIsland compared to sucrose.

In addition to the aforementioned assays, fermentation of cellobiose in the strains MSHR11, WU2, WCH211 and G54, which vary in their carriage of particular cellobiose PTSs (Table 6.1), was also investigated. The assay was performed as described above except glucose was used as the positive control and one A_{570} reading was taken at 20 h. Results are shown in Table 6.1. Data were analysed using unpaired *t*-tests. WCH211 fermented significantly more cellobiose than WU2 or G54 ($P = 0.0465$, 0.0459, respectively). However, it also fermented significantly more glucose ($P = 0.0009$, 0.0013, respectively), and the results may simply reflect better growth for this strain. Fermentation of cellobiose appeared to be less efficient than glucose for all four strains. The strain G54, which does not carry the putative cellobiose PTS described by McKessar and Hakenbeck (2007) [PTS 2], and has a large deletion in the final gene of the PTS 1 operon, as well as single bp insertions that shift the reading frame resulting in premature stop codons (Section 5.2.3), was no less capable of fermenting cellobiose than MSHR11 or WU2.

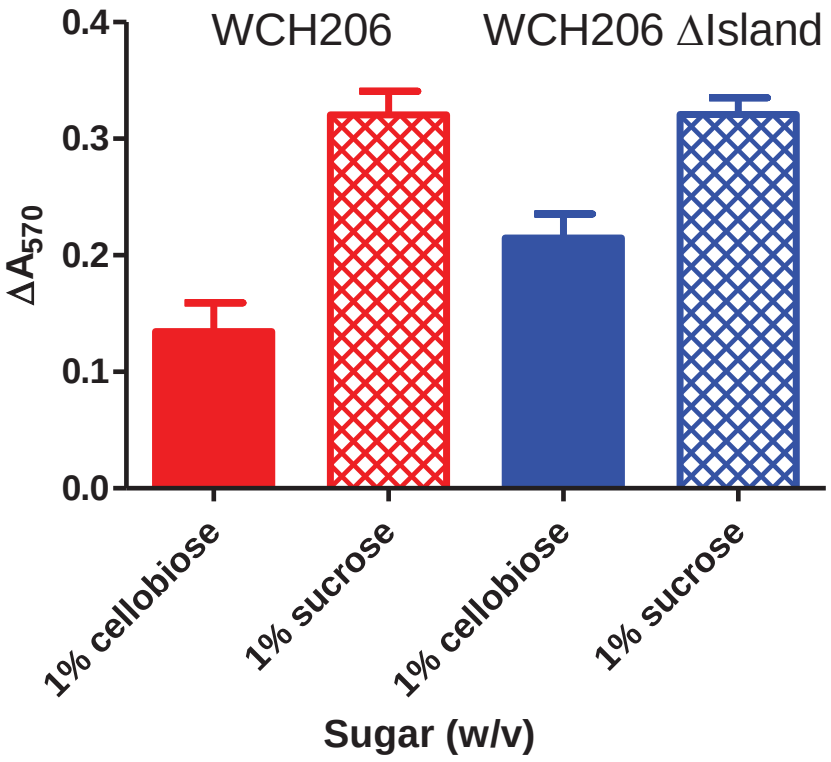
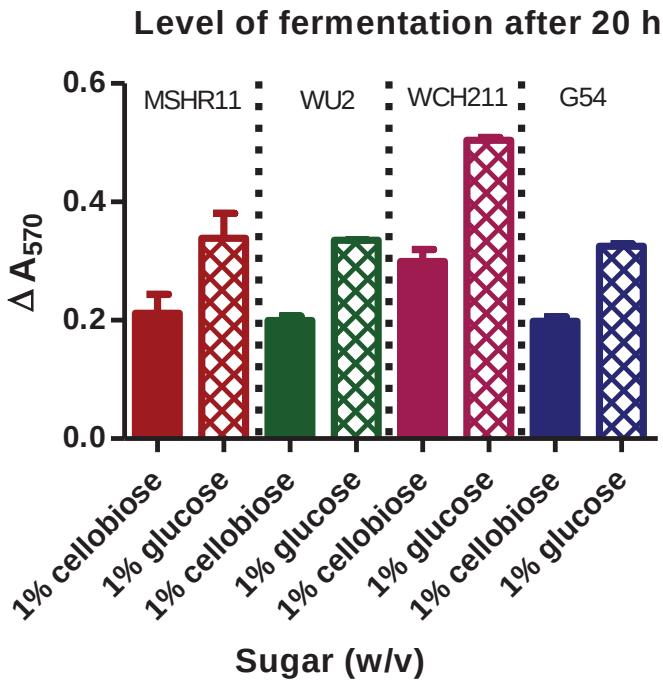


Figure 6.2. Level of fermentation after 20 h.

Fermentation was measured using a purple broth fermentation assay. Data represents two independent assays. Bacteria were grown in THY to A_{600} = 0.25 and then resuspended in purple broth supplemented with 1% (w/v) cellobiose or sucrose (Section 2.2.7). Purple broth cultures without sugar supplementation were used for the blanks. $\Delta A_{570} = A_{570}$ (0 h) - A_{570} (24 h).

Table 6.1. Distribution of PTS 1 and PTS 2, as well as the level of cellobiose/glucose fermentation over 24 h, in unrelated pneumococcal strains

	MSHR11	WU2	WCH211	G54
Serotype/group (ST)	3 (ST458)	3 (ST378)	11 (ST3020)	19F
PTS 1	+ (deletion of part of final IIC component)	-	+	+ (same deletion as MSHR11 plus 2 single bp insertions in the operon
PTS 2	+	+	-	-



Information regarding PTS distribution is from Sections 5.2.1 (Table 5.1), 5.2.2 and 5.2.3. The purple broth fermentation assay was performed as described in Section 2.2.7 with two samples prepared for each strain and sugar. Purple broth cultures without sugar supplementation were used for the blanks. $\Delta A_{570} = A_{570}$ (0 h) - A_{570} (24 h).

6.2.4 Growth

As deleting PTS 1 in WCH206 did not adversely affect the capacity to ferment cellobiose, growth in a defined medium that could be supplemented with cellobiose was not investigated. However, a growth curve was constructed for WCH206 and WCH206 ΔIsland in THY (Figure 6.3) as described in Section 2.2.7, except cultures were inoculated directly from BA plates. Interestingly, the wild-type strain grew slower and did not reach as high an optical density compared to the mutant.

Both MSHR5 and MSHR5 ΔIsland grew poorly in THY (result not shown); the growth seemed slow compared to WCH206 and WCH206 ΔIsland. Therefore, growth of these strains, as well as WCH206 and WCH206 ΔIsland was measured in BHI (Figures 6.4 and 6.5). MSHR5 ΔIsland grew slightly slower than MSHR5, but no other differences were observed.

6.2.5 Effect of mutation on biofilm formation

A 7 day biofilm assay was performed as described in Section 2.2.8. 0.5 × BHI was used because biofilms would not form in THY. The number of biofilm bacteria, as well as planktonic bacteria in the media, for triplicate samples on Days 1, 2, 4 and 7 was calculated and is shown in Figure 6.6 and 6.7. No statistically significant difference in the bacterial numbers for both biofilm and planktonic forms was seen between the wild-type and respective mutant (Mann Whitney *U*-test).

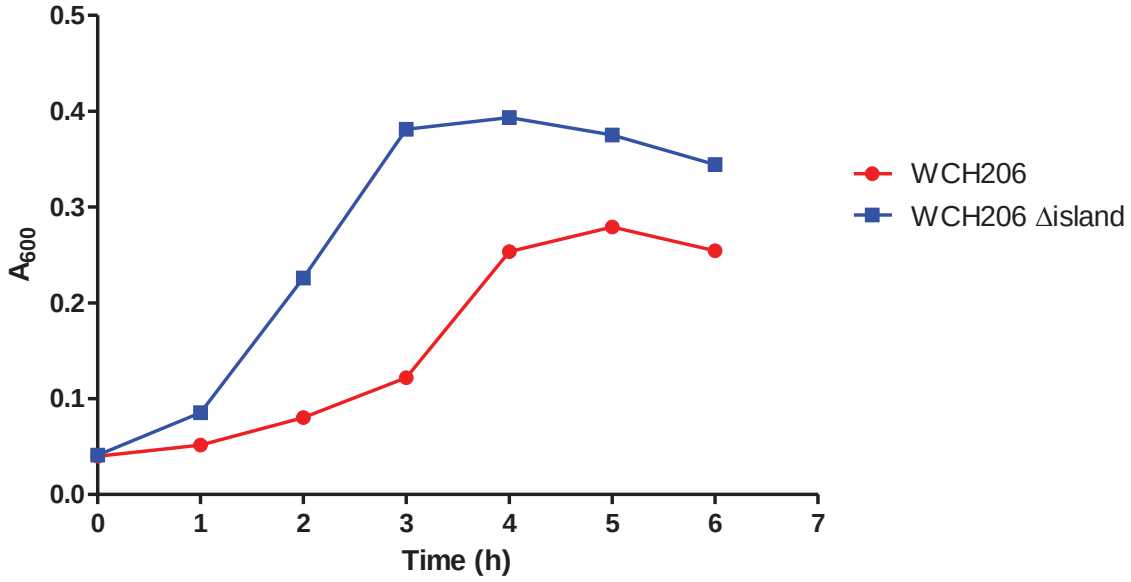


Figure 6.3. Growth of WCH206 and WCH206 ΔIsland in THY over a 6 h period.

A_{600} was measured using a Sherwood Colorimeter 257. A culture was split into four replicates for each strain and A_{600} readings were taken every hour over a 6 h period. Data are the mean of quadruplicate cultures (standard error is too small for error bars to be visible on the graph).

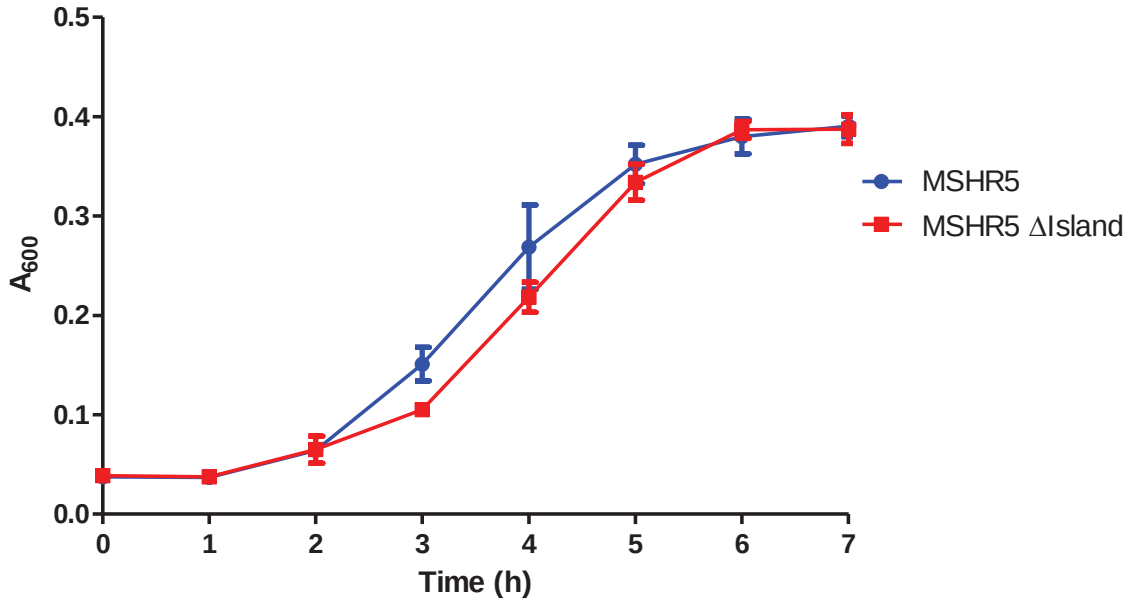


Figure 6.4. Growth of MSHR5 and MSHR5 ΔIsland in BHI over a 7 h period.

A_{600} measurements were taken every hour using a Sherwood Colorimeter 257. Data are from triplicate experiments, each which comprised a single culture for each strain.

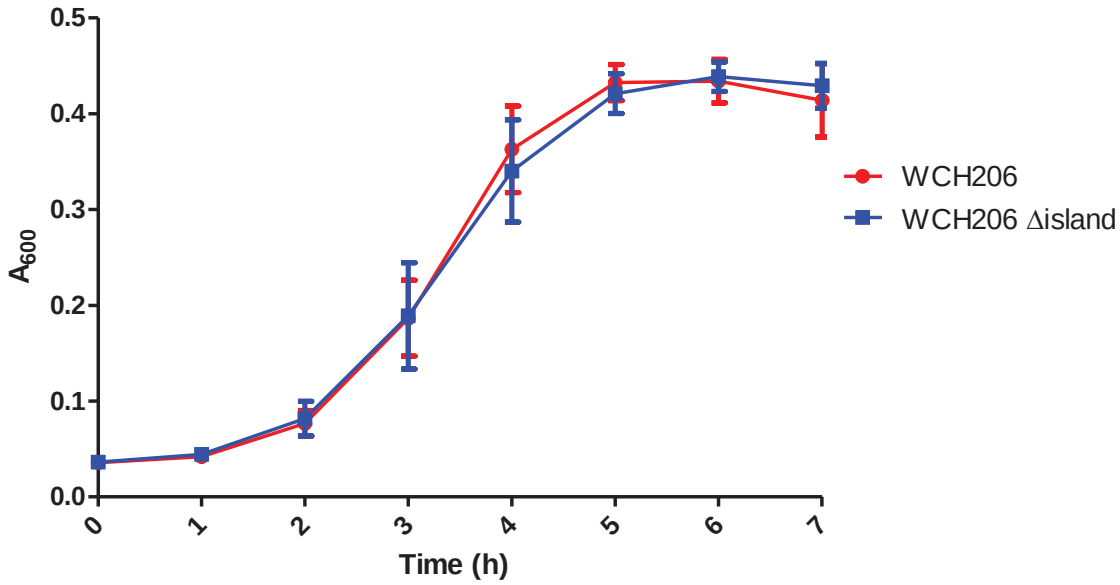


Figure 6.5. Growth of WCH206 and WCH206 ΔIsland in BHI over a 7 h period.

A_{600} measurements were taken every hour using a Sherwood Colorimeter 257. Data are from triplicate experiments, each which comprised a single culture for each strain.

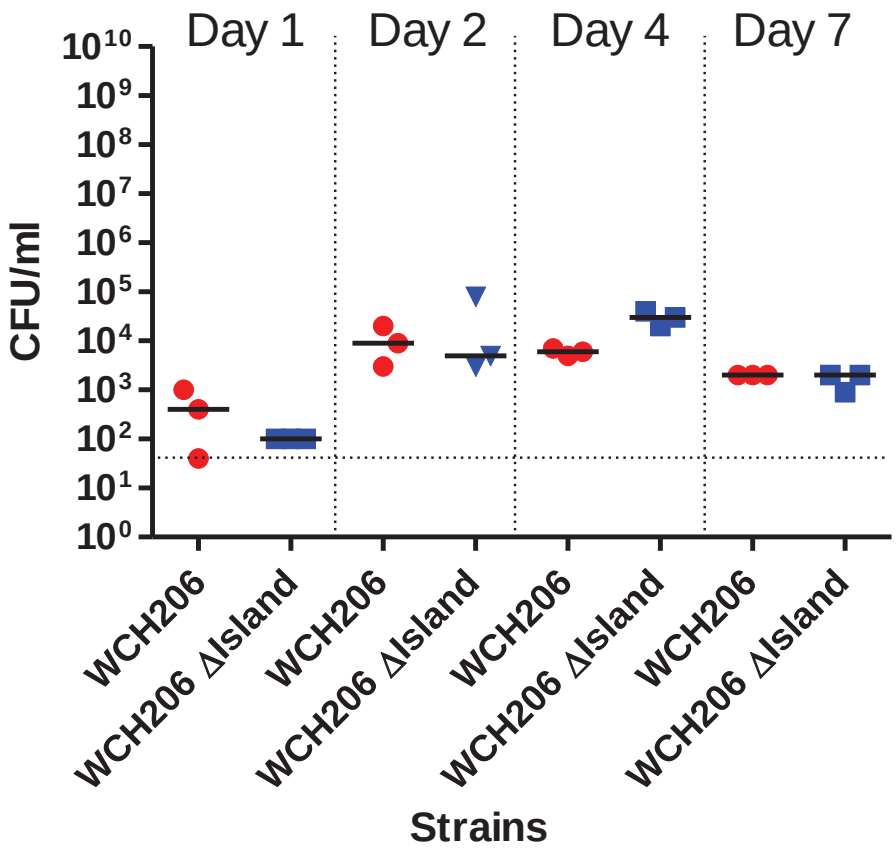


Figure 6.6. Number of biofilm bacteria.

WCH206 and WCH206 ΔIsland were grown in flat bottom tissue culture trays in 0.5 × BHI, and the numbers of bacteria that had formed biofilms were analysed after 1, 2, 4 and 7 days of incubation. The limit of detection is 40 CFU/ml.

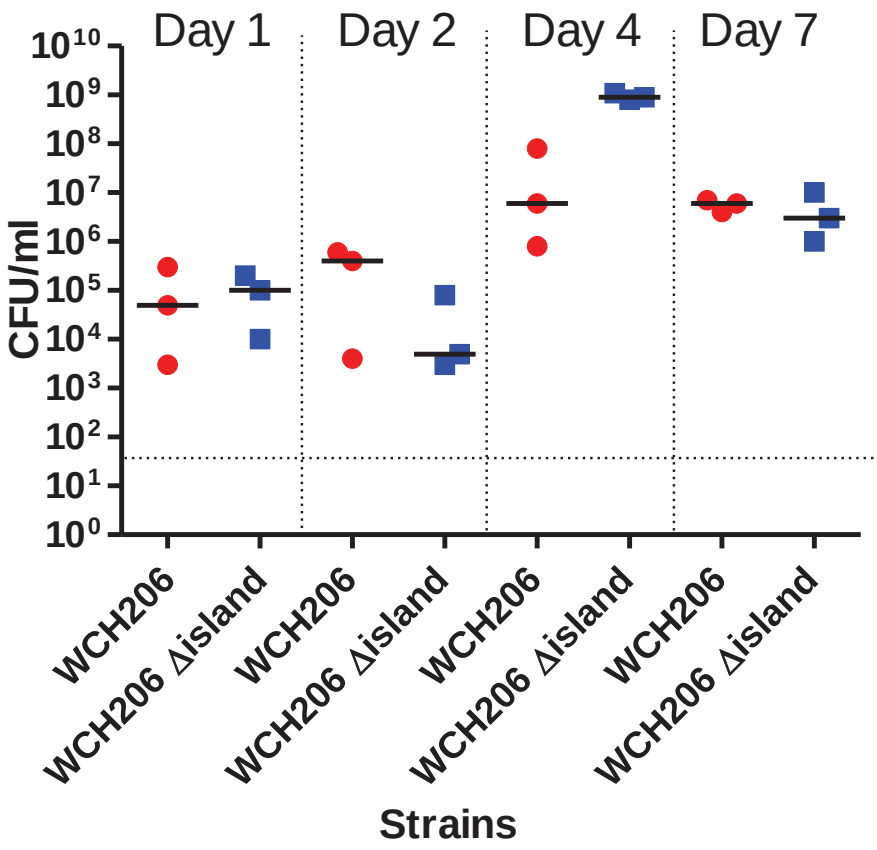


Figure 6.7. Number of planktonic bacteria.

WCH206 and WCH206 ΔIsland were grown in flat bottom tissue culture trays in 0.5 × BHI, and the numbers of bacteria in the media were analysed after 1, 2, 4 and 7 days of incubation. The limit of detection is 40 CFU/ml.

6.2.6 Virulence studies

The strain WCH206 had been demonstrated to be virulent in both the i.p. and pneumonia/sepsis murine models in Section 3.2.5.1. Therefore, the impact of replacing the PTS 1 genomic island with an erythromycin resistance cassette on virulence was assessed in these two models. The strain MSHR5 and its mutant MSHR5 ΔIsland were not investigated since MSHR5 was avirulent in these mouse models (Section 3.2.5.1).

6.2.6.1 Intraperitoneal virulence

The i.p virulence model was performed using 10 mice per group, as described in Section 2.12.1 with the results shown in Figure 6.8. Mice were inoculated with 1-4 $\times$ 10^4 CFU/mouse of either wild-type or mutant. All bacteria used in this experiment were checked and confirmed to be in the opaque phase. All mice in both groups died within 80 h and when the median survival times were compared using the Mann Whitney *U*-test (two tailed), no statistically significant difference was found between the mutant and wild-type.

6.2.6.2 Intraperitoneal competition study

Although no statistically significant difference was observed in the median survival times of WCH206 and WCH206 ΔIsland in the i.p. model, this might not resolve small differences in virulence. Therefore, the virulence/fitness of the mutant and wild-type were examined in an i.p competition assay, which was performed as described in Section 2.12.2. This involves challenge with a mixture of mutant and wild-type and the ratio of one to the other calculated at the time of challenge and at subsequent timepoints. Ten mice were inoculated with 2 $\times$ 10^3 CFU/mouse. The input ratio (mutant CFU/wild-type CFU) was 0.5887. Five mice were sacrificed 24 h post-challenge and four mice were sacrificed 48 h

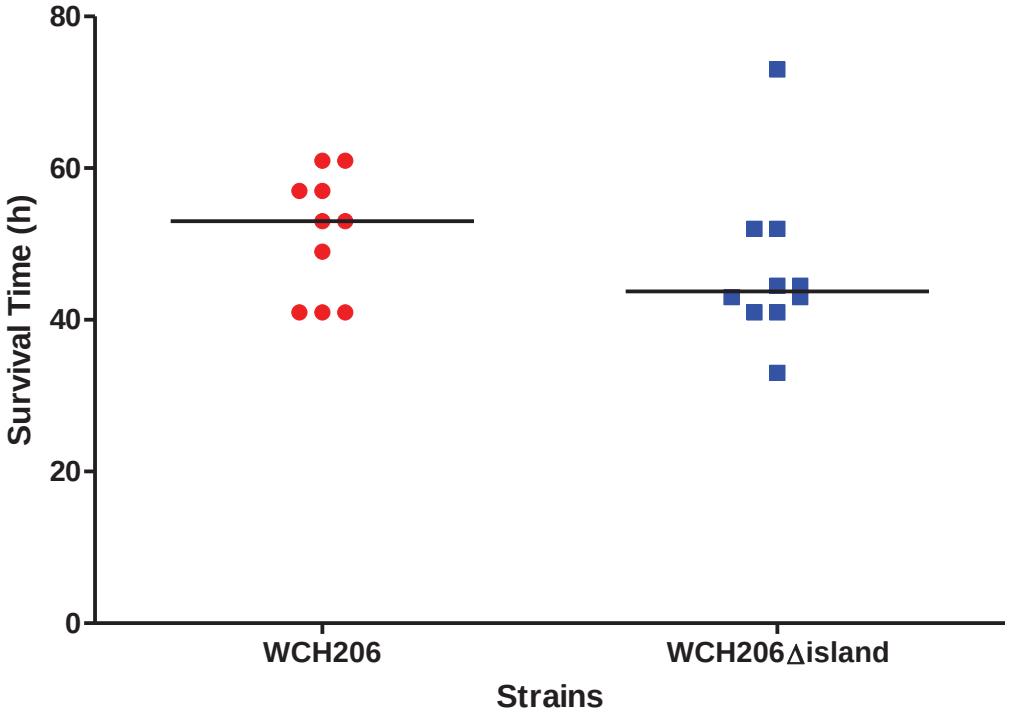


Figure 6.8. Survival time of mice following i.p. challenge.

Groups of 10 CD-1 mice were challenged through the i.p. route with either WCH206 or WCH206 ΔIsland. Retrospective plating of the inocula indicated 1.3×10^4 CFU/mouse and 3.6×10^4 CFU/mouse of WCH206 and WCH206 ΔIsland, respectively, had been administered. The median survival time is indicated by a black, horizontal line. The median survival time of the two groups was compared using the Mann Whitney *U*-test (two tailed) the *P* value obtained was greater than 0.05 indicating no significant difference exists between the groups.

post-challenge. A fifth mouse died just before the 4 mice at 48 h were euthanased and blood was taken from this mouse as well. The numbers of bacteria in the blood at the time of death is shown in Figure 6.9. The $\log_{10}$ competitive index (CI) was calculated for each timepoint, on the basis of CFU counts with or without antibiotics, as described in Section 2.12.2 (Figure 6.10). The neutral value is 0. Values less than 0 indicate that the mutant is less fit than the wild-type, while values more than 0 indicate the mutant is more fit than the wild-type. One sample *t*-tests were used to compare the mean $\log_{10}$CI of each group to the hypothetical value of 0. The mutant displayed statistically significant attenuation in the blood compared to the wild-type at 48 h post-challenge ($P = 0.0252$).

6.2.6.3 Pneumonia/sepsis model

The pneumonia/sepsis model was performed as described in Section 2.12.1 with results illustrated in Figure 6.11. Mice were inoculated with 1.2-1.3 $\times$ 10^7 CFU/mouse. Bacteria in the opaque phase were used. All mice challenged with WCH206 died within 115 h (median survival time = 59 h), while mice challenged with WCH206 ΔIsland died within 170 h (median survival time = 123.5 h), with the exception of one mouse, which survived the 21-day duration of the study. The median survival times were compared using the Mann Whitney *U*-test (two tailed) and there was a statistically significant increase in the survival time of WCH206 ΔIsland compared to WCH206 ($P<0.001$).

6.2.7 Competition colonisation study

The competition colonisation model was adapted from the colonisation model used in Section 3.2.5.2. The avirulent serotype 11A strain MSHR5 was also used in this study

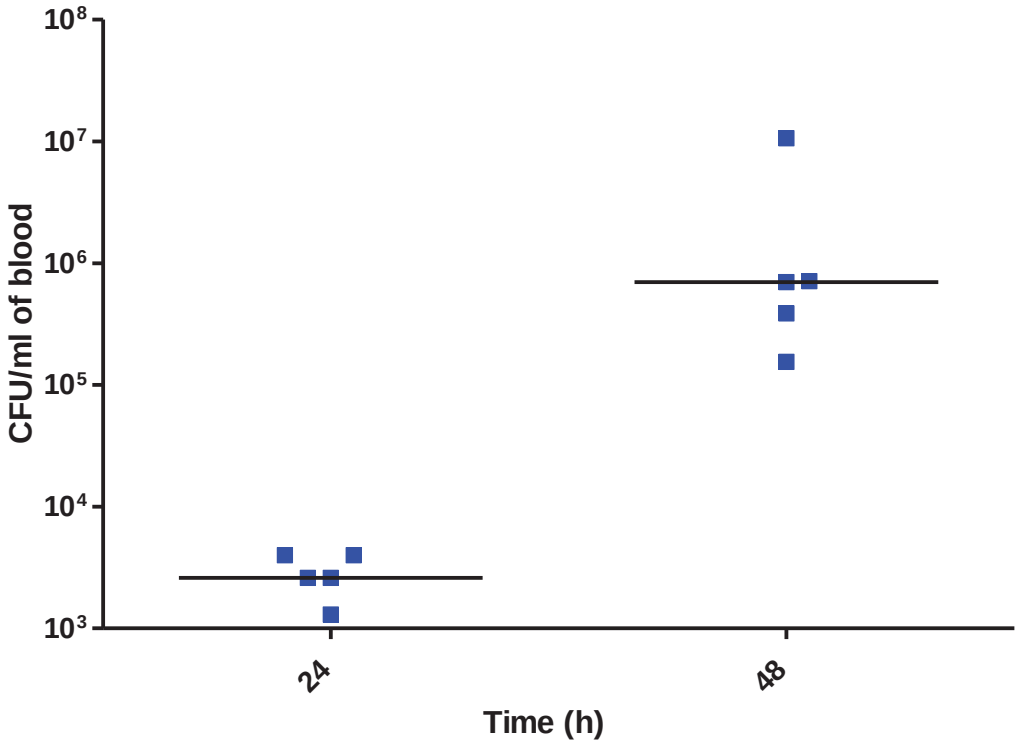


Figure 6.9. Total number of bacteria (mutant + wild-type) in the blood at 24 h and 48 h post-challenge in the i.p. competition study.

10 CD-1 mice were inoculated with 2×10^3 CFU/ml of WCH206 and WCH206 ΔIsland (initial input ratio 1:1.7) and the blood was taken from 5 mice at 24 h and 48 h post-challenge. The total CFU/ml (i.e. WCH206 + WCH206 ΔIsland) was calculated at the two timepoints. Median CFU/ml is indicated by a black, horizontal bar.

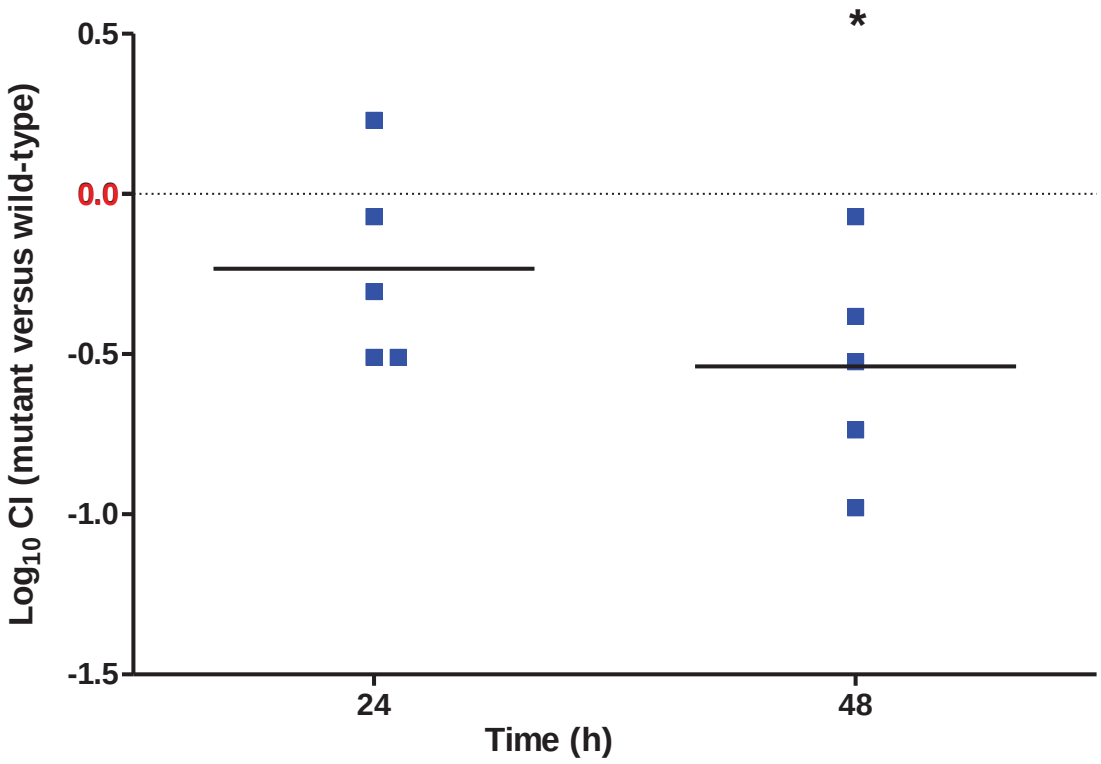


Figure 6.10. The log_{10} CI of the blood in the i.p. competition model at 24 h and 48 h post-challenge.

A group of 10 CD-1 mice were inoculated with 2×10^3 CFU/mouse of a mixture of WCH206 ΔIsland and WCH206 at a ratio of approximately 1:1.7. The ratio of mutant to wild-type was calculated in the blood for groups of 5 mice at 24 h and 48 h post-challenge in order to determine the log_{10} CI. Data were analysed using one sample *t*-tests (hypothetical value = 0).

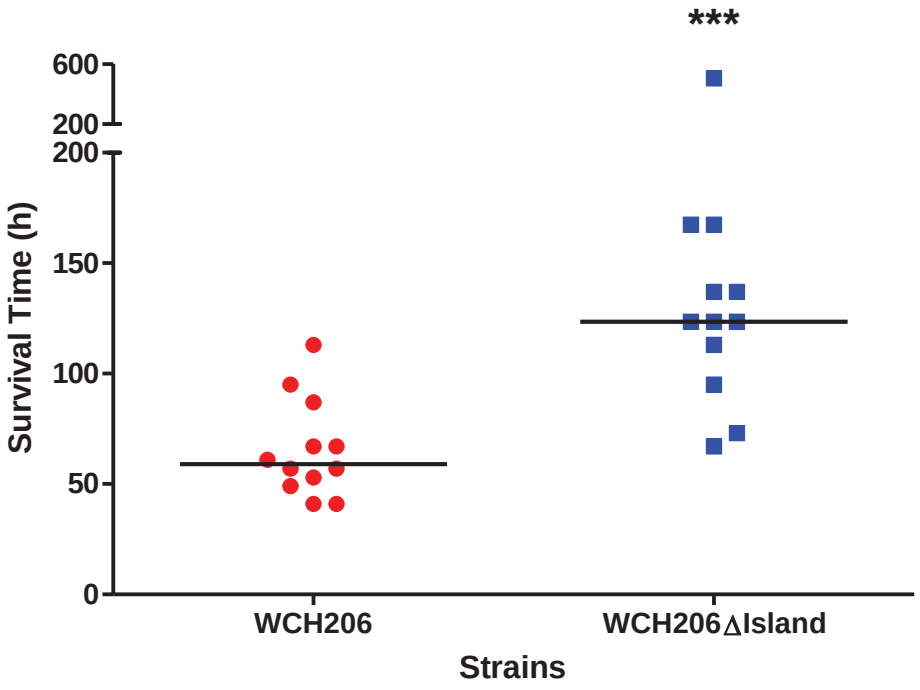


Figure 6.11. Survival time of mice challenged in the pneumonia/sepsis model.

Groups of 12 mice were inoculated with 1.2-1.3 $\times$ 10^7 CFU/mouse through the i.n. route under anaesthesia. Survival was monitored over a 504 h period. The median survival time is indicated by a black, horizontal line. The Mann Whitney *U*-test was used to compare the two median survival times (***, P < 0.001).

because it was found to colonise better than WCH206 in this model (Section 3.2.5.2). For both studies, transparent phase bacteria were used since they have been shown to be predominant in the nasopharynx compared to the opaque phase (Weiser *et al.*, 1994). Mice were inoculated with 2×10^7 CFU/mouse of WCH206/WCH206 ΔIsland and 1.5×10^6 CFU/mouse of MSHR5/MSHR5 ΔIsland. The input ratios (mutant:wild-type) were 4:1 and 1.3:1, respectively. The $\log_{10}$ CI for the nasal wash and nasopharyngeal tissue homogenate was calculated as described in Section 2.12.2 for Days 1, 2, 4 and 7 with results shown in Figure 6.12, along with the total numbers of bacteria in the nasopharynx. The other niches could not be assessed due to a large number of animals in which no bacteria were detected. In the nasopharynx there were some instances where there were low levels of infection and only the mutant or wild-type could be isolated resulting in CI values of infinity or 0, respectively. Therefore, in the case of no wild-type being present, the output ratio was defined as the number of mutant colonies plus 1. In the case of no mutant being present, the output ratio was the inverse of the following: number of wild-type plus 1. The one sample *t*-test was used to assess whether there was a significant difference between the mean $\log_{10}$CI and the null hypothesis value of 0.

In the WCH206 background, there was a statistically significant decrease in the fitness of the mutant compared to the wild-type in the nasal wash 24 h post-inoculation ($P<0.01$) and in the nasopharyngeal tissue 24 h and 96 h post-inoculation ($P<0.05$) [Figure 6.12]. Also the mutant MSHR5 ΔIsland demonstrated a statistically significant decrease in fitness compared to MSHR5 in the nasal wash at all 4 time points, with *P* values less than 0.001 on the first two days. However, for the nasopharyngeal tissue there was only a statistically significant decrease in fitness on day 7 ($P<0.05$).

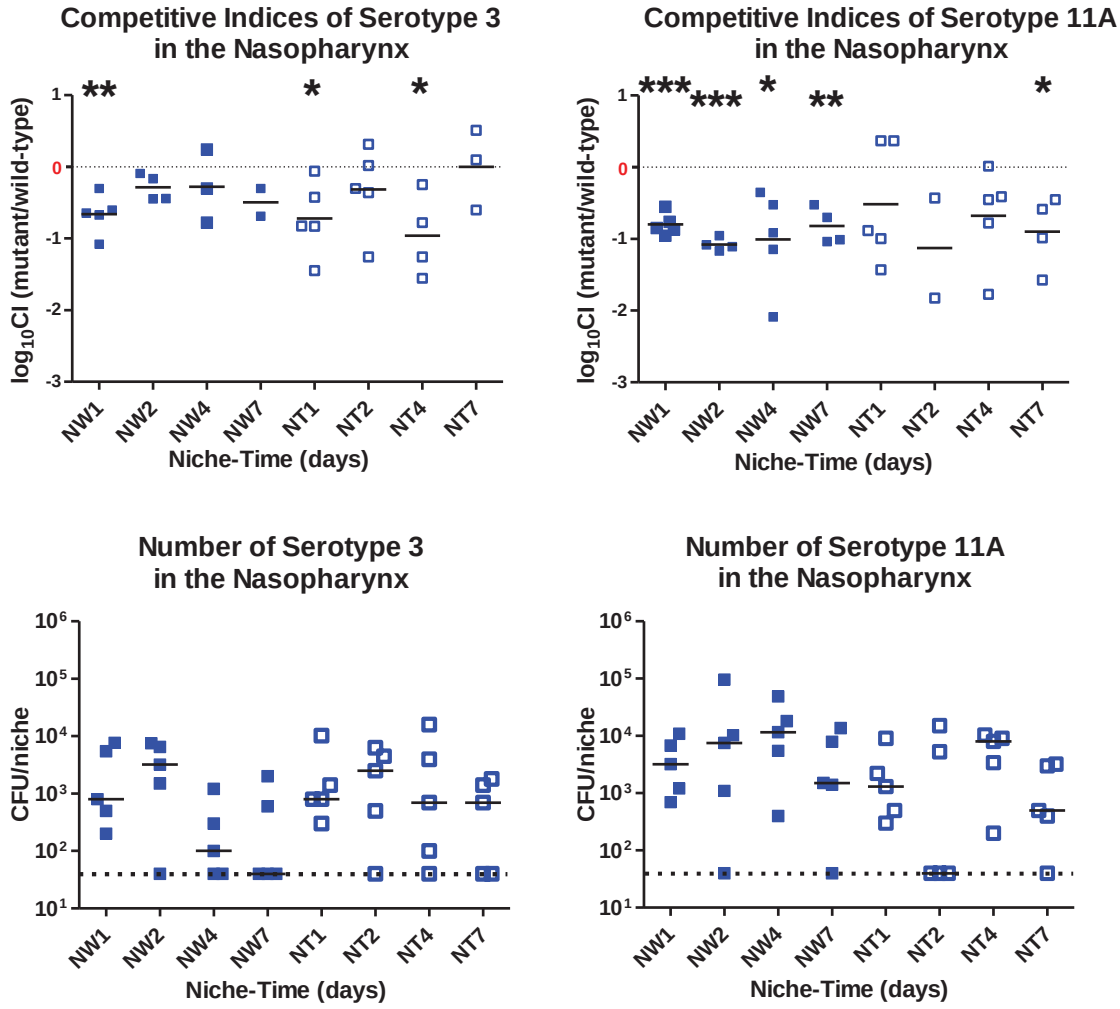


Figure 6.12. Log_{10} CI and number of bacteria in the nasopharynx at Days 1, 2, 4 and 7 post-challenge in the murine i.n. competition models for serotypes 3 and 11A.

Groups of 10 CD-1 mice were inoculated via the i.n. route without anaesthesia with 2 × 10^7 CFU/mouse of WCH206 and WCH206 ΔIsland (serotype 3) at a ratio of 1:4, or 1.5 × 10^6 CFU/mouse of MSHR5 and MSHR5 ΔIsland (serotype 11A) at a ratio of 9:7. The number of bacteria per niche and log_{10} CI in the nasal wash (NW) and nasopharyngeal tissue (NT) was calculated at Days 1, 2, 4 and 7 post-challenge. For log_{10} CI, the neutral value of 0 is indicated by a broken horizontal line (half point) and the mean is indicated by a solid, horizontal bar. One sample *t*-tests were used to compare the mean log_{10}CI of each group to the hypothetical value of 0 (*, *P* <0.05; **, *P* <0.01; ***, *P* <0.001). For the number of bacteria, the limit of detection is 40 organisms and is indicated by a thicker broken horizontal line; the median number of bacteria per niche is indicated by a solid, horizontal bar.

6.3 Mutational Analysis of Individual Components of the Genomic Island

6.3.1 Construction of mutants of strains WCH206 and MSHR5

Deletion replacement mutants were constructed for the ROK family protein, putative cellobiose PTS and sulfatase/sulfatase modifying factor by replacing the genes of interest with an erythromycin resistance gene. Mutants were created in both WCH206 and MSHR5, although the MSHR5 mutants were not ultimately used in any animal studies. The construct used to transform MSHR5 was created using overlap extension PCR (Figure 6.1), which was performed as described in Section 2.5.5. The erythromycin resistance cassette was amplified from the MSHR5 ΔIsland mutant constructed in Section 6.2.1 using the primers J214 and J215 (Table 2.2). The primers for the primary and secondary PCR reactions used to construct the overlap PCR product for the three mutants are shown in Table 6.2 with their sequence listed in Table 2.2.

The competent MSHR5 cells used in this study had a higher rate of transformability compared to WCH206 and mutants were readily obtained using overlap PCR product, which were confirmed by sequencing. Attempts were made to construct mutants in WCH206 using overlap PCR product, but WCH206 is less transformable than MSHR5 and no mutant in WCH206 had been obtained by the time all three mutants in MSHR5 had been confirmed by sequencing. It was therefore decided to amplify each erythromycin replacement mutation from the MSHR5 background mutants rather than use overlap PCR product. This is because a higher concentration of the desired construct would be obtained compared to overlap PCR, which would therefore increase the probability of obtaining a transformant. The primer pairs spg1706 F/mRNA spn02109 F, spg1706 F/sph1923 R, and sph1928 F/sph1922 R (Table 2.2) were used for ΔROK Ery^R, ΔPTS Ery^R and ΔSulfatase Ery^R, respectively. Following successful transformation, the erythromycin replacement

Table 6.2. Primer combinations used in overlap extension PCR to construct mutants.

Mutant	1º PCR	2º PCR
ΔROK	J214/sph1930 R vs. G54 spn02109 F	G54 spn02109 F vs. sph1932 F
	J215/sph1930 F vs. sph1932 F	
ΔPTS	J215/sph1925 F vs. sph1924aR	sph1924aR vs. sph1930 F
	J214/sph1929 R vs. sph1930 F	
ΔSulfatase	J215/sph1923 F vs. sph1919 R	sph1919 R vs. mRNA sph1925 PTS IIC R
	J214/sph1924 F vs. mRNA sph1925 PTS IIC R	

mutation was amplified from the erythromycin cassette-containing WCH206 transformant and used in a second transformation with a fresh aliquot of competent WCH206 cells. This was done to ensure that only the cassette was transferred from the MSHR5 background and not other non-contiguous chromosomal DNA. The second round transformants were confirmed by sequencing (Section 6.2.1) and used in the subsequent work.

6.3.2 Growth curve

Growth curves in BHI were constructed using the WCH206 and MSHR5 mutants, as well as wild-type strains, as performed in Section 6.2.3 with the exception that cultures were inoculated from a frozen glycerol stock for each strain rather than BA plates (Figure 6.13). WCH206 grew faster than the mutants, especially WCH206 ΔROK. This was surprising given there was no growth difference between WCH206 and WCH206 ΔIsland (Section 6.2.4). However, when WCH206 and WCH206 ΔROK were cultured for RNA analysis (Section 6.3.3), WCH206 ΔROK reached mid-exponential phase slightly faster than WCH206, although both were inoculated from BA into BHI with the same inoculation A_{600} (data not shown). MSHR5 ΔIsland had previously been shown to grow slightly slower than MSHR5 (Section 6.2.4), but the mutants of the individual components exhibited an identical growth rate to MSHR5, although MSHR5 ΔSulfatase grew to a slightly lower final A_{600} than the other 3 strains. In each experiment, the cultures were inoculated from the same frozen stock and therefore, cell viability, which cannot be determined by A_{600} measurement, may have influenced the results.

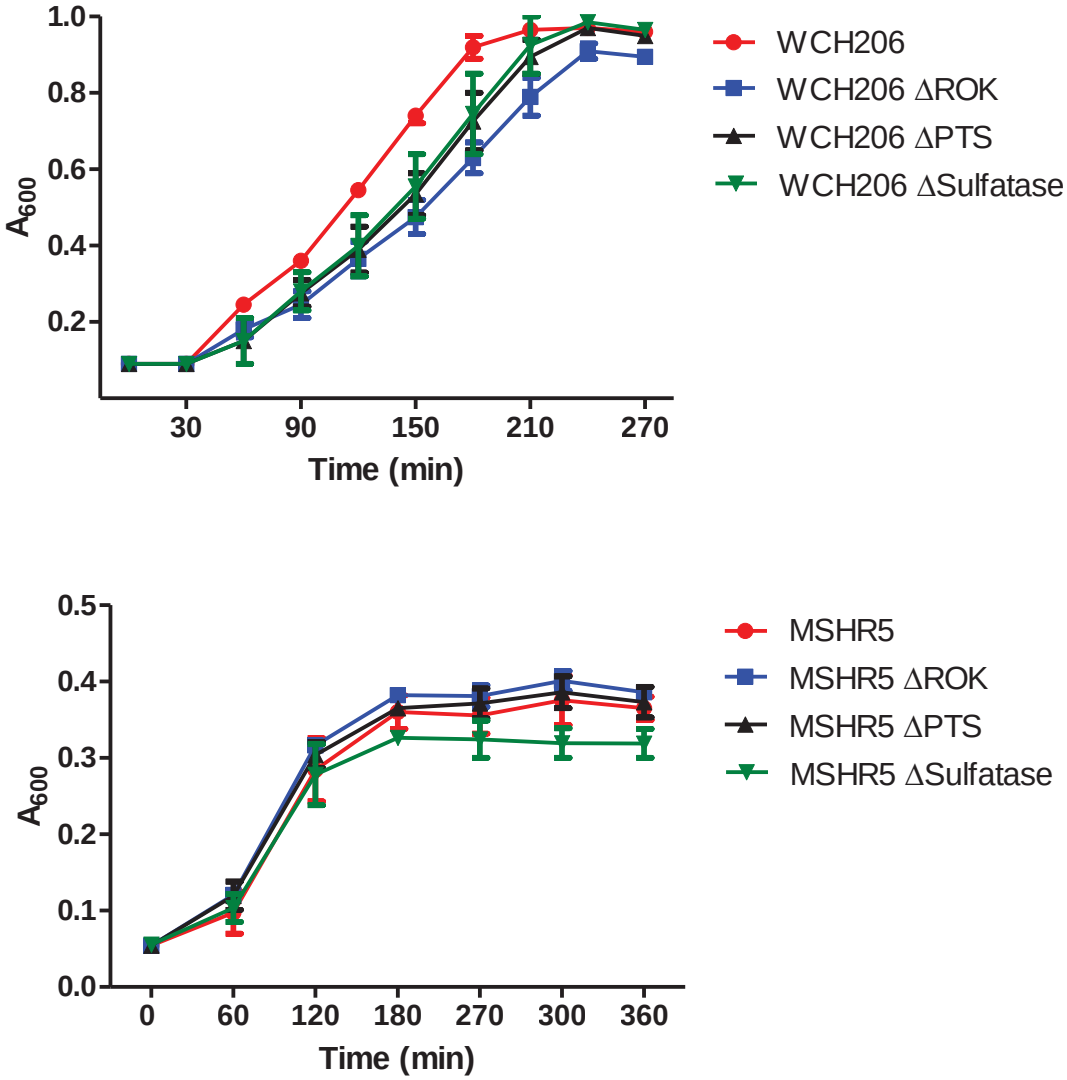


Figure 6.13. Growth in BHI of WCH206 and MSHR5 and corresponding island gene mutants over a 6 hour period.

A_{600} measurements were taken every half an hour for 6 hours. Readings for WCH206 were taken with an OD600 Diluphotometer (Implen, California, USA). However, the readings for MSHR5 were taken with a Sherwood Colorimeter 257.

6.3.3 Pneumonia/sepsis competition study

The pneumonia/sepsis competition model was performed in duplicate under the same conditions, which are described in Section 2.12.5 with the surface of the nasopharynx (nasal wash), nasopharyngeal tissue, ears, lungs and blood investigated. Transparent phase competent bacteria had been used to construct the mutants, but the pneumonia/sepsis model is performed using bacteria in the opaque phase and it had proved difficult to obtain opaque phase variants for some mutants *in vitro*. Therefore, all bacteria were passaged through mice as described in Section 2.2.9 to help isolate opaque phase variants prior to the experiment. The challenge dose was calculated retrospectively along with the input ratio of mutant to wild-type, both of which are listed in Table 6.3 for the duplicate experiments. 15 mice were inoculated per mutant and 5 mice were sacrificed from each group at 24 h, 48 h and 72 h following inoculation. The experiment was performed twice under the same conditions and the data from the two experiments combined. One mouse from the ΔROK, ΔSulfatase and ΔIsland groups had died by 72 h in the first experiment and so only four mice could be analysed. One additional mouse was used in the ΔROK and ΔPTS groups and as such there were 6 mice alive in these groups at 72 h post-challenge, which were all sacrificed.

In some niches, the mutant was not detectable, as was seen in Section 6.2.7. In these cases, due to the large amount of data in this experiment, the following nominal values were assigned as the output ratio: 10-20 wild-type colonies: 0.05, 21-50 wild-type colonies: 0.01, 51-100 wild-type colonies: 0.001. However, in niches where there were very high numbers of wild type bacteria, such as the blood, when no mutant was detected, a nominal value was calculated similar to Section 6.2.7. The blood/tissue homogenates had been serially diluted and each dilution spotted onto BA/BA-ery. The output ratio was calculated as 1/(1 + estimated CFU in the lowest dilution's spot on BA). Moreover, the

Table 6.3. Input ratio and challenge dose of the inocula from duplicate studies using the pneumonia/sepsis competition model.

	Input ratio (mutant/wild-type)		Challenge dose (CFU/mouse)	
	Experiment 1	Experiment 2	Experiment 1	Experiment 2
WCH206 ΔROK	2.3	1.2	2.3×10^7	3.7×10^7
WCH206 ΔPTS	0.5	0.5	2.3×10^7	3.2×10^7
WCH206 ΔSulfatase	1.2	2.1	2.8×10^7	3×10^7
WCH206 ΔIsland	1.3	1.6	2.1×10^7	3.6×10^7

Mice were challenged with a mixture of wild-type (WCH206) and mutant (WCH206 ΔROK, ΔPTS, ΔSulfatase or ΔIsland). The challenge dose represents the total number of bacteria (i.e. number of wild-type + number of mutant).

wild-type was not detected in some samples, and in these instances, the reciprocal of the aforementioned formula (i.e. 1 + the estimated number of mutants in the lowest dilution's spot on BA) was used as the output ratio. The log_{10}CI values were calculated as described in Section 2.12.2 and are shown in Figures 6.14 and 6.15. As with the competition studies described in Section 6.2, the one sample *t*-test was used to assess whether there was a significant difference between the mean log_{10}CI and the null hypothesis of 0.

The statistical analyses for all niches over the 72 h period are shown in Table 6.4. Only the lungs displayed statistically significant attenuation at all timepoints for all mutants. However, statistically significant reduction in mutant fitness was demonstrated in other niches for most mutants by 72 h. This included the ears where the ΔROK ($P<0.05$), ΔPTS ($P<0.01$) and ΔIsland ($P<0.01$) all demonstrated statistically significant attenuation. ΔPTS and ΔROK also had statistically significant attenuation in the ears at 24h and 48 h post-challenge, respectively. In every niche, ΔIsland was no more attenuated than the individual mutants as determined by unpaired *t*-tests.

In addition to the one sample *t*-tests, the mean and median log_{10}CI values of the nasal washes were compared to the ears for each mutant at 72 h using unpaired *t*-tests and Mann Whitney *U*-tests (two tailed), but only ΔIsland demonstrated a significant decrease in the ears compared to the nasal wash (P =0.0122 and 0.0451, respectively). Similarly, the lungs and blood were compared at 48 h and 72 h. On Day 2, there was a significant difference between the mean and median log_{10}CI values for ΔPTS (P =0.0142 and 0.0076 for the unpaired *t* test and Mann Whitney 2-tailed *U*-test, respectively), as well as the mean log_{10}CI values of the lungs and blood for ΔIsland (P =0.0241, unpaired *t*-test). On Day 3, there was a significant difference in the mean log_{10}CI value for ΔROK ($P<0.0001$, unpaired *t* test).

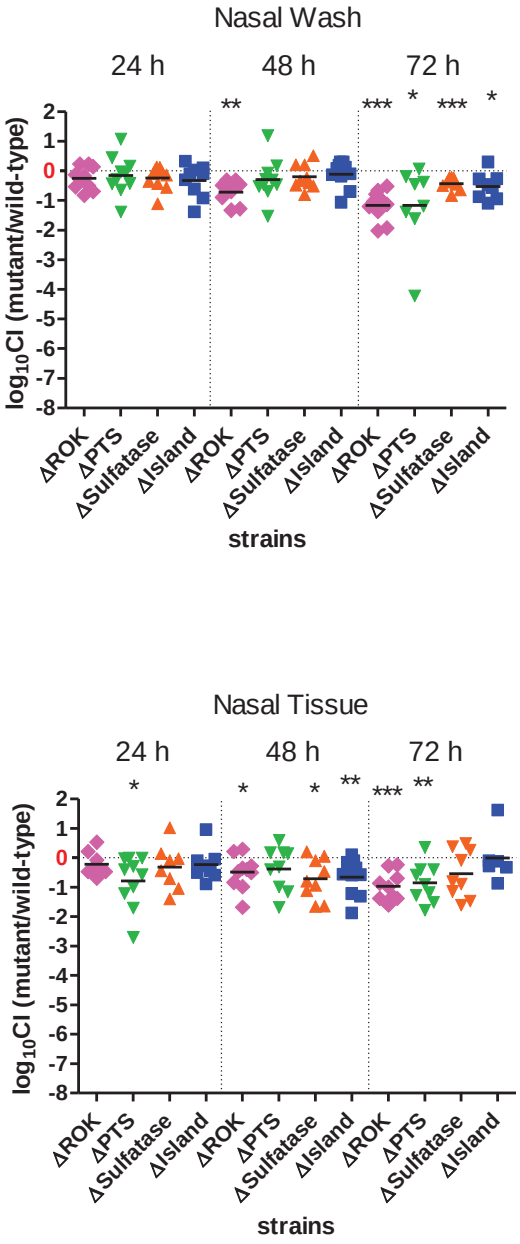


Figure 6.14. The log_{10} CI values in the nasopharynx over a 72 h period.

CD-1 mice were inoculated via the i.n. route without anaesthesia with a mixture of wild-type (WCH206) and mutant. Input ratios and challenge doses are shown in Table 6.6. Data have been pooled from duplicate experiments. The log_{10} CI values in the nasal washes (NW) and nasal tissue (NT) were calculated 24 h, 48 h and 72 h post-challenge. The neutral value 0 is indicated by a dotted, horizontal line and the mean is indicated by a solid, horizontal bar. One sample t-tests were used to compare the mean log_{10}CI of each group to the null hypothetical value of 0 (*, P<0.05; **, P<0.01; ***, P <0.001).

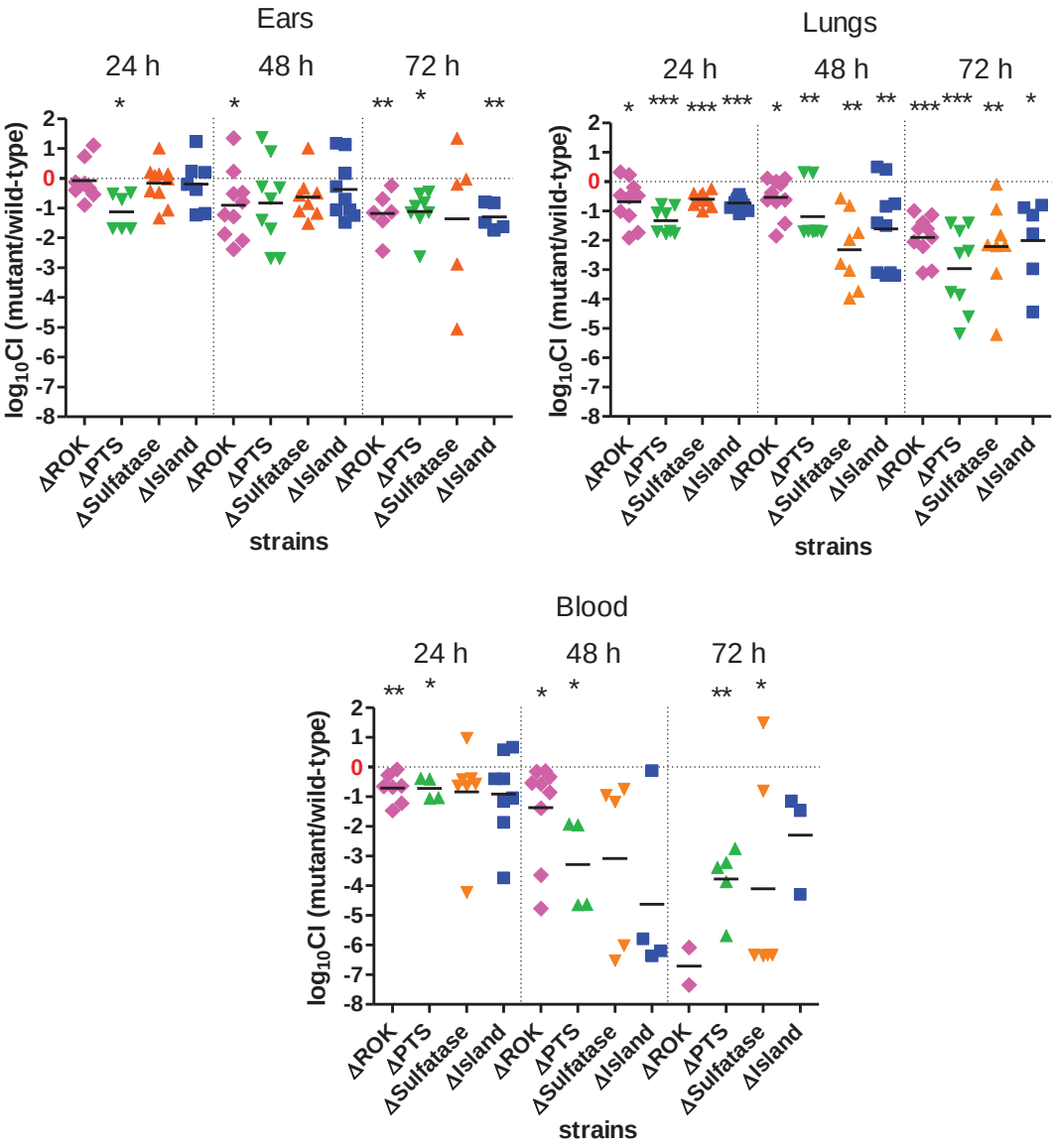


Figure 6.15. The log_{10} CI values in the ears, lungs and blood over a 72 h period.

CD-1 mice were inoculated via the i.n. route without anaesthesia with a mixture of wild-type (WCH206) and mutant. Input ratios and challenge doses are shown in Table 6.6. Data have been pooled from two duplicate experiments. The log_{10} CI values in the ears, lungs and blood were calculated 24 h, 48 h and 72 h post-challenge. The neutral value 0 is indicated by a dotted, horizontal bar and the mean is indicated by a solid, horizontal bar. One sample t-tests were used to compare the mean log_{10}CI of each group to the null hypothetical value of 0 (*, P <0.05; **, P <0.01; ***, P <0.001).

Table 6.4. Statistical analysis of the pneumonia/sepsis competition study.

Mutant	Time	Nasal Wash	Nasal Tissue	Ears	Lungs	Blood
ΔROK	24 h	NS	NS	NS	0.0184	0.0090
	48 h	0.0017	0.0238	0.0329	0.0298	0.0398
	72 h	< 0.0001	0.0004	0.0118	< 0.0001	NS
ΔPTS	24 h	NS	0.0202	0.0065	< 0.0001	0.0326
	48 h	NS	NS	NS	0.0084	0.0245
	72 h	0.0476	0.0050	0.0023	0.0002	0.0018
ΔSulfatase	24 h	NS	NS	NS	< 0.0001	NS
	48 h	NS	0.0156	NS	0.0013	NS
	72 h	0.0009	NS	NS	0.0044	0.0357
ΔIsland	24 h	NS	NS	NS	< 0.0001	NS
	48 h	NS	0.0085	NS	0.0069	NS
	72 h	0.0141	NS	0.0031	0.0190	NS

One sample *t*-tests were used to assess the in CI in various niches at 24 h, 48 h and 72 h post-challenge. *P* values are listed above with NS indicating a non significant value ($P > 0.05$).

The level of infection differed in the various niches of the mouse. For example, the lungs and blood had very high infection levels, while fewer bacteria were isolated from the ears. Figure 6.16 illustrates the median CFU in the various niches for the three timepoints.

6.3.4 RNA analysis of WCH206 ΔROK

In Section 6.3.3, WCH206 ΔIsland was no more attenuated than mutants of the individual components in any niche of the pneumonia/sepsis model. If the various gene functions of the island were interdependent, the attenuation of WCH206 ΔIsland would be predicted to be similar to the attenuation of the single mutants. Therefore, the results suggest the genes may contribute equally to the same virulence function. Nonetheless, since CIs of the four mutants were compared between different groups of mice, host-to-host variation may also have hidden subtle differences (Beuzon *et al.*, 2000; Lau *et al.*, 2001).

One possible example of gene function dependence in the island would be if the ROK family protein was involved in the regulation of the PTS 1 operon. For example, if it was required for the expression of the PTS 1 operon, deleting it would be expected to have the same impact as deleting the entire PTS 1 operon, assuming it did not regulate any other genes. Therefore, the impact of deletion of the ROK family protein on the expression of the PTS 1 operon and sulfatase gene was investigated by real-time RT-PCR (Section 2.10.3) using the primer pairs mRNA spd0016 F and R, mRNA sph1929 PTSIIC F and R, and mRNAsph1924 sulfatase F and R (Table 2.2).

Results are displayed in Figure 6.17. Data were analysed using the $2^{-\Delta\Delta C_T}$ method (Section 2.10.3). For statistically analysis, the $2^{-\Delta C_T}$ data were combined for duplicate

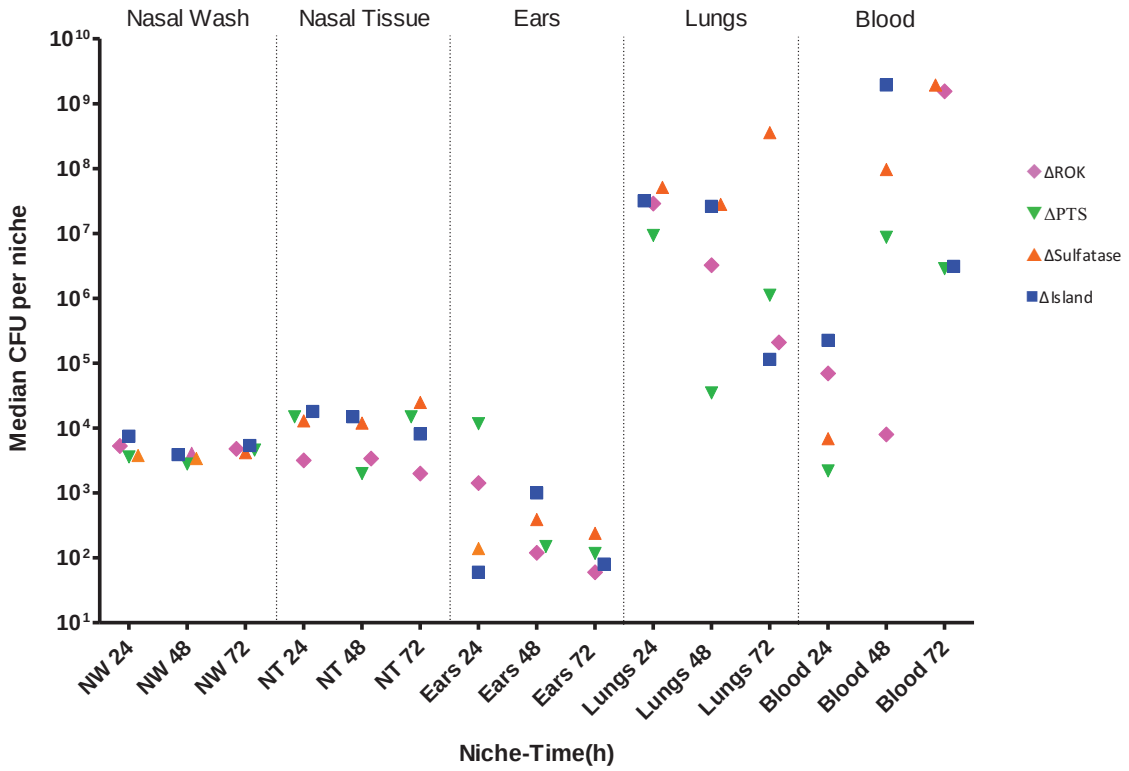


Figure 6.16. Median total CFU (wild-type and mutant) in the niches investigated in the pneumonia/sepsis competition study at each timepoint for mice with infections above the limit of detection. For blood, the number of bacteria is given as CFU/ml of blood. Mice with an infection below the limit of detection (40) were not included because a $\log_{10}$ CI could not be calculated for these animals. NW represents nasal wash and NT represents nasal tissue.

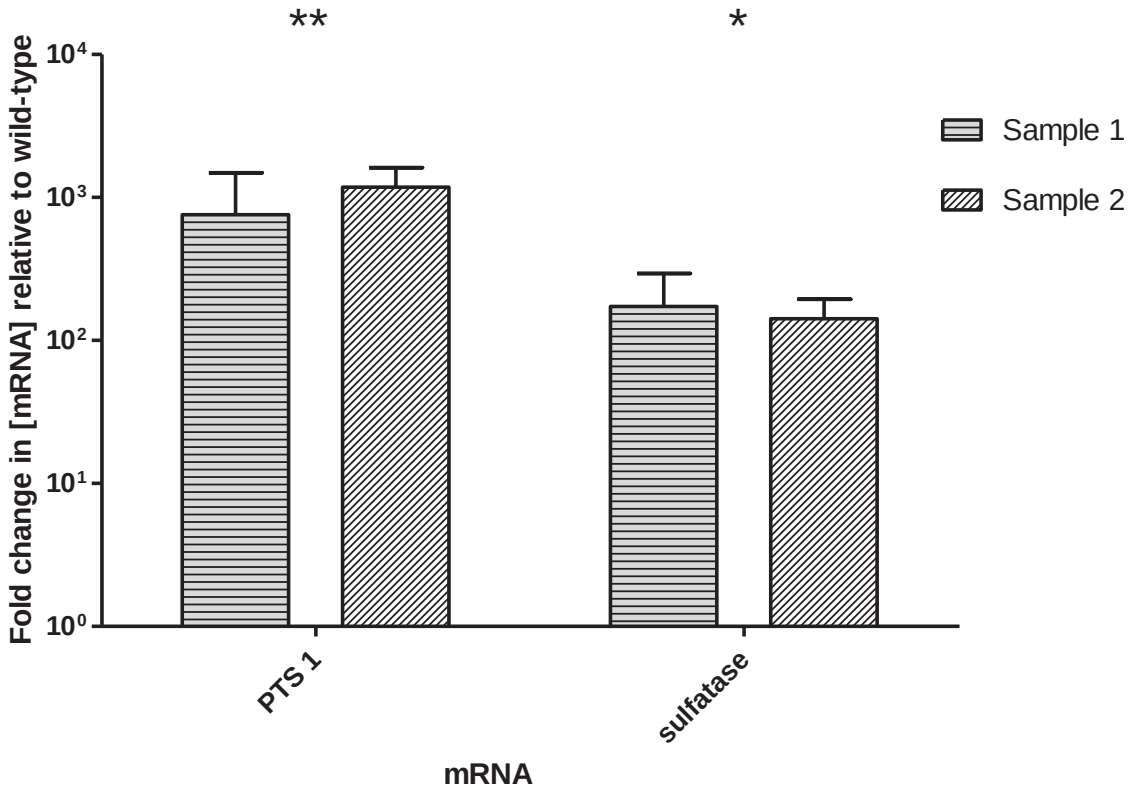


Figure 6.17. Expression of the PTS 1 operon and sulfatase gene *in vitro* by real-time RT-PCR in WCH206 ΔROK.

RNA preparations were obtained from duplicate cultures of WCH206 and WCH206 ΔROK grown in BHI to mid-exponential phase. Following RNA purification, mRNA was assayed by real-time RT-PCR as described in Section 2.10.3. Data represent the fold difference between [mRNA] present in WCH206 ΔROK relative to WCH206 and was calculated using the $2^{-\Delta\Delta C}{}_{T}$ method with 16S rRNA used as an internal control. Error bars represent the 95% confidence interval. Prior to this, the fold difference in the mRNA concentration of the PTS 1 operon and sulfatase gene in relation to 16s rRNA was calculated ($2^{-\Delta CT}$), and compared between WCH206 and WCH206 ΔROK using an unpaired *t*-test (*, P <0.05; **, P <0.01).

cultures, and the wild-type and mutant were compared using unpaired *t*-tests. WCH206 ΔROK demonstrated a statistically significant increase in the expression of both the PTS 1 operon and sulfatase gene (P = 0.0016, 0.0142, respectively). This indicates the ROK family protein of the PTS 1 island is involved in the repression of both the PTS 1 operon and sulfatase gene. However, it does not have the helix-turn-helix DNA binding motif which characterises the repressors of this family (Section 5.2.3), suggesting an alternative mechanism of repression, such as the phosphorylation of a separate repressor molecule.

6.4 Discussion

PTS 1 is part of a genomic island that contains other potential virulence factors. In order to establish a role for this island in virulence, it was deleted in the ST180 strain WCH206 and replaced with an erythromycin resistance cassette joined to the promoter from a serotype 2 capsule locus. Unfortunately, while numerous attempts were made to obtain a mutant in the ST458 strain MSHR17, which carries the island with a large deletion that includes the sulfatase, this could not be achieved. However, a mutant in the avirulent strain MSHR5, which is a serotype 11A isolate of ST662 from a remote Aboriginal community, was obtained.

Deleting the island did not result in any *in vitro* growth defect in either the WCH206 or MSHR5 backgrounds (Figures 6.3 and 6.4). In the WCH206 background, the mutant was shown to grow as well as the wild-type in both planktonic culture and biofilms using BHI (Figures 6.5 and 6.6). Furthermore, deleting the island did not have a negative effect on the ability to ferment cellobiose (Section 6.2.3). However, there are at least 4 different pneumococcal PTS that are annotated as putative cellobiose transporters (Section

5.2.2), including the two that are discussed in this study (PTS 1 and 2) and there may also be other systems capable of transporting cellobiose, which have not been annotated as such. Therefore, even though the system identified in Chapter 4 was deleted, other transport systems may have been compensating for its loss. Nevertheless, the data suggest glucose and sucrose are preferential carbohydrate sources. For glucose this is understandable, given it is a monosaccharide. Cellobiose has not been proven to be the substrate of these transporters and therefore, it is also possible that cellobiose is not the actual substrate of some or all of them, including PTS 1.

When the mutant and wild-type were compared using the pneumonia/sepsis and i.p. murine models, a statistically significant increase in survival time was observed for the group of mice inoculated with the mutant in the pneumonia/sepsis model, but not the i.p. model. However, when the mutant and wild-type were directly compared in the i.p. model by competitive challenge with a mixture of mutant and wild-type, the mutant was isolated from the blood less frequently than the wild-type after 48 h when compared to the initial input ratio. This difference in virulence was presumably too subtle to detect in the original direct i.p. virulence experiment. Furthermore, in the pneumonia-sepsis experiment, although the median survival time of the mutant group had increased, the mutant was still virulent and all but one mouse died.

In addition to the virulence studies, colonisation in the i.n. murine model was investigated. Both WCH206 and MSHR5 backgrounds were investigated because serogroup 11 strains were isolated from the nasopharynx at higher levels than serotype 3 in the i.n. study in Chapter 3 (Section 3.2.4.2). The WCH206 ΔIsland mutant was less competitive compared to the WCH206 wild type when examining the surface colonisation of the nasopharynx by nasal washes on Day 1. Furthermore, it was less competitive on

Days 1 and 4 in the deeper nasal tissue. Interestingly, for MSHR5 the mutant was less competitive than the wild-type on the surface of the nasopharynx at every timepoint and on Day 7, it was less competitive in the nasal tissue. The results suggested that the island could play a role in both colonisation and virulence and impart a competitive advantage on both invasive and noninvasive strains.

Based on the above results, mutants in the ROK family protein and PTS 1, as well as the sulfatase and its associated modifying factor, were constructed in WCH206 and a 72 h competitive study was conducted using the pneumonia/sepsis model. This suggested that the island is important to virulence in a variety of niches. In particular, in the lungs there was a statistically significant attenuation for all mutants at each timepoint. In the blood, where there are high numbers of bacteria in animals with sepsis, ΔPTS was significantly attenuated at all time points, but ΔROK showed significant attenuation only at 24 and 48 h post-challenge, while ΔSulfatase showed significant attenuation at 72 h. It is possible, indeed likely, that attenuation in the lungs could have influenced the development of sepsis, since reduced numbers of bacteria in the lungs may lead to reduced translocation across the alveolar barrier. However, in the i.p. competition study, which provides a more direct route to sepsis, ΔIsland demonstrated statistically significant attenuation by 48 h post-challenge. This indicates that deleting the island has an additional impact on sepsis over and above lung colonisation and translocation.

In the nasal tissue, ΔROK and ΔPTS showed significant attenuation at 72 h, as well as 48 h and 24 h, respectively. ΔSulfatase and ΔIsland showed significant attenuation only at 48 h. In the nasal wash and ears, WCH206 mutants did not typically show attenuation until 72 h post-challenge with the exception of ΔSulfatase, which did not show statistically significant attenuation in the ears at any timepoint.

The mean and median $\log_{10}$CI values of the ears and nasal washes were also compared to one another for each mutant using unpaired *t*-tests and Mann Whitney *U*-tests, respectively. Only ΔIsland demonstrated a significant difference between the two niches. The comparison was undertaken because it is not known if colonisation of the nasopharynx influences colonisation of the ears in the pneumonia/sepsis model. The mice inhaled the inoculum while under anaesthesia and it is possible that this may have caused a change in pressure that forced some of the inoculum into the ears, thereby inoculating the ears directly and independent of nasopharyngeal colonisation. However, if the bacteria colonised the nasopharynx before colonising the ears, it is possible that the lower levels of mutant isolated from the ears compared to the initial inoculum was simply due to attenuation in the nasopharynx. To directly assess fitness in the ears, a competition study using a surgical ear model where the challenge dose is inoculated directly into the ear cavity could be employed. However, such a model is not readily available. Regardless, whether it is due to decreased fitness in the ears or attenuation in the nasopharynx, the island still appeared to impart a competitive advantage in OM in addition to invasive disease.

The number of bacteria or their growth rate in a particular niche could also influence results. For example, in niches with a low number of bacteria, such as the ears, when no mutants are isolated, it is impossible for the nominal CI assigned to be as low as in niches with a high number of bacteria. Furthermore, it is expected that if any of the mutations caused *in vivo* growth defects, it would be more noticeable in niches where the bacteria grow to a high number.

A final important observation from this study is that in all niches, the $\log_{10}$ CI was not significantly lower for the mutant of the whole island compared to the individual

component mutants, which suggested that the genes may contribute similarly to the same virulence function. WCH206 ΔROK was examined further by real-time RT PCR to see if the ROK family protein had a role in regulating PTS 1, as well as the sulfatase. Both were upregulated in WCH206 ΔROK compared to the wild-type, indicating the involvement of the ROK family protein in repression. The control of the expression of sugar utilisation systems in prokaryotes is important for metabolic efficiency and thus optimal growth (Titgemeyer and Hillen, 2002).

Overall, the island containing PTS 1 is not essential, but confers a significant competitive advantage in both passive colonisation and active invasion.

Chapter 7

Whole Genome Analysis of Unrelated Serotype 3 Strains

7.1 Introduction

In Chapter 3, two unrelated STs (ST180 and ST458) were identified in a collection of serotype 3 middle ear isolates from the WCH and MSHR (Section 3.2.3). The genomes of isolates of these two STs were subsequently analysed by DNA microarray and PCR-based subtractive hybridisation (Sections 4.2 and 4.3). PCR-based subtractive hybridisation was employed to identify genes common to isolates of both ST180 and ST458, which were not represented on the laboratory's microarray slides (that is not present in the genomes of R6 or TIGR4) and could play a potential role in OM (Section 4.3 This approach identified a putative cellobiose PTS, but since PCR-based subtractive hybridisation has limitations in the ability to identify unique sequences, it was not known if there were additional genes that had been overlooked. Following the work of Chapter 4, two developments occurred that allowed further analysis of the genomes: a) an unfinished whole genome sequence of a serotype 3 strain of ST180 (OXC141) was released by the Sanger Institute (http://www.sanger.ac.uk/Projects/S_pneumoniae/INV104B_INV200_OXC141/), and b) the opportunity to determine whole genomes using next generation sequencing technology became available to the laboratory.

Two strains from this study (WCH206 and WCH207) are ST180 isolates of serotype 3 like OXC141 (Table 3.3). As mentioned in Section 3.3, ST180 has been found to be the dominant ST within serotype 3 in numerous studies conducted worldwide

(McEllistrem *et al.*, 2005; Brueggemann *et al.*, 2003; Isozumi *et al.*, 2008; Bozdogan *et al.*, 2004; Vestrheim *et al.*, 2008; Porat *et al.*, 2008). Furthermore, this ST has been isolated from children with AOM more frequently in the Unites States of America following the introduction of Prevenar™ (McEllistrem *et al.*, 2005). In addition to OXC141, another ST180 isolate of serotype 3 (SP3-BS71) has been sequenced and is one of the genomes included in the NCBI database, although only the amino acid sequences of the ORFs are available. Comparison of orthologous clusters of the genomes of OXC141 and SP3-BS71 indicates that they are very similar despite being isolated from two distinct geographic locations (the United Kingdom and the United States of America, respectively) (Hiller *et al.*, 2007). In addition, when two serotype 3 isolates of ST180 and ST1826 (single locus variant of ST180) were compared to one another by DNA microarray analysis using slides containing R6 and TIGR4 ORFs, they exhibited the same accessory gene pattern and only differed in the hybridisation of four genes overall, which were not from known accessory regions (Dagerhamn *et al.*, 2008; Blomberg *et al.*, 2009). The availability of the assembled ST180 genome online enabled the *in silico* identification of sequences absent in the genomes of D39 and TIGR4. In addition, the OXC141 genome was used as a template to analyse the genomes of strains MSHR17 (ST458) and WU2 (ST378) by resequencing.

7.2 *In silico* Analysis of the OXC141 Genome

The genome of the OXC141 was downloaded from the Sanger website (section 7.1). As assembly had not been finished, the genome may have contained errors, misassemblies or *E. coli* and vector contamination. The genome had been assembled into two contigs: spn_oxc-13go8.qlk (2035005 bp) and J28181-2c02.plk (1847 bp), which were briefly analysed using BLAST-N on the NCBI website (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The analysis revealed a number of genes that are present in the ST180 genome, but not in

either the D39 or TIGR4 genome sequences, in addition to the putative cellobiose PTS identified by PCR-based subtractive hybridisation (Section 4.3). This included an integrated prophage of approximately 34 kb in size that has been previously described in OXC141 and has 98.6% identity to the equivalent phage of strain SP3-BS71 (Romero *et al.*, 2009).

Another accessory region which could be related to published work was found, and is illustrated in Figure 7.1. This region is approximately 6.4 kb in size and encodes mostly hypothetical proteins, which have homologues in either *S. pneumoniae* 70585 or TIGR4, although not D39. However, NCBI BLAST-X analysis revealed that one of the encoded proteins (FtsK/SpoEIII family protein, Sp70585_1842 in *S. pneumoniae* 70585) has 30% identity to putative otitis media-associated H10 in Taiwan19F-14 (Spt0347, also annotated as hypothetical protein Sph2228 in Hungary 19A-6). OM-associated H10 was first identified in the PCR-based subtractive hybridisation study of Pettigrew and Fennie (2005), which used a serogroup 19 middle ear isolate as the tester strain and R6 as the driver strain. When a number of clinical isolates were analysed to determine its distribution, OM-associated H10 was seen to be present more often in middle ear isolates than isolates from other niches, although it was present in serogroup 19 more often than other serogroups (Pettigrew and Fennie, 2005). Later in the OM signature tagged mutagenesis study of Chen *et al.* (2007), a mutant of OM-associated H10 was attenuated in the middle ears of a chinchilla model.

Genes that appeared to be unique to isolates of ST180 were found. These sequences were analysed using the NCBI BLAST-X program due to the inclusion of the ST180 strain SP3-BS71 in the database. A large region of approximately 24 kb was identified, which included genes encoding D-alanine-D-alanine ligase and related ATP-grasp enzyme, enolase, branch-chain amino acid aminotransferase, probable transketolase, serine hydroxymethyl transferase, putative LPS biosynthesis related phosphoenol pyruvate

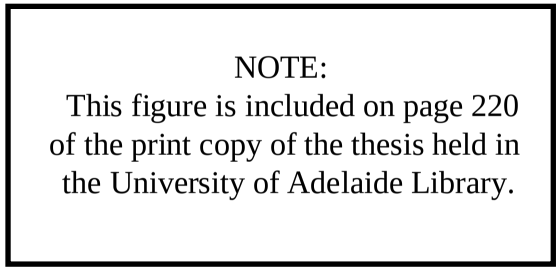


Figure 7.1. An accessory region over 6 kb in the OXC141

The genes of this accessory region have homologues in either *S. pneumoniae* 70585 (green font) or TIGR4 (brown font). The FtsK/SpoEIII family protein has 30% identity to otitis media-associated H10, which was identified in a serogroup 19 middle ear isolate by Pettigrew and Fennie (2005).

phosphomutase, seryl-tRNA synthetase-like protein, ATP-binding protein of an ATPase binding cassette (ABC) transporter, probable 6-phospho-beta-galactosidase, putative recombinase and five hypothetical proteins. This region is discussed in Section 7.3.2.3. In addition, a putative endonuclease gene that appears to be unique to ST180 was also identified in another location of the OXC141 genome.

Finally, two regions that are absent in both D39 and TIGR4 but present in G54 were identified. One of these was over 12 kb and encodes a putative L-fuculose-phosphate aldolase, two blood group antigen cleaving endo-β-galactosidases, α-galactosidase AgaN, F5/8 type C domain containing protein, α-L-fucosidase, an ABC transporter and a putative L-fuculose kinase (spg2099 to spg2108 in G54). These genes are also present in the strains 70585 and CGSP14, and are discussed in Section 7.3.2.7. The second region is over 2.5 kb and encodes a MerR-family transcriptional regulator (spg1815 in G54) together with a hypothetical protein, transposase gene and pseudogene, and is also carried by *S. pneumoniae* Taiwan19F-14. This region is discussed in Section 7.3.2.6.

7.3 Second Generation Sequencing

Resequencing exploits completed genomes by sequencing the genome of a strain of interest in fragments and then using the completed genome as a template to assemble the genome of the resequenced strain. Resequencing has been facilitated by the rapid growth in the number of fully sequenced genomes available, as well as the development of next generation (second generation, massively parallel) sequencing. Conventional genomic sequencing using the Sanger method (Sanger *et al.*, 1977) involves constructing genomic libraries in *E. coli*, and only limited sequences can be processed at a time. However, with next generation sequencing, libraries are constructed *in vitro* and millions of sequences can

be processed in parallel (Mardis, 2008).

Second generation sequencing was performed as described in Section 2.5.10. Genomic DNA of WU2 and MSHR17 was sequenced in fragments of about 36 bp by Geneworks using the Illumina Genome Analyzer *II*, which utilises sequencing-by-synthesis technology (Bentley *et al.*, 2008) and is illustrated in Figure 7.2. These fragments were then assembled against the OXC141 genome sequence and analysed using DNAstar Lasergene version 8 SeqMan Pro (Section 2.5.10). An updated sequence of the OXC141 genome was downloaded from the Sanger website for this purpose. This sequence consisted of two contigs: contig000329 (1847 bp) and contig000334 (2036.9 kb). The former contig consisted of sequence from a tyrosyl-tRNA synthetase (spd1926, spn2100) and penicillin binding protein 1B (spd1925, spn2099). Therefore, the latter contig was used for the alignments.

The average coverage depth was 14.68 times and 11.45 times for MSHR17 and WU2 respectively. The entire assembled sequence of over 2000 kb was scanned visually in house for each strain to identify gaps and regions of poor alignment, which could indicate deletions, insertions or sequence variations. Poor alignments include low depth of sequence coverage or sequences which do not appear to fully align. Examples of good and bad coverage are illustrated in Figures 7.3 and 7.4. For any regions containing poor alignments, files of a minimum of 150 bp (based on the reference sequence) were created with 671 and 908 files obtained for MSHR17 and WU2 respectively. All 1579 files were then analysed using NCBI BLAST-N. When no alignment was obtainable using BLAST-N, NCBI BLAST-X was used. The programs of NCBI were chosen over the programs of KEGG for several reasons. First, the way in which NCBI BLAST-N analyses the whole sequence and presents the data proved quicker to interpret than KEGG when investigating a massive volume of data. Second, a large amount of data was for intergenic sequences, which are not recognised by the KEGG BLAST-N program. Finally, NCBI BLAST-X

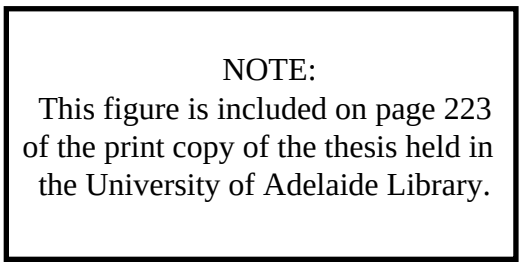


Figure 7.2. Next generation sequencing using the Illumina Genome Analyzer *II*.

DNA is randomly fragmented and adaptors are ligated to the 3' and 5' ends. The fragments are then attached to the surface of the reaction chamber (flow cell) and amplified by solid-phase bridge amplification. Each fragment forms a cluster of about a thousand copies and there are hundreds of millions of clusters in the reaction chamber. The templates are sequenced using four-colour sequencing-by-synthesis in which repeated cycles of single base extension are performed using fluorescently labelled reversible terminators to determine template sequences one base at a time. After polymerisation, the base that was added is determined for each cluster by capturing the emitted fluorescence following laser excitation. The fluorescent dye is removed and the 3' hydroxyl group exposed to allow the addition of the subsequent base.

This description has been adapted from Bentley *et al.* (2008) and the Illumina website (http://www.illumina.com). Pictures are from the Genome Resource Facility of the London School of Hygiene and Tropical Medicine (http://www.lshtm.ac.uk/itd/grf/sequencing.htm).

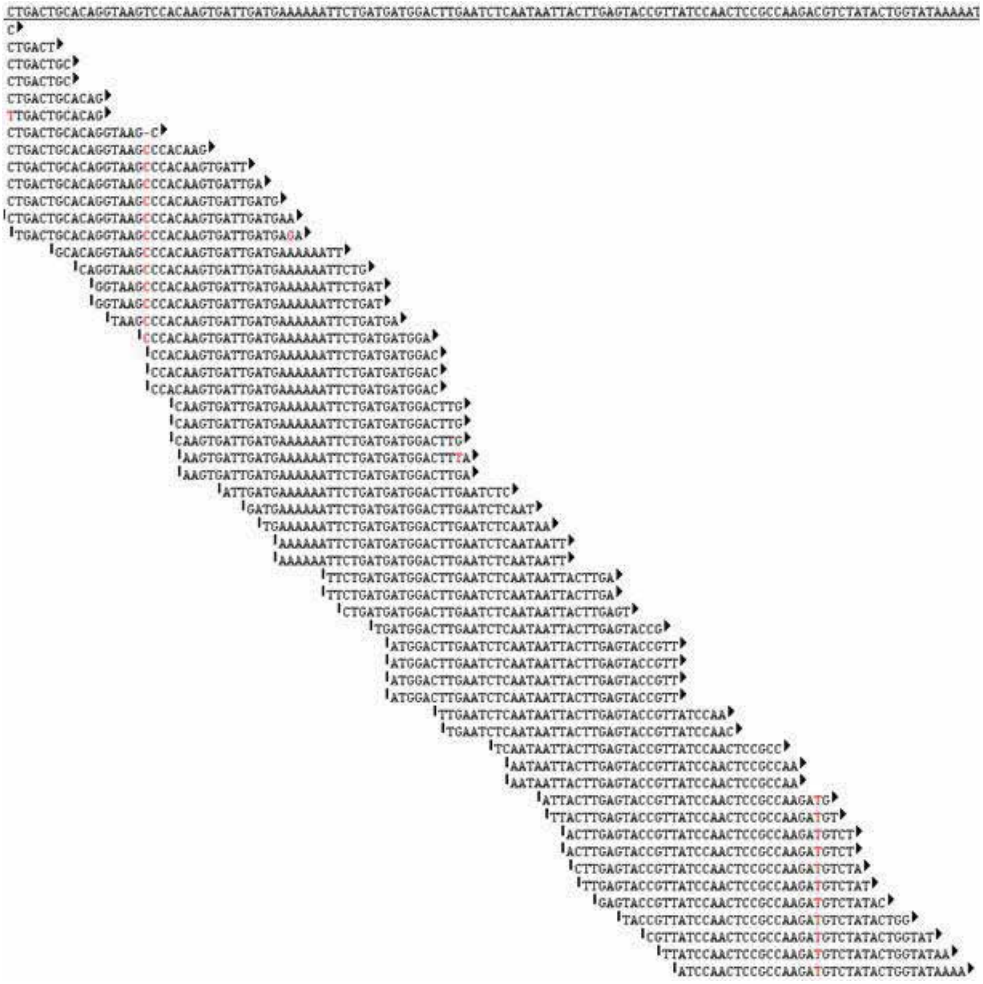


Figure 7.3. An example of good coverage.

This example has been taken from the alignment of WU2 with OXC141 in DNAstar Lasergene version 8 SeqMan. The sequence at the top is the reference genome OXC141. An Illumina Genome Analyser *II* was used to sequence the WU2 genome into 36 bp fragments, which were then aligned against the OXC141 genome. In this section of the alignment, the coverage is deep with no sequencing gaps or fragments that do not appear to truly align.

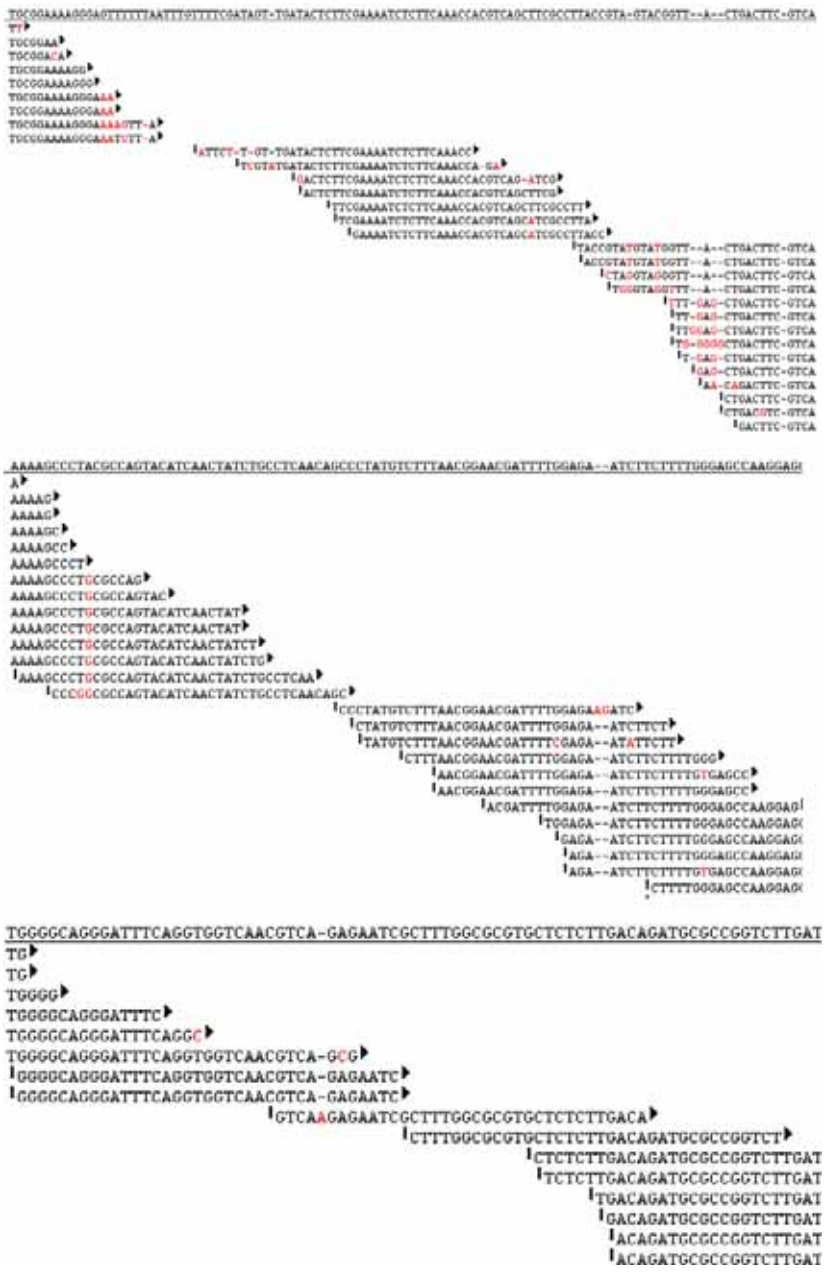


Figure 7.4. Examples of Poor Alignments.

This example has been taken from the alignment of WU2 with OXC141 in DNAstar Lasergene version 8 SeqMan like Figure 7.3. The sequence at the top of each example is the reference genome OXC141. An Illumina Genome Analyser *II* was used to sequence the WU2 genome into 36 bp fragments, which were then aligned against the OXC141 genome.

Poor alignments can take several forms and include sequences that overlap by only a few lines of sequence (bottom) or even single bp (middle), as well as sequences that do not appear to truly align with the OXC141 reference sequence (all). The genomic fragments aligned in the middle of the top example may be masking a larger deletion.

currently has more pneumococcal genomes available, including the amino acid sequences of the ORFs of SP3-BS71, than KEGG BLAST-X. After analysis, the sequences were then classified into two different groups for each strain: a) gaps under 100 bp or poor alignments, b) gaps over 100 bp. Any gene containing both a gap over 100 bp and another gap under 100 bp or poor overlap were classified into the latter group.

Genes affected by gaps under 100 bp or poor alignments are presented in Tables A1 and A2 of the Appendix for MSHR17 and WU2 respectively. Due to large amount of data, intergenic sequences, BOX elements, sequences that appeared more than once in the genome, genes labelled as hypothetical proteins in multiple strains (D39, R6, TIGR4, CGSP14, G54 and Hungary19A-6) and lone transporter components were not included. Moreover, sequences repeated in the genome were excluded because this is the most likely explanation for the gap or poor alignment rather than a genuine genetic difference. Sequences in the list (almost 500 in total) were then aligned using KEGG BLAST-N, which only aligns to coding sequences. This was done to confirm that the poor alignments, overlaps or gaps were within a recognised gene and to obtain the relevant identification for the genes in D39 and TIGR4. In this analysis, no genes where found which were absent in both D39 and TIGR4 and there were only ten incidences where a sequence was absent in one of these strains. Genes were grouped based on their assigned pathway category colours in KEGG (Table 7.1), for example, transporter, carbohydrate metabolism, etc.

Lists of gaps bigger than 100 bp are compiled in Tables 7.2 and 7.3 for MSHR17 and WU2 respectively and are listed according to the order in which they appeared in the Intergenic sequences and sequences repeated in the genome were not included. Genes were assigned identification using KEGG BLAST-N. For genes that were unique to ST180, the sequences were analysed using NCBI BLAST-X due to the inclusion of the SP3-BS71 genome in the database.

Table 7.1. Colour Code for KEGG Pathway Categories.

NOTE:
This table is included on page 227 of the print copy of the thesis held in the University of Adelaide Library.

This table was downloaded from http://www.genome.jp/kegg/docs/ko_color.html. The colour code was used for the figures of this chapter, as well as to group genes in the tables of the Appendix (Tables A1 and A2).

Table 7.2. Gaps of 100 bp or more in the MSHR17 consensus.

Gaps are in order of the genes appearance in the OXC141 reference genome. Genes which appear to be completely missing in both MSHR17 and WU2 are highlighted in blue. Gaps which have been confirmed by PCR in Section 7.3.1 are numbered in red. Where a gene is absent in both D39 and TIGR4, the ID of another strain is used where possible with strains represented by the following abbreviations as used by KEGG: spp- P1031, spv- Hungary19A-6, spw- CGSP14, snm- 70585, snt- Taiwan19F-14, spx-G54.

Δ size (bp)	Gene	D39 ID	TIGR4 ID	OTHER ID
34069	Integrase	-	-	spp:spp0028
	DNA polymerase III subunit epsilon	-	-	-
	Putative phage terminase, large subunit	-	-	spp:spp0060
	Phage portal protein, HK97 family	-	-	spp:spp0062
	Phage major capsid protein, HK97 family	-	-	spp:spp0064
	Prophage LambdaSa04, tail tape measure family	-	-	spp:spp0072
	PblB	-	-	spp:spp0075
	Lytic amidase	-	-	spp:spp0083
112	Hypothetical protein	spd0035	spn0029	
116	Phosphoribosylaminoimidazole-succinocarboxamide synthase	spd0051	spn0044	
114	Cell wall surface anchor	spd0080	spn0082	
1317	Hypothetical protein	spd0105	-	
	Hypothetical protein	-	-	spv:sph0209
	Hypothetical protein	-	-	spv:sph0210
190	ABC transporter (membrane protein)	spd0107	spn0110	
1986	Hypothetical protein	spd0116	spn0115	
	Hypothetical protein	spd0112	-	
4727	Hypothetical protein	spd0119	-	
	Conserved hypothetical protein	spd0120	-	
1530	PspA	spd0126	spn0117	

309	MutR	spd0144	spn0141	
158	Transporter, major facilitator family	spd0148	spn0145	
327	Branched chain amino acid permease	-	spn0146	
	Transmembrane protein	-	spn0147	
212	Hypothetical protein (chlorohydrolase)	spd1674	spn0172	
129	Dihydrofolate synthetise, putative	spd0183	spn0197	
183	Acetyltransferase, GNAT family	spd0189	spn0204	
355	Choline binding protein J, CbpJ	spd0357	spn0378	
556	Integrase	-	-	spp:spp0431
528	Hypothetical protein	spd0422	spn0470	
	Hypothetical protein	-	spn0471	
1454	Integrase/recombinase	spd0452	spn0506	
	Putative type I restriction modification system, S subunit, HsdS	spd0453	spn0505	
	Putative type I restriction modification system, S subunit	spd0450	spn0507	
598	Chaperone protein, DnaK	spd0460	spn0517	
401	Histidine kinase, BlpH	spd0469	spn0527	
2211	Immunity protein, BlpX	-	spn0544	
	Putative immunity protein, PncM	-	-	spp:spp0561
	Hypothetical protein	-	spn0543	
998	Acetyltransferase, GNAT family	spd0492	spn0566	
	Hypothetical protein	spd0493	-	
1333	Valyl-tRNA synthetase	spd0494	spn0568	
	Hypothetical protein	spd0495	-	
273/107	Serine protease, PrtA	spd0558	spn0641	
505	β-galactosidase precursor, BgaA	spd0562	spn0648	
4639[1]	Zinc metalloprotease, ZmpB	spd0577	spn0664	
121	Chorisimate binding enzyme	spd0578	spn0665	
287	Sodium-dependent transporter	spd0642	spn0737	
781	Acetyl transferase, GNAT family	spd0697	-	
139	Amino acid transporter, permease	spd0719	spn0823	
631	Hypothetical protein	-	-	-

506	Type I restriction modification	-	-	spv:sph0997
116/411	Competence protein, CelB	spd0844	spn0955	
151	Hypothetical protein	spd0920	spn1039	
21663[2]	Enolase	-	-	-
	Branch-chain amino acid aminotransferase	-	-	-
	Probable transketolase	-	-	-
	Serine hydroxymethyl transferase	-	-	-
	Hypothetical protein	-	-	-
	Hypothetical protein	-	-	-
	Hypothetical protein	-	-	-
	Putative LPS biosynthesis related phosphoenol pyruvate phosphomutase	-	-	-
	Hypothetical protein	-	-	-
	Seryl-tRNA synthetase-like protein	-	-	-
	Hypothetical protein	-	-	-
	ABC transporter, ATP-binding protein	-	-	
	Probable 6-phospho-β-galactosidase	-	-	-
	Putative recombinase	-	-	spw:spcg1217
1549	Immunoglobulin A1 protease	spd1018	spn1154	
1271	PTS , lactose specific, IIBC components	spd1047	spn1185	
522	Chromosome segregation protein	spd1104	spn1247	
1243	Putative endonuclease	-	-	-
117	Hypothetical protein	spd1939	-	
135	MucBP domain protein	spd1321	spn1492	
1296	Oligopeptide ABC transporter, oligopeptide-binding protein AliB	spd1357	spn1527	
132	Hypothetical protein	spd1491	spn1679	
523	Neuraminidase, NanA	spd1504	spn1693	
1478	PTS IIABC components, sucrose, Scr	-	-	spw:spcg1694
264	Prevent-host-death family protein	spd1551	spn1741	
215	Hypothetical protein	-	-	spp:spp1778
351	Ribosomal protein methyltransferase,	spd1573	spn1782	

6383[3]	Integrase/recombinase	-	-	snm: sp70585_1838
	Hypothetical protein CGSSp3BS71_03342	-	-	snm: sp70585_1840
	Hypothetical protein CGSSp3BS71_03347	-	-	snm: sp70585_1841
	FtsK/SpoEIII family protein	-	-	snm: sp70585_1842
	Hypothetical protein CGSSp3BS71_03372	-	spn1793	
	Hypothetical protein CGSSp3BS71_03367	-	-	snm: sp70585_1844
	Hypothetical protein CGSSp3BS71_03377		spn1794	
4522[4]	ABC transporter, substrate binding protein	-	spn1796	
	ABC transporter, permease protein	-	spn1797	
	ABC transporter, permease protein	-	spn1798	
2844	Sulfatase modifying factor	-	-	spv:sph1923
	Sulfatase	-	-	spv:sph1924
	PTS , cellobiose-specific IIC component	-	-	spv:sph1925
273	Iron ABC transporter, solute-binding protein, putative	spd1609	spn1826	
4618[5]	Alcohol dehydrogenase, Zn containing	spd1636	spn1855	
	Transcriptional regulator, MerR family	spd1637	spn1856	
	Cation efflux system protein	spd1638	spn1857	
	Transcriptional regulator, TetR family protein	spd1639	spn1858	
	Nicotinamide mononucleotide transporter, PnuC	spd1640	spn1859	
	ProWX, choline transporter (glycine betaine transport system permease)	spd1642	spn1860	
2670	MerR family transcriptional regulator	-	-	snt:spt1850

	Hypothetical protein	-	-	snt:spt1851
	Transposase	-	-	snt:spt1852
	Pseudogene	-	-	snt:spt1854
102	YdfG (short chain dehydrogenase/ reductase family protein)	spd1712	spn1909	
260	Response regulator, LytR/AlgR family	spd1718	spn1915	
470	Hypothetical protein	spd1746	spn1947	
307	ABC-type multidrug transporter,permease	spd1800	spn2002	
	ABC transporter, ATP-binding protein	spd1801	spn2003	
227	ABC transporter, ATP-binding protein	spd1801	spn2003	
	Hypothetical protein	spd1802	spn2004	
154	HAD superfamily hydrolase	spd1875	spn2064	
633	Methyltransferase small domain superfamily (or two hypothetical proteins)	spd1855 /1856	spn2046	
12448[6]	Blood group cleaving endo-β-galactosidase			spx:spg2100
	Blood group cleaving endo-β-galactosidase			spx:spg2101
	α-galactosidase, AgaN			spx:spg2102
	F5/8 type C domain containing protein			spx:spg2103
	α-L-fucosidase			spx:spg2014
	ABC transporter, ATP-binding permease protein			spx:spg2105
	ABC transporter, permease protein			spx:spg2106
	ABC transporter, binding protein			spx:spg2107
	L-fuculose kinase, FucK, putative			spx:spg2108
	L-fuculose -phosphate aldolase, FucA			spx:spg2099
155	Choline binding protein E, CbpE	spd0821	spn0930	

Table 7.3. Gaps of 100 bp or more in the WU2 consensus sequence.

Genes are in order of appearance in the OXC141 reference genome. Genes which appeared to be completely missing in both MSHR17 and WU2 are highlighted in blue. Gaps which have been confirmed by PCR in Section 7.3.1 are numbered in red. Where a gene is absent in both D39 and TIGR4, the ID of another strain is used where possible with strains represented by the following abbreviations as used by KEGG: spp- P1031, spv- Hungary19A-6, snm- 70585, spw- CGSP14, snt- Taiwan19F-14, spx-G54.

Δ size (bp)	Gene	D39 ID	TIGR4 ID	OTHER ID
34082	Integrase	-	-	spp:spp0028
	DNA polymerase III subunit epsilon	-	-	-
	Phage terminase, large subunit, putative	-	-	spp:spp0060
	Phage portal protein, HK97 family	-	-	spp:spp0062
	Phage major capsid protein, HK97 family	-	-	spp:spp0064
	Prophage LambdaSa04, tail tape measure family	-	-	spp:spp0072
	PblB	-	-	spp:spp0075
	Lytic amidase	-	-	spp:spp0083
289	Predicted repair protein	spd0031	spn0026	
119	Membrane protein	spd0040	spn0034	
6748	Hypothetical protein	spd0105	-	
	Hypothetical protein	-	-	spv:sph0209
	Hypothetical protein	-	-	spv:sph02110
	Bacteriocin	spd0106	spn0109	
	ABC transporter membrane-spanning permease - amino acid transport	spd0107	spn0110	
	Macrolide export ATP-binding/permease protein, MacB	spd0108	spn0111	
	ABC transporter solute-binding protein - amino acid transport	spd0109	spn0112	
	Argininosuccinate synthase, ArgG	spd0110	-	

	Argininosuccinate lyase, ArgH	spn0111	-	
6209	Hypothetical protein	spd0112	-	
	Transporter truncation	spd0113	-	
	Hypothetical protein	spd0114	-	
	Hypothetical protein	spd0115	-	
1416	PspA	spd0126	spn0117	
143	Glycoprotein endopeptidase	spd0134	spn0127	
196	Hypothetical protein	-	-	spv:sph0254
166	Acetyltransferase, GNAT family	spd0239	spn0255	
140	Dihydropteroate synthase, FolP (SulA)	spd0269	spn0289	
123	Choline binding protein	spd0345	spn0377	
999	Hypothetical protein	spd0391	spn0429	
	Hypothetical protein	spd0392	spn0430	
	Hypothetical protein	-	spn0431	
8197[1]	HsdM, type I restriction modification system, M subunit	spd0454	spn0509	
	HsdR, type I restriction modification system, R subunit	spd0455	spn0510	
175	Histidine kinase BlpH, putative (TCS13)	spd0469	spn0527	
1030	Acetyltransferase, GNAT family	spd0492	-	
	Hypothetical protein	spd0493	-	
512	Valyl-tRNA synthetase, ValS	spd0494	spn0568	
215	HIT family protein	spd0548	spn0628	
322	Serine protease, subtilase family PrtA	spd0558	spn0641	
158	Hypothetical protein	spd0565	spn0651	
180	Radical SAM enzyme, Cfr family	spd0669	spn0768	
781	Acetyltransferase, GNAT family	spd0697	-	
664	Hypothetical protein CGSSp3BS71_11901	-	-	-
2879	Prophage maintenance system killer protein	-	spn0889	spv:sph0995
	Integrase/recombinase	-	spn0890	
	Putative type I restriction-modification system, S subunit	-	spn0891	

161	Choline binding protein E, CbpE	spd0821	spn0930	
216	Hypothetical protein	-	spn0956	
124	Histidine triad protein B, PhtB	spd1037	spn1174	
24583[2]	D-alanine-D-alanine ligase and related ATP-grasp enzyme			snm: sp70585_1137
	Enolase	-	-	
	Branch-chain amino acid aminotransferase	-	-	
	Probable transketolase	-	-	
	Serine hydroxymethyl transferase	-	-	
	Hypothetical protein	-	-	
	Hypothetical protein	-	-	
	Hypothetical protein	-	-	
	Putative LPS biosynthesis related phosphoenol pyruvate phosphomutase	-	-	
	Hypothetical protein	-	-	
	Seryl-tRNA synthetase-like protein	-	-	
	Hypothetical protein	-	-	
	ABC transporter, ATP-binding protein)	spd0554	spn0636	
	Probable 6-phospho-beta-galactosidase	-	-	
	Recombinase, putative	-	-	spw:spcg1217
164/357/286	Immunoglobulin A1 protease	spd1018	spn1154	
334	Histidine triad protein B, PhtB	spd1037	spn1174	
1031	Potassium/ion channel protein	-	-	snt:spt1006
1243	Endonuclease, putative	-	-	-
136	Heat shock protein, HtpX	spd1138	spn1283	
9093	Hypothetical protein	-	-	spw: spCG1314
	Hypothetical protein	spd1174	-	
	Hypothetical protein	spd1175	spn1340	
	ABC transporter, ATP-binding protein	spd1176	spn1341	
	Drug efflux ABC transporter, ATP-binding/permease protein	spd1177	spn1342	

	Putative lantibiotic synthetase	spd1179	spn1344	
	CAAX amino terminal protease family protein	spd1180	spn1346	
	Hypothetical protein	spd1181	spn1347	
	Hypothetical protein	spd1182	spn1348	
	Hypothetical protein	spd1183	spn1349	
	Hypothetical protein	spd1184	spn1350	
	Hypothetical protein	-	spn1351	
	Hypothetical protein	spd1186	spn1353	
131	MucBP domain protein	spd1321	spn1492	
118	Oxidoreductase, DadA family	spd1433	spn1608	
537	Neuraminidase, NanA	spd1504	spn1693	
172	Ribosomal protein L11 methyltransferase	spd1573	spn1782	
153	Putative peptidase	spd1575	spn1784	
6415[3]	Integrase/recombinase	-	-	snm: sp70585_1838
	Hypothetical protein CGSSp3BS71_03342	-	-	snm: sp70585_1840
	Hypothetical protein CGSSp3BS71_03347	-	-	snm: sp70585_1841
	FtsK/SpoEIII family protein	-	-	snm: sp70585_1842
	Hypothetical protein CGSSp3BS71_03372	-	spn1793	
	Hypothetical protein CGSSp3BS71_03367	-	-	snm: sp70585_1844
	Hypothetical protein CGSSp3BS71_03377	-	spn1794	
9911	Hypothetical protein	-	-	spv:sph1921
	Hypothetical protein	-	-	spv:sph1922
	Sulfatase modifying factor	-	-	spv:sph1923
	Sulfatase	-	-	spv:sph1924
	PTS, cellobiose-specific IIC component	-	-	spv:sph1925
	PTS, cellobiose-specific IIA component	-	-	spv:sph1926
	PTS, cellobiose-specific IIB component	-	-	spv:sph1927

	Hypothetical protein	-	-	spv:sph1928
	PTS, cellobiose-specific IIC component	-	-	spv:sph1929
	ROK family protein	-	-	spv:sph1930
2673	MerR family transcriptional regulator	-	-	spt:snt1850
	Hypothetical protein	-	-	spt:snt1851
	Transposase	-	-	spt:snt1852
	Pseudogene	-	-	spt:snt1854
624	Methyltransferase small domain superfamily (or 2 hypothetical proteins)	spd1855 /1856	spn2046	
12446[4]	Blood group cleaving endo-β-galactosidase	-	-	spx:spg2100
	Blood group cleaving endo-β-galactosidase	-	-	spx:spg2101
	α-galactosidase, AgaN	-	-	spx:spg2102
	F5/8 type C domain containing protein	-	-	spx:spg2103
	α-L-fucosidase	-	-	spx:spg2014
	ABC transporter, ATP-binding/permease protein	-	-	spx:spg2105
	ABC transporter, permease protein	-	-	spx:spg2106
	ABC transporter, binding protein	-	-	spx:spg2107
	L-fuculose kinase, FucK, putative	-	-	spx:spg2108
	L-fuculose -phosphate aldolase, FucA	-	-	spx:spg2099
882	Regulator of fucose operon	spd1996	spn2168	
5136[5]	Major facilitator family transport protein	spd2007	-	
	GBSi1, group II intron, maturase	spd2008	-	
	Hypothetical protein CGSSp3BS71_10523	spd2009	-	
242	PspC	spd2017	spn2190	

7.3.1 PCR and sequence analysis of selected gaps

Eight sequencing gaps larger than 4.5 kb were selected to be analysed by PCR and sequencing as described in Sections 2.5.5 and 2.5.9. Three of these gaps are common to both MSHR17 and WU2. The gaps have been numbered in red font in Tables 7.2 and 7.3, and contain genes of a range of functions, including transporters and hydrolases. Genes unique to ST180, such as the region over 24 kb containing an enolase and numerous other genes (Section 7.2), were of particular interest. Although genes which are present in OXC141 but not MSHR17 will not be investigated further for a role in OM at this stage, the confirmation of their absence in the MSHR17 genome is important. The data from this section also explained why large sequencing gaps occurred.

Chromosomal DNA from the ST180 middle ear isolate WCH206 was used as a positive control in the PCR amplifications. Primers were designed approximately 100-300 bp up and downstream of the gap and are listed in Table 2.2. Therefore, true deletions were expected to produce a PCR product size of 200-500 bp. When PCR products were larger than the expected size, the first 1 kb was sequenced at either end. Gaps are listed in the order they appear in the OXC141 genome. Information from the KEGG website was used to construct the figures of this section, asides from genes that are unique to one of the serotype 3 strains.

7.3.2.1 Gap 1 of WU2

This gap was over 8 kb in size for the genes of HsdM (M subunit of a type I restriction modification system; spd0454, spn0509) and HsdR (R subunit of a type I restriction modification system; spd0455, spn0510). PCR amplification using the primers WU2 file229 up and down revealed that this was a true deletion.

7.3.2.2 Gap 1 of MSHR17

This gap was within the gene for the zinc metalloprotease gene ZmpB (spd0577, spn0664) and was approximately 4639 bp. After PCR amplification a product of over 4.5 kb was obtained using chromosomal DNA of strain MSHR17 with the primers 17 file198 up and down. When sequenced, this region was confirmed to be from *zmpB*, but it did not initially align to the OXC141 genome due to significant sequence variation.

7.3.2.3 Gap 2 of MSHR17 and WU2

This gap was over 21.5 kb and 24 kb in MSHR17 and WU2, respectively, and included genes annotated with the following functions: enolase, branch-chain amino acid aminotransferase, probable transketolase, serine hydroxymethyl transferase, hypothetical protein (aminotransferase class III in other species), hypothetical protein (nucleoside diphosphate kinase in other species), hypothetical protein (siderophore biosynthesis protein in other species), putative LPS biosynthesis related phosphoenol pyruvate phosphomutase, hypothetical protein (phosphopyruvate decarboxylase or thiamine pyrophosphate enzyme), seryl-tRNA synthetase-like protein, hypothetical protein (putative polyketide synthase in another species), ABC transporter, ATP-binding protein, probable 6-phospho-beta-galactosidase and a putative recombinase. In WU2 an additional gene encoding D-alanine-D-alanine ligase B (sp70585_1137 in *S. pneumoniae* 70585) was missing and therefore, this was a larger gap (Figure 7.5). This region appears unique to ST180 isolates and has been previously described in Section 7.2.

PCR amplification using the primer pairs 17 file340 up and down and WU2 file536 up and down for MSHR17 and WU2, respectively, identified product in both strains. The product in MSHR17 was approximately 1 kb and contained a transposase and

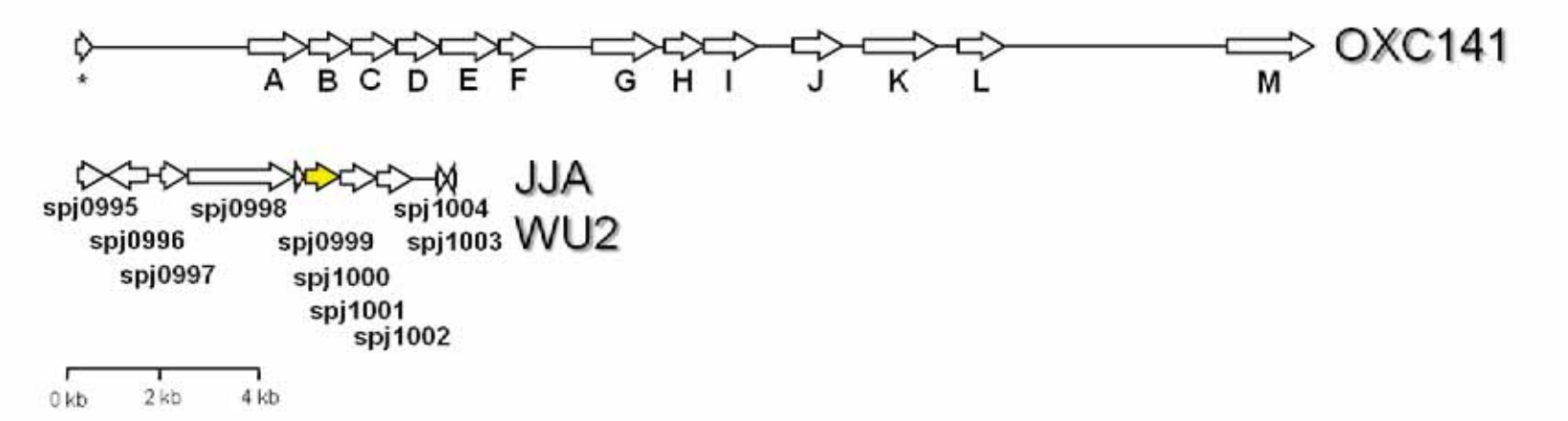


Figure 7.5. The region of PP1 that is unique to oxc141 and the genes that this sequence is believed to be replaced with in WU2, which are found in JJA.

A region over 21 kb in the OXC141 genome did not align with sequence from the strains MSHR17 and WU2. However, in WU2 this region is believed to be replaced by an incomplete lantibiotic system that is also present in the strain JJA and is approximately 7.5 kb.

A, enolase; B, branched-chain amino acid aminotransferase; C, probable transketolase; D= serine hydroxymethyltransketolase; E, hypothetical protein; F, hypothetical protein; G, hypothetical protein; H, putative LPS biosynthesis related phosphoenol pyruvate phosphomutase; I, hypothetical protein; J, hypothetical protein; K, hypothetical protein; L, ABC transporter, ATP-binding protein; M, methionyl-tRNA synthetase (recombinase). * indicates part of D-alanine-D-alanine ligase and related ATP-grasp enzyme, which is absent in WU2 but not MSHR17

The annotations of the incomplete lantibiotic system have been taken from the strain JJA. spj0995, chromosome segregation ATPase; spj0996, transcriptional regulator; spj0997, hypothetical protein; spj0998, lantibiotic mersacidin transporter; spj0999, hypothetical protein; spj1000, bacteriocin transport ATP-binding protein BcrA; spj1001, hypothetical protein; spj1002, hypothetical protein; spj1004,hypothetical protein; spj1003,hypothetical protein. ORFs are coloured according to their KEGG putative functional category as shown in Table 7.1.

aminoglycoside 6'-N-acetyltransferase (sph1149 in Hungary19A-6). An amplicon over 7 kb was obtained for WU2 and sequencing of the first 1 kb revealed a chromosome segregation ATPase (spd0938) and transcriptional regulator (spd0939). Sequencing of the last 1 kb revealed 2 hypothetical proteins (spn1065 and spd0951 respectively.). All four of these genes are present in *S. pneumoniae* JJA, which are annotated as spj0995, spj0996, spj1003 and spj1004, and there is 7.5 kb between the start codon of spj0995 and the stop codon of spj1004. The genes in this region, which are illustrated in Figure 7.5, are annotated as hypothetical proteins or components of an ABC lantibiotic permease and are also found in *S. pneumoniae* ATCC 700669 (Spain 23F-1). In the strain ATCC700669, this region has been described as an incomplete lantibiotic synthesis operon that is part of putative pathogenicity island 1 (PP1), and although it does not contain the gene for the structural lantibiotic prepeptide, the coding sequence for immunity is conserved (Croucher *et al.*, 2009). The size of the PCR product obtained from WU2 and the content detected by sequencing suggested that WU2 also carries the incomplete lantibiotic operon. This was subsequently confirmed by PCR analysis using primers designed for the incomplete lantibiotic operon (x4-y5, Table 2). Therefore this region in both MSHR17 and WU2 is not so much a deletion as a replacement of what is present in OXC141 with other genes.

7.3.2.4 Gap 3 of MSHR17 and WU2

PCR amplification using primers 17 file555 up and down confirmed that the genomes of both MSHR17 and WU2 contain a deletion of the accessory region that is present in OXC141 and encodes a protein with 30% identity to the putative OM-associated H10 gene described in Section 7.2. The sequencing gap was approximately 6.3 kb in size in both strains and contained the following genes in OXC141: integrase/recombinase phage associated protein (SP70585_1838 in the strain 70585), hypothetical protein

CGSSp3BS71_03342 (SP70585_1840), hypothetical protein CGSSp3BS71_03347 (SP70585_1841), FtsK/SpoEIII family protein (SP70585_1842), hypothetical protein CGSSp3BS71_03372 (spn1793 in TIGR4), hypothetical protein CGSSp3BS71_03367 and hypothetical protein CGSSp3BS71_03377 (spn1794). These genes have been previously described in Section 7.2 and are illustrated in Figure 7.1.

7.3.2.5 Gap 4 of MSHR17

This gap was approximately 4.5 kb in size and in OXC141 contains the genes for the substrate binding protein and two permeases of an ABC transporter (spn1796-1798 in TIGR4), as well as a LacI family sugar-binding transcriptional regulator (spn1799) [Figure 7.6], which are part of AR35 (Blomberg *et al.*, 2009). Although BLAST searches of these genes indicated that they are present in many other strains, including SP11-BS70 (a serogroup 11 isolate of ST62 like WCH213), ATCC700669 and Taiwan19F-14, they are not ubiquitous and are absent in some strains, including D39, G54 and Hungary19A-6. PCR amplification of this region using primers 17 file556 up and down revealed not a deletion but a product of over 3.6 kb in MSHR17 of which approximately the first and last kb was sequenced. The first 1 kb was found to encode two ABC transporter permeases (spd1583 and spd1584), although there was a 51 nt deletion within the permease gene homologous to spd1583. Sequencing of approximately the last 1 kb revealed a sugar-binding transcriptional regulator, LacI family protein (spd1586). These sequences are found in strains, such as D39 and G54, which do not possess the above mentioned transporter of OXC141 and TIGR4, such as D39 and G54. Instead they possess an alternative ABC transporter and transcriptional regulator (spd1583-1586), and it is believed that this 3.6 kb region in MSHR17 by analogy is the same as is found in D39 and G54 (Figure 7.6), apart from the 51 nt in-frame deletion. Both ABC transporters are

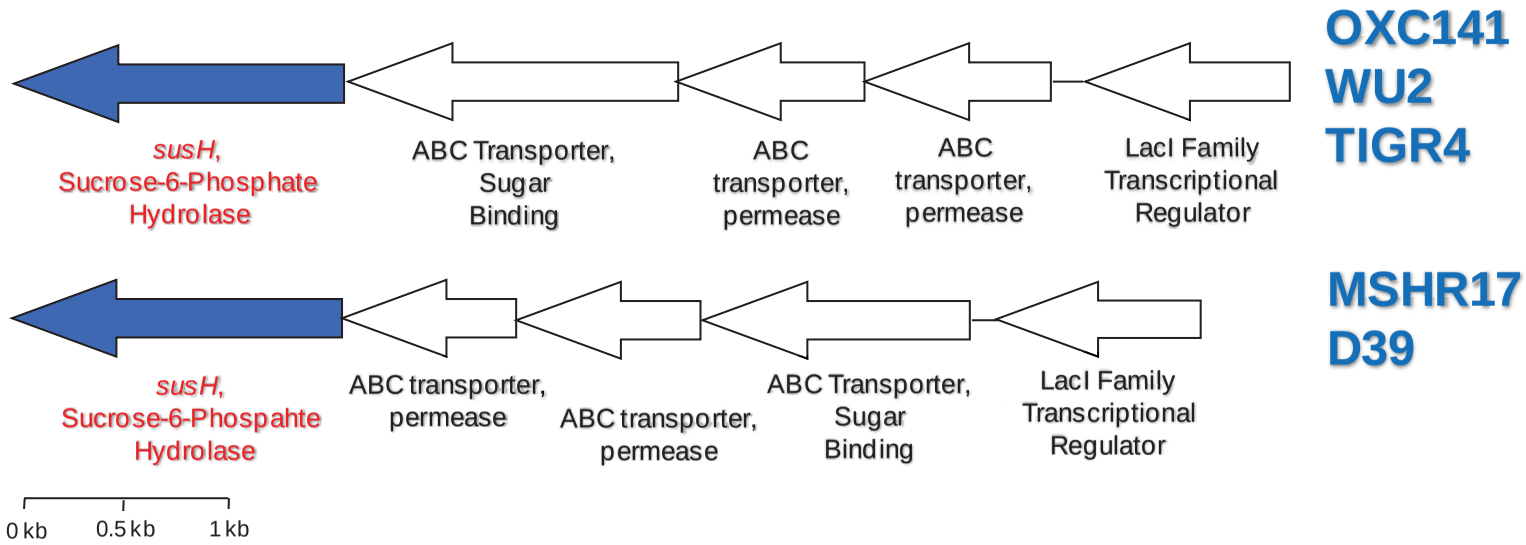


Figure 7. 6. Two alternative sucrose ABC permeases/ transcriptional regulators associated with the sucrose-6-phospahte kinase SusH

The gene *susH* (blue gene) is common to D39 (spd1583), TIGR4 (spn1795), OXC141, WU2 and MSHR17 (indicated by red writing). However, it is associated with two different ABC permeases. The strains OXC141, WU2 and TIGR4 (spn1796-1798) share the system illustrated at the top, but the strains MSHR17 and D39 (spd1583-1585) both possess the second permease. The sequence of the transcriptional regulator associated with these ABC permeases (spn1799, spd1586) also differs. ORFs are coloured according to their KEGG putative functional category as shown in Table 7.1.

adjacent to the sucrose-6-phosphate hydrolase SusH, which appears to be conserved amongst all strains. The ABC transporter of TIGR4 has been demonstrated to be involved in sucrose transportation (Iyer and Camilli, 2007).

7.3.2.6 Gap 5 of MSHR17

This sequencing gap was approximately 4.6 kb in size and included the following genes of OXC141: a zinc containing alcohol dehydrogenase (spd1636, spn1855), a transcriptional regulator of the MerR family (spd1637, spn1856), a cation efflux system protein (spd1638, spn1857), a transcriptional regulator of the TetR family (spd1639, spn1858), nicotinamide mononucleotide transporter PnuC (spd1640, spn1859) and choline transporter ProWX (or glycine betaine transport system permease protein; spd1642, spn1860) [Figure 7.7]. These genes are found in the eleven strains listed in KEGG, although spd1636 is annotated as two ORFs in G54.

A product over 6 kb was obtainable by PCR amplification with the primers 17 file568 up2 and down2 (Table 2.2). Sequencing of approximately the first 1 kb of the PCR product revealed part of the galaxies operon repressor GalR (spd1635, spn1854), as well as a region of approximately 300 bp that was not present in any other pneumococcal genome sequences, but showed up to 60% identity with an aldolase in other species, such as *Enterococcus faecium* and *Lactobacillus rhamnosus*. Sequencing of the last 900 bp of the PCR product revealed a transposase sequence. Due to a lack of additional information on this region, it was not possible to judge what other genes may have been present. Therefore, the remaining sequence of the PCR product was ascertained using the primers file 568 17 internal up-dn4 (Table 2.2). A putative galactose PTS, putative lactose PTS repressor LacR, and putative tagatose-6-phosphate kinase were identified in addition to the

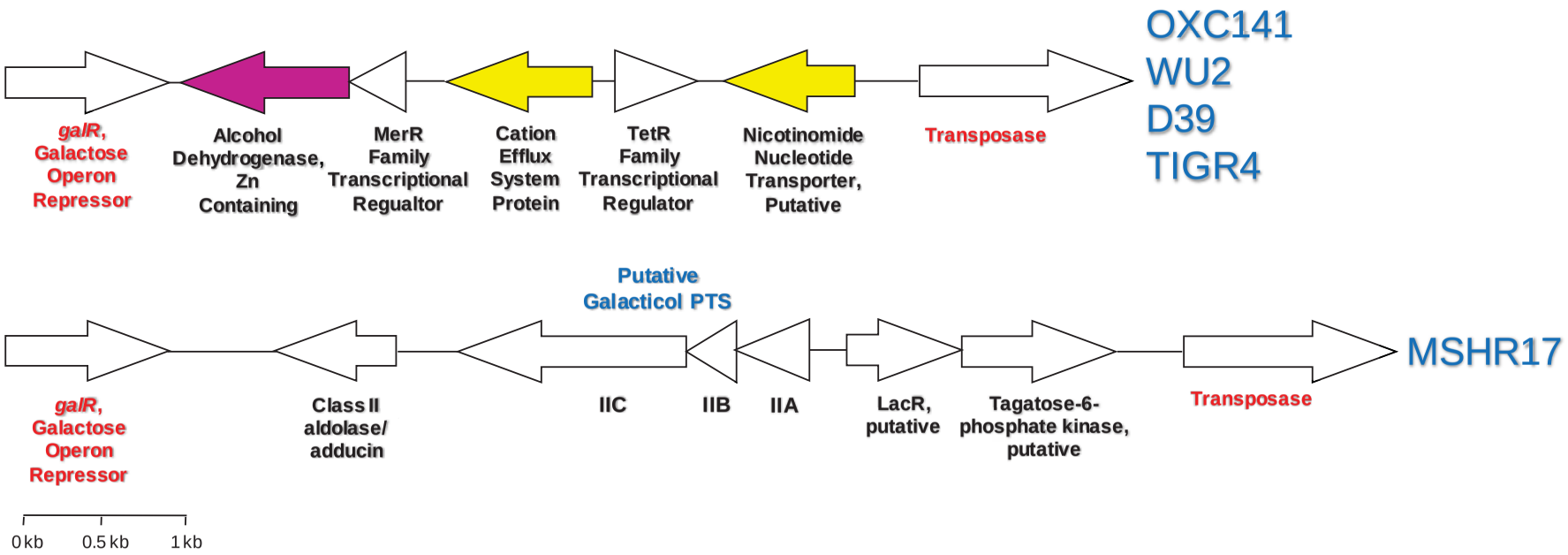


Figure 7.7. The region between the galactose operon repressor GalR and a transposase in oxc141 and MSHR17.

The gene of the galactose operon repressor GalR is common to OXC141, WU2, MSHR17, D39 (spd1635) and TIGR4 (spn1854). In between this gene and a transposase in OXC141, WU2, D39 and TIGR4 is an alcohol dehydrogenase, MerR family transcriptional regulator, cation efflux system protein, TetR family transcriptional regulator and a putative nicotinomade transporter (spd1636-1640, spn1855-1859) [approximately 4.6 kb]. However, in between the same gene and transposase in MSHR17 is a class II aldolase/adducin, components of a putative galacticol PTS, a putative lactose PTS repressor LacR and a putative tagatose-6-phosphate kinase [approximately 6 kb], which are homologous to genes in *Enterococcus casseliflavus* and *E. faecium*. ORFs are coloured according to their KEGG putative functional category as shown in Table 7.

aldolase. The predicted annotation of this region is illustrated in Figure 7.7. All proteins had homologues in *Enterococcus* and *Lactobacillus* spp.

7.3.2.7 Gap 6 of MSHR17/Gap 4 of WU2

This gap was believed to be over 12 kb in both MSHR17 and WU2 and contained the following genes in OXC141: L-fuculose -phosphate aldolase FucA (spg2099 in G54), blood group cleaving endo-α-galactosidase (spg2100), blood group cleaving endo-α-galactosidase (spg2101), α-galactosidase AgaN (spg2102), F5/8 type C domain containing protein (spg2103), α-L-fucosidase (spg2104), ABC transporter, ATP-binding/permease. protein (spg2105), ABC transporter, permease protein (spg2106), ABC transporter, sugar binding protein (spg2107) and putative L-fuculose kinase FucK (spg2108) [Figure 7.8]. Using the primer pair 17 file663 up and down (Table 2.2) amplicons over 8.5 kb were obtained for both MSHR17 and WU2, which appeared identical in size on an 0.8% TBE agarose gel.

Sequencing of the first 1 kb of the MSHR17 and WU2 amplicons revealed a fucolectin related protein (spd1987, spn2159), which is not present in G54. Sequencing of the last 1 kb revealed L-fuculose phosphate aldolase, FucA (spd1994, spn2166) and L-fuculose kinase, FucK (rhamnulokinase, spd1995, spn2167), which are also absent in G54. Based on this result and comparison of the genome maps of D39 and TIGR4 through KEGG, this region is also believed to harbour genes for fucose operon protein FucU (spd1993, spn2165), IIA component of a PTS (spd1992, spn2164), IIB component of a PTS (spd1991, spn2163), IIC component of a PTS (spd1990, spn2162), IID component of a PTS (spd1989, spn 2161) and a large secreted protein (spd1988, spn2160) [Figure 7.8], which are part of AR41 (Blomberg *et al.*, 2009). These genes are absent in 70585, CGSP14 and G54, which carry homologous genes to OXC141. Therefore, there appears to

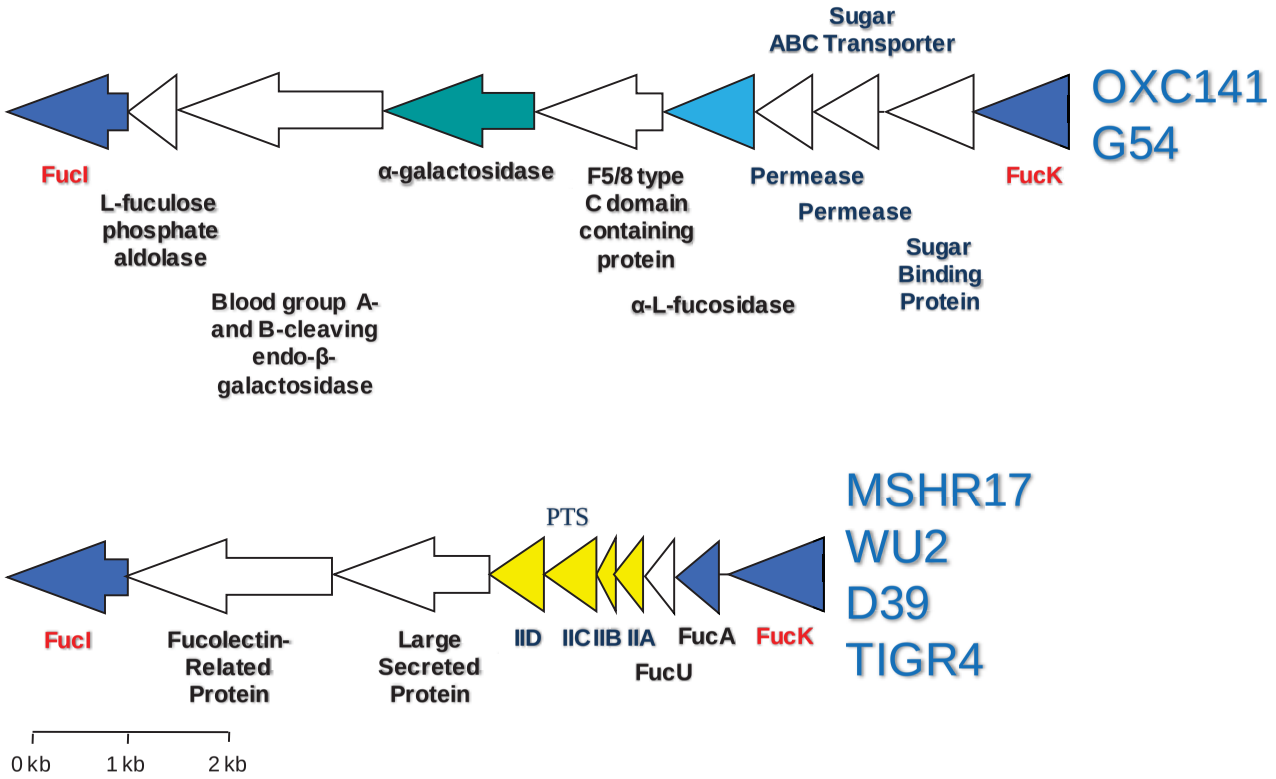


Figure 7.8. Alternative putative pneumococcal fucose utilisation systems.

The genes *fucI* (fucose operon protein) and *fucK* (L-fuculose kinase), which are written in red, are common to all strains. However, the sequence in between this region varies in OXC141 and G54 compared to D39 and TIGR4, although both regions appear to encode proteins for sugar utilisation. Of particular note are the sugar transporters, which are indicated in blue font. OXC141 and G54 carry an ABC transporter but the other strains carry a PTS. MSHR17 and WU2 are believed to carry the same genes as D39 and TIGR4. ORFs are coloured according to their KEGG putative functional category as shown in Table 7.1.

be two different variants of a sugar uptake system and associated enzymes with identical locations within the pneumococcal genome. This is similar to the fourth gap analysed in MSHR17 that revealed an alternative sucrose uptake system to OXC141 (Section 7.3.2.5).

7.3.2.8 Gap 5 of WU2

This gap was approximately 5.1 kb in size and in OXC141 contains the following genes: major facilitator family transport protein (spd2007), GBSi1, group II intron, maturase (spd2008) and hypothetical protein CGSSp3BS71_10523 (spd2009) [Figure 7.9]. PCR analysis using the primers WU2 file888 up2 and down 2 indicated that this was a true deletion in WU2.

7.3.3 Assembly and analysis of the "boneyards" sequences

The residual sequences that were unable to be matched to the reference sequence, which in this study was OXC141, formed the "boneyards" of MSHR17 and WU2. These boneyard sequences can be assembled and used to identify regions absent in OXC141 but present in the sequenced stains. Assembly of the "boneyard" sequences into contigs was performed by Geneworks due to the large number of files. The 50 largest "boneyard" contigs of both MSHR17 and WU2 (with median sizes of approximately 300 bp and 250 bp, respectively) were then analysed using the BLAST programs of KEGG and NCBI with the results presented in Tables 7.4 and 7.5, respectively. Although this is by no means an exhaustive list, this analysis gives a good indication of genes that are not carried by OXC141. Only about half of the sequences were found in both D39 and TIGR4 for MSHR17, while only about a third of sequences had homologues in D39 and TIGR4 for

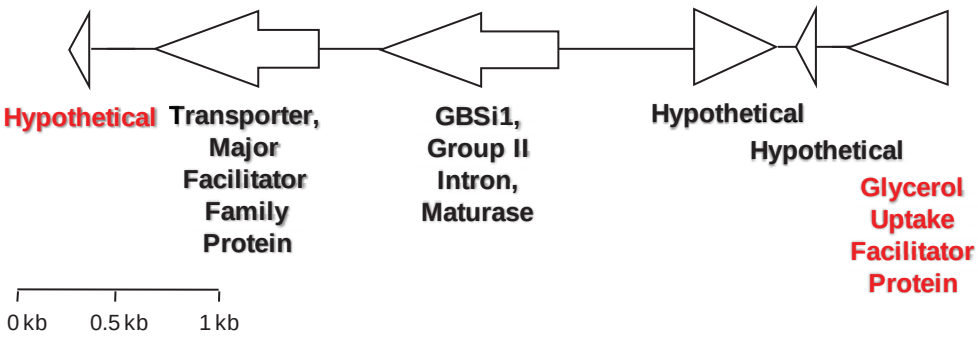


Figure 7.9. A region of the OXC141 genome that is approximately 5.1 kb and is deleted in the strain WU2.

Genes that are common to OXC141 and WU2 are indicated by red font. Genes that are carried by OXC141 but not WU2 are indicated by black font.

Table 7.4. Analysis of the 50 biggest contigs in the "boneyard" of MSHR17.

Contigs were analysed using the BLAST-N program of KEGG to assign function and ID. IDs are given for D39 and TIGR4 where possible but when a gene is absent in both strains, the ID from another pneumococcal genome is used where possible, with the strains abbreviated as follows: spw- CGSP14, snm- 70585, spv- Hungary19A-6, snt- Taiwan19F-14.

Contig Size (bp)	D39 ID	TIGR4 ID	Other ID	Gene
1269	-	-	spw: spcg1038	Transposase IS1239 (repeated in some strains)
520	-	-	snm: sp70585_2243	Sugar/sodium symporter
504	spd1109	spn1251		Endonuclease, putative
492	spd0604	spn0695		HesA/MoeB/Thi family protein
466	-	-	-	69% identity with hypothetical protein SsuiDRAFT_3900 *S. suis* 89/1591
447	-	-	-	(nt 6-96 intergenic)
438	-	-	snm: sp70585_2242	Sugar kinase, ribokinase family
385	spd0126	spn0117		PspA
366	spd0427	spn0477		6-phospho-beta-galactosidase
364	-	-	snm: sp70585_2233	β-N-acetylglucosaminidase
360	spd0604/ 0705	spn0696/ 0697		HesA/MoeB/ThiF family protein and pseudogene
353	spd0139	spn0136		Glycosyl transferase, group 2 family protein
352	spd0426	spn0476		PTS, lactose-specific IIA component, LacF-1
349	spd1987	spn2159		Fucolectin-related protein
347	spd0361	spn0395		Putative transcriptional regulator, MtlR
344	spd0361	spn0395		Putative transcriptional regulator, MtlR
341	spd1968	spn2139		Hypothetical protein

339	-	-	snm: sp70585_2238	Glucuronate isomerase
338	-	-	-	81% identity with hypothetical protein CLONEX_00984 in *Clostridium nexile* DSM 1787
332	spd0126	spn0117		PspA
315	-	spn0569		Hypothetical protein
314	-	-	snm: sp70585_2234	β-N-acetylglucosaminidase
313	-	-	-	57% identity with hypothetical protein SAP039A_043 *Staphylococcus aureus*
308			snm: sp70585_2233	β-N-acetylglucosaminidase/beta-glucosidase
306	spd1583/ 1584			ABC transporter, permease protein x 2
303	-	spn0566/ 0567		Acetyl transferase, GNAT family and Shikimate kinase
303			snm: sp70585_2241	β-glucuronidase (GUS)
303	spd1988	spn2160		Large secreted protein
297	spd1532	spn1722		PTS, IIABC components (sucrose)
294	-	-	-	Galacticol PTS, IIA component in *Lactobacillus* spp.
294	-	-	-	Intergenic
290			snm: sp70585_2241	β-glucuronidase (GUS)
289	- /spd0364	spn0398/ 0399		Hypothetical protein and pseudogene
285			snm: sp70585_2232	Pseudogene
283	spd0562	spn0648		Putative β-galactosidase precursor
282			snm: sp70585_2233	β-N-acetylglucosaminidase
282	-	-	spv:sph2134	Cell wall surface anchor family protein

278	-	-	-	Lactose PTS regulator, LacR, putative
279	spd0110	-		Argininosuccinate synthase
278	spd1586	-		Sucrose operon repressor
273			snm: sp70585_2245	Hypothetical protein
270	spd0357	spn0391		Choline binding protein F, CbpF
269	spd0363	spn0397		Mannitol-1-phosphate 5-dehydrogenase
267	spd0605	spn0697		ABC transporter ATP-binding protein - unknown substrate/ pseudogene
267	spd1585	-		ABC transporter, sugar-binding protein
267	spd0360	spn0394		PTS, mannitol-specific IIBC components
266	-	spn0518		Hypothetical protein
263	-	spn0536		Immunity protein, BlpL
262	-	-	snt:spt0472	L-iditol 2-dehydrogenase
259			snm: sp70585_2237	Mannonate dehydratase, UxuA

Table 7.5. Analysis of the 50 biggest contigs in the "boneyard" of WU2.

Contigs were analysed using the BLAST-N program of KEGG to assign function and ID. IDs are given for D39 and TIGR4 where possible but when a gene is absent in both strains, the ID from another pneumococcal strain is used where possible with the strains abbreviated as follows: spv- Hungary19A-6, spw- CGSP14, sjj-JJA , spp- P1031, snt- Taiwan19F-14, spx-G54, snm- 70585.

Contig Size (bp)	D39 ID	TIGR4 ID	Other ID	Gene
441	-	-	spv:sph2077	Transcriptional regulator, LacI family protein
440	-	-	spw:spcg1273	Glutamate dehydrogenase
417	-	-	sjj:spj0998	Lantibiotic export protein, putative
404	-	-	spv:sph2077	Transcriptional regulator, LacI family protein
401	spd1992	spn2164		PTS, IIA component (mannose)
386	(1-90 in spd1160)	-	sjj:spj0998	Lantibiotic export protein, putative
330			spv:sph2234	Phage integrase family site specific recombinase
319			spv:sph2072	ABC transporter, substrate-binding protein
291			sjj:spj1002	Hypothetical protein
289	-	-	-	-
288	-	-	-	First 225 bp shows 41% identity to Amino acid adenylation domain protein in *Bacillus thuringiensis* serovar *berliner*
283	-	spn1189	-	Hypothetical protein
281	Spd0939	-		Transcriptional activator, Rgg/GadR/MutR family protein
279	-	-	spv:sph2076	Hydrolase
277	-	-	spv:sph2071	Hypothetical protein
274	-	-	spp:spp1779	ABC-2 type transporter ; K09694

				lipooligosaccharide transport
273	spd1988	spn2160		Large secreted protein
271	-	-	-	Intergenic
261	-	-	spv:sph2073	Spermidine/putrescine import ATP-binding protein, PotA
258			spv:sph2077	Transcriptional regulator, LacI family protein
256			spv:sph2074	Binding-protein-dependent transport systems inner membrane component
254	spd1994	spn2166		L-fuculose phosphate aldolase, FucA
254	spd0146	spn0143		CAAX amino protease family
253			spv:sph2231	Ser/Thr protein phosphatase family protein
252	spd1109	spn1251		Endonuclease, putative
247			spv:sph0145	Hypothetical protein
247			spv:sph2075	MgtC/SapB transporter
243			spw:spcg1273	Glutamate dehydrogenase
243	spd1987	spn2159		Fucolectin-related protein
240	-	-	-	Glutamate decarboxylase in unnamed pneumococcal strain
240	spd1993	spn2165		Fucose operon FucU protein
235	spd0950	-		Transporter, major facilitator family protein
232	-	-	-	Portion of hypothetical protein spv:sph2228
229	-	-	-	-
227	-	-	spv:sph0143	Hypothetical protein
225	spd0548	spn0628		HIT family protein
224	-	-	-	Intergenic
224	-	-	-	First 22 bp in hypothetical protein spp:spp1891

223	-	-	sjj:spj1791/ 1792	ABC transporter ATP binding protein and permease protein
222	spd1987	spn2159		Fucolectin-related protein
220	spd0939			Transcriptional activator, Rgg/GadR/MutR family protein (shows more identity to *S. pyogenes* [87-89%] than other *S. pneumoniae* strains [84-86%])
219	-	-	-	-
216	-	-	sjj:spj1000	Lantibiotic transport ATP-binding protein, putative
215	-	-	spv:sph2236	Hypothetical protein
215	-	-	spv:sph2073	Spermidine/putrescine import ATP-binding protein, PotA
215	spd0500	spn0574		Hypothetical protein
214	-	-	spv:sph0142	XRE family transcriptional regulator
211	-	-	spv:sph2074	Binding-protein-dependent transport systems inner membrane component
211	spd1080	spn1222	-	Type II restriction endonuclease, putative
209	-	-	snm:sp70585_0574	Hypothetical protein

WU2. Several sequences did not have homology to a known gene in other pneumococcal genome sequences.

In Section 7.3.2.7, both MSHR17 and WU2 were found to carry an alternative putative fucose utilisation operon on the basis of PCR and sequence analysis. Genes from this region, which are annotated as spd1987 (fucolectin-related protein)–spd1995 (L-fuculose kinase FucK) in the genome of D39, appeared in the boneyards of both MSHR17 and WU2 with a larger number appearing in WU2's boneyard. In addition, components of the partial lantibiotic system believed to be carried by WU2 (Section 7.3.2.3), as well as components of the alternative sucrose transporter to the one present in OXC141 detected in MSHR17 (Section 7.3.2.5), were present in their respective "boneyards".

Accessory genes absent in the OXC141 genome but present in only a small number of strains in the databases of KEGG and NCBI were detected in MSHR17 and WU2, and are described in the order in which they were detected through "boneyard" analysis.

7.3.3.1 Accessory region 1 of MSHR17

In the "boneyard" of MSHR17, segments of genes from a large region of over 16 kb appeared, which could not be detected in any other pneumococcal strain besides 70585. The genes from this region show homology to genes from a similar sized region in *S. suis*. In *S. pneumoniae* 70585, this region (sp70585_2232 to sp70585_2245) is between a histidyl-tRNA synthetase (sp70585_2231, spd1950 in D39) and a transporter (sp70585_2246, spd1951) and contains two β-N-acetylglucosaminidases, a phosphoglycolate phosphatase, a D-mannonate oxidoreductase, a mannonate dehydratase, a glucuronate isomerase, a khg/kdpg (4-hydroxy-2-oxogluturate/ 2-dehydro-3-deoxy-phosphogluconate) aldolase, a transcriptional regulator of the GntR family, a β-glucuronidase (GUS), sugar kinase, ribokinase family, a sugar/sodium symporter, as well

as a pseudogene and hypothetical proteins (Figure 7.10). The hypothetical proteins are most likely transposase sequences and are found in other pneumococcal strains. The genetic content suggests that this region is involved in carbohydrate utilisation.

7.3.3.2 Accessory region 1 of WU2

The genes sph2071-2077 in the WU2 "boneyard" are part of an accessory region in Hungary19A-6 (Figure 7.11). Sph2072-2075 show approximately 50% identity to Sph1943-1946, indicating that there may be redundancy. Taiwan19F-14 carries this accessory region, but does not have homologues for sph2074 and sph2075, which encode a binding-protein-dependent transport system inner membrane component and MgtC/SapB transporter respectively.

7.3.3.3 Accessory region 2 of WU2

The second largest contig assembled from WU2's "boneyard" aligned with spcg1273 of CGSP14 and snt0922 of Taiwan19F-14, which are annotated as glutamate dehydrogenase GdhA and hypothetical protein respectively (Figure 7.12). The gene in Taiwan19F-14 is encoded on the sense strand and is 3578 bp long. The gene in CGSP14 is 3612 bp long and lies on the opposite DNA strand to its homologue in Taiwan19F-14. The orientation of the genes surrounding the glutamate dehydrogenase of Taiwan19F-14 is opposite to the homologues in CGSP14, as well as glutamate dehydrogenase. In strains that do not carry the glutamate dehydrogenase gene, such as D39, these homologues were in the same orientation as CGSP14 (Figure 7.12). Therefore, part of the genome appears to have been inverted in Taiwan19F-14.

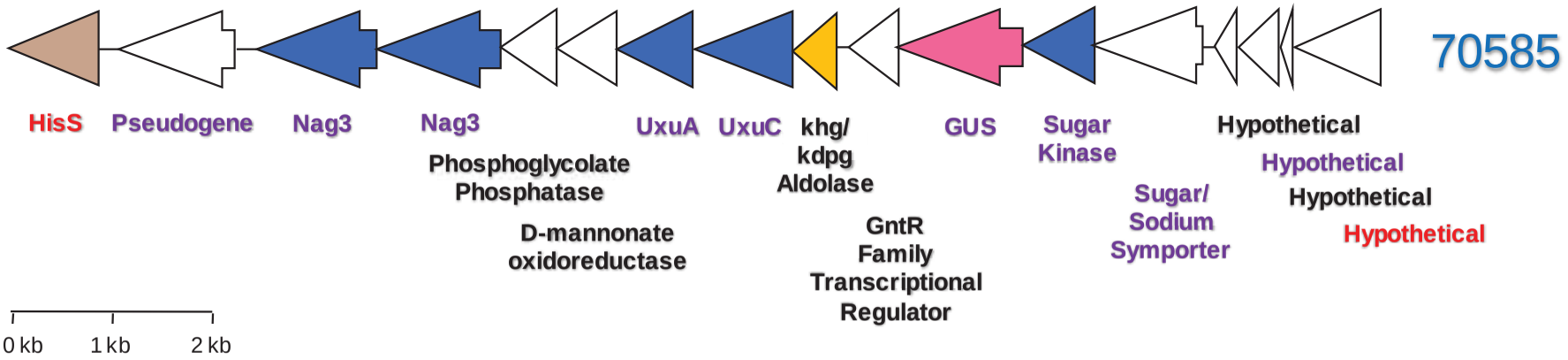


Figure 7.10. The unique Accessory Region of over 16 kb that is carried by *S. pneumoniae* 70585 (nt 2037813-2056376) and has genes that were detected in the boneyard of MSHR17.

The strain 70585 carries a unique accessory region between a histidyl-tRNA synthetase and transporter gene with most genes related to carbohydrate utilisation. However, many sequences from this accessory region were detected in the boneyard of MSHR17.

Genes that are common to the pneumococcal genomes submitted to KEGG are indicated by red font. Genes that were detected through the analysis of MSHR17's boneyard are indicated by purple font. HisS= histidyl tRNA-synthetase, Nag3= β-N-acetylglucosaminidase, UxuA= Mannonate dehydratase, UxuC=Glucuronate isomerise, GUS= β-glucuronidase. ORFs are coloured according to their KEGG putative functional category as shown in Table 7.1.

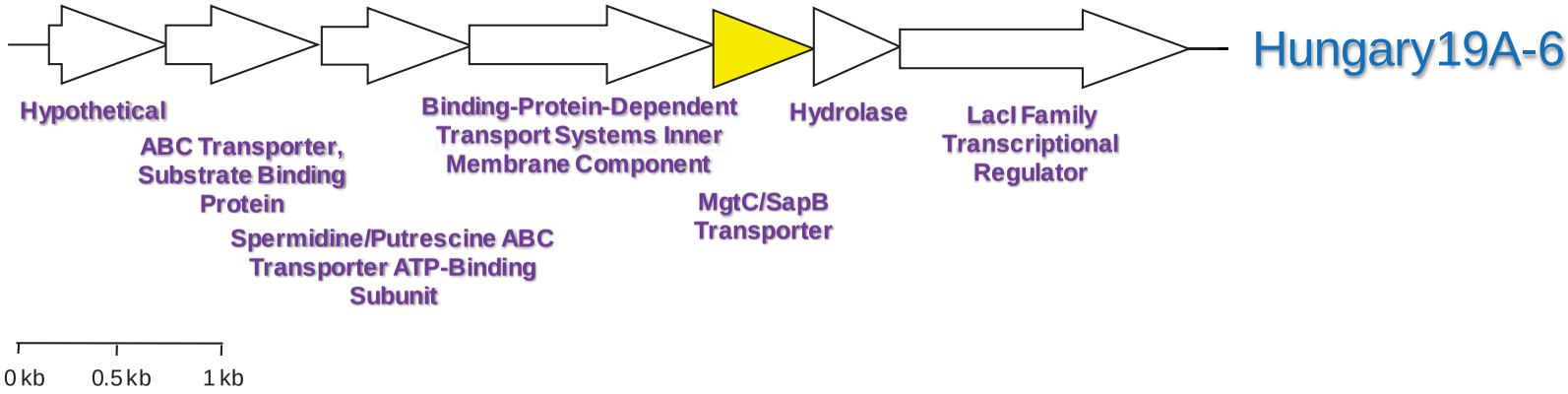


Figure 7.11. The Accessory Region of 8 kb in *S. pneumoniae* Hungary19A-6 (nt 1906905-1914939) which appears to be carried by WU2 but not OXC141.

The region is shown in Hungary19A-6 but all genes were detected in the boneyard of WU2, which is indicated by purple font. The strain Taiwan19F-14 carries a similar region but lacks the binding-protein-dependent transport system inner membrane component and MgtC/SapB transporter (yellow arrow). ORFs are coloured according to their KEGG putative functional category as shown in Table 7.1.

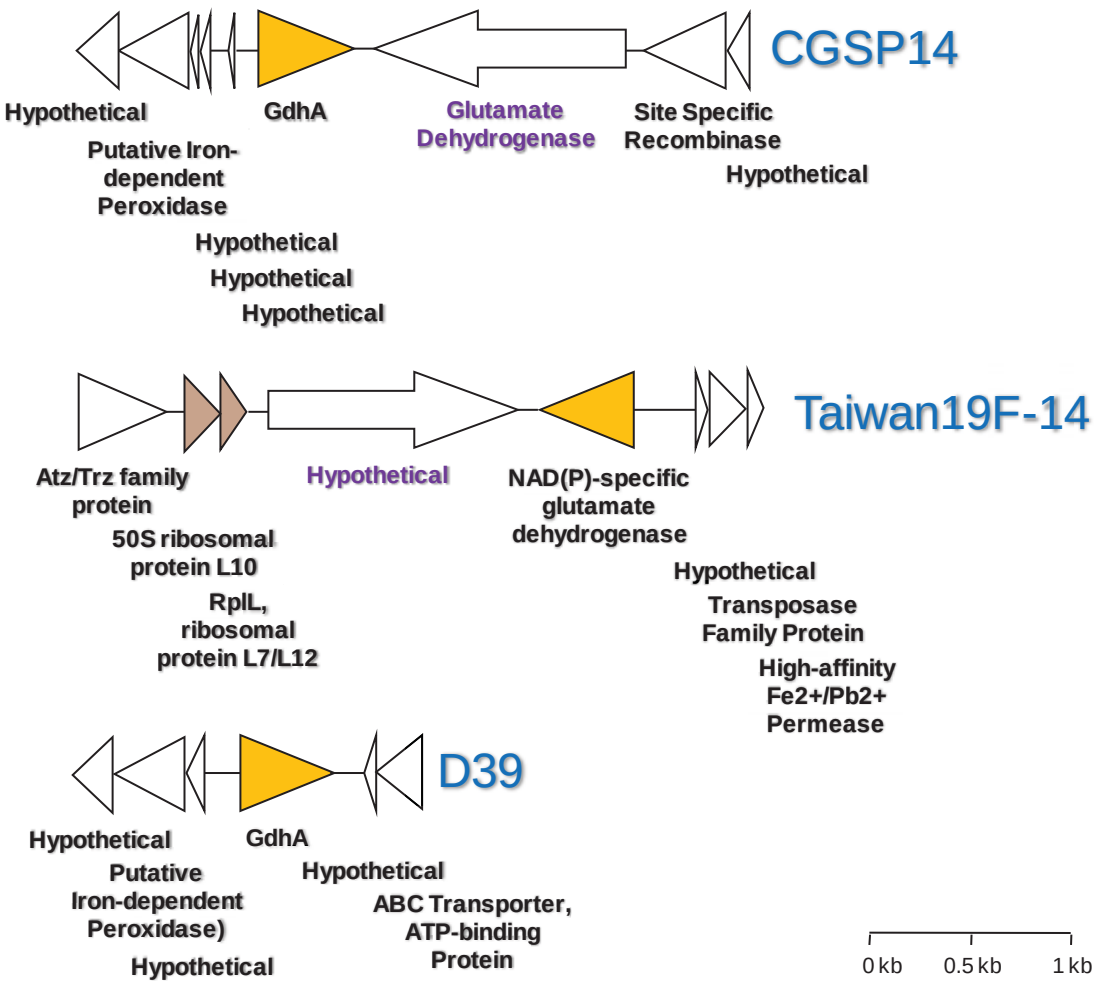


Figure 7.12. nt 1251637-1261179 in CGSP14, nt 867002-876693 in Taiwan19F-14 and nt 1184440-1189669 in D39.

Sequence from a gene carried by CGSP14 (annotated as glutamate dehydrogenase) and Taiwan19F-14 (annotated as a hypothetical protein) was identified in the "boneyard" of WU2 and these two genes are in purple font in the figure. The orientation of this gene varies in the two strains. The surrounding genes are also shown in this figure, as well as the equivalent region in the D39 genome, which does not have the putative glutamate dehydrogenase. ORFs are coloured according to their KEGG putative functional category as shown in Table 7.1.

7.3.3.4 Accessory region 3 of WU2

Three contigs demonstrated homology to the three genes sph2231, sph2236 and sph2234, which are part of a large accessory region in Hungary19A-6 (Figure 7.13). The genes sph2232 and sph2234-sph2237 have homologues in G54 as well. The accessory region is between sph2223 and sph2239 in the Hungary19A-6 genome and encodes hypothetical ORFs, as well as genes assigned functions, including an endonuclease/exonuclease/phosphatase family protein, Rep protein, metal dependent phosphohydrolase and Ser/Thr protein phosphatase family protein.

7.3.3.5 Accessory region 4 of WU2

One contig aligned with the gene spp1779, which is found in strain P1031 only. The adjacent genes spp1780 and spp1781 are also unique to P1031 and encode for a Nod-factor export ATP-binding protein I and transcriptional regulator of the ArsR family. However, sequences from these two genes were not detected in the 50 largest contigs of the WU2 "boneyard".

7.3.3.6 Accessory region 5 of WU2

Three contigs aligned with the genes sph0142, sph0143 and sph0145 of Hunagry19A-6. These genes are part of an accessory region that is adjacent to *comA* (sph0142-sph0147) and are also found in the strains JJA and 70585. One gene encodes for a transcriptional regulator (sph0142) but the other genes encode hypothetical proteins that show homology to ABC permeases in other species.

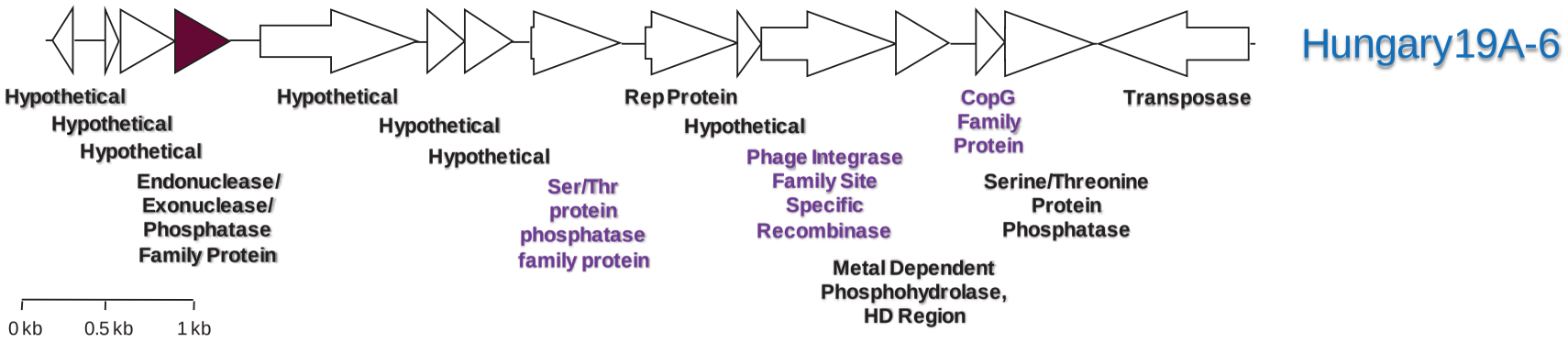


Figure 7.13. A 10.5 kb accessory region of Hungary19A-6 (sph2224-2238, nt 2044289-2054798) that has genes detected in the "boneyard" of WU2.

Some genes from this accessory region are carried by G54 as well. The genes that were detected in the "boneyard" of WU2 are in purple font. ORFs are coloured according to their KEGG putative functional category as shown in Table 7.1.

7.3.3.7 Accessory region 6 of WU2

One contig aligned with two genes that encode components of an ABC transporter (spj1791 and spj1792 in the strain JJA). These genes are part of an accessory region that is found in P1031, JJA and ATCC700669 (Figure 7.14). The permease in P1031 is labelled as a pseudogene and the sequence is almost identical, except 16 bp is missing. Although the nucleotide sequence is given in KEGG, the amino acid sequence is not due to the pseudogene being 878 nucleotides in length. It cannot be determined from the contig whether WU2 carries the permease or the pseudogene. Adjacent to these genes is a pseudogene, which is annotated as ribose/xylose/arabinose/galactoside ABC-type transport system, permease component in JJA, and differs in size across all three strains. This accessory region is between the trehalose operon repressor TreR and the oligopeptide transport ATP-binding protein AmiF. In strains that do not carry this accessory region, a pseudogene or a number of transposase sequences are found between this repressor and transporter but are encoded on the complimentary DNA strand (Figure 7.14).

7.3.3.8 Accessory region 7 of WU2

The start of the final contig aligned with the last 196bp and 138 bp of genes for hypothetical proteins in strains 70585 and Spain23F-1, which are 405 bp and 357 bp respectively. These hypothetical proteins are located between the genes for the molecular chaperone DnaK and chaperone protein DnaJ, which are annotated in the D39 genome as spd0460 and spd0461, respectively.

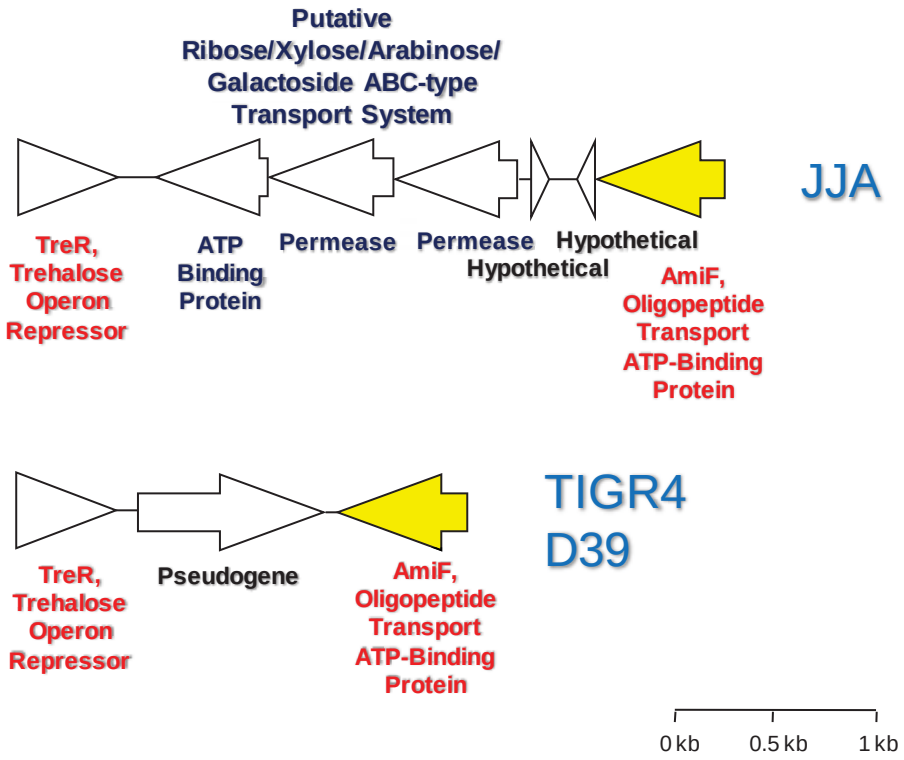


Figure 7.14. The variable region between *treR* and *amiF* in the pneumococcal genome.
The pneumococcal strains JJA, P1031 and ATCC70069 carry an accessory region between the genes for a trehalose operon repressor TreR and an oligopetide transporter AmiF. There is some difference between size and sequence. The region of JJA is shown in the figure (nt 1708206-1713023). In strains that do not possess the aforementioned accessory region, a pseudogene or transposase genes encoded on the complementary strand are found between *treR* and *amiF*. D39 and TIGR4 are illustrated as examples (D39: nt 1677776-1680949, TIGR4: nt 1788834-1792026). ORFs are coloured according to their KEGG putative functional category as shown in Table 7.1.

7.3.3.9 PCR analysis of the accessory region 1 of MSHR17

Genes from a 16 kb accessory region that was only detectable in *S. pneumoniae* 70585 through BLAST analysis were identified in the "boneyard" of MSHR17. However, not every gene found in 70585 was detected in MSHR17 through the analysis of 50 contigs. As MSHR17 is an OM isolate, it was decided to investigate this region further and confirm that it is absent in the ST180 middle ear isolate WCH206. PCR analysis of MSHR17 using the primer pairs spd1950F/snm2237R, snm2235F/snm2241R, and sp1951R/snm2240F (Table 2.2) produced amplicons of the expected size (approximately 8 kb, 6 kb and 6.5 kb, respectively) and indicated that not only is this accessory region the same size as 70585, but is found in the same location in the genome. WCH206 (ST180) was analysed using primers spd1950F and spd1951R and the size of the amplicon obtained (approximately 570 bp) indicated that it did not contain any genes between the histidyl-tRNA synthetase and transporter genes. There were also no genes in the OXC141 reference sequences between the synthetase and transporter. The MSHR17 alignment was reanalysed at this region, but there was no strong indication of the presence of an insertion apart from a poor alignment at the very end of the transporter gene (Figure 7.15).

7.4 Reconciliation of Genome Sequencing Data with Chapter 4 Data

7.4.1 Comparison to microarray data

In Chapter 4, the genome of the ST180 middle ear isolate WCH206 was compared to the genome of MSHR17 using DNA microarray analysis. The assembly of the "boneyard" of MSHR17 finally allowed the data from the next generation sequencing to be compared to this data. Tables 4.1 and 4.2 derived from the microarray data (genes that are present in MSHR17, but unidentifiable in WCH206) were compared to Table 7.3,

Figure 7.15. The poor alignment of the MSHR17 resequencing near the unique accessory region.

The large accessory region identified in MSHR17 through "boneyard" and PCR analysis was between a histidyl tRNA-transferase (annotated as spd1950 in D39) and a transporter gene (spd1951). The alignment of this region is shown above in DNAstar Lasergene version 8 SeqMan. The sequence at the top is the OXC141 reference genome, and the 36 bp sequence fragments are from the MSHR17 genome, which were generated by Geneworks using an Illumina Genome Analyser *II*. In this region of the alignment, there were no obvious signs of an insertion aside from a region of poor alignment near the transporter gene.

while Table 4.3, which was also derived from the microarray data (genes that are present in WCH206, but unidentifiable in MSHR17), was compared to Table 7.1 (large gaps in the alignment of MSHR17 with OXC141), as well as Table A1 of the Appendix (smaller gaps and poor alignments).

The DNA microarray analysis suggested that were 15 genes which were absent in WCH206 but present in MSHR17 (Tables 4.1 and 4.2). Five of these genes (spn2160, spn0477, spn0695, spn0697, and spn2159) appeared in the "boneyard" of MSHR17 (Tables 4.1 and 4.2). Interestingly, the genes spn0397 and spn2166 were also adjacent to genes suggested to be missing in WCH206 by microarray analysis (spn0695, spn0697, spn0477). Furthermore, two genes from the "boneyard" (spn0696, spn0476) had been suggested to be present in MSHR17 but absent in WCH206 on two out of four slides. Since the "boneyard" was not completely analysed (only the fifty largest contigs were analysed), genes present as part of smaller contigs may have been missed.

Three genes were suggested to be absent in MSHR17 but present in WCH206 through DNA microarray analysis. Two of these genes (spn0506 and spn1857) were part of large sequencing gaps of approximately 1.5 kb and 4.6 kb respectively in the MSHR17 alignment (Table 7.2). The 4.6 kb sequencing gap was analysed by PCR in Section 7.3.2.6 and found to contain a product variable in size and sequence to the analogous region in OXC141.

Microarray analysis did not detect all differences in homologous genes of R6 or TIGR4. For example, it did not detect the difference in the sucrose transporter (Section 7.3.2.5). When the microarray data were reanalysed, the genes spn1796-spn1799 and spr1618-spr1620 (homologous to spd1583-151586) were not detected on most microarray slides. MSHR17 has a 51 nt deletion within the gene homologous to spr1618, which may have been why it was not detected, although it is not known which part of the gene is represented on the microarray slide. The other genes may not have been detected due to

sequence variation. Otherwise, the microarray conditions used in this study may not have been optimal for the hybridisation of those particular sequences.

7.4.2 Comparison to PCR-based subtractive hybridisation data

PCR-based subtractive hybridisation was used in Chapter 4 to identify genes that were present in the genomes of ST180 and ST458 strains but not D39 and TIGR4 (Section 4.2). Only a putative cellobiose PTS was found, which was later shown to be part of a larger island in Chapter 5. The other genes of the island are also absent in D39 and TIGR4.

Analysis of the "boneyard" of MSHR17 in Section 7.3.3 revealed a number of genes that were absent in D39 and TIGR4 (Table 7.3), including a large 16 kb insert that had only previously been annotated in the pneumococcus in the strain 70585 (Section 7.3.3.1). However, as these sequences were present in the "boneyard", they were also absent in OXC141.

A number of sequences absent in D39 and TIGR4 were also detected in the OXC141 genome by BLAST analysis (Section 7.2). This included a >30 kb prophage, a unique endonuclease, a 2.5 kb region carrying a MerR family transcriptional regulator, a 6.4 kb region carrying a FtsK/SpoEIII family protein and a large 24 kb region that had only been annotated in another serotype 3 strain of ST180, SP3-BS71. However, when the MSHR17 genome was assembled, none of these genes were present apart from the D-alanine-D-alanine ligase from the large unique region over 24 kb (Table 7.1). The other genes from this unique region, as well as the 6.4 kb accessory region carrying the FtsK/SpoEIII family protein, were confirmed to be absent in MSHR17 by PCR analysis (Section 7.3.2) In addition, OXC141 carries the variant of a fucose operon found in D39 and TIGR4 (Section 7.2). However, the fucose operon carried by MSHR17 was found to be the same as D39 and TIGR4 (Section 7.3.2.7). Therefore, although a number of

sequences absent in D39 and TIGR4 have been identified in both ST180 and ST458; it appears that only the D-alanine-D-alanine ligase and island carrying the putative cellobiose PTS is common to both.

7.5 Discussion

The next generation sequencing agreed with the findings of Dagerhamn *et al.* (2008) that there can be vast genetic differences within a serotype. Furthermore the sequencing supported the results of the DNA microarray analysis and PCR-based subtractive hybridisation of Chapter 4. The "boneyard" of MSHR17 contained some genes that were identified as absent in WCH206 (ST180 like OXC141) but present in MSHR17 through DNA microarray analysis (spn2160, spn0477, spn0695, spn0697, spn0397 and spn2159). In addition, genomic fragments of MSHR17 did not align with homologues in the OXC141 genome of spn0506 and spn1857, which were detectable in WCH206 by microarray analysis but not MSHR17 (Tables 4.3 and 7.2).

PCR-based subtractive hybridisation was performed to identify genes present in isolates of both ST180 and ST458 which were absent in TIGR4 and D39 (Section 4.3). Only a putative cellobiose PTS (PTS 1) was identified. In this chapter, genes which were absent in D39 and TIGR4 but present in OXC141 (ST180) or MSHR17 (ST458) were identified in addition to PTS 1 and the genes from its associated island (Sections 7.2 and 7.3.3). However, apart from D-alanine-D-alanine ligase B from PP1 (Section 7.3.2.3), these genes were only found in one of the strains. Therefore, although PCR-based subtractive hybridisation missed some genes such as the gene for the ROK family protein adjacent to PTS 1 and D-alanine-D-alanine ligase B, it seems that there are not many accessory regions present in strains of ST180 and ST458 that are also absent in TIGR4 and D39. Although the accessory regions of OXC141 that are absent in MSHR17 may not be

essential for OM, they are still of interest because of the dominance of ST180 worldwide. Furthermore, the accessory regions of MSHR17 may reflect genetic requirements for the Australian Aboriginal population. For instance, children from remote Australian Aboriginal communities carry a higher number and variety of bacteria compared to Caucasian children (Waston *et al.*, 2006), and ST458 may be better adapted to this highly competitive environment than ST180.

Following analysis of the resequencing data, some of the large sequencing gaps in the resequenced genomes of MSHR17 and WU2 were analysed by PCR and sequencing. Apart from a gap in the *zmpB* gene of MSHR17 that was due to sequence variation, the gaps were either due to true deletions or differences in genetic content due to variability within the pneumococcal genome or the possession of alternative carbohydrate metabolism systems.

The second gap analysed in MSHR17 and WU2 (Section 7.3.2.3) appeared to be that of an accessory region over 20 kb in size and unique to OXC141 and SP3-BS71 that included an enolase, transketolase and aminotransferase amongst other genes. However, PCR products in this region were obtainable for MSHR17 and WU2. The product for MSHR17 was only 1 kb and contained a transposase and 6'N-acetyl transferase. On the other hand, the product for WU2 was 6 kb and contained an incomplete lantibiotic system. The incomplete lantibiotic system is part of pneumococcal pathogenicity island 1, which is a highly variable region of the pneumococcal genome that is located between spn1046 and spn1064 in the TIGR4 genome (Croucher *et al.*, 2009; Blomberg *et al.*, 2009; Brown *et al.*, 2004). Therefore, the unique genes within OXC141 and SP3-BS71 are part of this pathogenicity island.

The third gap analysed in MSHR17 and WU2 was of particular interest because it contained a gene with identity to the otitis media-associated H10 protein in Taiwan19F-14, which is also present in Hungary19A-6, annotated as hypothetical protein sph2228. OM-

associated H10 was identified through a PCR-based subtractive hybridisation study that used a serotype 19 middle ear isolate as the tester strain (Pettigrew and Fennie, 2005). This probe was then found to hybridise more frequently amongst middle ear isolates compared to carriage, blood and cerebrospinal fluid isolates, although it appeared to have a higher association with serotype 19 compared to other serotypes investigated (Pettigrew and Fennie, 2005). Furthermore, OM-associated H10 was identified in the signature tagged mutagenesis study of Chen *et al.* (2007), which screened a serotype 19F middle ear isolate for genes that were required for survival in the middle ear in a chinchilla OM model. OM-associated H10 was described as a gene with 74% identity to an unknown protein from *S. agalactiae* NEM316 and was next to a gene that displayed identity to a FtsK/SpoEIII family protein of *Bacillus subtilis* (Pettigrew and Fennie, 2005). However, in the sequence submitted to GenBank (AY845429), the sequence for a gene showing similarity to FtsK/SpoEIII family proteins was labelled as putative OM--associated H10. The amino acid sequence for putative otitis media-associated H10 shows almost 100% identity to Spt0647 and approximately 88% identity to Sph2228. In Hungary19A-6, the gene next to sph2228, which is also annotated as a hypothetical protein (sph2229), shows almost 90% identity to an unknown protein from *S. agalactiae*. A similar gene is not adjacent to spt0647 in Taiwan19F-14. Despite this confusion regarding which gene H10 is, both MSHR17 and WU2 did not harbour the region showing identity to putative otitis media-associated H10.

The other large gaps of particular interest were described in Sections 7.3.2.5 and 7.3.2.7 and demonstrated the tendency of the pneumococcus to possess alternative versions of carbohydrate metabolism and uptake systems. First, an alternative sucrose transporter system was identified in MSHR17 compared to OXC141 (Section 7.3.2.5). Iyer and Camilli (2007) reported that there are two variants of sucrose permeases associated with the hydrolase SusH, one which is present in TIGR4 and the other which is present in D39.

The system found in TIGR4, which is similar to the one found in OXC141 and WU2, is important for colonisation of the lung but the importance of the alternative system found in D39, which is believed to be similar to the system in MSHR17, in this niche has not been established yet (Iyer and Camilli, 2007). The authors reported a second sucrose metabolising system, the Scr system, which is found in a different region of the pneumococcal genome (spn1722-spn1725 in TIGR4) and is important for nasopharyngeal colonisation. Unlike the Sus system, the authors did not report variation in this system. However, a 1478 bp gap in the Scr PTS components was identified in the MSHR17 alignment (Table 7.2). This sequence of the OXC141 genome is not found in D39 or TIGR4, but is found in CGSP14 (spcg1694, 1962 bp). Therefore, it appears that there is variation in both sucrose metabolism systems of the pneumococcus.

In addition, the next generation sequencing identified an alternative fucose utilisation operon in MSHR17 and WU2 as compared to OXC141. The transporter is an ABC permease in OXC141 but a PTS in MSHR17 and WU2. The operon for fucose utilisation of TIGR4, which is believed to be the operon possessed by MSHR17 and WU2, has been demonstrated to be essential for invasive pneumococcal disease (IPD), but dispensable for asymptomatic carriage (Iyer and Camilli, 2007; Embry *et al.*, 2007). Moreover, the endo-β-galactosidases of the two operons have been to shown to hydrolyse the galactosyl-β-1,4-*N*-acetylglucosamine linkage found in type 2 carbohydrate blood group antigens. However, the hydrolase from TIGR4, which is identical to that in MSHR17 and WU2, has substrate specificity for Lewis^{y} antigen, while the hydrolase from the ST180 isolate SP3-BS71 is selective for blood-group A/B-antigens (Higgins *et al.*, 2009). The Lewis^{y} antigen is found in people of all bloodtypes, but blood-group A/B antigens are expressed on a wider range of tissues and therefore, the particular variant of hydrolase carried by a pneumococcal strain may influence which host cells it most effectively targets (Higgins *et al.*, 2009). All the sequenced pneumococcal strains lodged

in the KEGG database carried one of the variants of the fucose utilisation operon, which suggests that even if the hydrolase variant can influences disease outcome, the operon is still an important component of the pneumococcal genome.

Although a number of large sequencing gaps over 100 bp were detected in the next generation sequencing work, a larger number of genes containing poor alignments or small gaps were identified in MSHR17 and WU2 and are tabled in the Appendix. Of the 193 and 277 genes that were identified in MSHR17 and WU2 respectively, no gene was found to be absent in both D39 and TIGR4 with only a small number absent in either one of these strains. However, most of the small gaps and poor overlaps were in fact found in transposases, intergenic sequences or BOX elements. The large quantity of data highlights a significant disadvantage of resequencing without having an almost identical genome available on which to map the resequenced data, and this is especially obvious where the data consists of only short sequence reads of approximately 36 bp generated by the Illumina technology. Even the analysis of all the large gaps proved too time consuming for this study without considering the poor alignments and small gaps. A small gap may actually harbour an interesting insert or a poor alignment may mask an important deletion, but unfortunately there is too much data to be able to identify these. As with PCR-based subtractive hybridisation, resequencing works best with closely related strains. At the time this work was performed, only resequencing using 36 bp fragments was available to the laboratory. Now this procedure can be performed using 72 bp fragments and the fragments can be sequenced from both ends (paired-end read). Therefore, if the resequencing was repeated using these longer fragments, it is likely that fewer unsatisfactory or poor alignments would have occurred.

A large amount of data was generated for the “boneyard” as well, which was due to the high genetic variation between strains. More accessory regions absent in D39 and TIGR4 were detectable in WU2 compared to MSHR17, when the 50 largest contigs were

analysed for each strain, including components of the incomplete lantibiotic transport system of WU2 detected in Section 7.3.2.3. Although MSHR17 and WU2 both harbour regions that are absent in OXC141, these were broken up into different contigs in the "boneyard". If a 36 bp fragment of a region absent in OXC141 is present in another location of the OXC141 genome or an almost identical sequence, the fragment will be excluded from the "boneyard", which will affect "boneyard" assembly. This is another issue that could be resolved by resequencing using longer fragments. There were a number of genes within the "boneyards" that were previously identified or predicted to be present from the sequence analysis of large gaps in Section 7.3.1. However, not all genes or parts thereof in the "boneyards" were detected in Section 7.3.1, in particular, genes from an accessory region of approximately 16 kb in MSHR17. This accessory region was inserted between a histidyl tRNA-transferase gene and transporter gene but there was no obvious indication of an insertion in the sequencing apart from a poor alignment at the very end of the transporter gene (Figure 7.15). Therefore, although accessory genes can be detected by analysing large gaps within the resequencing, analysis of the "boneyard" is still of vital importance in identifying accessory regions that are not carried by the reference strain.

Next generation sequencing was found to have limitations during this study, but by investigating large sequencing gaps within the data and analysing the "boneyards", overall the resequencing has given a useful comparison of major accessory regions present in three unrelated serotype 3 strains.

Chapter 8

Final Discussion

8.1 Introduction

Otitis media is a major public health burden worldwide impacting both developed and developing nations. In Australia, children of remote Aboriginal communities suffer from unacceptably high levels of OM, which have not improved over the last three decades (Morris *et al.*, 2005). Not only is chronic suppurative otitis media (CSOM), which is normally rare in a developed country, present at ten times the level defined as a major public health issue by WHO (Coates *et al.*, 2002), but the children of these communities experience the highest rates of tympanic membrane perforations in the world (Leach and Morris, 2007). Long term infections of the middle ear can result in deafness and it has been estimated that 80% of children in these communities suffer from some degree of hearing impairment (Morris *et al.*, 2005). This has a long term negative effect on both learning and employment outcomes in these children, who are already disadvantaged without the added burden of OM.

8.1.1 Characterisation of clinical isolates from Aboriginal communities

At the start of this project, serotype 3 middle ear isolates from remote Aboriginal communities, as well as the general population, were obtained from MSHR and WCH (Table 2.1). In addition to serotype 3, serogroup 11 isolates were obtained from the same sources. Serotype 3 is commonly associated with middle ear infections worldwide (Prymula, 2009; Hausdorff, 2000, 2002), including in remote Aboriginal communities, whereas, serogroup 11 is rarely associated with OM in Aboriginal communities and is generally not as common as serotype 3 in OM elsewhere (Hausdorff, 2000, 2002). Data presented in this thesis show that the serotype 3 isolates used were highly virulent in both .i.p. and i.n. murine models of infection, while the serogroup 11 isolates that were tested were avirulent (Section 3.2.5.1).

To date, isolates from remote Aboriginal communities have been poorly characterised, and one of the aims of this project was to increase our molecular understanding of OM-associated strains from these communities at both the phenotypic and genotypic level (Section 1.5). Therefore characterisation of the MSHR isolates was undertaken. This included Western blot analysis, haemolysis assays, MLST typing and DNA microarray analysis (Sections 3.2.1, 3.2.2, 3.2.3 and 4.2). MLST typing of the serotype 3 middle ear isolates from the MSHR, as well as one serotype 3 nasopharyngeal isolate from this source, revealed they all belong to ST458 (Section 3.2.3). This ST was also identified in one of the serotype 3 isolates from the WCH; however, the other two serotype 3 isolates from this latter source belong to ST180. The other serotype 3 nasopharyngeal isolates from the MSHR, which were not typed, displayed identical protein profiles to the ST458 isolates when analysed by Western blot analysis (Section 3.2.1), which suggests they may be from the same or closely related ST. When the ST458 isolate

from WCH (WCH208) and three of the serotype 3 isolates from MSHR (MSHR12, 13, 15) were compared to one another by DNA microarray analysis using slides containing ORFs from R6 and TIGR4, no genomic differences could be detected between the strains (Section 4.2.2). The high degree of relatedness between the isolates of the MSHR was interesting since the isolates were from five different remote Aboriginal communities.

ST180 is generally the dominant ST within serotype 3 in developed countries (McEllistrem *et al.*, 2005; Isozumi *et al.*, 2008; Bozdogan *et al.*, 2004; Vestrheim *et al.*, 2008; Porat *et al.*, 2008; Brueggemann *et al.*, 2003) and only a few genetic differences have previously been observed between isolates of this lineage (Hiller *et al.*, 2007; Dagerhamn *et al.*, 2008). As such, it is of interest that ST458 is dominant within the serotype 3 population from the Aboriginal communities. It may be that genetically, ST458 is better suited to these communities. It also suggests that as well as serotype, the ST distribution within these communities may vary compared to the general population. However, only a small number of isolates were investigated and numerous additional strains would need to be sequence typed in order to draw a conclusion. In serogroup 11, two novel STs were identified (ST3020 for WCH211 and ST3021 for MSHR1), as well as the first report of an ST662 isolate (MSHR5) from this serogroup (Section 3.2.3). WCH213 was ST62, which reports from Europe and the USA suggests is the dominant ST type within serotype 11A (Gertz *et al.*, 2003; Hanage *et al.*, 2005; Sjostrom *et al.*, 2006) and isolates of this ST have also been isolated from OM patients (Hanage et al., 2004). However, Western blot analysis and MLST suggested ST62 does not dominate within serotype 11A in remote Aboriginal communities and the ST distribution within serotype 11A differs to the general population. Data provided by the MSHR suggests that serotype 11A is extremely rare in OM in remote Aboriginal communities and this could reflect differences in the genetic background of these serotype 11A isolates compared to other

communities in developed countries. However, the role of host factors is an important consideration as well.

8.1.2 Serotype, genetic background and disease

There are numerous studies indicating the importance of serotype and more recently ST in disease. The observation from the MSHR that serotype 3 was commonly isolated from the middle ears of children in remote Aboriginal communities, but serotype 11A was not, suggested that serotype could be important. However, the genetic background within these serotypes was not known at the start of this study.

As mentioned previously, two STs were identified within the serotype 3 isolates obtained for this study from the MSHR and WCH (Section 3.2.3), which indicated that serotype 3 isolates of different genetic backgrounds can cause OM. In this study, statistically significant reduced survival times were displayed between some serotype 3 isolates in the i.p. and pneumonia/sepsis mouse models (Section 3.2.5.1). In particular, the serotype 3 strain WU2, which is a different ST to the other isolates (ST378), was dramatically less virulent in the pneumonia/sepsis model, where it was the only serotype 3 strain that killed less than 50% of the mice within the experimental timeframe. However, in both models there were also some strains of the same ST that demonstrated a statistically significant difference in their median survival times. A report by Silva *et al.* (2006) also found that isolates of the same ST and serotype produced different levels of bacteremia in an i.p. mouse model. Therefore, it appears that there are other factors in addition to ST that influence disease. ST is based on the sequence of seven housekeeping genes, and as such does not give information about other genes, including virulence factors whose presence or sequence may differ between isolates of the same ST. Nevertheless,

isolates of the same ST have been shown to exhibit the same pattern of accessory regions when compared by DNA microarray analysis (Section 4.2.2, Dagerhamn *et al.*, 2008). Furthermore, PspA family type has been shown to have an ST association (Rolo *et al.*, 2009), and the data from this study suggests that this may be the case for other polymorphic virulence factors, such as NanA and PspC. However, small changes in gene content, such as the gain or loss of an accessory region, could impact virulence.

When four serogroup 11 isolates from unrelated STs were assessed in the same models, all isolates were found to be avirulent (Section 3.2.5.1). DNA microarray comparison between serotype 3 and serogroup 11 did not identify any distinguishing genes between the two serotype/groups (Section 4.2.4). Therefore, the capsule loci remain the only known distinguishing feature between serotype 3 and serogroup 11. In addition, the serogroup 11 strains also harbour PTS 1 , which was not present on the microarray slides (Section 5.2.1). Since the four unrelated isolates from serogroup 11 were avirulent, the capsule may very well be the reason for these strains' avirulence. For example, if serogroup 11 capsules confer low resistance to phagocytosis, this would certainly have a major impact on invasiveness unless it was compensated for by some other factors (Obert *et al.*, 2006).

In order to investigate the influence of serotype on disease, the serotype 11A strain MSHR5 was switched to serotype 3 (Section 3.3.1). The ability of this capsule switch mutant (MSHR5^3) to colonise in various niches of the pneumonia-sepsis model was not the same as the serotype 3 capsule donor strain WCH206 (Section 3.3.3). Not only was MSHR5^3 not isolated from the blood, but it was also less fit in the middle ear compared to WCH206. This suggests that the reason serotype 3 is commonly isolated from children suffering from OM in remote Aboriginal communities and serotype 11A is not cannot

simply be due to capsular serotype alone. The work of Kelly *et al.* (1994) also showed that switching a strain to serotype 3 did not necessary result in identical virulence to the serotype 3 donor strain in an i.p. murine model, and while some capsule switch mutants increased in virulence, others were attenuated, as described in Section 1.3.2.

8.1.3 Genomic analysis of serotype 3 isolates

Serotype 3 isolates of ST180 and ST458 have been analysed in this study using DNA microarray analysis, PCR-based subtractive hybridisation and second generation sequencing technology (Chapters 4 and 7). All of these methods have limitations, but by employing a variety of techniques, the likelihood of achieving the first aim of the project, which was to identify candidate proteins that might play a role in OM, was increased (Section 1.5). A putative cellobiose PTS system (PTS 1) was identified in ST180 and ST458 by PCR-based subtractive hybridisation (Section 4.3). This PTS is part of an accessory region that also includes a ROK family protein, sulfatase, sulfatase modifying factor and several hypothetical proteins (Section 5.2.3). However, the sulfatase, as well as the start of the sulfatase modifying factor and the end of the PTS operon, are absent in ST458 (Section 5.2.4).

Although only the putative cellobiose PTS was identified through PCR-based subtractive hybridisation, the next generation sequencing revealed that there are only a small number of genes that are absent in both D39 and TIGR4, but present in both ST180 and ST458 (Sections 7.2 and 7.3). These genes were not identified through PCR-based subtractive hybridisation due to limitations of this technique, such as the requirement of unique sequence to be on a restriction fragment from 200 bp to 1.2 kb in size (Agron *et al.*, 2002).

8.1.4 Distribution and mutational analysis of the PTS 1 island

The presence of the above-mentioned island was investigated in a number of pneumococcal isolates by PCR analysis and was extended by bioinformatic analysis of the growing number of fully sequenced pneumococcal genomes available online (Sections 5.2.1, 5.2.2 and 5.2.5). This related to the third specific aim of the project, which was to investigate the distribution of any potential OM-associated proteins identified (Section 1.5). The PTS 1 region was found to be carried in a wide range of isolates of various serotypes and disease potential, with the vast majority carrying the full island.

The island in the ST180 isolate WCH206 was deleted and replaced with an erythromycin resistance gene, which was attached to a promoter from D39's capsule locus to allow expression (Section 6.2.1). Unfortunately, although numerous attempts were made using a variety of transformation methods, a mutant in the ST458 middle ear isolate MSHR17 was not obtained (Section 6.2.2). Replacing the entire island in WCH206 with the aforementioned erythromycin resistance cassette did not adversely affect *in vitro* growth, including biofilm formation, or fermentation of cellobiose (Sections 6.2.3 and 6.2.4). Other putative cellobiose transporters have been annotated in the pneumococcal genomes, including one described by McKessar and Hakenbeck (2007) [PTS 2]. PTS 2 is also present in WCH206 (Section 5.2.1), and this may have compensated for the loss of the mutated one in this study (Section 5.2.1). Having said this, sucrose and glucose appear to be preferential carbohydrate sources (Section 6.2.3). Furthermore, cellobiose has not been established as the true substrate of the PTS described in this study or other putative cellobiose transporters annotated in the pneumococcal genomes.

The role of the island in pathogenesis was explored, in accordance with the fourth specific aim of this project (Section 1.5). Deleting the island in WCH206 resulted in a

longer survival time in the mouse pneumonia-sepsis model, but did not influence survival in the i.p. model (Sections 6.2.5.1 and 6.2.5.3). Although removal of the island did not affect *in vitro* growth, the fitness of WCH206 ΔIsland was also reduced compared to the wild-type in the nasopharynx, ears, lungs and blood in a competition study using the pneumonia-sepsis model (Section 6.3.3). The attenuation in the lungs could have influenced the ability to develop sepsis. However, when a mixture of wild-type and mutant was injected into the peritoneal cavity, which bypassed the requirement to enter the blood via the respiratory tract, the mutant displayed attenuation at 48 h post-challenge indicating that there is an effect in addition to impacting the passage into the blood (Section 6.2.6.2). Decreased fitness in the nasopharynx could also have an influence on the development of opportunistic invasive infections and OM. Decreased fitness in the i.n. model on the surface of the nasopharynx was also seen for MSHR5 ΔIsland, which was constructed using the same method as for WCH206 ΔIsland, but in an avirulent serotype 11A strain (Sections 6.2.1, 6.2.6). Therefore, the island is involved in multiple facets of pneumococcal pathogenesis. This is not surprising when considering the island is widespread and is found in strains with a variety of virulence potential. Mutants of the individual components were also constructed (WCH206 ΔROK, ΔPTS and ΔSulfatase) [Section 6.3.1], but showed a similar level of attenuation to WCH206 ΔIsland in the pneumonia-sepsis competition model [Section 6.3.3]. This suggested there may be some degree of dependence between the components, and the ROK family protein was shown to be involved in the repression of the PTS 1 operon and sulfatase gene (section 6.3.4).

8.1.5 Accessory regions and otitis media

In this study only one accessory region of interest, which contained a putative cellobiose PTS, was identified in ST180 and ST458, and even then there is a large deletion in ST458 isolates not present in the ST180 isolates. This island is present in a wide range of isolates with different virulence potential (with most possessing the full island) and gave a competitive advantage in additional niches apart from the ear. Numerous accessory regions have been identified in signature-tagged mutagenesis (STM) studies using murine invasive disease models, but it has been difficult to identify a specific region that is consistently found in invasive but not carriage isolates (Blomberg *et al.*, 2009; Polissi *et al.*, 1998; Lau *et al.*, 2001; Hava and Camilli, 2002). One region, AR27 (RD8a), has been shown to be preferentially associated with invasive isolates by Obert *et al.* (2006), but Blomberg *et al.* (2009) did not see this preferential association with their own invasive isolates. It is likely therefore, that pinpointing an accessory region that is associated with OM would be difficult as well.

Blomberg *et al.* (2009) found that their isolates with the highest invasive disease potential possessed the most sugar transporters, although their study had limitations due to its reliance on DNA microarray using slides only containing ORFs from TIGR4 and R6. At least some known sugar transporters were not included on the slide, including PTS 1, and others may not have been detected due to sequence variation. Nevertheless, it is an interesting idea and it could be that rather than possessing a single factor required for OM, bacteria need to possess a number of factors for causing OM, and even this may well depend on the particular strain.

The study of accessory regions is certainly an important area of pneumococcal research, although there appears to be no simple answer to the question of why some

strains cause disease and others do not. Two types of accessory regions were identified in this study. The first type is either present or absent in a pneumococcal strain, such as the islands containing PTS 1 and PTS 2. The second type of accessory region is replaced by an alternative accessory region, usually with a similar function. For example, in this study, accessory regions for sucrose and fucose metabolism in the ST180 strain OXC141 were replaced by alternative accessory regions in MSHR17, with genes which were assigned the same functions in MHSR17 and located in the same region of the genome (Sections 7.3.2.5 and 7.3.2.7). Furthermore, there are probably many more variable regions that remain to be discovered, including in well-characterised genomes like D39 and TIGR4. In this study, a 4.6 kb region encoding an alcohol dehydrogenase, transcriptional regulator and transport-related proteins was found to be absent in MSHR17 (Section 7.3.2.6). In its place are genes with similar functions (an aldolase, PTS and a regulator. The 4.6 kb region absent in MSHR17 is present in all 11 strains available through the KEGG website, including D39 and TIGR4, and this is probably why this accessory region was not identified previously. Furthermore, the genes identified appeared to be unique to MSHR17. Although these "variant" accessory regions encode genes with a similar function, they could be important in the context of differences in virulence. For example, the two alternative fucose metabolising accessory regions encode hydrolases, which both target type 2 blood group antigens, but display different substrate specificity (Higgins *et al.*, 2009). As one of the substrates is not found in people with bloodtype O but is expressed on a wider range of cells in people of other bloodtypes, this could have a real impact on virulence. The combination of accessory regions possessed by ST180 may explain why this ST tends to dominate in serotype 3 worldwide, but the combination of accessory regions of ST458 may make it better adapted for the microenvironmental niches in remote Aboriginal communities.

8.1.6 Summary and conclusion

Overall, this study has provided a detailed analysis of potential genetic requirements for OM pathogenesis within serotype 3. The capsule switch study of Section 3.3 clearly demonstrated that there are additional bacterial factors apart from the capsule that are important for fitness in the middle ear. However, serotype 3 bacteria from two unrelated backgrounds (ST180 and ST458), which were correspondingly shown to have numerous differences in the distribution of accessory regions, are both capable of infecting the middle ear. The analysis of isolates from these two STs using DNA microarray analysis, PCR-based subtractive hybridisation and next generation sequencing, identified only one accessory region to be further explored for a role in OM, and even then there was a major difference between the two STs due to a large deletion in ST458. However, the PTS 1 island in an ST180 isolate was shown to confer a competitive advantage on this strain in the ears, as well as in the nasopharynx, lungs and blood. The accumulation of multiple beneficial accessory regions, such as the one identified in this study, rather than the carriage of a single specific region seems likely to be important for disease capability in serotype 3. Furthermore, the repertoire of accessory regions that provide a selective advantage for the pneumococcus could be influenced by the host as well.

8.2 Future Directions

This study has provided some insight into pneumococcal OM but there is still much to be learnt about its pathogenesis. OM is a major health burden in both developed and developing countries and its continued study is of vital public concern.

In Chapter 3, a limited number of isolates were MLST typed. Numerous studies on ST distribution have been conducted in Europe and the USA, but there is very limited information on the distribution in Australia, including remote Aboriginal communities. Therefore, it is important to continue genotyping of pneumococcal clinical isolates, especially those from patients with invasive disease or OM. This is especially important in remote Aboriginal communities due to their high burden of disease, but MLST information obtained from the general population is also needed for comparative purposes.

Another study from Chapter 3 that would be of interest to continue would be the capsule switch work. Although the serotype 11A strain MSHR5 was converted to serotype 3, the reverse was not performed due to time constraints. MSHR5 is avirulent in the i.p. and pneumonia-sepsis model, but colonised better than the serotype 3 capsule donor strain WCH206 in the intranasal model. Therefore, it would be interesting to see what impact switching WCH206 to serotype 11A would have in these models. Serotype 11A is rarely isolated from the ears of Aboriginal children in remote communities and it would be of interest to see if the capsule switch had any effect on ear colonisation in the pneumonia-sepsis model. Although it is clear from the capsule switch study of Chapter 3 that serotype alone is not enough to determine disease profile, it does not mean that capsule serotype does not influence disease. For instance, if the serotype 11A capsule confers less resistance to phagocytosis than another capsular serotype, this could have an impact on virulence as previously stated (Obert *et al.*, 2006).

In the capsule switch study of Chapter 3, as well as the pneumonia-sepsis competition study of Chapter 6, bacteria were isolated from the ears in the pneumonia-sepsis model, in which mice are anaesthetised and then intranasally inoculated. There are also OM models where mice or chinchillas are surgically inoculated directly into the ear.

Chinchillas are not available in Australia, but the mouse model would be useful to establish in the laboratory. It could be used to assess the mutants constructed in Chapter 6, since bacteria were not isolated from the ears from 100% of the mice in the pneumonia/sepsis model and it was not known what influence levels of colonisation in the nasopharynx had. Furthermore, since the mice develop invasive disease and subsequently die, the duration of studies using the pneumonia/sepsis model is restricted to three days.

Despite limitations of the pneumonia/sepsis model, it would be useful for bioluminescence studies. The laboratory has access to a Xenogen IVIS 200 Imaging System, which allows bioluminescence imaging in living animals. Consequently, the colonisation of bacteria in variation murine niches, including the ears, could be visualised in real-time following the introduction of a *lux* plasmid into the bacteria. Unfortunately, this was not achievable during this study due to time constraints. The advantage this system has is that colonisation can be visualised in the same animals over the course of the study, eliminating animal to animal variation. Therefore, it could give a better understanding of the kinetics of colonisation in the various niches. Furthermore, plasmids producing bioluminescence of different wavelength could also conceivably be generated and used in competition studies.

Finally, in this study, it was not possible to construct mutants in the ST458 isolate MSHR17. However, the expression of PTS 1 could be measured *in vivo* by real-time RT-PCR using the pneumonia/sepsis model and investigating the same niches as analysed in the competition study of Section 6.3.3. Expression in WCH206 (ST180) would be investigated in order to compare the two middle ear isolates, and it would also be worthwhile to measure the expression of PTS 2. Finally, in a step to ascertain the true substrate specificity of PTS 1, as well as PTS 2, it would be interesting to measure their

expression in defined media supplemented with various carbohydrates, including cellobiose. Collectively, these experiments should give further insight into the role of particular regions of the pneumococcal genome in pathogenesis.

Appendix

Table A1. Genes from OXC141 that had sequencing gaps under 100 bp or poor alignments in the consensus sequence with MSHR17.

GENE	D39 ID	TIGR4 ID
FOLDING, SORTING, DEGRADATION		
Glutamine synthetase, type 1, GlnA	spd0448	spn0502
SsrA-binding protein, SmpB	spd0863	spn0976
Protease maturation protein, putative	spd0868	spn0981
Peptide deformylase, Def-2	spd1381	spn1549
Histidine kinase TCS11	spd1799	spn2001
Competence protein CglB	spd1862	spn2052
TRANSLATION		
Leucyl-tRNA synthetase, LysS	spd0238	spn0254
Ribosomal small subunit pseudouridine synthase A, RsuA	spd0260	spn0280
Aspartyl/glutamyl-tRNA amidotransferase, subunit B	spd0396	spn0436
Aspartyl/glutamyl-tRNA amidotransferase, subunit A	spd0397	spn0437
Ribonuclease III family protein	spd0516	spn0592
Ribonuclease Z, Rnz	spd0586	spn0674
Methyltransferase, small domian, putative	spd0735	spn0841
Glycyl-tRNA synthetase beta chain, GlyS	spd1304	spn1474
tRNA CCA-pyrophosphorylase (polyA polymerase family protein)	spd1386	spn1554
tRNA pseudoridine synthase A	spd1424	spn1599
Arginyl-tRNA synthetase, ArgS	spd1905	spn2078
CARBOHYDRATE METABOLISM		
Glycosyl transferase, group 2 family protein	spd0139	spn0136

Anaerobic ribonucleoside-triphsopahte reductase activating protein (part)	spd0190	spn0205
Glycosyl hydrolase family protein 31	spd0285	spn0312
Mannose-6-phosphate isomerase, ManA	spd0641	spn0736
6-phosphofructokinase	spd0789	spn0896
Competence protein CelA	spd0843	spn0954
Glycosyl transferase CpoA	spd0960	spn1075
Glycogen biosynthesis protein, GlgD	spd1007	spn1123
Enolase/phosphopyruvate hydralase	spd1012	spn1128
6-phospho-beta-galactosidase	spd1046	spn1184
Phosphomannomutase, Pgm	spd1326	spn1498
Fructokinase, ScrK	spd1531	spn1721
Galactose-1-phosphate uridylyltransferase GalT	spd1633	spn1852
Competence protein CglA	spd1863	spn2053
N-acetylglucosamine-6-phosphate deacetylase, NagA	spd1866	spn2056
L-fucose isomerase	spd1986	spn2158
METABOLISM OF COFACTORS AND VITAMINS		
Dihydropteroate synthase, SulA	spd0269	spn0289
Thymidylate synthase	spd0581	spn0669
Hydroxyethylthiazole kinase, ThiM	spd0623	spn0717
Hydroxylethylthiazole kinase, putative	spd0630	spn0724
MutT/nudix family protein	spd0715	spn0817
5,10-methylene-tetrahydrofolate dehydrogenase (methenyl tetrahydrofolate cyclohydrolase)	spd0721	spn0825
Cysteine desulfurase (aminotransferase class V)	spd0776	spn0880
Thiamine biosynthesis (tRNA modification protein ThiI)	spd0777	spn0881
Nicotinate phopshoribosylransferase, PncB	spd1251	spn1421
Pyridoxine biosynthesis	spd1297	spn1468
Phosphomethylpyrimidine nad kinase (tRNA pseudoridine synthase A)	spd1423	spn1598
7,8-dihydro-8-oxoguanine-triphosphatase (mutT/nudix family protein)	spd1574	spn1783
Phopshopantetheine adenyltransferase, CoaD	spd1766	spn1968
Thiamine pyrophosphokinase	spd1779	spn1982
GLYCAN BIOSYNTHESIS AND METABOLISM		
Phospho-N-acetylmuramoyl-pentapeptide-transferase, MraY	spd0307	spn0337

UDP-N-acetylmuramoyl-L-alanyl-D-glutamate sythetase, MurD	spd0598	spn0688
UDP-N-acetylmurate-alanine ligase, MurC	spd1349	spn1521
UDP-N-acetylmuramyl-tripeptide synthetase, MurE	spd1359	spn1530
Mur ligase protein	spd1416	spn1589
UDP-N=acetylmuramoylalanine-D-glutamylrlysine-D-alanyl, MurF	spd1483	spn1670
N-acetyl-beta-D-glucosaminidase	spd1969	spn2141
α-L-fucosidase FucA, putative	spd1974	spn2146
CELLULAR PROCESSES		
Cell division protein FtsH, putative	spd0013	spn0013
tRNA uridine 5-carboxymethylaminomethyl modification enzyme, GidA	spd0129	spn0120
DNA translocase FtsK	spd0774	spn0878
tRNA uridine 5-carboxymethyl aminomethyl modification enzyme (Gid protein), TrmFO	spd0833	spn0943
XENOBIOTICS BIODEGRADATION AND METABOLSIM		
23s rRNA (uracil-5-)-methyltransferase RumA	spd0914	spn1029
REPLICATION AND REPAIR		
Holliday junction DNA helicase RuvB	spd0241	spn0259
DNA polymerase III, alpha subunit, gram-positive type	spd0254	spn0274
Exinuclease ABC, C subunit, UvrC	spd0538	spn0618
ATP-dependent DNA helicase, PcrA	spd0973	spn1087
ATP-dependent RNA helicase, DEAD/DEAH box family	spd1413	spn1586
Exodeoxyribonulcease, ExoA	spd1626	spn1845
AMINO ACID METABOLISM		
Acetylornithine deacetylase (succinyl-diaminopimelate desuccinylase)	spd0152	spn0150
γ-glutamyl phosphate reductase, ProA	spd0823	spn0932
Pyrroline-5-carboxylate reductase, ProC	spd0824	spn0933
Acetoin dehydrogenase, E1 component, beta subunit, putative	spd1027	spn1163
O-acetylhomoserine sulfhydrylase, CysD	spd1073	spn1214
Glucose-6-phospahte dehydrogenase Zwf	spd1100	spn1243
Glutamate dehydrogenase	spd1158	spn1306
Shikimate kinase	spd1204	spn1370
Shikimate dehydrogenase	spd1210	spn1376
ATP-dependent DNA helicase RecG	spd1507	spn1697

Peptidase family M20/M25/M40	spd1922	spn2096
Arginine deiminase, ArcA	spd1975	spn2148
ENERGY METABOLISM		
Phosphoenolpyruvate carboxylase	spd0953	spn1068
ATP synthase F1, alpha subunit, AtpA	spd1337	spn1510
Mn-dependent inorganic pyrophosphatase PpaC	spd1363	spn1534
NUCLEOTIDE METABOLISM		
Phosphoribosylaminoimidazole carboxylase ATPase subunit, PurK	spd0060	spn0054
Phosphopentomutase, DeoB	spd0724	spn0829
Ribonucleoside-diphosphate reductase beta subunit, NrdF	spd1043	spn1180
Thioredoxin-disulphide reductase, TrxB	spd1287	spn1458
UNASSIGNED		
Phosphorylase Pnp/Udp family	spd0074	spn0075
LysM domain protein	spd0104	spn0107
Metallo-beta-lactamase superfamily	spd0130	spn0121
Glycoprotein endopeptidase	spd0134	spn0127
Transcriptional regulator MutR	spd0144	spn0141
CAAX amino terminal protease family	spd0147	spn0144
Ribonucleotide reductase (NrdI family protein)	spd0160	spn0158
Response regulator, LytR/AlgR family	spd0163	spn0161
Lipoprotein, putative	spd0184	spn0198
Transcirptional regulator	spd0231	spn0247
Alkaline amylopullanase, spuA	spd0250	spn0268
Alcohol dehydrogenase, AdhA/AdhP	spd0265	spn0285
CAAX amino protease family	spd0268	spn0288
Endo alpha-N-acetylgalactosaminidase	spd0335	spn0368
Choline binding protein C, CbpC	spd0345	spn0377
Choline binding protein G (CbpG)	spd0356	spn0390
Macrophage infectivity potentiator protein	spd0373	spn0409
Peptide chain release factor, PrfC	spd0399	spn0439
Membrane protein YdcD	spd0432	spn0482
PAP2 family protein	spd0438	spn0489
Endo-β-N-acetylglucosaminidase, putative	spd0444	spn0498

Heat-inducible transciption repressor HrcA	spd0458	spn0515
BlpT fusion protein	spd0466	spn0524
Bacteriolysin BlpC (peptide pheromone)	spd0470	spn0528
Immunity protein BlpY	spd0473	spn0545
BlpZ, fusion	spd0474	spn0546
Hemerythrin HHE cation binding domain subfamily, putative	spd0488	spn0562
Leucine-rich protein	spd0517	spn0593
ABC transporter membrane-spanning permease - glutamine transport	spd0528	spn0607
Ser/Thr protein phosphatase family (metallophosphoesterase)	spd0539	spn0619
Nitroreductase	spd0541	spn0622
Dipetidase PepV	spd0542	spn0623
Peptidase, M42	spd0547	spn0627
HIT family protein	spd0548	spn0628
Thioredoxin family protein	spd0572	spn0659
Excalibur domain-containing protein	spd0579	spn0667
GTP-binding protein HflX	spd0584	spn0672
Sensor histidine kinase CiaH	spd0702	spn0799
Carbamoylphosphate synthase, large subunit, CarB	spd0781	spn0885
Type I restriction modification system, M subunit	spd0782	spn0886
Type I restriction modification system, S subunit, putative	spd0783	spn0887
Hypothetical (part of big deletion)	-	spn0888
Prophage maintenance system killer protein	-	spn0889
Type I restriction modification system, R subunit	spd0784	spn0892
X-prolyl-dipetidyl-peptidase, PepX	spd0787	spn0894
S1 RNA binding domain protein (putative transcirptional regulator)	spd0802	spn0908
Carboxynor spermidine decarboxylase, NspC	spd0813	spn0920
Tetrapyrrole methylase family protein	spd0828	spn0938
Membrane protein, putative	spd0846	spn0958
Ribonuclease R (exoribonuclease, VacB/Rnb family), Rnr	spd0862	spn0975
Oligoendopeptidase F, PepF	spd0866	spn0979
Peptidase T	spd0894	spn1008
Citrulline cluster-linked protein	spd0898	spn1012
tRNA modification GTPase TrmE	spd0902	spn1016

Peptide chain relase factor, PrfA	spd0906	spn1020
Zeta toxin, putative	spd0931	spn1051
Pullulanase, type I, putative	spd1002	spn1118
Recombinase helicase AddA/ exonuclease RexA	spd1016	spn1152
Ribosomal biogenesis GTPase RbgA (YlqF)	spd1019	spn1155
VicX protein	spd1083	spn1225
NTP pyrophosphohydrolase	spd1092	spn1235
Cof family protein	spd1103	spn1246
CTP:phosphocholine cytidylyl transferase LicC	spd1123	spn1267
LicB	spd1124	spn1268
Alcohol dehydrogenase, Zn-dependent, putative	spd1126	spn1270
Chromosome condensation/camphor resistance protein, CrcB	spd1150	spn1295
NADPH-depedemt FMN reductase	spd1301	spn1471
Oxidoreductase, putative	spd1302	spn1472
Bacteriocin transport accessory protein	spd1327	spn1499
Oxidoreductase (NADPH-dependent FMN reductase)	spd1375	spn1546
Cation efflux family protein	spd1384	spn1552
ATP-dependentClp protease, ATP-binding subunit ClpX	spd1399	spn1569
GTP-sensing transcriptional pleiotropic repressor CodY	spd1412	spn1584
Oxidoredustase, DadA family	spd1433	spn1608
Cation transport ATPase (E1-E2 family protein)	spd1436	spn1623
Cadmium resistance transporter CadD, putative	spd1438	spn1625
YGGT family protein, putative	spd1476	spn1663
Nitroreductase family protein	spd1520	spn1710
Primosomal protein DnaI	spd1521	spn1711
3 OH-3methylglutaryl CoA reductase	spd1536	spn1726
Membrane protein	spd1540	spn1730
7,8-dihydro-8-oxoguanine-triphosphatase	spd1574	spn1783
Peptidase, M50 family protein	spd1575	spn1784
LacI family sugar-binding transcriptional regulator	spd1605	spn1821
Membrane protein, putative (squalene desaturase in other species)	spd1717	spn1914
ABC transporter (transcriptional regulator MarR family in Hungary 19A_6)	spd1721	spn1918
Type II restriction modification system regulatory protein	-	spn1936

DNA-entry nuclease, putative	spd1762	spn1964
Cmp-binding factor , Cbf1	spd1777	spn1980
Bacteriocin-associated integral membarne protein	spd1785	spn1988
Dicarboxylate carrier protein	spd1827	spn2017
Pyrrolidone-carboxylate peptidase	spd1870	spn2060
Transcriptional regulator, MarR family	spd1873	-
LysM domain protein	spd1874	spn2063
MATE efflux family protein	spd1876	spn2065
ROK family protein	spd1970	spn2142
Cell wall surface anchor family (putative α-1,2-mannosidase)	spd1973	spn2145
Major facilitator family transport protein	spd2007	-
PspC	spd2017	spn2190
MEMBRANE TRANSPORT		
PTS system, N-acetylgalactosamine-specific IIA	spd0293	spn0321
PTS system, N-acetylgalactosamine-specific IIB	spd0295	spn0323
Glutamine transport ATP-binding protein GlnQ	spd0531	spn0610
Sodium/hydrogen exchanger protein	spd0569	spn0655
MATE efflux family protein	spd1029	spn1166
Membrane protein (sodium:dicarboxylate symporter family protein, putative	spd1563	spn1753
SIGNAL TRANSDUCTION		
Sensor histidine kinase TCS07	spd0157	spn0155
Response regulator TCS09	spd0574	spn0661
Anthranilate synthase component I	spd1602	spn1817

Genes have been grouped based on their assigned pathway category colours in KEGG (Table 7.1).

Table A2. Genes from OXC141 that had sequencing gaps under 100 bp or poor alignments in the consensus sequence with WU2.

GENE	D39 ID	TIGR4 ID (G54 ID)
BIOSYNTHESIS OF POLYKETIDES AND NONRIBOSOMAL PROTEINS		
Transketolase, RecP-1	-	spn1615
FOLDING, SORTING AND DEGRADATION		
Signal peptidase I, LepB	spd0367	spn0402
Signal peptidase (Spase) II, LpsA	spd0819	spn0928
Cell division protein FtsY	spd1101	spn1244
SpoIIIJ family protein or mebrane protein OxaA2	spd1773	spn1975
Pyruvate formate-lyase 1-activiating enzyme, PflA	spd1774	spn1976
TRANSLATION		
tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase, TrmU	spd0127	spn0118
Ribosomal protein s19, RpsS	spd0197	spn0213
Leucyl-tRNA synthetase, LysS	spd0238	spn0254
Prolyl-tRNA synthetase, ProS	spd0246	spn0264
Ribosomal small subunit pseudouridine synthase A, RsuA	spd0260	spn0280
Ribosome-binding factor A, RbfA	spd0483	spn0557
Phenylalanine tRNA synthetase beta subunit, PheT	spd0506	spn0581
Phenylalanyl-tRNA synthetase, beta subunit	spd0506	spn0581
Cysteinyl-tRNA synthetase, CysS	spd0515	spn0591
tRNA delta(2)-isopentenyl pyrophosphate transferase, MiaA	spd0583	spn0671
tRNA (quanine-N1-)methyltransferase, TrmD	spd0679	spn0779
Methyltransferase, small domian, putative	spd0735	spn0841
Sua5/YciO/YrdC/YwlC family protein	spd0908	spn1022
Alanyl-tRNA synthetase, AlaS	spd1216	spn1383
S-adenosylmethionine:tRNA ribosyltransferase-isomerase, QueA	spd1247	spn1416
SpoU rRNA methylase family, YjfA	spd1771	spn1973
Glutamyl-tRNA synthetase, GltX	spd1896	spn2069
Arginyl-tRNA synthetase, ArgS	spd1905	spn2078

Aspartyl-tRNA synthetase, AspS	spd1941	spn2114
Tryptophanyl-tRNA synthetase, TrpS	spd2056	spn2229
CARBOHYDRATE METABOLISM		
Aldose-1-epimerase, GalM	spd0071	spn0066
Glycosyl transferase, group 1 family protein	spd0138	spn0133
Glycosyl transferase, group 2 family protein	spd0139	spn0136
Anaerobic ribonucleoside-triphsopahte reductase activating protein (part)	spd0190	spn0205
Glycosyl hydrolase family protein 31	spd0285	spn0312
6-phospho-beta-glucosidase	spd0503	spn0578
Glucose kinase, Gki	spd0580	spn0668
Mannose-6-phosphate isomerase, ManA	spd0641	spn0736
Competence protein CelA	spd0843	spn0954
Competence protein CelB	spd0844	spn0955
Glycogen biosynthesis protein, GlgD	spd1007	spn1123
6-phospho-beta-galactosidase	spd1046	spn1184
Tagatose 1,6-diphosphate aldolase, LacD	spd1050	spn1190
Phosphomannomutase, Pgm	spd1326	spn1498
Fructokinase, ScrK	spd1531	spn1721
Sucrose-6-phopshate hydrolase	spd1582	spn1795
Galactokinase, GalK	spd1634	spn1853
Competence protein CglA	spd1863	spn2053
N-acetylglucosamine-6-phosphate deacetylase, NagA	spd1866	spn2056
CELLULAR PROCESSES		
tRNA uridine 5-carboxymethylaminomethyl modification enzyme, GidA	spd0129	spn0120
Rod shape determining protein RodA (cell cycle protein, FtsW/RodA/SpoVE family)	spd0706	spn0803
tRNA uridine 5-carboxymethyl aminomethyl modification enzyme (Gid protein), TrmFO	spd0833	spn0943
Cell division protein FtsW, putative	spd0952	spn1067
Chromosome segregation protein Smc, putative	spd1104	spn1247
Jag protein (SpoIIIJ-associated), putative	spd1849	spn2040
LIPID METABOLISM		
Cardiolipin synthase	spd0185	spn0199

Phosphomevalonate kinase, MvaK2	spd0348	spn0383
Acyltransferase family protein	spd1437	spn1624
(NAD(P)+) glycerol-3-phosphate dehydrogenase	spd1918	spn2091
α-glycerophosphate oxidase (glycerol-3-phosphateoxidase), GlpO	spd2012	spn2185
Glycerol kinase, GlpK	spd2013	spn2186
METABOLISM OF COFACTORS AND VITAMINS		
β-galactosdiase, BgaC	spd0065	spn0060
Riboflavin biosynthesis protein, RibD	spd0169	spn0178
Dihydrofolate synthetase FolC, putative	spd0183	spn0197
5,10-methylenetetrahydrofolate reductase, MetF	spd0511	spn0586
Chorismate binding enzyme (para-aminobenzoate synthase)	spd0578	spn0665
Thymidylate synthase	spd0581	spn0669
Hydroxylethylthiazole kinase, putative	spd0630	spn0724
Phosphomethylpyrimidine kinase, ThiD	spd0632	spn0726
5,10-methylene-tetrahydrofolate dehydrogenase (methenyl tetrahydrofolate cyclohydrolase)	spd0721	spn0825
Pantothenate kinase, CoaA	spd0733	spn0839
Methyltransferase, HemK family	spd0907	spn1021
Serine hydroxymethyltransferase, GlyA	spd0910	spn1024
Riboflavin kinase/flavin adenine dinucleotide synthase (RibF, RibC)	spd0994	spn1110
NH(3)-dependent NAD(+) synthetase, NadE	spd1250	spn1420
7,8-dihydro-8-oxoguanine-triphosphatase (mutT/nudix family protein)	spd1574	spn1783
Phopshopantetheine adenyltransferase, CoaD	spd1766	spn1968
Threonine synthase, ThrC	spd1877	spn2066
GLYCAN BIOSYNTHESIS AND METABOLISM		
β-N-acetylhexosaminidase	spd0063	spn0057
UDP-N-acetylmuramoyl-L-alanyl-D-glutamate sythetase, MurD	spd0598	spn0688
UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)pyrophosphoryl-undecaprenol N-acetylglucosamine transferase, MurG	spd0599	spn0689
UDP-N-acetylmuramyl-tripeptide synthetase, MurE	spd1359	spn1530
Neuraminidase A, NanA	spd1504	spn1693
N-acetyl-beta-D-glucosaminidase	spd1969	spn2141

XENOBIOTICS DEGRADATION AND METABOLISM		
23s rRNA (uracil-5-)-methyltransferase RumA	spd0914	spn1029
NOL1/NOP2/sun family	spd1233	spn1402
Acylphosphatase, AcyP	spd1772	spn1974
REPLICATION AND REPAIR		
Transcription-repair coupling factor, Mfd	spd0006	spn0006
DNA polymerase III, alpha subunit, gram-positive type	spd0254	spn0274
DNA polymerase III, alpha subunit, gram-positive type	spd0254	spn0274
Ribonuclease H III, RnhC	spd0368	spn0403
Single-stranded-DNA-specific exonuclease RecJ	spd0532	spn0611
Exinuclease ABC, C subunit, UvrC	spd0538	spn0618
DNA polymerase III, delta subunit	spd0666	spn0765
DNA polymerase III, alpha subunit, DnaE	spd0788	spn0895
Formamidopyrimidine-DNA glycosylase, MutM	spd0858	spn0970
DNA primase, DnaG	spd0957	spn1072
ATP-dependent DNA helicase, PcrA	spd0973	spn1087
NAD-dependent DNA ligase, LigA	spd1001	spn1117
Endonuclease III, Nth	spd1135	spn1279
ATP-dependent RNA helicase, DEAD/DEAH box family	spd1413	spn1586
ATPase, AAA family, RarA	spd1580	spn1790
Exodeoxyribonulcease, ExoA	spd1626	spn1845
Competence induced protien Ccs50 (RmuC family protein)	spd1778	spn1981
Primase homolog, putative	spd1787	spn1990
AMINO ACID METABOLISM		
Adenylsuccinate synthetase, PurA	spd0024	spn0019
L-serine dehydratase, iron-sulfur-dependent, alpha subunit	spd0102	spn0105
Argininosuccinate synthase, ArgG	spd0110	spn0113
Argininosuccinate lyase, ArgH	spd0111	-
Glutamine-fructose-6-phosphate transaminase (isomerizing), GlmS	spd0248	spn0266
5-methyltetrahydropteroyl triglutamate--homocysteine methyltransferase	spd0510	spn0585
S-adenosylmethionine synthetase, central domain, MetK	spd0664	spn0762
Glutathione-disulfide reductase, Gor	spd0685	spn0784
Spermidine synthase, SpeE	spd0811	spn0918

γ-glutamyl phosphate reductase, ProA	spd0823	spn0932
Pyrroline-5-carboxylate reductase, ProC	spd0824	spn0933
Aspartate-semialdehyde dehydrogenase, Asd	spd0900	spn1013
Glycerate kinase	spd1011	spn1126
O-acetylhomoserine sulfhydrylase, CysD	spd1073	spn1214
Carbamoyl-phosphate synthase, large subunit	spd1131	spn1275
Glutamate dehydrogenase	spd1158	spn1306
Homoserine kinase, ThrB	spd1194	spn1360
ATP-dependent DNA helicase RecG	spd1507	spn1697
Alanine racemase, Alr	spd1508	spn1698
2,3,4,5-tetrahydropyridine-2-craboxylate aminotransferase	spd1923	spn2097
Arginine deiminase, ArcA	spd1975	spn2148
D-alanine-poly(phosphoribitol) ligase subunit 1, DltA	spd2005	spn2176
ENERGY METABOLISM		
Cation-transporting ATPase, E1-E2 family protein	spd0635	spn0729
Ribose-5-phosphate isomerase A	spd0723	spn0828
ATP synthase F1, alpha subunit, AtpA	spd1337	spn1510
NUCLEOTIDE METABOLISM		
Phosphoribosyl aminoimidazole-succinocarboxamide synthase	spd0051	spn0044
Phospohribosyl formylglycinamide synthetase, putative	spd0052	spn0045
Phosphoribosylamine--glycine ligase, PurR	spd0058	spn0051
Phosphoribosylaminoimidazole carboxylase catalytic subunit, PurE	spd0059	spn0053
Adenylate kinase, Adk	spd0214	spn0231
Uracil phosphoribosyltransferase, Upp	spd0649	spn0745
Phosphopentomutase, DeoB	spd0724	spn0829
Pyrimidine-nucleoside phosphorylase, Pdp	spd0736	spn0842
Pyruvate kinase Pyk, putative	spd0790	spn0897
Thymidylate kinase, Tdk	spd0825	spn0935
Dihydroorotase, PyrC	spd1030	spn1167
Cytidylate kinase, Cmk	spd1428	spn1603
TRANSCRIPTION		
Transcription antiterminator protein NusB	spd0393	spn0433

UNASSIGNED		
Chromosomal replication initiator protein, DnaA	spd0001	spn0001
DNA repair protein RadA	spd0029	-
Fatty acid/phospholipid synthesis protein, PlsX	spd0043	spn0037
Competence protein, ComB	spd0050	spn0043
Transcriptional regulator, GntR family	spd0064	spn0058
Sugar isoamerase domain protein A, AgaS	spd0070	spn0065
Phosphorylase Pnp/Udp family	spd0074	spn0075
Cell wall surface anchor protein	spd0080	spn0082
Bacteriocin, putative	spd0106	spn0109
ABC transporter membrane spanning permease- amino acid transport	spd0107	spn0110
Macrolide export ATP-binding/permease protein MacB	spd0108	spn0111
ABC transporter solute binding protein- amino acid transport	spd0109	spn0112
Metallo-β-lactamase superfamily	spd0130	spn0121
Competence induced protein, Ccs1	-	spn0123
Transcriptional regulator MutR	spd0144	spn0141
Ribonucleotide reductase (NrdI family protein)	spd0160	spn0158
Alkaline amylopullanase, spuA	spd0250	spn0268
DNA inducible protein	spd0256	spn0276
Alcohol dehydrogenase, AdhA/AdhP	spd0265	spn0285
Xanthine/uracil permease family protein	spd0267	spn0287
Glucan 1, 6 alpha glucosidase, DexB	spd0311	spn0342
Endo-α-N-acetylgalactosaminidase	spd0335	spn0368
Response regulator	spd0344	spn0376
Histidine kinase TCS03	spd0351	spn0386
Response regulator TCS03	spd0352	spn0387
Choline binding protein G (CbpG)	spd0356	spn0390
Macrophage infectivity potentiator protein	spd0373	spn0409
Dihydroxyacetone kinase family protein	spd0403	spn0443
DNA-damage inducible protein P/ DNA polymerase IV	spd0419	spn0458
Membrane protein YdcD	spd0432	spn0482
Endo-β-N-acetylglucosaminidase, putative	spd0444	spn0498

Chaperone protein DnaJ	spd0461	spn0519
BlpT fusion protein	spd0466	spn0524
Bacteriolysin BlpC (peptide pheromone)	spd0470	spn0528
Immunity protein BlpY	spd0473	spn0545
Translation initiation factor IF-2	spd0482	spn0556
Hemerythrin HHE cation binding domain subfamily, putative	spd0488	spn0562
Dipetidase PepV	spd0542	spn0623
Peptidase, M42	spd0547	spn0627
D-alanyl-D-alanine carboxypeptidase	spd0549	spn0629
Zinc metalloprotease ZmpB	spd0577	spn0664
Elongation factor Tu family	spd0593	spn0681
Lactate oxidase, LctO	spd0621	spn0715
ATP-dependent Clp protease proteolytic subunit, ClpP	spd0650	spn0746
Metal dependent hydrolase	spd0662	spn0760
Hydrolase, alpha/beta fold family, putative	spd0677	spn0777
Biotin synthase/BioY family, putative	spd0684	spn0783
Protease	spd0704	spn0801
DNA polymerase III, epsilon subunit (ATP-depenent helicase DinG)	spd0705	spn0802
ATP dependnet Clp protease, ATP-binding subunit, ClpE	spd0717	spn0820
YjeF homolog, C-terminus, putative	spd0722	-
Hemolysin related protein	spd0729	spn0834
DNA topology modulation protein FlaR, putative	spd0731	spn0837
Topoisomerase IV, subunit B, ParE	spd0746	spn0852
FeS assembly protein, SufD	spd0763	spn0868
Etserase superfamily, putative	spd0778	spn0882
Carbamoylphosphate synthase, large subunit, CarB	spd0781	spn0885
Type I restriction modification system, M subunit	spd0782	spn0886
Type I restriction modification system, S subunit, putative	spd0783	spn0887
Type I restriction modification system, R subunit	spd0784	spn0892
Saccharopine dehydrogenase, Lys	spd0812	spn0919
N-carbomoylputrescine amidase, AguB	spd0815	spn0922
Cof family protein	spd0816	spn0923
Membrane protein, putative	spd0846	spn0958

Dihdroorotate dehydrogenase electron transfer subunit, PyrK	spd0851	spn0963
GTP-binding protein Era	spd0857	spn0969
Ribonuclease R (exoribonuclease, VacB/Rnb family), Rnr	spd0862	spn0975
O-methyltransferase	spd0867	spn0980
Histidine triad protein E, PhtE	spd0892	spn1005
tRNA modification GTPase TrmE	spd0902	spn1016
Zeta toxin, putative	spd0931	spn1051
CBS domain containing cytosolic protein	spd0969	spn1083
Hydrolase, haloacid dehalogenase-like family	spd1034	spn1171
Transcriptional regulator, biotin repressor family	spd1093	spn1234
MutT/nudix family	spd1094	spn1236
Cof family protein	spd1102	spn1245
Cof family protein	spd1103	spn1246
LicB-family membrane protein	spd1124	spn1268
LemA family protein (cytoplasmic membrane protein)	spd1139	spn1284
Cof family protein (phosphatase YidA)	spd1146	spn1291
Chromosome condensation/camphor resistance protein, CrcB	spd1150	spn1295
DHH subfamily 1 protein	spd1153	spn1298
Chlorohydrolase/amidohydrolase family/ Atz?Trz family protein	spd1190	spn1356
Phosphotransferase LicD3	spd1201	spn1367
Psr protein (transcriptional regulator, LytR family)	spd1202	spn1368
ATP-dependent RNA helicase, DEAD/DEAH box family protein	spd1312	spn1483
SNF2 family protein (helicase)	spd1351	spn1523
Cof family protein (peptidyl-prolyl cis-trans isomerase, cyclophilin type)	spd1367	spn1538
Apsartate aminotransferase, ApsC	spd1373	spn1544
DegV family protein	spd1388	spn1557
Methyltransferase, putative	spd1408	spn1578
Pyridine nucleotide-disulfide oxidoreductase, MerA	spd1415	spn1588
SAM-dependent methyltransferase (Bcl-2 family protein)	spd1435	spn1610
Nitroreductase family protein	spd1520	spn1710
HAD family phosphatase, YqeG	spd1560	spn1750
Oligoendopeptidase F	spd1571	spn1780
7,8-dihydro-8-oxoguanine-triphosphatase	spd1574	spn1783

Transcritpional regulator, TetR family	spd1639	spn1858
Glutamyl aminopeptidase, PepA	spd1647	spn1865
RecA regulator RecX	spd1705	spn1902
Oxidoreductase, short chain dehydrogenase/reductase family	spd1712	spn1909
Thioredoxin family protein	spd1714	spn1911
Membrane protein, putative (squalene desaturase in other species)	spd1717	spn1914
MATE efflux family protein DinF	spd1738	spn1938
Cmp-binding factor , Cbf1	spd1777	spn1980
Bacteriocin-associated integral membarne protein	spd1785	spn1988
Acyltransferase family protein	spd1867	spn2057
Pyrrolidone-carboxylate peptidase	spd1870	spn2060
LysM domain protein	spd1874	spn2063
MATE efflux family protein	spd1876	spn2065
Arginine repressor, ArgR	spd1904	spn2077
Transcriptional regulator	spd1916	spn2090
Twin-arginine translocation pathway signal protein	spd1972	spn2144
Cell wall surface anchor family(putative α-1,2-mannosidase)	spd1973	spn2145
Choline binding protein D, CbpD	spd2028	spn2201
MEMBRANE TRANSPORT		
Competence protein, ComA	spd0049	spn0042
CorA like magnesium transporter protein	spd0175	spn0185
Sodium:alanine symporter family protein	spd0372	spn0408
Trk family potassium uptake	spd0429	spn0479
Chaperone protein DnaK	spd0460	spn0517
ABC transporter membrane-spanning permease - glutamine transport	spd0529	spn0608
ABC transporter substrate-binding protein - glutamine transport/major cell binding factor	spd0530	spn0609
Glutamine transport ATP-binding protein GlnQ	spd0531	spn0610
Sodium/hydrogen exchanger protein	spd0569	spn0655
Sodium-dependent transporter	spd0642	spn0641
PTS system, II ABC, Exp5	spd0661	spn0758
Fructose specific PTS system IIABC	spd0773	spn0877
PTS system II ABC	spd1047	spn1185

Drug efflux ABC transporter, ATP-binding/permease protein	spd1177	spn1342
Membrane protein (sodium:dicarboxylate symporter family protein, putative	spd1563	spn1753
Nicotinamide mononucleotide transporter PnuC, putative	spd1640	spn1859
Choline transporter, ProWX	spd1642	spn1860
SIGNAL TRANSDUCTION		
Histidine kinase TCS08	spd0082	spn0084
Sensor histidine kinase TCS07	spd0157	spn0155
Sensor histidine kinase TCS09	spd0575	spn0662
Anthranilate synthase component I	spd1602	spn1817
Competence protein, ComD	spd2064	spn2236

Genes have been grouped based on their assigned pathway category colours in KEGG (Table 7.1).

Bibliography

Aaberge, I. S., Eng, J., Lermark, G. and Lovik, M. (1995) Virulence of *Streptococcus pneumoniae* in mice: a standardized method for preparation and frozen storage of the experimental bacterial inoculum. *Microb. Pathog.*, **18:** 141-152.

Adhin, M. R. and van Duin, J. (1990) Scanning model for translational reinitiation in eubacteria. *J. Mo.l Biol.,* **213:** 811-818.

Agron, P. G., Macht, M., Radnedge, L., Skowronski, E. W., Miller, W. and Andersen, G. L. (2002) Use of subtractive hybridization for comprehensive surveys of prokaryotic genome differences. *FEMS Microbiol. Lett.*, **211:** 175-182.

Akopyants, N. S., Fradkov, A., Diatchenko, L., Hill, J. E., Siebert, P. D., Lukyanov, S. A., Sverdlov, E. D. and Berg, D. E. (1998) PCR-based subtractive hybridization and differences in gene content among strains of *Helicobacter pylori*. *Proc. Natl. Acad. Sci. U. S. A.*, **95:** 13108-13113.

Alanee, S. R., McGee, L., Jackson, D., Chiou, C. C., Feldman, C., Morris, A. J., Ortqvist, A., Rello, J., Luna, C. M., Baddour, L. M., Ip, M., Yu, V. L. and Klugman, K. P. (2007) Association of serotypes of *Streptococcus pneumoniae* with disease severity and outcome in adults: an international study. *Clin. Infect. Dis.*, **45:** 46-51.

Allegrucci, M., Hu, F. Z., Shen, K., Hayes, J., Ehrlich, G. D., Post, J. C. and Sauer, K. (2006) Phenotypic characterization of *Streptococcus pneumoniae* biofilm development. *J. Bacteriol.*, **188:** 2325-2335.

Allegrucci, M. and Sauer, K. (2007) Characterization of colony morphology variants isolated from *Streptococcus pneumoniae* biofilms. *J. Bacteriol.*, **189:** 2030-2038.

Allegrucci, M. and Sauer, K. (2008) Formation of *Streptococcus pneumoniae* non-phase-variable colony variants is due to increased mutation frequency present under biofilm growth conditions. *J. Bacteriol.*, **190:** 6330-6339.

Alloway, J. L. (1932) The transformation *in vitro* of R pneumococci into S forms of different specific types by the use of filtered pneumococcus extracts. *J. Exp. Med*., 55, 91-99.

Ardanuy, C., Fenoll, A., Berron, S., Calatayud, L. and Linares, J. (2006) Increase of the M phenotype among erythromycin-resistant *Streptococcus pneumoniae* isolates from Spain related to the serotype 14 variant of the Spain9V-3 clone. *Antimicrob. Agents Chemother.*, **50:** 3162-3165.

Ardanuy, C., Tubau, F., Pallares, R., Calatayud, L., Dominguez, M. A., Rolo, D., Grau, I., Martin, R. and Linares, J. (2009) Epidemiology of invasive pneumococcal disease among adult patients in Barcelona before and after pediatric 7-valent pneumococcal conjugate vaccine introduction, 1997-2007. *Clin. Infect. Dis.*, **48:** 57-64.

Armstrong, R. R. (1931) A swift and simple method for deciding pneumococcal "type". *Br. Med. J.*, **1:** 214-215.

Austrian, R. (1981) Pneumococcus: the first one hundred years. *Rev. Infect. Dis.*, **3:** 183-189.

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 deoxyribonucleic acid fraction isolated from *pneumococcus* type III. *J. Exp. Med.*, **79:** 137-158.

Bautz, E. K. and Reilly, E. (1966) Gene-specific messenger RNA: isolation by the deletion method. *Science*, **151:** 328-330.

Beall, B., McEllistrem, M. C., Gertz, R. E., Jr., Wedel, S., Boxrud, D. J., Gonzalez, A. L., Medina, M. J., Pai, R., Thompson, T. A., Harrison, L. H., McGee, L. and Whitney, C. G. (2006) Pre- and postvaccination clonal compositions of invasive pneumococcal serotypes for isolates collected in the United States in 1999, 2001, and 2002. *J. Clin. Microbiol.*, **44:** 999-1017.

Bennett, K. E., Haggard, M. P., Silva, P. A. and Stewart, I. A. (2001) Behaviour and developmental effects of otitis media with effusion into the teens. *Arch. Dis. Child.*, **85:** 91-95.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Keira Cheetham, R., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., Irving, L. J., Karbelashvili, M. S., Kirk, S. M., Li, H., Liu, X., Maisinger, K. S., Murray, L. J., Obradovic, B., Ost, T., Parkinson, M. L., Pratt, M. R., Rasolonjatovo, I. M., Reed, M. T., Rigatti, R., Rodighiero, C., Ross, M. T., Sabot, A., Sankar, S. V., Scally, A., Schroth, G. P., Smith, M. E., Smith, V. P., Spiridou, A., Torrance, P. E., Tzonev, S. S., Vermaas, E. H., Walter, K., Wu, X., Zhang, L., Alam, M. D., Anastasi, C., Aniebo, I. C., Bailey, D. M., Bancarz, I. R., Banerjee, S., Barbour, S. G., Baybayan, P. A., Benoit, V. A., Benson, K. F., Bevis, C., Black, P. J., Boodhun, A., Brennan, J. S., Bridgham, J. A., Brown, R. C., Brown, A. A., Buermann, D. H., Bundu, A. A., Burrows, J. C., Carter, N. P., Castillo, N., Chiara, E. C. M., Chang, S., Neil Cooley, R., Crake, N. R., Dada, O. O., Diakoumakos, K. D., Dominguez-Fernandez, B., Earnshaw, D. J., Egbujor, U. C., Elmore, D. W., Etchin, S. S., Ewan, M. R., Fedurco, M., Fraser, L. J., Fuentes Fajardo, K. V., Scott Furey, W., George, D., Gietzen, K. J., Goddard, C. P., Golda, G. S., Granieri, P. A., Green, D. E., Gustafson, D. L., Hansen, N. F., Harnish, K., Haudenschild, C. D., Heyer, N. I., Hims, M. M., Ho, J. T., Horgan, A. M., *et al.* (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, **456:** 53-59.

Berman, S. (1995) Otitis media in developing countries. *Pediatrics*, **96:** 126-131.

Beuzon, C. R., Meresse, S., Unsworth, K. E., Ruiz-Albert, J., Garvis, S., Waterman, S. R., Ryder, T. A., Boucrot, E. and Holden, D. W. (2000) *Salmonella* maintains the integrity of its intracellular vacuole through the action of SifA. *EMBO J.*, **19:** 3235-3249.

Black, S., Shinefield, H., Fireman, B., Lewis, E., Ray, P., Hansen, J. R., Elvin, L., Ensor, K. M., Hackell, J., Siber, G., Malinoski, F., Madore, D., Chang, I., Kohberger, R., Watson, W., Austrian, R. and Edwards, K. (2000) Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children. Northern California Kaiser Permanente Vaccine Study Center Group. *Pediatr. Infect. Dis. J.*, **19:** 187-195.

Blomberg, C., Dagerhamn, J., Dahlberg, S., Browall, S., Fernebro, J., Albiger, B., Morfeldt, E., Normark, S. and Henriques-Normark, B. (2009) Pattern of accessory regions and invasive disease potential in *Streptococcus pneumoniae*. *J. Infect. Dis.*, **199:** 1032-1042.

Bogaert, D., Veenhoven, R. H., Sluijter, M., Wannet, W. J., Rijkers, G. T., Mitchell, T. J., Clarke, S. C., Goessens, W. H., Schilder, A. G., Sanders, E. A., de Groot, R. and Hermans, P. W. (2005) Molecular epidemiology of pneumococcal colonization in response to pneumococcal conjugate vaccination in children with recurrent acute otitis media. *J. Clin. Microbiol.* , **43:** 74-83.

Bozdogan, B., Bogdanovich, T., Kosowska, K., Jacobs, M. R. and Appelbaum, P. C. (2004) Macrolide resistance in *Streptococcus pneumoniae*: clonality and mechanisms of resistance in 24 countries. *Curr. Drug Targets Infect. Disord.,* **4:** 169-176.

Brandileone, M. C., Andrade, A. L., Teles, E. M., Zanella, R. C., Yara, T. I., Di Fabio, J. L. and Hollingshead, S. K. (2004) Typing of pneumococcal surface protein A (PspA) in *Streptococcus pneumoniae* isolated during epidemiological surveillance in Brazil: towards novel pneumococcal protein vaccines. *Vaccine*, **22:** 3890-3896.

Bricker, A. L. and Camilli, A. (1999). Transformation of a type 4 encapsulated strain of *Streptococcus pneumoniae*. *FEMS Microbiol. Lett.*, **172:** 131-5.

Briles, D. E., Nahm, M., Schroer, K., Davie, J., Baker, P., Kearney, J. and Barletta, R. (1981) Antiphosphocholine antibodies found in normal mouse serum are protective against intravenous infection with type 3 *Streptococcus pneumoniae*. *J. Exp. Med.*, **153:** 694-705.

Briles, D. E., Tart, R. C., Swiatlo, E., Dillard, J. P., Smith, P., Benton, K. A., Ralph, B. A., Brooks-Walter, A., Crain, M. J., Hollingshead, S. K. and McDaniel, L. S. (1998) Pneumococcal diversity: considerations for new vaccine strategies with emphasis on pneumococcal surface protein A (PspA). *Clin. Microbiol. Rev.*, **11:** 645-657.

Brown, J. S., Gilliland, S. M., Spratt, B. G. and Holden, D. W. (2004) A locus contained within a variable region of pneumococcal pathogenicity island 1 contributes to virulence in mice. *Infect. Immun.*, **72:** 1587-1593.

Bruckner, R., Nuhn, M., Reichmann, P., Weber, B. and Hakenbeck, R. (2004) Mosaic genes and mosaic chromosomes-genomic variation in *Streptococcus pneumoniae*. Int *J. Med. Microbiol.*, **294:** 157-168.

Brueggemann, A. B., Griffiths, D. T., Meats, E., Peto, T., Crook, D. W. and Spratt, B. G. (2003) Clonal relationships between invasive and carriage *Streptococcus pneumoniae* and serotype- and clone-specific differences in invasive disease potential. *J. Infect. Dis.*, **187:** 1424-1432.

Brueggemann, A. B. and Spratt, B. G. (2003) Geographic distribution and clonal diversity of *Streptococcus pneumoniae* serotype 1 isolates. *J. Clin. Microbio.*, **41:** 4966-4970.

Brueggemann, A. B., Pai, R., Crook, D. W. and Beall, B. (2007) Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. *PLoS Pathog*., **3:** e168.

Burnaugh, A. M., Frantz, L. J. and King, S. J. (2008) Growth of *Streptococcus pneumoniae* on human glycoconjugates is dependent upon the sequential activity of bacterial exoglycosidases. *J. Bacteriol.*, **190:** 221-230.

Butler, J. C. (2004) Epidemiology of pneumococcal disease. In *The Pneumococcus* (Ed, Tuomanen, E. I.) ASM Press, Washington D.C., pp. 148-168.

Cartee, R. T., Forsee, W. T. and Yother, J. (2005) Initiation and synthesis of the *Streptococcus pneumoniae* type 3 capsule on a phosphatidylglycerol membrane anchor. *J. Bacteriol.*, **187:** 4470-4479.

Cerca, N., Jefferson, K. K., Oliveira, R., Pier, G. B. and Azeredo, J. (2006) Comparative antibody-mediated phagocytosis of *Staphylococcus epidermidis* cells grown in a biofilm or in the planktonic state. *Infect. Immun.*, **74:** 4849-4855.

Chen, Y. Y., Yao, S. M., Chou, C. Y., Chang, Y. C., Shen, P. W., Huang, C. T., Su, H. P. and Li, S. Y. (2006) Surveillance of invasive *Streptococcus pneumoniae* in Taiwan, 2002-2003. *J. Med. Microbiol.*, **55:** 1109-1114.

Chen, H., Ma, Y., Yang, J., O'Brien, C. J., Lee, S. L., Mazurkiewicz, J. E., Haataja, S., Yan, J. H., Gao, G. F. and Zhang, J. R. (2007) Genetic requirement for pneumococcal ear infection. *PLoS ONE*, **3:** e2950.

Chen, C. J., Lin, C. L., Chen, Y. C., Wang, C. W., Chiu, C. H., Lin, T. Y. and Huang, Y. C. (2009) Host and microbiologic factors associated with mortality in Taiwanese children with invasive pneumococcal diseases, 2001 to 2006. *Diagn. Microbiol. Infect. Dis.*, **63:** 194-200.

Coates, H. L., Morris, P. S., Leach, A. J. and Couzos, S. (2002) Otitis media in Aboriginal children: tackling a major health problem. *Med .J. Aust.*, **177:** 177-178.

Costerton, J. W. (1999) Introduction to biofilm. *Int. J. Antimicrob. Agents*, **11:** 217-221; discussion 237-219.

Cripps, A. W., Otczyk, D. C. and Kyd, J. M. (2005) Bacterial otitis media: a vaccine preventable disease? *Vaccine*, **23:** 2304-2310.

Crook, D. W., Brueggemann, A. B., Sleeman, K. S. and Peto, T. E. A. (2004) Pneumococcal Carriage. In *The Pneumococcus* (Ed, Tuomanen, E. I.) ASM Press, Washington, DC, pp. 136-147.

Croucher, N. J., Walker, D., Romero, P., Lennard, N., Paterson, G. K., Bason, N. C., Mitchell, A. M., Quail, M. A., Andrew, P. W., Parkhill, J., Bentley, S. D. and Mitchell, T. J. (2009) Role of conjugative elements in the evolution of the multidrug-resistant pandemic clone *Streptococcus pneumoniae* Spain23F ST81. *J. Bacteriol.*, **191:** 1480-1489.

Crum, N. F., Barrozo, C. P., Chapman, F. A., Ryan, M. A. and Russell, K. L. (2004) An outbreak of conjunctivitis due to a novel unencapsulated *Streptococcus pneumoniae* among military trainees. *Clin. Infect. Dis.,* **39:** 1148-1154.

Cundell, D. R., Weiser, J. N., Shen, J., Young, A. and Tuomanen, E. I. (1995) Relationship between colonial morphology and adherence of *Streptococcus pneumoniae*. *Infect. Immun.*, 63, 757-761.

Cutts, F. T., Zaman, S. M., Enwere, G., Jaffar, S., Levine, O. S., Okoko, J. B., Oluwalana, C., Vaughan, A., Obaro, S. K., Leach, A., McAdam, K. P., Biney, E., Saaka, M., Onwuchekwa, U., Yallop, F., Pierce, N. F., Greenwood, B. M. and Adegbola, R. A. (2005) Efficacy of nine-valent pneumococcal conjugate vaccine against pneumonia and invasive pneumococcal disease in The Gambia: randomised, double-blind, placebo-controlled trial. *Lancet*, **365:** 1139-1146.

Dagerhamn, J., Blomberg, C., Browall, S., Sjostrom, K., Morfeldt, E. and Henriques-Normark, B. (2008) Determination of accessory gene patterns predicts the same relatedness among strains of *Streptococcus pneumoniae* as sequencing of housekeeping genes does and represents a novel approach in molecular epidemiology. *J. Clin. Microbiol.*, **46:** 863-868.

Damoiseaux, R. A. (2005) Antibiotic treatment for acute otitis media: time to think again. *CMAJ*, **172:** 657-658.

Dave, S., Carmicle, S., Hammerschmidt, S., Pangburn, M. K. and McDaniel, L. S. (2004) Dual roles of PspC, a surface protein of *Streptococcus pneumoniae*, in binding human secretory IgA and factor H. *J. Immunol.*, **173:** 471-477.

Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D. and Siebert, P. D. (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. USA*, **93:** 6025-6030.

Dierks, T., Miech, C., Hummerjohann, J., Schmidt, B., Kertesz, M. A. and von Figura, K. (1998) Posttranslational formation of formylglycine in prokaryotic sulfatases by modification of either cysteine or serine. *J. Biol. Chem.*, **273:** 25560-25564.

Dillard, J. P., Vandersea, M. W. and Yother, J. (1995) Characterization of the cassette containing genes for type 3 capsular polysaccharide biosynthesis in *Streptococcus pneumoniae*. *J. Exp. Med.*, **181:** 973-983.

Ehrlich, G. D., Veeh, R., Wang, X., Costerton, J. W., Hayes, J. D., Hu, F. Z., Daigle, B. J., Ehrlich, M. D. and Post, J. C. (2002) Mucosal biofilm formation on middle-ear mucosa in the chinchilla model of otitis media. *JAMA*, **287:** 1710-1715.

Embry, A., Hinojosa, E. and Orihuela, C. J. (2007) Regions of Diversity 8, 9 and 13 contribute to *Streptococcus pneumoniae* virulence. *BMC Microbiol.*, **7:** 80.

Enright, M. C. and Spratt, B. G. (1998) A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. *Microbiology*, **144:** 3049-3060.

Enright, M. C. and Spratt, B. G. (1999) Multilocus sequence typing. *Trends Microbiol.*, **7:** 482-487.

Eskola, J., Kilpi, T., Palmu, A., Jokinen, J., Haapakoski, J., Herva, E., Takala, A., Kayhty, H., Karma, P., Kohberger, R., Siber, G. and Makela, P. H. (2001) Efficacy of a pneumococcal conjugate vaccine against acute otitis media. *N. Engl. J. Med.*, **344:** 403-409.

Evans, R. C. and Holmes, C. J. (1987) Effect of vancomycin hydrochloride on *Staphylococcus epidermidis* biofilm associated with silicone elastomer. *Antimicrob. Agents Chemother.*, **31:** 889-894.

Fedson, D. S. and Scott, J. A. (1999) The burden of pneumococcal disease among adults in developed and developing countries: what is and is not known. *Vaccine*, **17**(Suppl 1): S11-18.

Fireman, B., Black, S. B., Shinefield, H. R., Lee, J., Lewis, E. and Ray, P. (2003) Impact of the pneumococcal conjugate vaccine on otitis media. *Pediatr. Infect. Dis. J.*, **22:** 10-16.

Fletcher, M. A. and Fritzell, B. (2007) Brief review of the clinical effectiveness of PREVENAR against otitis media. *Vaccine*, **25:** 2507-2512.

Forbes, M. L., Horsey, E., Hiller, N. L., Buchinsky, F. J., Hayes, J. D., Compliment, J. M., Hillman, T., Ezzo, S., Shen, K., Keefe, R., Barbadora, K., Post, J. C., Hu, F. Z. and Ehrlich, G. D. (2008) Strain-specific virulence phenotypes of *Streptococcus pneumoniae* assessed using the Chinchilla laniger model of otitis media. *PLoS ONE*, **3:** e1969.

Forrest, J. M., McIntyre, P. B. and Burgess, M. A. (2000) Pneumococcal disease in Australia. *Commun. Dis. Intell.*, **24:** 89-92.

Forsee, W. T., Cartee, R. T. and Yother, J. (2000) Biosynthesis of type 3 capsular polysaccharide in *Streptococcus pneumoniae*. Enzymatic chain release by an abortive translocation process. *J. Biol. Chem.*, **275:** 25972-25978.

Freid, V. M., Makuc, D. M. and Rooks, R. N. (1998) Ambulatory health care visits by children: principal diagnosis and place of visit. *Vital Health Stat.*, **13:** 1-23.

Garcia, P., Gonzalez, M. P., Garcia, E., Lopez, R. and Garcia, J. L. (1999a) LytB, a novel pneumococcal murein hydrolase essential for cell separation. *Mol Microbiol.*, **31:** 1275-1281.

Garcia, P., Paz Gonzalez, M., Garcia, E., Garcia, J. L. and Lopez, R. (1999b) The molecular characterization of the first autolytic lysozyme of *Streptococcus pneumoniae* reveals evolutionary mobile domains. *Mol. Microbiol.*, **33:** 128-138.

Gates, G. A. (1996) Cost-effectiveness considerations in otitis media treatment. *Otolaryngol. Head Neck Surg.*, **114:** 525-530.

Gertz, R. E., Jr., McEllistrem, M. C., Boxrud, D. J., Li, Z., Sakota, V., Thompson, T. A., Facklam, R. R., Besser, J. M., Harrison, L. H., Whitney, C. G. and Beall, B. (2003) Clonal distribution of invasive pneumococcal isolates from children and selected adults in the United States prior to 7-valent conjugate vaccine introduction. *J. Clin. Microbiol.*, **41:** 4194-4216.

Gillis, R. J., White, K. G., Choi, K. H., Wagner, V. E., Schweizer, H. P. and Iglewski, B. H. (2005) Molecular basis of azithromycin-resistant *Pseudomonas aeruginosa* biofilms. *Antimicrob. Agents Chemother.*, **49:** 3858-3867.

Gonzales, R., Malone, D. C., Maselli, J. H. and Sande, M. A. (2001) Excessive antibiotic use for acute respiratory infections in the United States. *Clin. Infect. Dis.*, **33:** 757-762.

Gram, C. (1884) Ueber die isolirte Farbung der Schizomyceten. *Schnitt Äund Trockenpraparaten.. Fortschr. Med.*, **2:** 185-189.

Griffith, F. (1928) The significance of pneumococcal types. *J. Hyg.,* **27:** 113–159.

Gristina, A. G., Hobgood, C. D., Webb, L. X. and Myrvik, Q. N. (1987) Adhesive colonization of biomaterials and antibiotic resistance. *Biomaterials*, **8:** 423-426.

Grodzicker, T. and Zipser, D. (1968) A mutation which creates a new site for the re-initiation of polypeptide synthesis in the *z* gene of the *lac* operon of *Escherichia coli. J. Mol. Biol.*, **38:** 305-314.

Gurskaya, N. G., Diatchenko, L., Chenchik, A., Siebert, P. D., Khaspekov, G. L., Lukyanov, K. A., Vagner, L. L., Ermolaeva, O. D., Lukyanov, S. A. and Sverdlov, E. D. (1996) Equalizing cDNA subtraction based on selective suppression of polymerase chain reaction: cloning of Jurkat cell transcripts induced by phytohemaglutinin and phorbol 12-myristate 13-acetate. *Anal. Biochem.*, **240:** 90-97.

Hall-Stoodley, L., Hu, F. Z., Gieseke, A., Nistico, L., Nguyen, D., Hayes, J., Forbes, M., Greenberg, D. P., Dice, B., Burrows, A., Wackym, P. A., Stoodley, P., Post, J. C., Ehrlich, G. D. and Kerschner, J. E. (2006) Direct detection of bacterial biofilms on the middle-ear mucosa of children with chronic otitis media. *JAMA*, **296:** 202-211.

Hall-Stoodley, L., Nistico, L., Sambanthamoorthy, K., Dice, B., Nguyen, D., Mershon, W. J., Johnson, C., Hu, F. Z., Stoodley, P., Ehrlich, G. D. and Post, J. C. (2008) Characterization of biofilm matrix, degradation by DNase treatment and evidence of capsule downregulation in *Streptococcus pneumoniae* clinical isolates. *BMC Microbiol.*, **8:** 173.

Hammerschmidt, S., Tillig, M. P., Wolff, S., Vaerman, J. P. and Chhatwal, G. S. (2000) Species-specific binding of human secretory component to SpsA protein of *Streptococcus pneumoniae* via a hexapeptide motif. *Mol. Microbiol.*, **36:** 726-736.

Hanage, W. P., Auranen, K., Syrjanen, R., Herva, E., Makela, P. H., Kilpi, T. and Spratt, B. G. (2004) Ability of pneumococcal serotypes and clones to cause acute otitis media: implications for the prevention of otitis media by conjugate vaccines. *Infect. Immun.*, **72:** 76-81.

Hanage, W. P., Kaijalainen, T. H., Syrjanen, R. K., Auranen, K., Leinonen, M., Makela, P. H. and Spratt, B. G. (2005) Invasiveness of serotypes and clones of *Streptococcus pneumoniae* among children in Finland. *Infect. Immun.*, **73:** 431-435.

Hanage, W. P., Kaijalainen, T., Saukkoriipi, A., Rickcord, J. L. and Spratt, B. G. (2006) A successful, diverse disease-associated lineage of nontypeable pneumococci that has lost the capsular biosynthesis locus. *J. Clin. Microbiol.*, **44:** 743-749.

Hanage, W. P., Huang, S. S., Lipsitch, M., Bishop, C. J., Godoy, D., Pelton, S. I., Goldstein, R., Huot, H. and Finkelstein, J. A. (2007) Diversity and antibiotic resistance among nonvaccine serotypes of *Streptococcus pneumoniae* carriage isolates in the post-heptavalent conjugate vaccine era. *J. Infect. Dis.*, **195:** 347-352.

Hanson, S. R., Best, M. D. and Wong, C. H. (2004) Sulfatases: structure, mechanism, biological activity, inhibition, and synthetic utility. *Angewandte Chemie International Ed.*, **43:** 5736-5763.

Harboe, Z. B., Thomsen, R. W., Riis, A., Valentiner-Branth, P., Christensen, J. J., Lambertsen, L., Krogfelt, K. A., Konradsen, H. B. and Benfield, T. L. (2009) Pneumococcal serotypes and mortality following invasive pneumococcal disease: a population-based cohort study. *PLoS Med.*, **6:** e1000081.

Hardy, G. G., Magee, A. D., Ventura, C. L., Caimano, M. J. and Yother, J. (2001) Essential role for cellular phosphoglucomutase in virulence of type 3 *Streptococcus pneumoniae*. *Infect. Immun.*, **69:** 2309-2317.

Hausdorff, W. P., Bryant, J., Kloek, C., Paradiso, P. R. and Siber, G. R. (2000) The contribution of specific pneumococcal serogroups to different disease manifestations: implications for conjugate vaccine formulation and use, part II. *Clin. Infect. Dis.*, **30:** 122-140.

Hausdorff, W. P., Yothers, G., Dagan, R., Kilpi, T., Pelton, S. I., Cohen, R., Jacobs, M. R., Kaplan, S. L., Levy, C., Lopez, E. L., Mason, E. O., Jr., Syriopoulou, V., Wynne, B. and Bryant, J. (2002) Multinational study of pneumococcal serotypes causing acute otitis media in children. *Pediatr. Infect. Dis. J.*, **21:** 1008-1016.

Hava, D. L. and Camilli, A. (2002) Large-scale identification of serotype 4 *Streptococcus pneumoniae* virulence factors. *Mol. Microbiol.*, **45:** 1389-1406.

Havarstein, L. S., Coomaraswamy, G. and Morrison, D. A. (1995) An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci. USA.*, **92:** 11140-11144.

Heidelberger, M. and Avery, O. T. (1923) The soluble specific substance of pneumococcus. *J. Exp. Med.*, **38:** 73-79.

Heidelberger, M. (1927) Immunologically specific polysaccharides. *Chem. Rev.*, **3:** 403-423.

Henegariu, O., Heerema, N. A., Dlouhy, S. R., Vance, G. H. and Vogt, P. H. (1997) Multiplex PCR: critical parameters and step-by-step protocol. *Biotechniques*, **23:** 504-511.

Henriques, B., Kalin, M., Ortqvist, A., Olsson Liljequist, B., Almela, M., Marrie, T. J., Mufson, M. A., Torres, A., Woodhead, M. A., Svenson, S. B. and Kallenius, G. (2000) Molecular epidemiology of *Streptococcus pneumoniae* causing invasive disease in 5 countries. *J. Infect. Dis.*, **182:** 833-839.

Henriques-Normark, B., Blomberg, C., Dagerhamn, J., Battig, P. and Normark, S. (2008) The rise and fall of bacterial clones: *Streptococcus pneumoniae*. *Nature Rev.*, **6:** 827-837.

Higgins, M. A., Whitworth, G. E., El Warry, N., Randriantsoa, M., Samain, E., Burke, R. D., Vocadlo, D. J. and Boraston, A. B. (2009). "Differential recognition and hydrolysis of host carbohydrate-antigens by *Streptococcus pneumoniae* family 98 glycoside hydrolases." *J Biol Chem.*, **284:** 26161-26173.

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. *J. Bacteriol.*, **189:** 8186-8195.

Hollingshead, S. K., Becker, R. and Briles, D. E. (2000) Diversity of PspA: mosaic genes and evidence for past recombination in *Streptococcus pneumoniae*. *Infect. Immun.*, **68:** 5889-5900.

Hollingshead, S. K., Baril, L., Ferro, S., King, J., Coan, P. and Briles, D. E. (2006) Pneumococcal surface protein A (PspA) family distribution among clinical isolates from adults over 50 years of age collected in seven countries. *J. Med. Microbiol.*, **55:** 215-221.

Horton, R. M., Ho, S. N., Pullen, J. K., Hunt, H. D., Cai, Z. and Pease, L. R. (1993) Gene splicing by overlap extension. *Methods Enzymol.*, **217:** 270-279.

Huang, S. S., Hinrichsen, V. L., Stevenson, A. E., Rifas-Shiman, S. L., Kleinman, K., Pelton, S. I., Lipsitch, M., Hanage, W. P., Lee, G. M. and Finkelstein, J. A. (2009) Continued impact of pneumococcal conjugate vaccine on carriage in young children. *Pediatrics*, **124:** e1-11.

Iannelli, F., Oggioni, M. R. and Pozzi, G. (2002) Allelic variation in the highly polymorphic locus *pspC* of *Streptococcus pneumoniae*. *Gene*, **284:** 63-71.

Isozumi, R., Ito, Y., Ishida, T., Hirai, T., Ito, I., Maniwa, K., Hayashi, M., Kagioka, H., Hirabayashi, M., Onaru, K., Tomioka, H., Tomii, K., Gohma, I., Osawa, M., Imai, S., Takakura, S., Iinuma, Y., Chin, K., Ichiyama, S. and Mishima, M. (2008) Molecular characteristics of serotype 3 *Streptococcus pneumoniae* isolates among community-acquired pneumonia patients in Japan. J Infect Chemother, 14, 258-261.

Iyer, R. and Camilli, A. (2007). Sucrose metabolism contributes to *in vivo* fitness of *Streptococcus pneumoniae*. *Mol. Microbiol.*, **66:** 1-13.

Janulczyk, R., Iannelli, F., Sjoholm, A. G., Pozzi, G. and Bjorck, L. (2000) Hic, a novel surface protein of *Streptococcus pneumoniae* that interferes with complement function. *J. Biol. Chem.*, **275:** 37257-37263.

Jefferies, J. M., Johnston, C. H., Kirkham, L. A., Cowan, G. J., Ross, K. S., Smith, A., Clarke, S. C., Brueggemann, A. B., George, R. C., Pichon, B., Pluschke, G., Pfluger, V. and Mitchell, T. J. (2007) Presence of nonhemolytic pneumolysin in serotypes of *Streptococcus pneumoniae* associated with disease outbreaks. *J. Infect. Dis.*, **196:** 936-944.

Jefferson, K. K., Goldmann, D. A. and Pier, G. B. (2005) Use of confocal microscopy to analyze the rate of vancomycin penetration through *Staphylococcus aureus* biofilms. Antimicrob. Agents Chemother., 49, 2467-2473.

Kauffmann, F., Morch, E. and Schmith, K. (1940) On the serology of the pneumococcus-group. *J. Immunol*., **39:** 397-426.

Kelly, T., Dillard, J. P. and Yother, J. (1994) Effect of genetic switching of capsular type on virulence of *Streptococcus pneumoniae*. *Infect. Immun.*, **62:** 1813-1819.

Kim, J. O. and Weiser, J. N. (1998) Association of intrastrain phase variation in quantity of capsular polysaccharide and teichoic acid with the virulence of *Streptococcus pneumoniae*. *J. Infect. Dis*., **177:** 368-377.

King, S. J., Whatmore, A. M. and Dowson, C. G. (2005) NanA, a neuraminidase from *Streptococcus pneumoniae*, shows high levels of sequence diversity, at least in part through recombination with *Streptococcus oralis*. *J. Bacteriol.*, **187:** 5376-5386.

Kirkham, L. A., Jefferies, J. M., Kerr, A. R., Jing, Y., Clarke, S. C., Smith, A. and Mitchell, T. J. (2006) Identification of invasive serotype 1 pneumococcal isolates that express nonhemolytic pneumolysin. *J. Clin. Microbiol.*, **44:** 151-159.

Klein, J. O. (2000) The burden of otitis media. *Vaccine*, **19**(Suppl 1): S2-8.

Knecht, J. C., Schiffman, G. and Austrian, R. (1970) Some biological properties of Pneumococcus type 37 and the chemistry of its capsular polysaccharide. *J. Exp. Med.*, **132:** 475-487.

Kozak, M. (1983) Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. *Microbiol. Rev.*, **47:** 1-45.

Krause, V. L., Reid, S. J. and Merianos, A. (2000) Invasive pneumococcal disease in the Northern Territory of Australia, 1994-1998. *Med. J. Aust.*, **173**(Suppl): S27-31.

Kubba, H., Pearson, J. P. and Birchall, J. P. (2000) The aetiology of otitis media with effusion: a review. *Clin.Otolaryngol. Allied Sci.*, **25:** 181-194.

Kunin, C. M. (1993) Resistance to antimicrobial drugs-a worldwide calamity. *Ann. Intern. Med.*, **118:** 557-561.

Lacks, S., and Hotchkiss, R.D. (1960) A study of the genetic material determining an enzyme activity in pneumococcus. *Biochim Biophys Acta* **39**: 508–517.[

Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227:** 680-685.

Landgrebe, J., Dierks, T., Schmidt, B. and von Figura, K. (2003) The human SUMF1 gene, required for posttranslational sulfatase modification, defines a new gene family which is conserved from pro- to eukaryotes. *Gene*, **316:** 47-56.

Larsen, T. and Fiehn, N. E. (1996) Resistance of *Streptococcus sanguis* biofilms to antimicrobial agents. *APMIS*, **104:** 280-284.

Lau, G. W., Haataja, S., Lonetto, M., Kensit, S. E., Marra, A., Bryant, A. P., McDevitt, D., Morrison, D. A. and Holden, D. W. (2001) A functional genomic analysis of type 3 *Streptococcus pneumoniae* virulence. *Mol. Microbiol.*, **40:** 555-571.

Leach, A. J. and Morris, P. S. (2007) The burden and outcome of respiratory tract infection in Australian and Aboriginal children. *Pediatr. Infect. Dis. J.*, **26:** S4-7.

Lee, J., Lee, H. J., Shin, M. K. and Ryu, W. S. (2004) Versatile PCR-mediated insertion or deletion mutagenesis. *Biotechniques*, **36:** 398-400.

Lehmann, D., Arumugaswamy, A., Elsbury, D., Finucane, J., Stokes, A., Monck, R., Jeffries-Stokes, C., McAullay, D., Coates, H. and Stanley, F. J. (2008) The Kalgoorlie Otitis Media Research Project: rationale, methods, population characteristics and ethical considerations. *Paediatr. Perinat. Epidemiol.*, **22:** 60-71.

Leid, J. G., Shirtliff, M. E., Costerton, J. W. and Stoodley, A. P. (2002) Human leukocytes adhere to, penetrate, and respond to *Staphylococcus aureus* biofilms. *Infect. Immun.*, **70:** 6339-6345.

Leimkugel, J., Adams Forgor, A., Gagneux, S., Pfluger, V., Flierl, C., Awine, E., Naegeli, M., Dangy, J. P., Smith, T., Hodgson, A. and Pluschke, G. (2005) An outbreak of serotype 1 *Streptococcus pneumoniae* meningitis in northern Ghana with features that are characteristic of *Neisseria meningitidis* meningitis epidemics. *J. Infect. Dis.*, **192:** 192-199.

Li-Korotky, H. S., Banks, J. M., Lo, C. Y., Zeng, F. R., Stolz, D. B., Swarts, J. D. and Doyle, W. J. (2008) Interaction of pneumococcal phase variation and middle ear pressure/gas composition: an in vitro model of simulated otitis media. *Microb. Pathog.*, **45:** 201-206.

Li-Korotky, H. S., Lo, C. Y., Zeng, F. R., Lo, D. and Banks, J. M. (2009) Interaction of phase variation, host and pressure/gas composition: Pneumococcal gene expression of PsaA, SpxB, Ply and LytA in simulated middle ear environments. *Int. J. Pediatr. Otorhinolaryngol.*. **73:** 1417-1422.

Livak, K. J. and Schmittgen, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25:** 402-408.

Llull, D., Munoz, R., Lopez, R. and Garcia, E. (1999) A single gene (*tts*) located outside the cap locus directs the formation of *Streptococcus pneumoniae* type 37 capsular polysaccharide. Type 37 pneumococci are natural, genetically binary strains. *J. Exp. Med.*, **190:** 241-251.

Lock, R. A., Paton, J. C. and Hansman, D. (1988) Purification and immunological characterization of neuraminidase produced by *Streptococcus pneumoniae*. *Microb. Pathog.*, **4:** 33-43.

Long, J. P., Tong, H. H., Shannon, P. A. and DeMaria, T. F. (2003) Differential expression of cytokine genes and inducible nitric oxide synthase induced by opacity phenotype variants of *Streptococcus pneumoniae* during acute otitis media in the rat. *Infect. Immun.*, **71:** 5531-5540.

Long, J. P., Tong, H. H. and DeMaria, T. F. (2004) Immunization with native or recombinant *Streptococcus pneumoniae* neuraminidase affords protection in the chinchilla otitis media model. *Infect. Immun.*, **72:** 4309-4313.

Lynch, J. P., 3rd and Zhanel, G. G. (2005) Escalation of antimicrobial resistance among *Streptococcus pneumoniae*: implications for therapy. *Semin. Respir. Crit. Care. Med.*, **26:** 575-616.

Mackenzie, G. A., Carapetis, J. R., Leach, A. J. and Morris, P. S. (2009) Pneumococcal vaccination and otitis media in Australian Aboriginal infants: comparison of two birth cohorts before and after introduction of vaccination. *BMC Pediatr.*, **9:** 14.

Macleod, C. M. and Kraus, M. R. (1950) Relation of virulence of pneumococcal strains for mice to the quantity of capsular polysaccharide formed *in vitro*. *J. Exp. Med.*, **92:** 1-9.

Magee, A. D. and Yother, J. (2001) Requirement for capsule in colonization by *Streptococcus pneumoniae*. *Infect. Immun.*, **69:** 3755-3761.

Majcherczyk, P. A., Langen, H., Heumann, D., Fountoulakis, M., Glauser, M. P. and Moreillon, P. (1999) Digestion of *Streptococcus pneumoniae* cell walls with its major peptidoglycan hydrolase releases branched stem peptides carrying proinflammatory activity. *J. Biol. Chem.*, **274:** 12537-12543.

Mardis, E. R. (2008). The impact of next-generation sequencing technology on genetics. *Trends Genet.*, **24**: 133-41.

Marsh, R. L., Smith-Vaughan, H., Beissbarth, J., Hare, K., Kennedy, M., Wigger, C., Mellon, G., Stubbs, E., Gadil, J. R., Pettit, A., Mackenzie, G., Tipakalippa, P., Morris, P. S. and Leach, A. J. (2007) Molecular characterisation of pneumococcal serotype 16F: Established predominant carriage and otitis media serotype in the 7vPCV era. *Vaccine*, **25:** 2434-2436.

Martin, M., Turco, J. H., Zegans, M. E., Facklam, R. R., Sodha, S., Elliott, J. A., Pryor, J. H., Beall, B., Erdman, D. D., Baumgartner, Y. Y., Sanchez, P. A., Schwartzman, J. D., Montero, J., Schuchat, A. and Whitney, C. G. (2003) An outbreak of conjunctivitis due to atypical *Streptococcus pneumoniae*. *N. Engl. J. Med.*, **348:** 1112-1121.

McEllistrem, M. C., Adams, J., Mason, E. O. and Wald, E. R. (2003) Epidemiology of acute otitis media caused by *Streptococcus pneumoniae* before and after licensure of the 7-valent pneumococcal protein conjugate vaccine. *J. Infect. Dis.*, **188:** 1679-1684.

McEllistrem, M. C., Adams, J. M., Patel, K., Mendelsohn, A. B., Kaplan, S. L., Bradley, J. S., Schutze, G. E., Kim, K. S., Mason, E. O. and Wald, E. R. (2005) Acute otitis media due to penicillin-nonsusceptible *Streptococcus pneumoniae* before and after the introduction of the pneumococcal conjugate vaccine. *Clin. Infect. Dis.*, **40:** 1738-1744.

McKessar, S. J. and Hakenbeck, R. (2007) The two-component regulatory system TCS08 is involved in cellobiose metabolism of *Streptococcus pneumoniae* R6. *J. Bacteriol.*, **189:** 1342-1350.

Mehta, A. J., Lee, J. C., Stevens, G. R. and Antonelli, P. J. (2006) Opening plugged tympanostomy tubes: effect of biofilm formation. *Otolaryngol. Head Neck Surg.*, **134:** 121-125.

Miech, C., Dierks, T., Selmer, T., von Figura, K. and Schmidt, B. *(1998) Arylsulfatase from Klebsiella pneumoniae* carries a formylglycine generated from a serine. *J. Biol. Chem.*, **273:** 4835-4837.

Morona, J. K., Morona, R. and Paton, J. C. (2006) Attachment of capsular polysaccharide to the cell wall of *Streptococcus pneumoniae* type 2 is required for invasive disease. *Proc. Natl. Acad. Sci. USA*, **103:** 8505-8510.

Morris, P. S., Leach, A. J., Silberberg, P., Mellon, G., Wilson, C., Hamilton, E. and Beissbarth, J. (2005) Otitis media in young Aboriginal children from remote communities in Northern and Central Australia: a cross-sectional survey. *BMC Pediatr.*, **5:** 27.

Moscoso, M., Garcia, E. and Lopez, R. (2006) Biofilm formation by *Streptococcus pneumoniae*: role of choline, extracellular DNA, and capsular polysaccharide in microbial accretion. *J. Bacteriol.*, **188:** 7785-7795.

Mothibeli, K. M. (2007) Molecular epidemiology of South African serotype 3 and serotype 19A pneumococcal isolates. In Faculty of Science, University of the Witwatersrand, Johannesburg.

Munoz-Elias, E. J., Marcano, J. and Camilli, A. (2008) Isolation of *Streptococcus pneumoniae* biofilm mutants and their characterization during nasopharyngeal colonization. *Infect. Immun.*, **76:** 5049-5061.

Neufeld, F. (1902) Ueber die Agglutination der Pneumokokken und uber die Theorien der Agglutination. *Zeitschrift fur Hygiene Infektionskrankheiten*, **40:**54-72.

Nickel, J. C., Ruseska, I., Wright, J. B. and Costerton, J. W. (1985) Tobramycin resistance of *Pseudomonas aeruginosa* cells growing as a biofilm on urinary catheter material. *Antimicrob. Agents Chemother.*, **27:** 619-624.

Obert, C., Sublett, J., Kaushal, D., Hinojosa, E., Barton, T., Tuomanen, E. I. and Orihuela, C. J. (2006) Identification of a candidate *Streptococcus pneumoniae* core genome and regions of diversity correlated with invasive pneumococcal disease. *Infect. Immun.*, **74:** 4766-4777.

Obert, C. A., Gao, G., Sublett, J., Tuomanen, E. I. and Orihuela, C. J. (2007) Assessment of molecular typing methods to determine invasiveness and to differentiate clones of *Streptococcus pneumoniae*. *Infect. Genet. Evol.*, **7:** 708-716.

O'Brien, M. A., Prosser, L. A., Paradise, J. L., Ray, G. T., Kulldorff, M., Kurs-Lasky, M., Hinrichsen, V. L., Mehta, J., Colborn, D. K. and Lieu, T. A. (2009) New vaccines against otitis media: projected benefits and cost-effectiveness. *Pediatrics.*, **123:** 1452-1463.

Oggioni, M. R., Trappetti, C., Kadioglu, A., Cassone, M., Iannelli, F., Ricci, S., Andrew, P. W. and Pozzi, G. (2006) Switch from planktonic to sessile life: a major event in pneumococcal pathogenesis. *Mol. Microbiol.*, **61:** 1196-1210.

Ogunniyi, A. D., Giammarinaro, P. and Paton, J. C. (2002) The genes encoding virulence-associated proteins and the capsule of *Streptococcus pneumoniae* are upregulated and differentially expressed in vivo. *Microbiology*, **148:** 2045-2053.

Ogunniyi, A. D., Paton, J. C., Kirby, A. C., McCullers, J. A., Cook, J., Hyodo, M., Hayakawa, Y. and Karaolis, D. K. (2008) c-di-GMP is an effective immunomodulator and vaccine adjuvant against pneumococcal infection. *Vaccine*, **26:** 4676-4685.

Ostroff, S. M. (1999) Continuing challenge of pneumococcal disease. *Lancet*, **353:** 1201-1202.

Overweg, K., Pericone, C. D., Verhoef, G. G., Weiser, J. N., Meiring, H. D., De Jong, A. P., De Groot, R. and Hermans, P. W. (2000) Differential protein expression in phenotypic variants of *Streptococcus pneumoniae*. *Infect. Immun.*, **68:** 4604-4610.

Palmu, A. A., Verho, J., Jokinen, J., Karma, P. and Kilpi, T. M. (2004) The seven-valent pneumococcal conjugate vaccine reduces tympanostomy tube placement in children. *Pediatr. Infect. Dis. J.*, **23:** 732-738.

Parker, D., Soong, G., Planet, P., Brower, J., Ratner, A. J. and Prince, A. (2009) The NanA neuraminidase of *Streptococcus pneumoniae* is involved in biofilm formation. *Infect. Immun.*, **77:** 3722-30

Pasteur, L. and Chamberland, M. (1881) Note sur la maladie nouvelle provoquée par la salive d'un enfant mort de la rage. *Bull. Acad. Méd.*, **10:** 94-103.

Paton, J. C., Andrew, P. W., Boulnois, G. J. and Mitchell, T. J. (1993) Molecular analysis of the pathogenicity of *Streptococcus pneumoniae*: the role of pneumococcal proteins. *Annu. Rev. Microbiol.*, **47:** 89-115.

Paton, J. C. (2004) New Pneumococcal Vaccines: Basic Science Developments. In *The Pneumococcus* (Ed, Tuomanen, E. I.) ASM Press, Washington, DC, pp. 382-402.

Perez-Trallero, E., Marimon, J. M., Ercibengoa, M., Vicente, D. and Perez-Yarza, E. G. (2009) Invasive *Streptococcus pneumoniae* infections in children and older adults in the north of Spain before and after the introduction of the heptavalent pneumococcal conjugate vaccine. *Eur. J. Clin. Microbiol. Infect. Dis.*, **28:** 731-738.

Pettigrew, M. M. and Fennie, K. P. (2005) Genomic subtraction followed by dot blot screening of *Streptococcus pneumoniae* clinical and carriage isolates identifies genetic differences associated with strains that cause otitis media. *Infect. Immun.*, 73, 2805-2811.

Pettigrew, M. M., Fennie, K. P., York, M. P., Daniels, J. and Ghaffar, F. (2006) Variation in the presence of neuraminidase genes among *Streptococcus pneumoniae* isolates with identical sequence types. *Infect. Immun.*, **74: 3360-3365.**

Polissi, A., Pontiggia, A., Feger, G., Altieri, M., Mottl, H., Ferrari, L. and Simon, D. (1998) Large-scale identification of virulence genes from *Streptococcus pneumoniae*. *Infect. Immun.*, **66:** 5620-5629.

Porat, N., Arguedas, A., Spratt, B. G., Trefler, R., Brilla, E., Loaiza, C., Godoy, D., Bilek, N. and Dagan, R. (2004) Emergence of penicillin-nonsusceptible *Streptococcus pneumoniae* clones expressing serotypes not present in the antipneumococcal conjugate vaccine. *J. Infect. Dis.*, **190:** 2154-2161.

Porat, N., Soley, C., Marengolciene, M. M., Greenberg, D., Givon-Lavi, N., Trefler, R., Arguedas, A. and Dagan, R. (2008) An international serotype 3 clone causing pediatric noninvasive infections in Israel, Costa Rica, and Lithuania. *Pediatr. Infect. Dis. J.*, **27:** 709-712.

Prudhomme, M., Attaiech, L., Sanchez, G., Martin, B. and Claverys, J. P. (2006) Antibiotic stress induces genetic transformability in the human pathogen *Streptococcus pneumoniae*. *Science*, **313:** 89-92.

Prymula, R. (2009) Re: Global serotype distribution among *Streptococcus pneumoniae* isolates causing otitis media in children: potential implications for pneumococcal conjugate vaccines. *Vaccine*, **27:** 4739-4740.

Quin, L. R., Carmicle, S., Dave, S., Pangburn, M. K., Evenhuis, J. P. and McDaniel, L. S. (2005) In vivo binding of complement regulator factor H by *Streptococcus pneumoniae*. *J. Infect. Dis.*, **192:** 1996-2003.

Rapola, S., Jantti, V., Haikala, R., Syrjanen, R., Carlone, G. M., Sampson, J. S., Briles, D. E., Paton, J. C., Takala, A. K., Kilpi, T. M. and Kayhty, H. (2000) Natural development of antibodies to pneumococcal surface protein A, pneumococcal surface adhesin A, and pneumolysin in relation to pneumococcal carriage and acute otitis media. *J. Infect. Dis.*, **182:** 1146-1152.

Reichmann, P., Konig, A., Linares, J., Alcaide, F., Tenover, F. C., McDougal, L., Swidsinski, S. and Hakenbeck, R. (1997) A global gene pool for high-level cephalosporin resistance in commensal *Streptococcus* species and *Streptococcus pneumoniae*. *J. Infect. Dis.*, **176:** 1001-1012.

Ring, A., Weiser, J. N. and Tuomanen, E. I. (1998) Pneumococcal trafficking across the blood-brain barrier. Molecular analysis of a novel bidirectional pathway. *J. Clin. Invest.*, **102:** 347-360.

Rodgers, G. L., Arguedas, A., Cohen, R. and Dagan, R. (2009) Global serotype distribution among *Streptococcus pneumoniae* isolates causing otitis media in children: potential implications for pneumococcal conjugate vaccines. *Vaccine*, **27:** 3802-3810.

Rolo, D., Ardanuy, C., Fleites, A., Martin, R. and Linares, J. (2009) Diversity of pneumococcal surface protein A (PspA) among prevalent clones in Spain. *BMC Microbiol.*, **9:** 80.

Romero, P., Croucher, N. J., Hiller, N. L., Hu, F. Z., Ehrlich, G. D., Bentley, S. D., Garcia, E. and Mitchell, T. J. (2009) Comparative genomic analysis of ten *Streptococcus pneumoniae* temperate bacteriophages. *J. Bacteriol.*, **191:** 4854-4862.

Rosenow, C., Ryan, P., Weiser, J. N., Johnson, S., Fontan, P., Ortqvist, A. and Masure, H. R. (1997) Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of *Streptococcus pneumoniae*. *Mol. Microbiol.*, **25:** 819-829.

Sagerstrom, C. G., Sun, B. I. and Sive, H. L. (1997) Subtractive cloning: past, present, and future. *Annu. Rev. Biochem.*, **66:** 751-783.

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

Sarabhai, A. and Brenner, S. (1967) A mutant which reinitiates the polypeptide chain after chain termination. *J. Mol. Biol.*, **27:** 145-162.

Sardiello, M., Annunziata, I., Roma, G. and Ballabio, A. (2005) Sulfatases and sulfatase modifying factors: an exclusive and promiscuous relationship. *Hum Mol Genet.*, **14:** 3203-3217.

Serrano, I., Melo-Cristino, J., Carrico, J. A. and Ramirez, M. (2005) Characterization of the genetic lineages responsible for pneumococcal invasive disease in Portugal. *J. Clin. Microbiol.*, **43:** 1706-1715.

Serrano, I., Melo-Cristino, J. and Ramirez, M. (2006) Heterogeneity of pneumococcal phase variants in invasive human infections. *BMC Microbiol,*, **6:** 67.

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. *Infect. Immun.*, **70:** 7161-7164.

Shaper, M., Hollingshead, S. K., Benjamin, W. H., Jr. and Briles, D. E. (2004) PspA protects *Streptococcus pneumoniae* from killing by apolactoferrin, and antibody to PspA enhances killing of pneumococci by apolactoferrin [corrected]. *Infect. Immun.*, **72:** 5031-5040.

Shayegani, M., Parsons, L. M., Gibbons, W. E., Jr. and Campbell, D. (1982) Characterization of nontypable *Streptococcus pneumoniae*-like organisms isolated from outbreaks of conjunctivitis. *J. Clin. Microbiol.*, **16:** 8-14.

Shemesh, M., Tam, A. and Steinberg, D. (2007) Differential gene expression profiling of Streptococcus mutans cultured under biofilm and planktonic conditions. *Microbiology*, **153:** 1307-1317.

Shen, K., Gladitz, J., Antalis, P., Dice, B., Janto, B., Keefe, R., Hayes, J., Ahmed, A., Dopico, R., Ehrlich, N., Jocz, J., Kropp, L., Yu, S., Nistico, L., Greenberg, D. P., Barbadora, K., Preston, R. A., Post, J. C., Ehrlich, G. D. and Hu, F. Z. (2006) Characterization, distribution, and expression of novel genes among eight clinical isolates of *Streptococcus pneumoniae*. *Infect. Immun.*, **74:** 321-330.

Shouval, D. S., Greenberg, D., Givon-Lavi, N., Porat, N. and Dagan, R. (2006) Site-specific disease potential of individual *Streptococcus pneumoniae* serotypes in pediatric invasive disease, acute otitis media and acute conjunctivitis. *Pediatr. Infect. Dis. J.*, **25:** 602-607.

Silva, N. A., McCluskey, J., Jefferies, J. M., Hinds, J., Smith, A., Clarke, S. C., Mitchell, T. J. and Paterson, G. K. (2006) Genomic diversity between strains of the same serotype and multilocus sequence type among pneumococcal clinical isolates. *Infect. Immun.*, **74:** 3513-3518.

Singleton, R. J., Hennessy, T. W., Bulkow, L. R., Hammitt, L. L., Zulz, T., Hurlburt, D. A., Butler, J. C., Rudolph, K. and Parkinson, A. (2007) Invasive pneumococcal disease caused by nonvaccine serotypes among Alaska native children with high levels of 7-valent pneumococcal conjugate vaccine coverage. *JAMA*, **297:** 1784-1792.

Sjostrom, K., Spindler, C., Ortqvist, A., Kalin, M., Sandgren, A., Kuhlmann-Berenzon, S. and Henriques-Normark, B. (2006) Clonal and capsular types decide whether pneumococci will act as a primary or opportunistic pathogen. *Clin. Infect. Dis.*, **42:** 451-459.

Slinger, R., Chan, F., Ferris, W., Yeung, S. W., St Denis, M., Gaboury, I. and Aaron, S. D. (2006) Multiple combination antibiotic susceptibility testing of nontypeable *Haemophilus influenzae* biofilms. *Diagn. Microbiol. Infect. Dis.*, **56:** 247-253.

Smalley, J. W., Dwarakanath, D., Rhodes, J. M. and Hart, C. A. (1994) Mucin-sulphatase activity of some oral streptococci. *Caries Res.*, 28, 416-420.

Smith-Vaughan, H., Byun, R., Halpin, S., Nadkarni, M. A., Jacques, N. A., Hunter, N., Morris, P. S. and Leach, A. J. (2008) Interventions for prevention of otitis media may be most effective if implemented in the first weeks of life. *Int. J. Pediatr. Otorhinolaryngol.*, **72:** 57-61.

Soong, G., Muir, A., Gomez, M. I., Waks, J., Reddy, B., Planet, P., Singh, P. K., Kaneko, Y., Wolfgang, M. C., Hsiao, Y. S., Tong, L. and Prince, A. (2006) Bacterial neuraminidase facilitates mucosal infection by participating in biofilm production. *J. Clin. Invest.*, **116:** 2297-2305.

Sorensen, U. B., Henrichsen, J., Chen, H. C. and Szu, S. C. (1990) Covalent linkage between the capsular polysaccharide and the cell wall peptidoglycan of *Streptococcus pneumoniae* revealed by immunochemical methods. *Microb. Pathog.*, **8:** 325-334.

Southern, E. M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.*, **98:** 503-517.

Sternberg, G. (1885) The pneumonia-coccus of Friedlander (*Micrococcus Pasteuri*, Sternberg). *Am. J. Med. Sci.*, **90:** 106-123.

Stubbs, E., Hare, K., Wilson, C., Morris, P. and Leach, A. J. (2005) *Streptococcus pneumoniae* and noncapsular *Haemophilus influenzae* nasal carriage and hand contamination in children: a comparison of two populations at risk of otitis media. *Pediatr. Infect. Dis. J.*, **24:** 423-428.

Swiatlo, E., Brooks-Walter, A., Briles, D. E. and McDaniel, L. S. (1997) Oligonucleotides identify conserved and variable regions of *pspA* and *pspA*-like sequences of *Streptococcus pneumoniae*. *Gene*, **188:** 279-284.

Talkington, D. F., Voellinger, D. C., McDaniel, L. S. and Briles, D. E. (1992) Analysis of pneumococcal PspA microheterogeneity in SDS polyacrylamide gels and the association of PspA with the cell membrane. *Microb. Pathog.*, **13:** 343-355.

Tettelin, H., Nelson, K. E., Paulsen, I. T., Eisen, J. A., Read, T. D., Peterson, S., Heidelberg, J., DeBoy, R. T., Haft, D. H., Dodson, R. J., Durkin, A. S., Gwinn, M., Kolonay, J. F., Nelson, W. C., Peterson, J. D., Umayam, L. A., White, O., Salzberg, S. L., Lewis, M. R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A. M., Utterback, T. R., Hansen, C. L., McDonald, L. A., Feldblyum, T. V., Angiuoli, S., Dickinson, T., Hickey, E. K., Holt, I. E., Loftus, B. J., Yang, F., Smith, H. O., Venter, J. C., Dougherty, B. A., Morrison, D. A., Hollingshead, S. K. and Fraser, C. M. (2001) Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science*, **293:** 498-506.

Titgemeyer, F., Reizer, J., Reizer, A. and Saier, M. H., Jr. (1994) Evolutionary relationships between sugar kinases and transcriptional repressors in bacteria. *Microbiology*, **140**(Pt 9), 2349-2354.

Titgemeyer, F. and Hillen, W. (2002) Global control of sugar metabolism: a gram-positive solution. *Antonie Van Leeuwenhoek*, **82:** 59-71.

Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E. F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H. G., Glodek, A., McKenney, K., Fitzegerald, L. M., Lee, N., Adams, M. D., Hickey, E. K., Berg, D. E., Gocayne, J. D., Utterback, T. R., Peterson, J. D., Kelley, J. M., Cotton, M. D., Weidman, J. M., Fujii, C., Bowman, C., Watthey, L., Wallin, E., Hayes, W. S., Borodovsky, M., Karp, P. D., Smith, H. O.,

Fraser, C. M. and Venter, J. C. (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature*, **388:** 539-547.

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. *Infect. Immun.*, **68:** 921-924.

Tong, H. H., Weiser, J. N., James, M. A. and DeMaria, T. F. (2001) Effect of influenza A virus infection on nasopharyngeal colonization and otitis media induced by transparent or opaque phenotype variants of *Streptococcus pneumoniae* in the chinchilla model. *Infect. Immun.*, **69:** 602-606.

Tonnaer, E. L., Rijkers, G. T., Meis, J. F., Klaassen, C. H., Bogaert, D., Hermans, P. W. and Curfs, J. H. (2005) Genetic relatedness between pneumococcal populations originating from the nasopharynx, adenoid, and tympanic cavity of children with otitis media. *J. Clin. Microbiol.*, **43:** 3140-3144.

Towbin, H., Staehelin, T. and Gordon, J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. *Natl. Acad. Sci. USA*, **76:** 4350-4354.

Trappetti, C., Kadioglu, A., Carter, M., Hayre, J., Iannelli, F., Pozzi, G., Andrew, P. W. and Oggioni, M. R. (2009) Sialic acid: a preventable signal for pneumococcal biofilm formation, colonization, and invasion of the host. *J. Infect. Dis.*, **199:** 1497-1505.

Tu, A. H., Fulgham, R. L., McCrory, M. A., Briles, D. E. and Szalai, A. J. (1999) Pneumococcal surface protein A inhibits complement activation by *Streptococcus pneumoniae*. *Infect. Immun.*, **67:** 4720-4724.

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:** 3553-3560.

van Ginkel, F. W., McGhee, J. R., Watt, J. M., Campos-Torres, A., Parish, L. A. and Briles, D. E. (2003) Pneumococcal carriage results in ganglioside-mediated olfactory tissue infection. *Proc. Natl. Acad. Sci. USA*, **100:** 14363-14367.

Vela Coral, M. C., Fonseca, N., Castaneda, E., Di Fabio, J. L., Hollingshead, S. K. and Briles, D. E. (2001) Pneumococcal surface protein A of invasive *Streptococcus pneumoniae* isolates from Colombian children. *Emerg. Infect. Dis.*, **7:** 832-836.

Vestrheim, D. F., Hoiby, E. A., Aaberge, I. S. and Caugant, D. A. (2008) Phenotypic and genotypic characterization of *Streptococcus pneumoniae* strains colonizing children attending day-care centers in Norway. *J. Clin. Microbiol.*, **46:** 2508-2518.

Waite, R. D., Struthers, J. K. and Dowson, C. G. (2001) Spontaneous sequence duplication within an open reading frame of the pneumococcal type 3 capsule locus causes high-frequency phase variation. *Mol. Microbiol.*, **42:** 1223-1232.

Waite, R. D., Penfold, D. W., Struthers, J. K. and Dowson, C. G. (2003) Spontaneous sequence duplications within capsule genes *cap8E* and *tts* control phase variation in *Streptococcus pneumoniae* serotypes 8 and 37. *Microbiology*, **149:** 497-504.

Walters, M. C., 3rd, Roe, F., Bugnicourt, A., Franklin, M. J. and Stewart, P. S. (2003) Contributions of antibiotic penetration, oxygen limitation, and low metabolic activity to tolerance of *Pseudomonas aeruginosa* biofilms to ciprofloxacin and tobramycin. *Antimicrob. Agents Chemother.*, **47:** 317-323.

Watson, D. A., Musher, D. M., Jacobson, J. W. and Verhoef, J. (1993) A brief history of the pneumococcus in biomedical research: a panoply of scientific discovery. *Clin. Infect. Dis.*, **17:** 913-924.

Watson, K., Carville, K., Bowman, J., Jacoby, P., Riley, T. V., Leach, A. J. and Lehmann, D. (2006) Upper respiratory tract bacterial carriage in Aboriginal and non-Aboriginal children in a semi-arid area of Western Australia. *Pediatr. Infect. Dis. J.*, **25:** 782-790.

Watson, M., Brett, M., Brown, M., Stewart, M. G. and Warren, S. (2007) Pneumococci responsible for invasive disease and discharging ears in children in Sydney, Australia. *J. Med. Microbiol.*, **56:** 819-823.

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. *Infect. Immun.*, **62:** 2582-2589.

Whatmore, A. M., Barcus, V. A. and Dowson, C. G. (1999) Genetic diversity of the streptococcal competence (*com*) gene locus. *J. Bacteriol.*, **181:** 3144-3154.

White, P., Hermansson, A., Svanborg, C., Briles, D. and Prellner, K. (1999) Effects of active immunization with a pneumococcal surface protein (PspA) and of locally applied antibodies in experimental otitis media. *ORL J Otorhinolaryngol Relat Spec.*, **61:** 206-211.

WHO (1999) Pneumococcal vaccines. WHO position paper. *Wkly Epidemiol. Rec.*, **74:** 177-183.

Wu, H. Y., Virolainen, A., Mathews, B., King, J., Russell, M. W. and Briles, D. E. (1997) Establishment of a *Streptococcus pneumoniae* nasopharyngeal colonization model in adult mice. *Microb. Pathog.*, **23:** 127-137.

Amendments

Chapter 1

- Page 1, line 1: replace “gram positive” with “Gram positive”.

Chapter 2

- Page 38, Table 2.1, Row 3: Replace “□*cpsAB ery*^{r}” with “Δ*cpsAB ery*^{r}”.
- Page 43, line 7: delete the space between “sulphuric acid/” and “sodium tetraborate”.
- Page 45, line 5: delete the space between the bracket and the full stop.

Chapter 3

- Page 77, line 16: replace the sentence “Although serotype 11A...” with “Although serotype 11A OM has been reported in other communities, including one study from Sydney, Australia, serotype 3 generally has a higher association with OM worldwide.” References remain unchanged.
- Page 78, line 7: delete “isolates” after “at the MSHR”.
- Page 81, line 4: after the word “dimers” add “, but could also reflect post-translational modification or covalent association with another protein”.
- Page 96, Figure 3.6: replace “either 5 x 10^2 CFU (serotype 3 strains) or 1 x 10^7 CFU (serogroup 11 strains) of bacteria” with “2 × 10^2 CFU of WU2, 3 × 10^2 CFU of WCH206, 3 × 10^2 CFU of WCH207, 3 × 10^2 CFU of MSHR11, 2.6 × 10^2 CFU of MSHR17, 1 × 10^7

CFU of WCH211, 9.7×10^6 CFU of WCH213, 1.8×10^7 CFU of MSHR1 or 9.9×10^6 CFU of MSHR5".

- Page 97, Figure 3.7: replace "1.5 – 4 x 10^7 CFU of the strains indicated" with "2×10^7 CFU of WU2, 1.5×10^7 CFU of WCH206, 1.5×10^7 CFU of WCH207, 1.5×10^7 CFU of MSHR11, 1.3×10^7 CFU of MSHR17, 2×10^7 CFU of WCH211, 5×10^7 CFU of WCH213, 3×10^7 CFU of MSHR1 or 2×10^7 CFU of MSHR5".

- Page 99, line 16: replace "serotype3" with "serotype 3".

Chapter 4

- Page 122, line 14: replace "genes" with "sequences".

- Page 144, line 15: replace "because visible" with "became visible".

Chapter 5

- Page 156, line 13: replace "publically" with "publicly".

- Page 168, line 3: replace "componentof" with "component of".

- Page 169, line 2: add "Finally, it is also possible that these observations in the genomes of TIGR4 and G54 are simply due to sequencing error." to the end of the paragraph."

- Page 172, line 18: replace "indivually" with "individually".

- Page 172, line 21: delete "was generated using RT-PCR."

- Page 174, line 3: Replace "IICcomponet" with "IIC component".

- Page 176, line 5: delete the full stop following *E. coli*.

Chapter 6

- Page 201, paragraph 2: replace the final sentence with "The slope of the growth rate during exponential phase was calculated for each culture and compared between strains of the same serotype using the Kruskal-Wallis test with Dunn's correction as described by the

tutorial of Wang and Bushman (2007). There was no statistically significant difference between strains ($P > 0.05$)."

- Page 203, line 23: replace "BA/BA-ery" with "BA and BA-ery".

Chapter 7

- Page 218, line 9: after "respectively)" add "and possess 45 pair-unique genes".

- Page 219, line 2: insert ", which have been annotated in the SP3-BS71 genome by Hiller *et al.* (2007)," after "(Section 4.3)".

- Page 221, line 14: delete the sentence starting "This region is...". Replace with "A visual representation of the comparison of the OXC141 genome with D39, which highlights genes unique to the ST180 genome and other regions of interest from the subsequent sequencing of MSHR17 and WU2, was constructed using the Artemis Comparison Tool (Carver *et al.*, 2005) following the BLAST analysis and is in the Appendix (Figure A1).

- Page 226, line 21: Insert "OXC141 genome." between "the" and "Intergenic".

-Page 239, line 18: add to the end of the paragraph "In addition to the enolase of the region above, OXC141 appears to possess a second enolase (annotated as phosphopyruvate hydrolase in SP3-BS71), which is not ubiquitous but is found in D39 and TIGR4. The sequencing suggests it is present in MSHR17 and WU2. However, this gene is listed in Table A1 and therefore, there may be sequence variation in MSHR17."

- Page 244, line 14: replace "galaxies" with "galactose".

Appendix

Figure A1 on the following page should be included before Table A1.

Bibliography

The following references should be added to the bibliography:

Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.A., Barrell, B.G. and Parkhill, J. (2005) ACT: the Artemis Comparison Tool. *Bioinformatics*, **21:**3422-3.

Tsuprun, V., Cureoglu, S., Schachern, P.A., Ferrieri, P., Briles, D.E., Paprella, M.M., and Juhn, S.K. (2008) Role of pneumococcal proteins in sensorineural hearing loss during otitis media. *Otol. Neurotol.*, **29:** 1056-1060

Wang , G.P. and Bushman, F.D. (2006) A statistical method for comparing viral growth curves. *J. Virol. Methods*, **135:**118-123

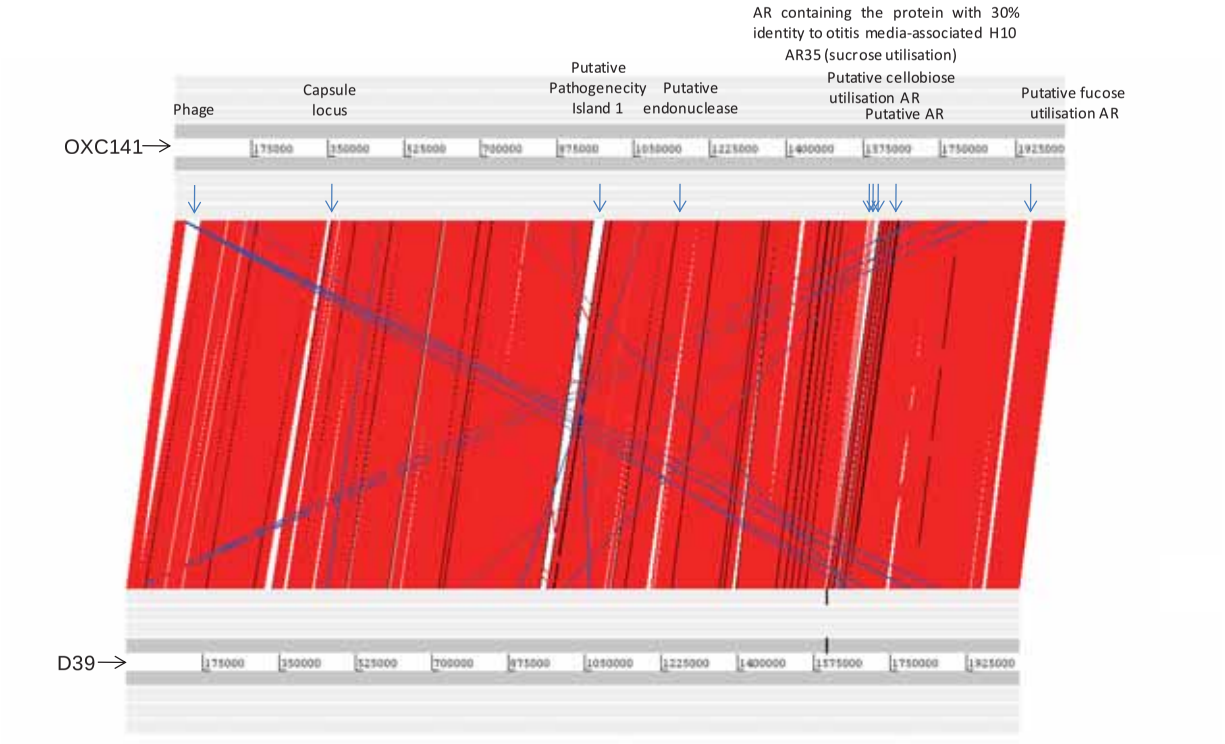


Figure A1. Comparison of the genomes of *S. pneumoniae* OXC141 and D39.

The genomes were compared using the Artemis Comparison Tool (Carver *et al.*, 2005) with the highest scoring matches shown. Regions of interest in the OXC141 genome from the work in Chapter 7 are indicated by arrows, as well as the serotype 3 capsule locus and the AR containing the putative cellobiose PTS identified in Chapter 4. The putative endonuclease, as well as genes of the phage, putative pathogenecity island and the AR containing the protein with 30% identity to otitis media-associated H10, are unique to ST180 (Sections 7.2, 7.3). D39 and MSHR17 contain alternative variants of the sucrose utilisation AR (Sections 7.2 and 7.3). The putative AR contains is from spd1636-spd1642 in the D39 genome. It is present in OXC141, TIGR4 and WU2 but not MSHR17. D39, TIGR4, WU2 and MSHR17 contain alternative variants of the putative fucose utilisation AR (Sections 7.2 and 7.3).