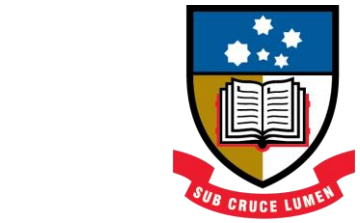
Characterisation of *Streptococcus pneumoniae* Opacity Phase Variation



THE UNIVERSITY of ADELAIDE

Melissa Hui Chieh Chai, BBiomedSc (Hon), MSc

A thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy from the University of Adelaide

February 2016

Research Centre for Infectious Diseases
Department of Molecular and Cellular Biology
The University of Adelaide
Adelaide, S.A., Australia

Table of Contents

List of Figures

List of Tables

Abstract

Streptococcus pneumoniae (the pneumococcus) is a common nasopharyngeal commensal in humans that can invade deeper tissues causing a range of diseases, resulting in significant global morbidity and mortality. Any given strain of *S. pneumoniae* can undergo phase variation between two different colony phenotypes, termed opaque (O) and transparent (T), as determined by their morphology on clear, solid media. In animal models, the O form is more virulent and commonly isolated from normally sterile sites, such as the blood, lungs and brain, compared to its T counterpart, which is more likely to reside in the nasopharynx. To date, the mechanism involved in the switch between opacity phenotypes is not known and there is discordance in the literature in terms of the precise molecular differences between them. This project set out to comprehensively characterise phenomic (2D-DIGE), transcriptomic (DNA microarray) and genomic (IonTorrent sequencing) differences between opacity variants in three unrelated pneumococcal strains (D39, serotype 2; WCH16, serotype 6A; and WCH43, serotype 4).

Proteomic, transcriptomic and genomic analyses revealed strain-specific differences associated with phase variation, but no single difference was consistent across the three strains tested. Nevertheless, there were examples of proteins and genes belonging to the same putative functional groups that were regulated similarly in both the O and T phase between strains. One example was the upregulation of proteins and genes related to genetic competence in the O variants. However, mutagenesis of one such gene, encoding the competence stimulating peptide receptor (*comD*), did not alter opacity phenotype. There were also inconsistencies between genes identified as differentially expressed at an mRNA and protein level, such as genes involved in cell division, and amino acid biosynthesis and acquisition. These genes were upregulated at an mRNA level in the T variants, but no such upregulation in protein expression was identified by proteomic analysis. At a genomic level, all single nucleotide polymorphisms identified by IonTorent sequencing were independently verified. However, the insertion/deletions, particularly those associated with homopolymeric tracts were all found to be false call-outs, which is a limitation of this sequencing technology.

The role of epigenetic changes mediated by genetic rearrangements in a Type I restriction-modification (RM) system on opacity phenotype was also investigated. These rearrangements result in switching between six alternative DNA methylation site specificities impacting on genomic methylation patterns. Examination of six "locked" mutants (SpnD39IIIA-F) with monospecific DNA methylation patterns, indicated that there was an epigenetic impact on colony opacity phenotype and virulence. Importantly, these genetic rearrangements at the Type I RM locus also occurred during experimental infection. However, there were inconsistencies between the opacity phenotypes of locked mutants and the RM allele distribution in wild-type D39O and D39T. Thus, RM allele switching cannot fully account for colony opacity phase variation in pneumococci.

This study identified that pneumococcal phase variation is a complex, multifactorial phenomenon and different strains employ alternative mechanisms to attain opacity phenotypes. Furthermore, epigenetic changes impact pneumococcal opacity morphology and pathogenicity. Hence, the role of epigenetic factors in phase variation and pathogenesis should be investigated in future studies.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution 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. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis, when deposited in the University Library, being made 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 Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

_____

Melissa Hui Chieh Chai

……/……/……

Abbreviations

°C	degrees Celsius
µg	microgram/s
µg/ml	microgram/milliliter
µl	microlitre/s
2D-DIGE	two-dimensional differential gel electrophoresis
A	adenine
A_{340}	absorbance at 340 nm
A_{520}	absorbance at 520 nm
A_{550}	absorbance at 550 nm
A_{600}	absorbance at 600 nm
Adc	zinc ABC transporter
AdhE	alcohol dehydrogenase
AdoMet	S-adenosyl methionine
ADP	adenosine diphosphate
AGRF	Australian Genome Research Facility Ltd
AliA	an oligopeptide ABC transporter
ami	aminopterin resistance operon
APC	adelaide Proteomics Centre
APIA	acid-phenol:chloroform:isoamly alcohol
ATP	adenosine triphosphate
BA	blood agar
BBB	blood-brain barrier
Blp	bacteriocin protein
bp	base pairs
C	cytosine
C+Y	semi-synthetic medium
CAP	community-acquired pneumococcal (infections)
Cbp	choline-binding protein
$CD4^+$	cluster of differentiation 4

CFU	colony forming units
Cgl	competence protein
ChoP	pneumococcal cell-wall phosphorylcholine
CiaRH	two-component signal-transducing system
cibA	competence induced bacteriocin A gene
CNS	central nervous system
Com	competence protein
comDE	two-component signal-transducing system, competence
cps	capsular biosynthesis genes
CPS	capsular polysaccharide
CR	coding repeat
CreX	Cre tyrosine recombinase
CRP	C-reactive protein
CSF	cerebrospinal fluid
CSP	competence-stimulating peptide
Dam	deoxyadenosine methlytransferase
dexB	glucan 1,6-alpha-glucosidase
divIVA	cell division initiation gene
DNA	deoxyribonucleic acid
DnaJ	chaperone protein DnaJ
DnaJ/K	molecular chaperone
DOC	sodium deoxycholate
ECM	extracellular matrix
EDTA	ethylene-diamine-tetra-acetic-acid disodium salt
EF-G	elongation factor G
EF-Ts	elongation factor Ts
emPAI	exponentially modified protein abundance index
Erm/Ery	erythromycin
Fba	fructose-bisphosphate aldolase
FDR	false discovery rate
fH	factor H

FRR/EF-4	ribosome-recycling factor
fucK	putative L-fuculose kinase
FusA	translation elongation factor G
g/L	grams per litre
G	guanine
g	relative centrifugal force
G-3-P	D-glyceraldehyde-3-phosphate
galU	UTP-Glc-1-phosphate uridylyltransferase
GAPDH	glycerolaldhyde-3-phosphate dehydrogenase
Gent	gentamicin
GlcNAc	N-acetyl-glucosamine
GlpO	Alpha-glycerophosphate oxidase
GMP reductase/*guaC*	guanosine 5'monophosphate oxidoreductase
Gor	glutathione reductase
GroEL	heat shock protein
GroES	heat shock protein
h	hour/s
hbMEC	human brain microvascular endothelial cells
Hic	factor H-binding inhibitor of complement
HIV	human immunodeficiency virus
Hyl	hylauronate lyase
hsd	host-specificity-determinant gene
hsdM	host-specificity-determinant gene for modification
hsdR	host-specificity-determinant gene for restriction
hsdS	host-specificity-determinant gene for sequence specificity
i.n.	intranasal
i.p.	intraperitoneal
i.v.	intravenous
IEF	isoelectric focusing
IL-1	interleukin-1
IL-12	interleukin-12

ileS	isoleucyl-tRNA synthetase
IMP	inosine phosphate
INDEL	insertion/deletion
INF-γ	interferon γ
IPD	invasive pneumococcal disease
IPS	internal pool standard
IPTG	isopropyl-β-D-thio-galactopyranoside
ISP	Ion Sphere™ Particles
kb	kilobase/s
kDA	kilodalton/s
KEGG	Kyoto Encyclopedia of Genes and Genomes
kg	kilogram/s
l	litre/s
LB	Luria Bertani broth
LC-ESI-IT MS	liquid chromatography electrospray ionisation ion-trap mass spectrometry
LctO	lactate oxidase
Ldh	lactate dehydrogenase
lgtA	beta-*N*-acetylglucosaminyltransferase
Liv	branched-chain amino acid ABC transporter
LPS	lipopolysaccharide
Lrp	leucine response regulatory protein
LuxS	S-ribosylhomocysteine lyase
lysM	lysin motif
LytA	N-acetylmuramoyl-L-alanine amidase
M	molar
Mal	maltose operon transcriptional protein
Mb	megabase/s
MDR	multi-drug-resistant
mg	milligram/s
mg/ml	milligram/millilitre

min	minute/s
ml	millilitre/s
mM	millimolar
Mod	methyltransferase
mRNA	messenger RNA
MurC	UDP-N-aceytylmuramate—alanine ligase
NAD	β-Nicotinamide adenine dinucleotide hydrate
NanA	neuraminidase A
ng	nanogram/s
ng/ml	nanogram/millilitre
nm	nanometers
nt	nucleotide/s
Nrd	anaerobic ribonucleoside triphosphate reductase
O	opaque variant
OD	optical density
OE-PCR	overlap-extension polymerase chain reaction
Opa	opacity outer-membrane proteins
ORF	open reading frame
p.s.i.	per square inch
PacBio	Pacific BioSciences
PAFr	platelet-activating factor receptor
PAGE	polyacrylamide gel electrophoresis
Pap	pyelonephritis-associated pilus
PbcA	C3-binding protein A
PBMC	human peripheral blood mononuclear cells
PBP2B	penicillin-binding protein 2B
PBS	phosphate buffered saline
PcpA	pneumococcal choline-binding protein A
PCR	polymerase chain reaction
PCV13	pneumococcal conjugate vaccine, 13 serotypes
PCV7	pneumococcal conjugate vaccine, 7 serotypes

PECAM-1	platelet endothelial cell adhesion molecule-1
Pfl	pyruvate formate lyase/formate acetyltransferase
PGM	personal genome machine
pgm	phosphoglucomutase
Pht	polyhistidine triad
PiaA	pneumococcal iron uptake
PIgR	polymeric Ig receptor
Pit	pneumococcal iron transport
PiuA	pneumococcal iron uptake
PlgR	polymeric Ig receptor
Ply	pneumolysin
PMT	photomultiplier
PNSP	penicillin-non-susceptible *S. pneumoniae*
PotD	spermidine/putrecine ABC transporter
Ppc	phosphoenolypyruvate carboxylase
PPI-1	pneumococcal pathogenicity island 1
PpmA	putative proteinase maturation protein A
PPSV23	pneumococcal polysaccharide vaccine, 23-valent
PrsP	pneumococcal serine-rich protein
PsaA	pneumococcal surface adhesin A
PspC	pneumococcal surface protein C
PurA	adenylosuccinate synthetase
Pyk	pyruvate kinase
qRT-PCR	quantitative real-time PCR
qRT-PCR	real-time reverse transcription polymerase chain reaction
radC	DNA repair gene
RE	restriction endonuclease
RM	restriction-modification
RNA	ribonucleic acid
rpsA	ribosomal protein S1 gene
RT-PCR	reverse transcription polymerase chain reaction

s	second/s
SB	serum broth
SDH	streptococcal surface dehydrogenase
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyarcylamide gel electrophoresis
sIgA	secretory component of immunoglobulin A
SMRT	single-molecule real-time
SNPs	single nucleotide polymorphisms
SpsA	secretory IgA-binding protein
SpxB	pyruvate oxidase
sRNA	short non-coding RNA
SsbB	single-stranded DNA-binding protein
SSM	slipped-strand mispairing
T	transparent variant
T	thymine
TA	teichoic acid
TAE	tris-acetate EDTA buffer
TBE	tris, borate and EDTA
THY	todd-Hewitt broth supplemented with yeast extract
THY+catalase	THY supplemented with catalase
TLR	toll-like receptor
TNF	tumour necrosis factor
TRD	target recognition domain
TSB	tryptic soy broth
TTBS	tris-buffered saline containing Tween 20
U	unit
U/ml	unit/millilitre
UDP-Glc	UDP-glucose
v/v	volume per volume
w/v	weight per volume
WT	wild-type

wzy	capsular biosynthesis gene
ZMW	zero-mode waveguide

Acknowledgements

I would like to firstly thank my principal supervisor, Professor James Paton for the opportunity to work on this project under his supervision and guidance. Thank you for the support, patience and advice throughout the course of my studies. Thank you also to Dr. David Ogunniyi for his guidance and helping me brainstorm ideas pertaining to my project, especially during the crucial first few months of my studies. I would also like to extend my gratitude to both past and present members of the Paton Laboratory, in particular, Dr. Richard Harvey, Dr. Lauren McAllister and Dr. Claudia Trappetti for your constant encouragement and letting me pick their brains to get through the last leg of my journey into finishing up the write-up. To Dr. Danny Wilson, Dr. Adrienne Paton, Cathy Scougall, Zarina Amin, Dr. Layla Mahdi, Dr. Melanie Higgins, Dr. Austen Chen, Dr. Charlie Plumptre, Dr. Catherine Hughes, Dr. Trisha Rogers, Dr. Hui Wang, Dr. Adam Potter, Stephanie Philp, Sarah Sims and Rory Leahy for contributing to the vibrancy of the place throughout my time in the Paton Laboratory. Also, to Dr. Florian Weiland, Mark Van der Hoek and Rosalie Kenyon for their advice and technical skills relating to certain areas of my project.

Thank you to the badminton folks, especially, Michael, Des and Paul, who have taken me in when I first arrived in Adelaide. Thank you for all the laughter, fun, pain but most of all, your company both on and off court. You have all made settling into Adelaide a much easier transition. To Agnes and Catherine, even though we are miles apart, I will always appreciate the friendship and chats (and the reminder about all things Kiwiana) about anything and everything. I look forward to many more adventures together.

A huge and special thanks goes out to my family, for their constant encouragement, support and someone to talk to, when I needed it most, no matter the time. Also, lots of love goes out to Rebecca and her family, especially the little girlies and their random FaceTime calls. I guess being a night owl helps with the time difference! Love you all lots.

Finally, to anyone that I may have not named (you know who you are), thank you for being part of my journey and making it such an enjoyable one.

Chapter 1: INTRODUCTION

1.1 The Pneumococcus

Streptococcus pneumoniae (the pneumococcus) is a significant cause of human mortality and morbidity and is a leading cause of diseases such as pneumonia (Madeddu *et al.*, 2010; McCullers & Tuomanen, 2001; Ochoa *et al.*, 2010), otitis media (Arguedas *et al.*, 2010), meningitis (Ochoa *et al.*, 2010; Rueda *et al.*, 2010) and bacteraemia (Rueda *et al.*, 2010). The pneumococcus is a Gram-positive, encapsulated, alpha haemolytic diplococcus, and is almost exclusively a human pathogen. Virtually all strains of *S. pneumoniae* are enveloped by a layer of capsular polysaccharide (CPS) that exists as more than 90 serologically and structurally distinct types (Bentley et al., 2006). The capsule is an important feature of the pneumococcus, as it masks highly immunogenic cell surface molecules and limits accessibility of deposited host complement components to the host immune cells (Hyams *et al.*, 2012) (Figure 1.1).

1.1.1 Burden of Disease

S. pneumoniae infection is most common in the young (under the age of five years) and the elderly (over 65 years). Its main ecological niche is the human nasopharynx, where it remains asymptomatic in more than 30% of children and up to 10% of the adult population (Rueda *et al.*, 2010). However, it has the ability to spread form the nasopharynx to other sites, causing invasive pneumococcal disease (IPD) (that is, invasion of the bacterium into normally sterile body sites such as blood or cerebrospinal fluid) in a proportion of the carriers (Lynch & Zhanel, 2010). It has been reported that worldwide, an estimated 1.6 million people, about 60% of which are children less than five years old, die of IPD annually (Lynch & Zhanel, 2010). The risk factors for nasopharyngeal carriage of pneumococci in children include being less than five years old, having young siblings, poverty, malnutrition, poor access to medical care and being in crowded areas such as day care centres or school (Harboe *et al.*, 2009; Lynch & Zhanel, 2010). In adults, the risk

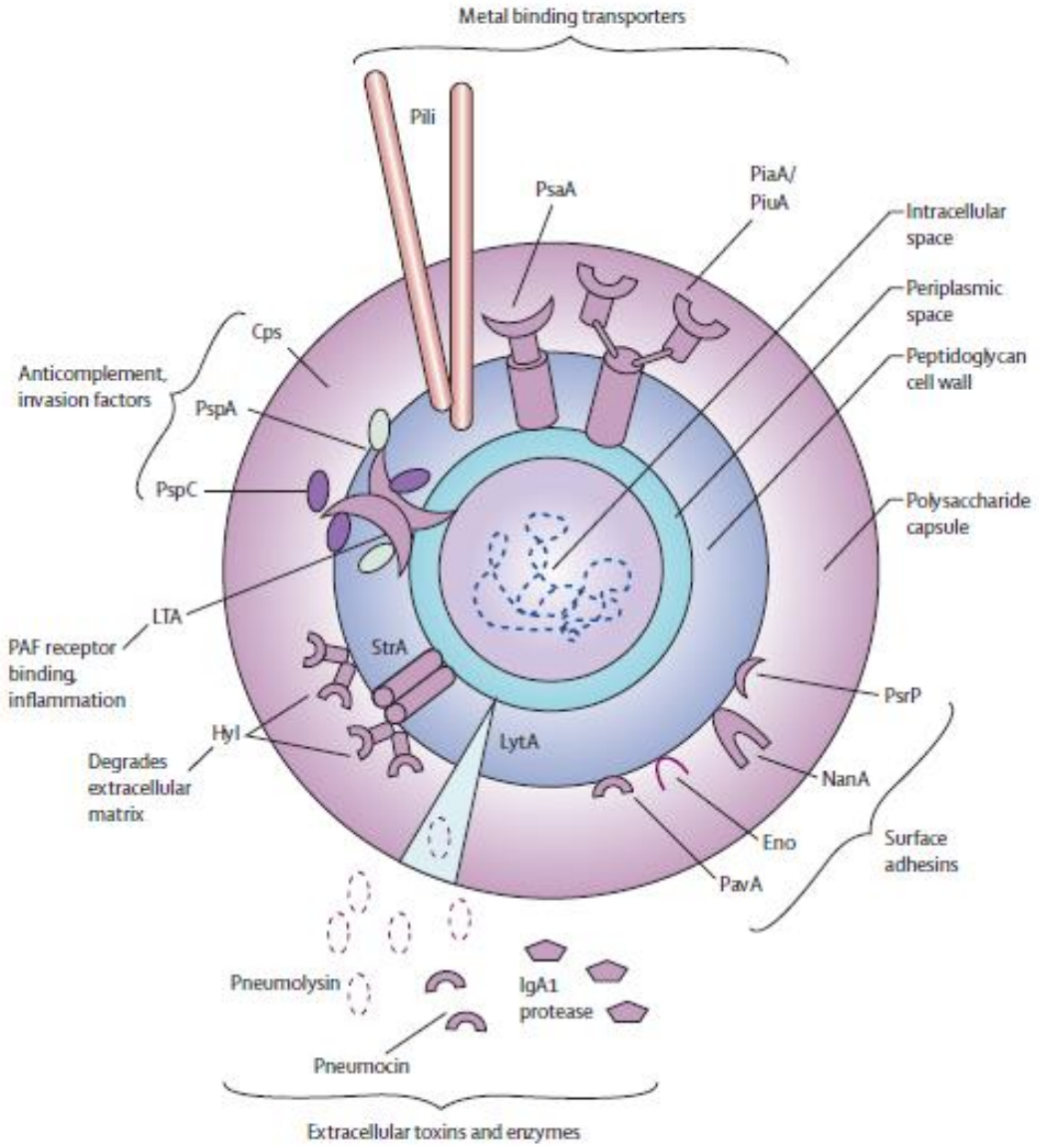


Figure 1.1. Virulence factors of *S. pneumoniae.*

A schematic representation of the cellular location of key pneumococcal virulence factors. Abbreviations that are not defined in the figure are: PsaA (pneumococcal surface antigen A), PiaA/PiuA (pneumococcal iron acquisition/uptake), PsrP (pneumococcal serine-rich repeat protein), NanA (neuraminidase A), Eno (enolase), PavA (pneumococcal adherence and virulence factor A), Hyl (hylauronate lyase), PspC (pneumococcal surface protein C), PspA (pneumococcal surface protein A), Cps (capsular polysaccharide). Figure is reproduced with permission from van der Poll and Opal (2009) (Rightslink licence number 3692260646220).

factors for nasopharyngeal carriage include cigarette smoking, asthma, being over 65 years old, being immune-compromised and the presence of an acute respiratory infection (Lynch & Zhanel, 2010). Harboe et al. (2009) found that of 19,000 IPD cases in Denmark, the mortality rate in those less than five years old was 5%; for 5-65 year-olds it was 14%, and for over 65 year-olds it was 24%, indicating that age and co-morbidities can influence mortality rates. However, the mortality rate of young children in developing countries was 10-40%, which may partly be due to poorer healthcare systems, more limited access to antimicrobials and the presence of co-morbidities such as HIV infection and malnutrition (Harboe et al., 2009; Madeddu et al., 2010). Geographical region and ethnicity have a role in community-acquired pneumococcal (CAP) infections. For example, indigenous peoples of Alaska and the Canadian arctic, American Indians, African Americans, Australian Aborigines, Maoris in New Zealand and Bedouins of Israel are more predisposed to CAP infections than others (Lynch & Zhanel, 2010).

Other respiratory diseases such as influenza can be complicated by secondary pneumococcal infections, particularly pneumonia, resulting in increased deaths. This is exemplified by the Spanish influenza pandemic of 1918-19; a significant proportion of the fatalities attributed to the flu were actually due to secondary bacterial, particularly pneumococcal, infection (Morens *et al.*, 2008). More recently, in a study of 128 patients during the 2009 H1N1 swine flu pandemic, 33% of cases involved bacterial co-infection, and in 62% of these, *S. pneumoniae* was the aetiological agent (Cilloniz *et al.*, 2012). Human immunodeficiency virus (HIV)-infected individuals are also at high risk of developing pneumococcal infection and this is exacerbated by cigarette smoking, intravenous drug abuse, immune deficiency, older age and alcohol abuse (Cilloniz *et al.*, 2012; Madeddu *et al.*, 2010).

1.1.2 Antimicrobial resistance

Prior to the 1970s, *S. pneumoniae* was almost invariably susceptible to penicillin and most other antimicrobial agents. However, following that, there was a steady increase in the detection of penicillin-non-susceptible *S. pneumoniae* (PNSP), followed by resistance to macrolides and finally, the appearance of multi-drug-resistant (MDR) *S. pneumoniae*

(Lynch & Zhanel, 2010). PNSP was first described in Australia in 1967, followed by reports from New Guinea in 1974, South Africa in 1977 and Spain in 1979 (Linares *et al.*, 2010). By the 1980s, there was worldwide distribution of PNSP. However, different geographic regions have varying degrees of antimicrobial resistance. For example, by 2007, 26% of pneumococci were PNSP in the United States of America (Gertz *et al.*, 2010) and by 2008 in Lima, Peru, 22.8% of children under the age of 24 months were found to have PNSP (Ochoa *et al.*, 2010), whereas Kempft *et al*. (2010) showed that by 2007, 38.2% of *S. pneumoniae* isolates were PNSP in France. The common use of antimicrobials in the early 1980s almost certainly had a role in the emergence and rapid rise in MDR strains of *S. pneumoniae* (Goossens *et al.*, 2007; Pallares *et al.*, 1987), which now contributes to the problem of management of pneumococcal diseases. Currently, depending on hospitals, the recommended treatments for pneumococcal infection include a β-lactam, macrolide or fluoroquinolone, either alone or in certain combinations (Caballero & Rello, 2011).

1.2 Pneumococcal Vaccines

Although antimicrobials are available to treat pneumococcal infections in most parts of the world, the rapid rise in the prevalence of antibiotic-resistant pneumococci means that alternative methods of management of the disease are required, particularly preventative methods such as vaccination (Black *et al.*, 2000). The earliest widely available vaccine against the pneumococcus was targeted at the CPS of 14 of the most commonly isolated *S. pneumoniae* serotypes in the USA in 1977 (Broome & Facklam, 1981), which covered about 70% of pneumococcal isolates. This was increased to a 23-valent vaccine, PPSV23 (Pneumovax 23, manufactured by Merck) in 1983, covering over 80% of isolates in the USA (Robbins *et al.*, 1983). Furthermore, Robbins *et al.* (1983) also showed that there was cross-reactivity between some closely-related serotypes (e.g. 6A and 6B), only one of which was included in PPSV23. PPSV23 is recommended for adults >65 years and younger patients with conditions that predispose them to pneumococcal infection, including asplenia, HIV, sickle cell anemia and chronic cardiovascular disease, to name a few (ACIP, 1997). However, as PPSV23 elicits a T-cell-independent immune response, it

is ineffective in infants due to their immature immune systems. The lack of immunogenicity of polysaccharides in children under the age of two years was demonstrated by the lack of significant increases in antibody titres, particularly for the serotypes that most commonly cause paediatric IPD (Douglas *et al.*, 1983). Furthermore, the mean antibody titres fell for most of the vaccine serotypes after a booster six months later (Douglas *et al.*, 1983). Although the 23-valent vaccine is effective among young adults, the poor clinical efficacy of PPSV23 among the very young has also been demonstrated in the elderly. Despite both young and elderly adults producing similar concentrations of anti-CPS antibodies, there was a decrease in the efficacy of the antibody obtained from the elderly, which could be partly attributed to a reduced functional antibody due to a senescent immune system (Westerink *et al.*, 2012).

Efforts to overcome the problems associated with the polysaccharide vaccines led to the production of conjugate vaccines, which induce a T-cell-dependent response, producing long-lived immunoglobulins and memory B cells. The earliest, most extensively used heptavalent pneumococcal conjugate vaccine (PCV7) only protects against seven serotypes which were most prevalent in the USA and Europe at the time of its licensure in 2000, namely types 4, 6B, 9V, 14, 18C, 19F and 23F (Black *et al.*, 2000). Conjugate vaccines covering 13 serotypes (PCV13, Prevenar 13®, manufactured by Pfizer) have since been licensed, but still have low coverage, considering the number of identified pneumococcal capsular serotypes (>90). Prevenar 13® is conjugated to a non-toxic diphtheria protein (cross-reactive material [CRM_{197}]) and has six extra serotypes (serotypes 1, 3, 5, 6A, 7F and 19A) in addition to those covered by PCV7, which cause 2.5-20% of IPD in Europe (Duggan, 2010). Vaccination of children with PCV7 reduced the incidence of disease covered by the serotypes, even among the unvaccinated due to a herd immunity effect (Black *et al.*, 2000; Black *et al.*, 2006). However, the elimination of one particular serotype from their ecological niche by the vaccine may give rise to occupation by pre-existing non-vaccine serotypes, potentially altering the serotype distribution of disease-causing strains rather than reducing the overall incidence of disease (Weinberger *et al.*, 2011). As the pneumococcus is naturally transformable, serotype-replacement could also occur via genetic recombination that results in the switching of CPS biosynthesis genes

from a vaccine serotype and replacement with those of a non-vaccine serotype (Croucher *et al.*, 2011). Hence, a highly invasive strain covered by the vaccine can now express a different capsular serotype not covered by the vaccine. For example, in USA, between 1998 and 2004, there was a reduction in PCV7 serotypes but an increase in serotypes 3, 15, 19A, 22F and 33F in those <5 and ≥65 years, with 19A being the dominant serotype (Hicks *et al.*, 2007). Furthermore, a study conducted in France of patients (6-24 months old) had a reduction of PCV7 carriage serotypes from 44.5% to 1.2% over a nine year period (2001-2010) (Cohen *et al.*, 2015). The same study also found that the increase of serotype 19A from 2001 (8.6%) to 2010 (15.8) was reduced to 1.2% in 2014, after the introduction of PCV13 in 2010. Another set-back to conjugate vaccines is that they are costly, so it is more difficult for them to be widely used in developing countries, where pneumococcal diseases are most prevalent and mortality is greatest (Brueggemann *et al.*, 2007; Hicks *et al.*, 2007; Spratt & Greenwood, 2000). However, through organisations such as the GAVI Alliance, these vaccines are currently provided to third world countries at a much reduced cost (up to 90% lower than that of doses supplied in the USA and Europe) (GAVI, 2013). It has been recommended that adults over 65 years use both PPV23 and PCV13 (Tomczyk *et al.*, 2014) and that children at risk of developing pneumococcal infections to receive the 13-valent conjugate vaccine (Mirsaeidi & Schraufnagel, 2014). The latter also does not interfere with other paediatric vaccines, and thus can be administered concurrently (Duggan, 2010).

Due to the limitations of both conjugated and unconjugated CPS vaccines, there is a need to develop alternatives that have better coverage of the disease-causing strains that also elicit a strong immune response and are more affordable. One such alternative could be the use of proteins associated with virulence of *S. pneumoniae* (Briles *et al.*, 2003; Ogunniyi *et al.*, 2000; Rosenow *et al.*, 1997; Schachern *et al.*, 2014). These protein vaccines would include proteins that are common to all pneumococcal serotypes and elicit a strong immune response in infants. These proteins can be expressed at a high level using modern recombinant expression systems, enabling production at a relatively low cost and they also can be conjugated to a CPS (Paton *et al.*, 1991). Pneumococcal proteins that have been demonstrated to confer protection in mice include non-toxic derivatives of the toxin

pneumolysin (Ply), pneumococcal surface protein A (PspA), pneumococcal surface protein C (PspC; also known as choline binding protein A [CbpA]), polyhistidine triad (Pht) proteins, autolysin (LytA) (Feldman & Anderson, 2014) as well as an oligopeptide ABC transporter (AliA) (Ogunniyi *et al.*, 2012) and alpha-glycerophosphate oxidase (GlpO) (Mahdi *et al.*, 2012), to name a few.

Another alternative to CPS or subunit-based vaccine approaches is the use of whole, killed, unencapsulated pneumococci. The whole cell vaccine is effective against pneumococcal colonisation as it relies on $CD4^+$ T cells and the cytokine IL-17, eliciting a response to a large range of cell surface antigens (Lu *et al.*, 2008; Malley *et al.*, 2005). On the other hand, protection against invasive disease elicited by the whole cell pneumococcal vaccine is mediated by serum antibodies rather than T cells (Lu *et al.*, 2010). Therefore, whole cell vaccine is another method that can elicit effective protective immunity against both nasopharyngeal colonisation as well as invasive disease, by mimicking the development of natural immunity in humans (Malley & Anderson, 2012).

1.3 Pathogenesis of Pneumococcal Disease

1.3.1 Pneumococcal colonisation

The first step in development of pneumococcal disease is the colonisation of the human nasopharynx (Figure 1.2). Pneumococcal infection is acquired mainly through airborne droplets containing the bacterium and transferred from person to person, as there are no known animal reservoirs or insect vectors. *S. pneumoniae* is part of the commensal flora of the upper respiratory tract and shares this niche with other potentially pathogenic bacteria including *Hemophilus influenzae*, *Neisseria meningitidis*, *Staphylococcus aureus*, and other haemolytic streptococci (Faden *et al.*, 1990). The nasopharyngeal flora is established in the first months of life (Faden *et al.*, 1997) and it has a large turnover of colonising species and serotypes that occupy the niche during the early years, but a balance is usually reached. For example, some other α-haemolytic streptococci can inhibit colonisation by *S. pneumoniae*, *H. influenzae*, *S. aureus* and *M. catarrhalis* (Faden *et al.*, 1990; Ghaffar *et al.*, 1999). In turn, *N. meningitidis* and *S. pneumoniae* can form a positive relationship initially,

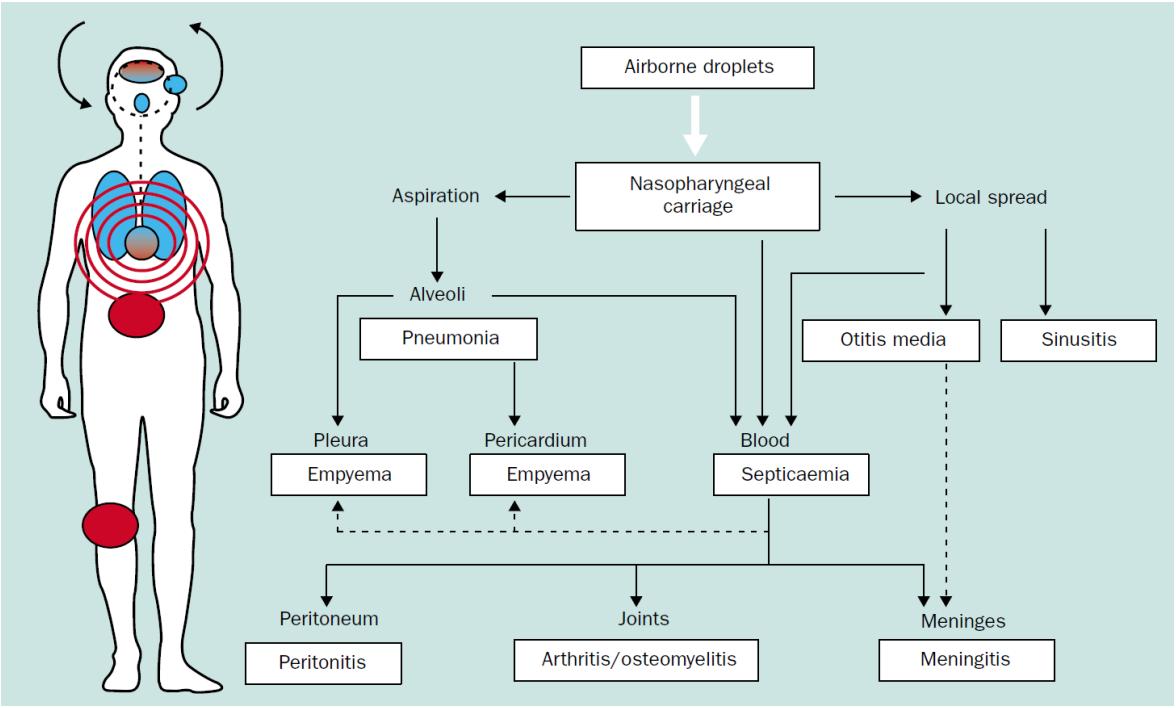


Figure 1.2. Progression of pneumococcal disease.

Pneumococcal colonisation begins in the nasopharynx when pneumococcal aerosols are inhaled. This may lead to clearance of the bacterium or asymptomatic carriage. Invasion of through the epithelial barrier in susceptible hosts leads to the spread of the bacterium into the lungs and bloodstream, causing pneumonia and septicaemia, respectively. Alternatively, progression to otitis media is common in young children. The pneumococcus can also cause inflammation in other parts of the body such as the pleura, pericardium, peritoneum, joints or sinuses. The most serious disease develops when the pneumococcus traverses the blood-brain barrier to cause meningitis. Figure is reproduced with permission from Bogaert *et al*. (2004) (Rightslink licence number 3506211191485).

but H_2O_2 production by *S. pneumoniae* can then interfere with the growth of *H. influenzae*, *M. catarrhalis* and *N. meningitidis* (Pericone *et al.*, 2000). Although most carriers of *S. pneumoniae* remain asymptomatic, in a small proportion of individuals, the organism is able to resist the innate defenses of the host and progress to infection of deeper tissues. Risk factors discussed in Section 1.1.1 also aid in this progression. Carriers are unaware or unaffected by the bacterium and this is believed to be an important feature as carriers aid in the dissemination of the bacterium. The carrier may unknowingly spread the organism to another individual who may be more susceptible to pneumococcal infections. Hence, in order to prevent dissemination of *S. pneumoniae* in the community, vaccine strategies are often focused on preventing nasopharyngeal colonisation in addition to invasive disease.

Colonisation by the pneumococcus requires interaction of the pneumococci with epithelial cells. The pneumococcal neuraminidase NanA may contribute to this by reducing the viscosity of the mucous layer by cleaving N-acetylneuraminic acid from glycosylated host proteins (for example, mucin) (King *et al.*, 2004). This increases exposure of N-acetyl-glucosamine (GlcNAc) on the epithelial cell surface allowing for adherence of the pneumococci to this sugar (Andersson *et al.*, 1983). This binding is thought to occur via non-specific, physiochemical interactions (Swiatlo *et al.*, 2002). During progression to infection, host-inflammatory factors such as interleukin 1 (IL-1) and tumour necrosis factor (TNF) are induced, changing the type and number of receptors on target epithelial and endothelial cells (Marriott *et al.*, 2012). One of the changes in response to cytokine stimulation of the host epithelial cells is upregulation of the platelet-activating-factor receptor (PAFr) to which pneumococcal cell-wall phosphorylcholine (ChoP) binds (Cundell *et al.*, 1995a). There is also an increase in the binding of choline-binding protein A (CbpA) to exposed sialic acid and lacto-*N*-neotetraose on eukaryotic cells (Rosenow *et al.*, 1997). Together, these events induce internalisation and translocation of the pneumococci through the respiratory epithelium and the vascular endothelium. Furthermore, the pneumococcus is able to exist in two phases, opaque (O) and transparent (T), a term used to describe their colonial morphology on a clear agar plate (discussed in Section 1.6). T variants are typically favoured in the nasopharynx, whereas, the O variants are typically isolated from deeper tissues (Weiser *et al.*, 1994). The involvement of

cytokines in facilitating adherence was demonstrated by the fact that the variant that preferentially colonises the nasopharynx has enhanced adherence after cytokine stimulation with IL-1 and TNF (Cundell *et al.*, 1995b). On the other hand, the ability of the O counterpart to adhere to buccal cells was not affected by cytokine stimulation.

1.3.2 Progression to disease

When a pathogen crosses the barrier from the nasopharynx to normally sterile sites such as the lungs, blood and the cerebrospinal fluid (CSF), it results in an invasive infection. One of the ways in which the host innate immune system reacts to invading pneumococci is through activation of the complement system, resulting in bacterial opsonisation and promotion of neutrophil chemotaxis, facilitating bacterial clearance. Although normally thought to occur principally via the classical and alternate pathways of complement deposition (Brown *et al.*, 2002b), it has been recently suggested that the lectin pathway also has a role in host-mediated clearance of the pneumococcus (Ali *et al.*, 2012). Recognition of the pneumococcal lipoteichoic acid, Ply and bacterial DNA in the host also initiates an inflammatory response via Toll-like receptor (TLR)-2, TLR-4 and TLR-9, respectively (Koppe *et al.*, 2012; Malley *et al.*, 2003). Over time, the innate immune responses via $CD4^+$ T cells are replaced by an adaptive immune response from the generation of antibodies against the pneumococcal antigens (Wilson *et al.*, 2014).

While the capsule aids the pneumococcus by preventing opsonisation (Kung *et al.*, 2014), other virulence factors such as Ply impairs ciliary beating frequency in the lungs impairing clearance and also damaging the alveolar barrier (Maus *et al.*, 2004), facilitating further penetration of the pneumococcus into the bloodstream (Steinfort *et al.*, 1989; Zysk *et al.*, 2001). Presence of the bacterium in the bloodstream allows for subsequent crossing of the blood-brain barrier (BBB) that separates the blood from the cerebrospinal fluid, into the brain compartment. The ability of the pneumococcus to enter the brain may lie in its ability to cause inflammation around endothelial cells; attachment of the pneumococci may cause damage to the microvascular endothelial walls, revealing an entry point for *S. pneumoniae* and recruitment of inflammatory cells to these areas, causing meningitis. One mechanism for attachment to the BBB endothelium employed by the pneumococcus

involves binding of CbpA to the laminin receptor (Orihuela *et al.*, 2009). Another protein thought to be contribute is GlpO, a protein that catalyses the oxidation of α-glycerophosphate to dihydroxyacetone phosphate, with concomitant H_2O_2 production (Claiborne, 1986). GlpO may be involved in the translocation from the blood into the brain, as mice challenged with a *glpO*-deficient strain had significantly less translocation to the brain compartment, but there was no difference in the number of bacteria in the blood when compared to mice challenged with the wild-type (WT) parent strain (Mahdi *et al.*, 2012). Furthermore, GlpO had a cytotoxic effect on human brain microvascular cells *in vitro* (Mahdi *et al.*, 2012). Recently, it has been found that brain microvasculature endothelium expressing platelet endothelial cell adhesion molecule-1 (PECAM-1) and PIgR (polymeric Ig receptor) interacts with blood-borne *S. pneumoniae* via pneumococcal adhesins, in particular, interaction between CbpA and PIgR (Iovino *et al.*, 2014; Zhang *et al.*, 2000). Binding to and penetrating the BBB endothelium provides a major mechanism by which the pneumococcus can invade the central nervous system (CNS) from the bloodstream. However, it has also been shown the pneumococcus can transmigrate directly from the nasal cavity to the CNS, without apparent bacteraemia, via interaction with the gangliosides on the olfactory nerves (van Ginkel *et al.*, 2003). Furthermore, clinical ear isolate of a *S. pneumoniae* strain have also been shown the traverse to the brain from the nasopharynx with minimal impact on the lungs and blood (Amin *et al.*, 2015)

During pneumococcal progression from carriage to disease, the pneumococcus is able to regulate gene expression in a niche-specific manner (Mahdi *et al.*, 2008; Orihuela *et al.*, 2004). Typically, the genes upregulated in the nasopharynx or on epithelial cells are cell surface genes whereas those upregulated in the blood are usually genes that would contribute to evasion of host clearance mechanisms, such as the production of the capsule (LeMessurier *et al.*, 2006; Mahdi *et al.*, 2008; Orihuela *et al.*, 2004).

1.4 *S. pneumoniae* Virulence Factors

The virulence factors of *S. pneumoniae* important for its survival in the host are numerous and diverse, and may include factors that promote colonisation, adaptation to nutrient availability or deprivation, evasion of the immune system or eliciting an inflammatory

response. Discussion of a substantial array of virulence factors characterised to date can be found in reviews by Kadioglu *et al.* (2008) and Mitchell & Mitchell (2010). Better known pneumococcal virulence factors include the capsule, Ply, choline-binding proteins (Cbps), ABC transporters, LPXTG-anchored proteins and the Pht family proteins, and these are discussed briefly below.

1.4.1 Pneumococcal capsule

The capsule forms the outermost layer of *S. pneumoniae*, is about 200-400 nm thick (Skov Sorensen *et al.*, 1988) and as mentioned previously, over 90 serologically distinct pneumococcal CPS have been identified to date (Bentley *et al.*, 2006). The capsule has been the principal target in the development of pneumococcal vaccines, and it is a crucial virulence factor for *S. pneumoniae*, due to its anti-phagocytic activity (Jonsson *et al.*, 1985), ability to limit mechanical removal by the mucus [thus aiding colonisation (Nelson *et al.*, 2007)], restriction of autolysis and reduction of the effects of antimicrobials (Mitchell & Mitchell, 2010).

There are two mechanisms of pneumococcal capsule biosynthesis - Wzy-dependent and synthase-dependent. In Wzy-dependent capsule biosynthesis, the pneumococcal capsule biosynthesis locus (*cps*) is located in the chromosome between *dexB* and *aliA*. The 5' region of the *cps* locus comprises four highly conserved genes (*cpsA-D; wzg, wzh, wzd, wze*) which are involved in regulation of chain length and attachment of CPS to the cell wall. These genes are common to all known serotypes, except serotypes 3 and 37. The central region of the *cps* locus comprises serotype-specific genes encoding glycosyltransferases, polymerisation (Wzy) and transport (Wzx or "flippase"). The 3' region is not unique to any given serotype and comprises genes encoding synthesis of nucleotide disphospho-sugar precursors and sugar modification enzymes (e.g. O-acetylases) (Morona *et al.*, 1997; Paton *et al.*, 1997; Tettelin *et al.*, 2001). CPS synthesis commences with assembly of the serotype-determining oligosaccharide repeat units on the inner side of the cell membrane by the glycosyltransferases. This is initiated by CpsE, an integral membrane protein, which attaches the first sugar (galactose or glucose) to a lipid carrier. Completed repeat units are then translocated by the aptly named Wzx flippase to

the external surface of the membrane (Islam & Lam, 2013). The Wzy polymerase then transfers the repeat unit to the growing lipid-linked CPS chain, which is finally ligated to the cell wall by CpsA (Eberhardt *et al.*, 2012).

In contrast, serotypes 3 and 37, which have simpler CPS structures, are synthesized by an alternative, much simpler, synthase-dependent mechanism. In serotype 3, the 3'-end of the *cps* locus contains several truncated genes, and only two intact genes are required for CPS synthesis. These are *cps3D* (*ugd*), which oxidises UDP-glucose to form UDP-glucuronic acid, and *cps3C* (*wchE*) encoding a synthase which is responsible for the initiation, polymerization and transport of the polymer (Arrecubieta *et al.*, 1995; Dillard *et al.*, 1995; Kelly *et al.*, 1994). On the other hand, the only gene required for serotype 37 CPS synthesis is the *tts* gene encoding the synthase, but this is located elsewhere on the chromosome (Llull *et al.*, 1999). Unlike Wzy-dependent capsules, which are covalently linked to the peptidogylcan, these capsules are attached to the cell via phosphotidylglycerol or interactions with the synthase. This may account for the large mucoid colony phenotype seen in serotypes 3 and 37 (Knecht *et al.*, 1970), a feature also seen in *cpsC* mutants of the Wzy-dependent serotype 2 strain D39 with defects in cell wall attachment (Byrne *et al.*, 2011).

The capsule has several properties that aid evasion of opsonophagocytosis, including the strong negative charge of the majority of CPS serotypes at physiological pH. Indeed, the more charged the CPS, the more resistant the respective serotype is to killing by neutrophils (Lee *et al.*, 1991; Li *et al.*, 2013). The capsule also forms a barrier that impairs opsonisation via C3b/iC3b by inhibiting the conversion of C3b bound to the bacterial surface to iC3b (Hyams *et al.*, 2012). The capsule is able to mask the subcapsular proteins and other cell wall components from being recognised by host antibodies. It also inhibits classical complement pathway activity by restricting binding of immunoglobulin G (IgG) and C-reactive protein (CRP) to the pneumococci and thus are less susceptible to phagocytosis by neutrophils (Hyams *et al.*, 2010a). CPS can also modulate the immune response by suppressing the release of pro-inflammatory cytokines such as CXCL8 (IL-8) and IL-6, reducing clearance of pneumococci from the upper respiratory tract (Kung *et al.*, 2014). Additionally, the negative charge of the capsule can also aid in colonisation by

reducing mucus-mediated clearance, allowing persistence of the bacterium in the nasopharynx. Unencapsulated pneumococci were more likely to remain attached to the luminal mucus compared to their encapsulated counterparts, with the latter more likely to adhere to the epithelial surface enabling more stable colonisation (Nelson *et al.*, 2007). In addition to evading to host immune responses, the capsule can also have a protective effect against lysis-causing antibiotics such as penicillin and vancomycin (Fernebro *et al.*, 2004).

CPS serotype has a role in determining the virulence of pneumococcal strains, as resistance to complement-mediated immunity can vary between serotypes, independent of genetic make-up (Hyams *et al.*, 2010b; Melin *et al.*, 2010b). In a study where the capsule of different serotypes were replaced onto a TIGR-4 strains, the mutants reacted differently to complement deposition and neutrophil phagocytosis (Hyams *et al.*, 2010b). Mutants with increased opsonisation also showed reduced virulence in a mouse sepsis model, suggesting that the capsule may have a role in the pathogenicity variation between serotypes.

1.4.2 Pneumolysin

Ply is a 53 kDa cholesterol-dependent, pore-forming toxin expressed by almost all *S. pneumoniae* clinical isolates (Benton *et al.*, 1997). A recent study has shown that its interaction with target cells commences with binding to glycan (particularly sialyl Lewis X) receptors (Shewell *et al.*, 2014), followed by interaction with membrane cholesterol and oligomerisation of up to 50 Ply molecules to form a trans-membrane pore approximately 30 nm in diameter (Morgan *et al.*, 1995). Ply lacks a signal peptide or any other export signal; originally it was thought to only be released with the aid of autolysins such as LytA (Berry *et al.*, 1989a), although it was later shown that it can be released independently of LytA in at least some strains (Balachandran *et al.*, 2001). Although the precise mechanism of Ply secretion/export remains uncertain, more recent studies have shown that it associates with the bacterial cell wall compartment. Price and Camilli (2009) showed that Ply is non-covalently attached to the cell wall, while still possessing haemolytic acitivity and is accessible to extracellular protease. Furthermore, the study showed that localisation of Ply

to the cell wall is independent of LytA, as Ply was still localised to the cell wall in a *lytA* mutant.

Ply has multiple functions, including pore-formation (cytolytic activity) (Tilley *et al.*, 2005), complement activation (Paton *et al.*, 1984) and interaction with TLR4 (Malley *et al.*, 2003). Furthermore, Ply can activate NADPH oxidase in human neutrophils and cause oxygen radicals to be produced in an intracellular compartment rather than being excreted extracellularly or to the phagosome, where the invading pneumococci would have been localised (Martner *et al.*, 2009). Ply has a role in bacteraemia (Benton *et al.*, 1995), pneumonia (Berry *et al.*, 1989b), and deafness due to pneumococcal meningitis (Wellmer *et al.*, 2002). Ply is able to reduce human ciliary beating and damage the respiratory epithelium *in vitro* (Feldman *et al.*, 1990), contributing to development of pneumococcal pneumonia. In murine models of bacteraemia, both Ply-deficient and Ply-expressing pneumococcal strains have high numbers during initial 10-12 hours post-infection. This level is maintained at 10^6-10^7 cfu/ml by mice infected with Ply-deficient strains for several days, whereas mice infected with Ply-expressing pneumococci had elevated numbers (10^9-10^{10} cfu/ml) at the time of death of 24-28 hours post-infection (Benton *et al.*, 1995). Hence, in the absence of Ply, chronic bacteraemia can develop, as opposed to acute sepsis seen with Ply-expressing pneumococci. Ply also aids the spread of the bacterium from the lungs into the bloodstream (Berry *et al.*, 1989b; Kadioglu *et al.*, 2002). On the other hand, there is debate as to whether Ply has a role in inflammation associated with meningitis. An earlier report suggested that there was little involvement of Ply in pneumococcal meningitis in a rabbit model (Friedland *et al.*, 1995). However, subsequent studies found that Ply was an important contributing factor (Braun *et al.*, 2002; Wellmer *et al.*, 2002). For example, Braun *et al.* (2002) found that the loss of Ply or H_2O_2 production led to a decrease in the condensation and apoptosis in the neurons of the hippocampus, compared to the wild-type. It was believed that neuronal cell death was due to the ability of both Ply and H_2O_2 to increase intracellular Ca^{2+} thus inducing the production of apoptosis-inducing factor which then leads to mitochondrial damage and cell death.

1.4.3 Choline-binding proteins

Pneumococcal choline-binding proteins (Cbps) are cell-surface proteins which are anchored to the cell wall by binding to phosphorylcholine moieties on either the teichoic acid (TA) or lipoteichoic acid, via a choline-binding domain consisting of 2 to 10 repeats of a 20 amino acid sequence (Garcia *et al.*, 1998; Giffard & Jacques, 1994; Wren, 1991). Cbps have multiple roles in pneumococcal colonisation and invasion, including adherence, autolysis and inhibition of complement activation. Currently, over 15 Cbps have been described, including PspA, PspC/CbpA and LytA (Perez-Dorado *et al.*, 2012).

The complement system is an important part of the innate immune defense, whereby an invading organism is rapidly identified and destroyed. PspA offers protection from complement-mediated clearance by inhibiting complement activation (Ren *et al.*, 2004a; Ren *et al.*, 2004b; Tu *et al.*, 1999); it also protects against the bactericidal effects of apolactoferrin (Shaper *et al.*, 2004). PspA-knockout mutants have enhanced killing by apolactoferrin compared to their wild-type counterparts and antibodies to PspA also facilitate killing of the pneumococcus (Shaper *et al.*, 2004). The importance of cell surface proteins such as PspA in pneumococcal virulence was first demonstrated by Briles *et al.* (1988), where *pspA* mutants had attenuated virulence in a mouse model of disease, and that two of the three strains tested became completely avirulent. Additionally, replication of PspA mutants were also impaired in the blood and brain of infected mice (Briles *et al.*, 1988; Ricci *et al.*, 2013).

CbpA has been named independently by various groups according to its multiple functions/properties: pneumococcal surface protein C (PspC), *S. pneumoniae* secretory IgA-binding protein (SpsA), C3-binding protein A (PcbA) and Factor H (fH)-binding inhibitor of complement (Hic) (Cheng *et al.*, 2000; Hammerschmidt *et al.*, 1997; Iannelli *et al.*, 2002; Janulczyk *et al.*, 2000; Rosenow *et al.*, 1997). In most strains, CbpA is attached to the cell wall via its C-terminal choline-binding domain. However, some pneumococci, such as the type 3 strain A66, produce a variant form (Hic) which has the same N-terminal region and biological functions, but is attached to the cell wall via an LPXTG motif in its C-terminal region, rather than via a choline-binding domain (Janulczyk *et al.*, 2000). CpbA

prevents complement-mediated opsonisation by binding to fH (Dave *et al.*, 2004; Quin *et al.*, 2005). It also binds to the secretory component of immunoglobulin A (sIgA) and can therefore facilitate the translocation of the bacterium across the respiratory epithelium by interaction with the pIgR (Dave *et al.*, 2004; Hammerschmidt *et al.*, 1997; Zhang *et al.*, 2000). Both PspA and CbpA contribute to colonisation by facilitating binding of *S. pneumoniae* to the epithelial cell surface of the human nasopharynx (Mitchell & Mitchell, 2010), and a mutant strain lacking both PspA and CbpA was shown to have significantly reduced ability to colonise the nasopharynx of mice (Ogunniyi *et al.*, 2007).

The major pneumococcal autolysin LytA is an amidase that cleaves the N-acetyl-muramoyl-L-alanine bond of the pneumococcal peptidyglycan and this action leads to cell lysis (Howard & Gooder, 1974). Mutations of this gene attenuated the virulence of *S. pneumoniae* in mouse and rat challenge models (Berry *et al.*, 1989a; Ng *et al.*, 2002). Previously, the contribution of LytA to virulence was thought to be due to its role in release of cell-associated toxins such as Ply and pro-inflammatory cell wall fragments (Berry *et al.*, 1989a; Kadioglu *et al.*, 2008). More recently however, LytA has been demonstrated to contribute to virulence through complement evasion that is independent of Ply (Ramos-Sevillano *et al.*, 2015); a *lytA* mutant had increased C3b, CRP and C1q binding to the surface of the pneumococcus compared to a *ply* mutant, but these parameters were markedly increased in a *lytA* and *ply* double mutant, confirming the roles of both LytA and Ply in complement resistance. LytA appears to inhibit complement deposition by recruiting the fluid-phase downregulators C4BP and fH, neither of which are involved in Ply-mediated complement inhibition (Ramos-Sevillano *et al.*, 2015). The release of bacterial products into the host caused by LytA-mediated autolysis also appears to inhibit phagocytosis by eliciting fewer phagocyte-activating cytokines from human peripheral blood mononuclear cells (PBMC) than intact pneumococci, thereby indirectly protecting unlysed pneumococci from phagocytosis (Martner *et al.*, 2009). Furthermore, increased pro-phagocytic cytokines such as TNF, interferon γ (INF-γ) and IL-12 were induced by LytA-deficient mutants in PBMC compared to the WTs (Martner *et al.*, 2009).

1.4.4 Transport and sequestration of nutrients in the host

Bacterial ABC transporters are an important class of membrane transporters involved in the import and export a variety of cargos, some of which are critical for the survival of the bacterium. One such example is PsaABC. PsaA, the solute binding component of an ABC transporter with specificity for manganese, in particular, has been recognised as a major pneumococcal virulence factor. Deletion of any of the genes in the *psaBCA* locus has massive effects on the virulence of *S. pneumoniae*, and PsaA-deficient mutants are unable to survive *in vitro* unless the medium is supplemented with manganese (Dintilhac *et al.*, 1997; McAllister *et al.*, 2004). As manganese is a co-factor for superoxide dismutase, mutants lacking *psaA* are hypersensitive to oxidative stressors such as superoxide and H_2O_2 (Tseng *et al.*, 2002). PsaA is also highly expressed during infection (Orihuela *et al.*, 2004) and may have an indirect role in adherence to epithelial cells (Berry & Paton, 1996; Novak *et al.*, 1998).

Other cation uptake ABC transporters that are involved in pneumococcal virulence include the iron transporters, Piu, Pia and Pit and the zinc transporters, AdcA and AdcAII (Bayle *et al.*, 2011; Brown *et al.*, 2001; Brown *et al.*, 2002a). Strains with mutations in *pit* have impaired growth in medium containing an iron chelator, increased sensitivity to an iron-dependent antibiotic and impaired virulence in a mouse model (Brown *et al.*, 2002a). Additionally, these iron transporters have site specific effects on virulence, whereby the loss of *pit* impacts on systemic virulence, whereas the loss of *piu* affects pneumococcal pneumonia (Brown *et al.*, 2002a).

AdcA and AdcAII are distinct zinc-specific binding proteins which both function through the same permease (AdcBC); both are required for efficient zinc transport *in vitro* and *in vivo* (Bayle *et al.*, 2011). Other ABC transporters with roles in murine models of disease include the polyamine transporter PotD (Ware *et al.*, 2006), glutamine transporters (Basavanna *et al.*, 2009; Hartel *et al.*, 2011), branched chain amino acid transporter Liv (Basavanna *et al.*, 2009), sugar transporters (Basavanna *et al.*, 2009; Iyer & Camilli, 2007; Marion *et al.*, 2011a), and the Ami and Ali oligopeptide uptake transporters (Kerr *et al.*, 2004). The loss of a functional ABC transporter(s) may lead to reduced virulence due to

insufficient nutrient uptake, especially under stressful conditions, poor bacterial growth, increased sensitivity to oxidative stress and/or effects on gene regulation (McCluskey *et al.*, 2004; Ogunniyi *et al.*, 2010).

1.4.5 Neuraminidase A

Neuraminidase A is an example of a pneumococcal surface protein, which is anchored to the peptidoglycan through recognition of the C-terminal sequence LPXTG by a sortase peptidase. Neuraminidases cleave the terminal sialic acid residues from glycolipids, glycoproteins and oligosaccharides on cell surfaces and in bodily fluids (Camara *et al.*, 1994; King *et al.*, 2004). In *S. pneumoniae*, there are three genes encoding for neuraminidase (*nan*A, *nan*B and *nan*C), but only *nanA* is conserved in all pneumococcal strains (Pettigrew *et al.*, 2006); *nanB* and *nanC* are present in 96% and 51%, respectively (Berry *et al.*, 1996; Pettigrew *et al.*, 2006). The cleavage of sialic acid residues by NanA may cause direct damage to the host or may unmask potential binding sites for the pneumococcus, facilitating colonisation. For example, the loss of sialic acid in a Chinchilla model of otitis media allowed *S. pneumoniae* to progress from the nasopharynx up the Eustachian tube to the middle ear (Linder *et al.*, 1994). Other resident upper respiratory tract pathogens such as *N. meningitidis* and *H. influenzae* can attach sialic acid to the terminal region of their lipopolysaccharide (LPS), thereby mimicking host structures and avoiding recognition by the immune system (Estabrook *et al.*, 1997; Soong *et al.*, 2006). Thus, expression of NanA by pneumococci may attenuate this competitive advantage by removing the sialic acid mask (Shakhnovich *et al.*, 2002). Additionally, the free sialic acid residues are an important source of carbon and energy for the growth of the pneumococci, particularly in the upper respiratory tract, where free carbohydrates are not widely available (King *et al.*, 2006; Marion *et al.*, 2011b). NanA is also proposed to have a role in meningitis, as *nanA*-negative *S. pneumoniae* strains have defective adherence to and internalisation into human brain microvascular endothelial cells (hbMEC) (Uchiyama *et al.*, 2009). However, the sialidase activity alone is not responsible for this activity. Rather, the Laminin G-like domain of NanA is thought to be responsible for *S. pneumoniae*-hbMEC interaction, which may lead to subsequent bacterial penetration through the BBB

(Banerjee *et al.*, 2010; Uchiyama *et al.*, 2009). In addition, NanA has been reported to have a role in pneumococcal biofilm formation (Parker *et al.*, 2009) and is upregulated when *S. pneumoniae* is grown under biofilm conditions (Oggioni *et al.*, 2006). Free sialic acid also enhances pneumococcal biofilm formation *in vitro* and may facilitate the transition of *S. pneumoniae* from the nasopharynx into the lower respiratory tract through upregulation of neuraminidase production and the virulence gene activator *mgrA* (Trappetti *et al.*, 2009).

1.4.6 Polyhistidine triad (Pht) proteins

Polyhistidine triad proteins, also known as pneumococcal histidine triad proteins, are a group of surface-exposed antigens characterised by the presence of five or six histidine triad (HxxHxH) motifs (Adamou *et al.*, 2001; Plumptre *et al.*, 2012; Rioux *et al.*, 2010). This family has four members in *S. pneumoniae* (PhtA, PhtB, PhtD and PhtE), which range in size from 96 to 116 kDa (Adamou *et al.*, 2001; Hamel *et al.*, 2004; Zhang *et al.*, 2000). The expression of each gene is monocistronic, with the exception of *phtD*, which is co-transcribed in an operon with *ccdA* (involved in the biogenesis of cytochrome c), spr0904 (has homology to thioredoxin), *yfnA* (putative amino acid transporter) and *adcAII*. Pht proteins have been implicated in adherence to epithelial cell lines (Khan & Pichichero, 2012), defense against complement deposition (Melin *et al.*, 2010a; Ogunniyi *et al.*, 2009) and scavenging of zinc ions (Plumptre *et al.*, 2014b; Rioux *et al.*, 2010).

Pht proteins have been demonstrated to contribute to adherence to host epithelial cells both *in vitro* and *in vivo.* Pneumococcal mutants deficient in either *phtD* or *phtE* demonstrated reduced adherence to Detroit 562 (nasopharyngeal) and A549 (type II pneumocyte) cells, but adherence was significantly increased in an IPTG-induced recombinant PhtD- or PhtE-expressing *E. coli* strain, where surface expression of PhtD or PhtE was confirmed using flow cytometry (Khan & Pichichero, 2012). The earliest evidence for the involvement of Pht proteins *in vivo* was a signature-tagged mutagenesis (STM) screen which showed reduced fitness of *phtA*, *phtB* and *phtD* mutants in a lung infection model (Hava & Camilli, 2002). Pneumococcal mRNA recovered from the nasopharynx of infected mice indicated higher relative expression of *pht* genes compared to pneumococci harvested from the lungs and blood, thus hinting that *pht* genes have a role

in pneumococcal colonisation (Ogunniyi *et al.*, 2009). Additionally, *phtABDE*-deficient mutants were severely attenuated in the nasopharynx, lungs and blood in a murine intranasal or intraperitoneal challenge model (Ogunniyi *et al.*, 2009; Plumptre *et al.*, 2013), indicating that Pht proteins have a role in sepsis as well as colonisation.

Factor H is a negative regulator of the alternative pathway of complement activation (Giannakis *et al.*, 2001) and the Pht proteins have been shown to recruit fH and decrease the deposition of complement protein C3 on the pneumococcal cell surface (Ogunniyi *et al.*, 2009). However, it should be noted that compared to another fH surface protein, CbpA, Pht has a much lesser role (Melin *et al.*, 2010a). In addition to adherence to epithelial cells and recruitment of fH, pneumococcal Pht proteins also has a role in metal ion homeostasis, particularly zinc acquisition, and this is regulated through AdcR (Ogunniyi *et al.*, 2009); a quadruple *phtABDE* mutant was unable to grow in a synthetic culture medium unless Zn^{2+} or Mn^{2+} was added (Rioux *et al.*, 2010). Recently, it has been proposed that Pht proteins act as low affinity 'sinks' for Zn^{2+} which is then imported into the cell via AdcAII (Plumptre *et al.*, 2014a).

1.5 Biofilms

Biofilms have been defined by Costerton *et al.* (1999) as "microbially-derived sessile communities characterised by cells that are irreversibly attached to the substratum or interface or to each other, are embedded in a matrix of extracellular polymeric substances that they have produced, and exhibit an altered phenotype with respect to growth rate and gene transcription". Biofilms can contribute to bacterial persistence as aggregated cells can exhibit an enhanced survival in these environments compared to non-aggregated bacterial populations. The ability of *S. pneumoniae* to form biofilms on host mucosal surfaces, such as that in the nasopharynx, lungs and middle ear, is an important event in the pathogenesis of pneumococcal disease (Bakaletz, 2012; Hall-Stoodley *et al.*, 2006; Hall-Stoodley & Stoodley, 2009; Shak *et al.*, 2013). Currently, two biofilm formation mechanisms [competence-stimulating peptide (CSP)- based and LuxS (S-ribosyl-homocysteine lyase)-dependent] have been identified in the pneumococcus. Both of these involve the use of a small signalling molecule in a mechanism termed quorum sensing (QS). In the CSP-based

system, the ability to form biofilm is dependent on a functional competence system, as exogenous CSP was able to stimulate stable biofilm formations after 24 h *in vitro* growth (Oggioni *et al.*, 2006; Trappetti *et al.*, 2009; Trappetti *et al.*, 2011b; Trappetti *et al.*, 2011c). Furthermore, Oggioni *et al.* (2006) showed that biofilm-grown pneumococci preferentially induced pneumonia and meningitis, sites that are associated with increased expression of competence genes. On the other hand, the study also found that planktonic cells preferentially induced sepsis, a site associated with low expression of competence genes. Additionally, *S. pneumoniae* in the nasopharynx was shown to have upregulated expression of competence genes and downregulated expression of *cps* genes, and that this niche provides an environment that is optimal for genetic exchange and transformation (Marks *et al.*, 2012). The other mechanism for biofilm formation, involves LuxS, which converts S-ribosyl-homocysteine into cysteine, creating the by-product autoinducer-2 (AI-2) (Trappetti *et al.*, 2011c). The importance of LuxS in pathogenesis was demonstrated whereby a *luxS*-knockout mutant was less invasive than the parent strain (Stroeher *et al.*, 2003), while a derivative over-expressing *luxS* had enchanced biofilm formation (Trappetti *et al.*, 2011c). Trappetti *et al.* (2011b) also found that competence genes and the murein hydrolase *cbpD* were associated with *luxS* expression and that these were linked to pneumococcal biofilm formation. The pneumococcus also has the ability to form an extracellular matrix (and biofilms) and that this phenomenon is exclusive to the O variant (Trappetti *et al.*, 2011b). Furthermore, Trappetti *et al.* (2011b) showed that only mice infected with pneumococci of the O phenotype were able to cause infection of the lungs and brain of mice, sites that are normally associated with biofilm-derived pneumococci. This further implicates the importance of O variants during pneumococcal infections.

1.6 Colony Opacity Phase Variation

Pneumococcal phase variation was first observed by Weiser *et al.* (1994), and was described as a difference in colony opacity when observed under oblique, transmitted light on a transparent medium. The three phenotypes included T, semi-transparent (an intermediate phase) and O (Figure 1.3). When T phenotype pneumococci were sub-cultured *in vitro*, a small portion were able to spontaneously switch to the O phenotype at a

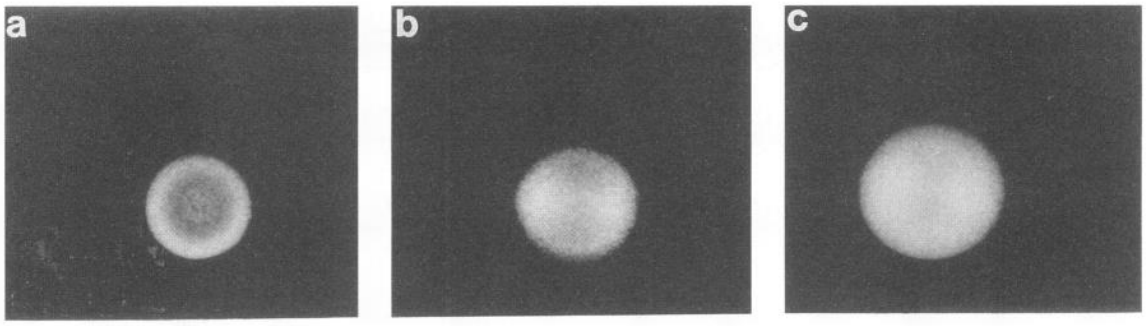


Figure 1.3. Opacity phase variant colonies.

Individual colonies of serotype 9V to show (a) transparent, (b) semi-transparent and (c) opaque colonies grown on a clear medium and viewed under oblique, transmitted illumination. Figure is reproduced from Weiser *et al*. (1994) with permission.

frequency of between 10^{-3} and 10^{-6} per generation (depending on strain) and vice-versa (Kim & Weiser, 1998; Weiser *et al.*, 1994). The relevance of phase variation to the pathogenesis of pneumococcal disease is supported by the preferential isolation of T or O variants of the same strain from the nasopharynx and blood, respectively, of patients with invasive disease. In animal models (mice and rats), T variants are also preferentially isolated form the nasopharynx, whereas O variants of the same strain are more often isolated from the blood or tissues (Briles *et al.*, 2005; Weiser *et al.*, 1994). In otitis media, O phenotype pneumococci were isolated at ten-fold higher frequency compared to the T phenotype of the same strain (Long *et al.*, 2003). The ability to adapt to one niche versus another by reversible phenotypic switching presumably provides a selective advantage to the bacterium. However, the molecular mechanism that underlies this phase variation phenomenon remains unknown.

1.6.1 Molecular comparisons of pneumococcal phase variants

The undoubted importance of phase variation to pathogenesis has prompted many studies aimed at molecular characterisation of the phenomenon. Capsule to TA ratio and expression levels of pyruvate oxidase (SpxB), putative proteinase maturation protein A (PpmA), elongation factor Ts (EF-Ts), LytA, NanA, Hyl, PspA, CbpA, pneumococcal choline-binding protein A (PcpA), PsaA and Ply have all been reported to be affected by switching of phenotypes in *S. pneumoniae* (Kim & Weiser, 1998; Li-Korotky *et al.*, 2009; Li-Korotky *et al.*, 2010; Mahdi *et al.*, 2008; Overweg *et al.*, 2000; Weiser *et al.*, 1996). The findings of these studies are summarised in Table 1.1.

Transparent pneumococci are reported to have thicker cell walls relative to the amount of CPS, compared with their O counter-part, which have thinner cell walls relative to capsule, favouring evasion of the host immune system and survival in the bloodstream or deeper host tissues (Hammerschmidt *et al.*, 2005; Kim & Weiser, 1998). TA is a major component of the cell wall and the transparent variant appears to have two- to four-fold more immunodetectable TA than its opaque counterpart (Kim & Weiser, 1998). Furthermore, the higher virulence of opaque *S. pneumoniae* was associated with an increase in expression of CPS and a decrease in expression of TA, compared with the T

Table 1.1. Comparison of protein and gene expression between O/T pairs *in vitro*.

Protein/gene/ structure	Serotype (strain)	Upregulated phenotype [a]	Method of quantification	Reference
CPS : TA ratio	6A, 6B, 18C	O	ELISA	(Kim & Weiser, 1998)
Chain length	9V (p6)	ns	Electron microscopy	(Weiser *et al.*, 1994)
SpxB	9V (p10), 6B (p314)	T	2D-gel/MS	(Overweg *et al.*, 2000)
spxB	6A	ns	qRT-PCR	(Li-Korotky *et al.*, 2009)
	2 (D39), 4 (WCH43), 6A (WCH16)	ns	qRT-PCR	(Mahdi *et al.*, 2008)
PpmA	9V (p10), 6B (p314)	T	2D-gel/MS	(Overweg *et al.*, 2000)
EF-Ts	9V (p10), 6B (p314)	O	2D-gel/MS	(Overweg *et al.*, 2000)
LytA	R6 (2), p45 (18C)	T	Colony immunoblot/ Western analysis	(Weiser *et al.*, 1996)
lytA	6A	ns	qRT-PCR	(Li-Korotky *et al.*, 2009)
ply	6A	ns	qRT-PCR	(Li-Korotky *et al.*, 2009)
	6A (WCH16)	O	qRT-PCR	(Mahdi *et al.*, 2008)
	2 (D39), 4 (WCH43)	ns	qRT-PCR	(Mahdi *et al.*, 2008)
psaA	6A	ns	qRT-PCR	(Li-Korotky *et al.*, 2009)
	4 (WCH43)	T	qRT-PCR	(Mahdi *et al.*, 2008)
	2 (D39), 6A (WCH16)	ns	qRT-PCR	(Mahdi *et al.*, 2008)
nanA	6A	T	qRT-PCR	(Li-Korotky *et al.*, 2010)
	6A (WCH16)	O	qRT-PCR	(Mahdi *et al.*, 2008)
	2 (D39), 4 (WCH43)	ns	qRT-PCR	(Mahdi *et al.*, 2008)

Continues on following page

Table 1.1 continued

Protein/gene	Serotype (Strain)	Upregulated phenotype [ab]	Method of quantification	Reference
pspA	6A	T	qRT-PCR	(Li-Korotky *et al.*, 2010)
	2 (D39), 4 (WCH43), 6A (WCH16)	ns	qRT-PCR	(Mahdi *et al.*, 2008)
cbpA	6A	O	qRT-PCR	(Li-Korotky *et al.*, 2010)
	6A (WCH16)	T	qRT-PCR	(Mahdi *et al.*, 2008)
	2 (D39), 4 (WCH43)	ns	qRT-PCR	(Mahdi *et al.*, 2008)
hylA	6A	T	qRT-PCR	(Li-Korotky *et al.*, 2010)
pcpA	2 (D39), 4 (WCH43)	T	qRT-PCR	(Mahdi *et al.*, 2008)
	6A (WCH16)	ns	qRT-PCR	(Mahdi *et al.*, 2008)
adcR	2 (D39), 4 (WCH43), 6A (WCH16)	ns	qRT-PCR	(Mahdi *et al.*, 2008)
cbpD	2 (D39), 4 (WCH43), 6A (WCH16)	ns	qRT-PCR	(Mahdi *et al.*, 2008)
cbpG	2 (D39), 4 (WCH43), 6A (WCH16)	ns	qRT-PCR	(Mahdi *et al.*, 2008)
cpsA	2 (D39), 4 (WCH43), 6A (WCH16)	ns	qRT-PCR	(Mahdi *et al.*, 2008)
piaA	2 (D39), 4 (WCH43), 6A (WCH16)	ns	qRT-PCR	(Mahdi *et al.*, 2008)

[a] 'ns' indicates that there was no significant difference between O and T expression of the protein/gene
[b] 'O' indicates opaque variant and 'T' indicates transparent variant

phenotype (Kim & Weiser, 1998). One advantage of the variation in CPS expression for pneumococcal pathogenesis could be that the thinner capsule of the T strain would allow for closer contact with epithelial cells, facilitating colonisation of the nasopharynx, due to enhanced exposure of cell-surface proteins. On the other hand, O variants, with their thicker capsules, are more resistant to opsonophagocytic clearance and have greater systemic virulence. Changes in oxygen availability can affect the production of CPS; O variants cultured under anaerobic conditions produced more CPS than those grown under normal atmospheric conditions, whereas CPS production by the T counterpart remained low in both environments (Weiser *et al.*, 2001). This appears to reflect the conditions in the niche in which the respective variants are preferred, as described previously. The difference in CPS production is attributed to the inhibitory effect of oxygen on tyrosine phosphorylation of CpsD (Weiser *et al.*, 2001). Although differences in the amount of CPS exist between the O and T phenotypes, this is not solely responsible for the colony morphological differences observed *in vitro*, as unencapsulated strains can also exhibit colony opacity variation (Weiser *et al.*, 1994).

In *S. pneumoniae*, the release of H_2O_2 during aerobic growth is largely due to SpxB, as a *spxB*-deficient mutant is unable to produce H_2O_2 (Spellerberg *et al.*, 1996). Unlike some other species such as *E. coli*, *S. pneumoniae* lacks catalase and thus, endogenously generated H_2O_2 can accumulate. H_2O_2 kills the pneumococcus by inactivating essential cellular enzymes and depleting ATP pools (Pericone *et al.*, 2003). T variants express higher levels of SpxB (Overweg *et al*., 2000) and demonstrate earlier autolysis on agar surfaces (Weiser *et al.*, 1994). This may cause a faster depletion in colony biomass, which in turn may cause a more transparent colonial morphology. Furthermore, the deletion of *spxB* resulted in a more mucoid phenotype (Carvalho *et al.*, 2013; Pericone *et al.*, 2002; Ramos-Montanez *et al.*, 2008). The increase in expression of this protein in the T phase reported by Overweg *et al*. (2000) suggests that this may play a role in the colonisation of *S. pneumoniae* in the nasopharynx. Indeed, the *spxB*-deficient mutant was shown to have a decreased ability to colonise the nasopharynx in a rabbit model (Spellerberg *et al.*, 1996). However, although SpxB has been shown to be upregulated in protein quantification, to

date, there is no significant difference in *spxB* transcription has been reported, at least *in vitro* (Li-Korotky et al., 2010; Mahdi et al., 2008).

Using proteomics, including two-dimensional gel electrophoresis (2D-GE) and mass spectrometry, Overweg *et al.* (2000) also found that the proteins PpmA and EF-Ts were differentially expressed in the two phenotypic variants. PpmA expression is higher in the T phenotype, suggesting that it may be contribute to the adherence of *S. pneumoniae* to host epithelial cells, either through the maturation of surface proteins involved in adherence, or indirectly, through the activation of proteases or other proteins. The increase in the expression of EF-Ts by the O variant could account for the increase in virulence, as expression of EF-Ts by *Coxiella burnetii* contributes to its survival in harsh environments such as the phagolysosome, potentially due to higher metabolic activity (Seshadri *et al.*, 1999).

Autolysis of *S. pneumoniae* begins during the stationary phase of growth, through the activation of the major amidase, LytA. Opaque variants undergo spontaneous autolysis at a slower rate than that of their T counterparts (Saluja & Weiser, 1995) and this could be due to lower levels of LytA activation (Weiser *et al.*, 1996). However, LytA did not contribute to the ability of the T strain to colonise the nasopharynx, as a *lytA*$^{-}$ mutant strain and its *lytA*$^{+}$ parent did not vary significantly in their ability to colonise infant rat nasopharynx (Weiser *et al.*, 1996). Nevertheless, increased LytA expression and a greater degree of cellular autolysis, could contribute to the more translucent and flatter appearance of colonies on clear media, a defining characteristic of the T phenotype.

There are several examples where comparisons of the expression levels of virulence-associated genes, including *nanA*, *pspA*, *cbpA* and *psaA*, between O and T variants have yielded inconsistent findings (see Table 1.1). For example, genes shown to be upregulated in the O variant in one study may be upregulated in the T variant in another. Alternatively, within a study, genes differentially regulated between O and T variants in one strain were not in another. This discordance has greatly complicated attempts to associate particular gene expression traits with colony opacity phenotype.

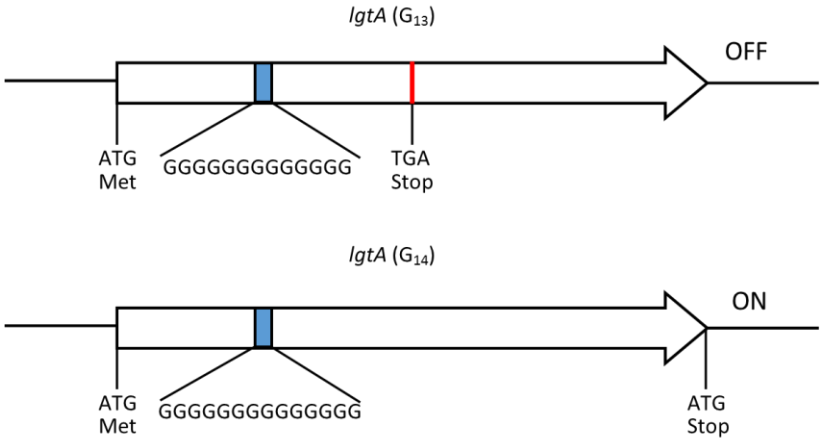
1.6.2 Genetics of phase variation

1.6.2.1 Genetic basis of phase variation in other bacteria

Several phase variation mechanisms have been described for other bacteria. One such mechanism is by slipped-strand mispairing (SSM) during replication. SSM causes changes in the length of short DNA sequence repeats or homopolymeric tracts, which may result in translational frame-shift mutations, thereby switching the expression of the encoded protein "ON" or "OFF". For example, in *N. meningitidis lgtA* gene has a homopolymeric tract of guanine (G) bases near the 5' end of the ORF (Jennings *et al.*, 1999) (Figure 1.4A). *lgtA* encodes a glycosyltransferase involved in the biosynthesis of lacto-*N*-neotetraose, one of the alternate terminal structures displayed on meningococcal LPS. SSM resulting in variation in a single G base (for example, G_{13} instead of G_{14}) results in an inactive *lgtA* gene and thus the absence of lacto-*N*-neotetraose expression (Jennings *et al.*, 1999). Furthermore, lacto-*N*-neotetraose is a sialic acid acceptor and the presence of such phase variable alterations in terminal LPS structure may assist in immune evasion during infection, favouring sialylated strains (Estabrook *et al.*, 1997). Hence, mutations in the homopolymeric tract of *lgtA* may also lead to switching from a sialylated (L3 type) to a unsialylated (L8 type) variant and vice versa (Jennings *et al.*, 1999). The switching of flagellar motility in *Helicobacter pylori* is another example where SSM has a role. Studies by Josenhans *et al*. (2000) demonstrated that a strain with nine cytosine bases in the *fliP* gene was non-flagellated and non-motile. In contrast, one that has eight C-bases was motile and through screening of about 50 000 colonies of the strain with nine C-bases, they found that a motile revertant could be isolated, proving that the switch is reversible.

Another example of SSM is the hydrophobic variable length region encoding the repetitive pentamer sequence (5'-CTCTT-3') of opacity proteins (Opa) of *Neisseria gonorrhoeae* (Stern *et al.*, 1986; Stern & Meyer, 1987), which belong to a family of outer-membrane proteins that confer adherence to various human cell types. The coding repeat (CR) sequence of the pyrimidine pentamer is variable in length and codes for the hydrophobic core of the Opa leader peptide. Depending on the number of CR units present, the reading frame of the *opa* structural gene would be either in-frame or out-of-frame, thus

A



B

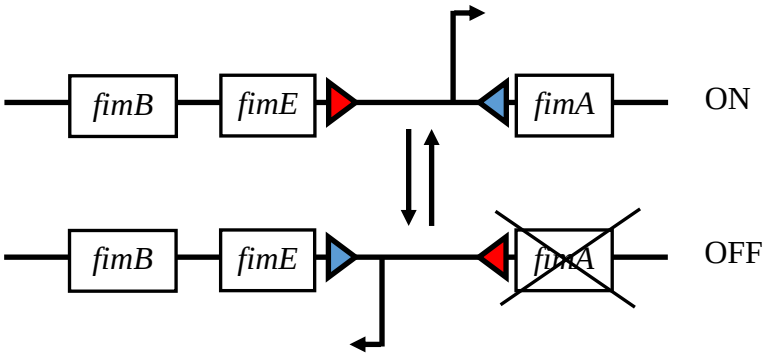


Figure 1.4. Phase variation due to SSM and DNA inversion.

(A) One terminal structure of meningococcal LPS, controlled in part by *lgtA*, is subject to phase variation. SSM of the homopolymeric tract of G bases from 14 to 13 switches its expression from on to off. This is because the deletion of a base in the homopolymeric tract leads to a frameshift, resulting in a premature stop codon. This in turn leads to a truncated or non-functional protein (Jennings *et al.*, 1999). (B) The phase variable type 1 fimbrial expression of *E. coli* is encoded by *fimA.* The invertible repeats (indicated by triangles) between *fimE* and *fimA* contains the promoter region for the latter gene; the orientation of these inverted repeats determines the "ON" or "OFF" expression of *fimA* (Klemm, 1986; McClain *et al.*, 1991). Image adapted from van der Woude and Baumler (2004). Both images are not drawn to scale.

affecting the translation of the Opa protein. For example, 6, 9 or 12 CR units would place the protein sequence in-frame, producing an intact and functional Opa protein, whereas, 4 or 8 CR units would introduce a frame-shift and loss of functional Opa expression.

Another mechanism whereby bacteria can exhibit phase variation is via DNA inversion, where the orientation of an invertible element either promotes expression of a certain gene or restricts it. This mechanism is employed by *E. coli* in the expression of *fimA* (type 1 fimbriae), which is important in the attachment of the bacterium to the host cells during bacterial colonization (Schilling *et al.*, 2001). The promoter region of *fimA* is positioned within an invertible element, such that its orientation determines whether the gene in the "ON" (translated) or "OFF" (untranslated) state (Klemm, 1986; McClain *et al.*, 1991). Hence, the invertible elements act as switches by either removing or adding a promoter upstream of the coding region of the gene (Figure 1.4B).

In contrast to examples where changes in simple tandem repeats alter the expression of a single gene, phase variation of the type III DNA methyltransferase, encoded by *mod* genes of *H. influenzae, N. meningitis, N. gonorrhoeae* and *H. pylori* (Srikhanta *et al*., 2005; Srikhanta *et al*., 2009; Srikhanta *et al*., 2011) can coordinate the random switching of multiple genes. This system has been termed a "phasevarion" (phase-variable regulon) and was initially demonstrated in *H. influenzae* using microarray analysis comparing WT strains (expressing *modA1*) and a *modA1* knock-out mutant. This revealed that *modA1* controlled the expression of 15 genes, 7 of which were upregulated in the mutant strain (Srikhanta *et al.*, 2005). These genes encode the outer membrane protein Opa (orthologous to the Opa adhesin of *Neisseria*) and the heat shock proteins HtpG, GroES, GroEL, DnaK, DnaJ and hypothetical protein HI1456. Srikhanta *et al.* (2005) used a *H. influenzae opa*::*lacZ* fusion construct to demonstrate the direct impact of differential methylation on target gene expression; when *modA1* is in-frame white colonies were observed, but when *modA1* is out-of-frame blue colonies were observed. Similarly, in *N. meningitidis* and *N. gononorrhoeae*, an active or inactive *mod* gene contributed to the expression of multiple genes, including those known to be important for iron acquisition, protection against oxidative stress and antimicrobial susceptibility (Srikhanta *et al.*, 2009). Furthermore, multiple phase variable *mod* alleles, characterised by differences in their DNA recognition

domain, can exist in a single organism (Srikhanta *et al.*, 2009; Srikhanta *et al.*, 2011). This would provide a mechanism whereby a pathogenic bacterium could easily and quickly adapt to a particular niche.

1.6.2.2 Pneumococcal gene and genome variation

The genome of *S. pneumoniae* consists of 2.0-2.2 million base pairs and over 2000 genes (Choi *et al.*, 2014; Tettelin *et al.*, 2001). *S. pneumoniae* is naturally transformable, meaning that they can easily take-up and incorporate DNA fragments from closely-related bacterial species that occupy the same niches (Johnsborg & Havarstein, 2009). This would confer a selective advantage on the bacterium, as it is able to evolve upon selection by surrounding stresses. An example of this is resistance to penicillin; the penicillin-binding protein 2B gene (*pbp2B*) from penicillin-sensitive and penicillin-resistant strains were compared and found to have extensive sequence divergence, which impacted on affinity for the antibiotic. The genetic differences were attributed to acquisition of fragments of this gene from *S. mitis*, a commensal found in the human nasopharynx (Dowson *et al.*, 1989). Furthermore, as mentioned previously, *S. pneumoniae* has the ability to evolve by switching CPS locus (Brueggemann *et al.*, 2007; Wyres *et al.*, 2013), potentially enabling escape from CPS-based vaccines.

The types of DNA re-arrangement described above could be responsible for phase variation of colony opacity in *S. pneumoniae*. However, to date there is no direct evidence to support this. In the Paton Laboratory, McKessar (2003) investigated potential phase variable genes containing repeat elements (either non-trimeric tandem repeats or homopolymeric tracts) based on analysis of the *S. pneumoniae* TIGR-4 genome (Tettelin *et al.*, 2001). Although a number of such elements were identified, DNA sequence analysis of these regions from O and T variants of *S. pneumoniae* D39 revealed no evidence of any correlation between SSM and O/T phenotype.

1.6.3 Frequency of phase variation

In *N. meningitidis*, a mathematical model system has been used to demonstrate that an increased rate of phase switching within a specified “contingency loci” increase the ability

of the bacterium to colonise a range of host environments and thus increase the invasive potential of the bacterium (Meyers *et al.*, 2003). The term contingency locus refers to a region of hypermutable DNA (i.e. susceptible to SSM) that is able to mediate high-frequency, stochastic, heritable, genotypic switching. This suggests that a single bacterium is capable of invading, replicating and surviving in the blood to cause invasive disease, and that subsequent survival and proliferation in the blood is due to within-host evolution. This idea has been supported by studies carried out in *H. influenzae* (Margolis & Levin, 2007; Moxon & Murphy, 1978).

Saluja and Weiser (1995) used transformation experiments to investigate the genetic basis for phase variation in *S. pneumoniae*. Genomic DNA of an O variant with a high frequency of phase switching was used to transform a T recipient such that it expressed an O phenotype. By screening the DNA library of a strain with high switching frequency, they found that the presence of a stem-loop structure (BOX A and C element) located between *glpF* and ORF3 was not required for phase variation *per se*, but was associated with an increased frequency of phase variation (up to three-logs higher). Box elements are repetitive intergenic sequences and were proposed to have a role in the regulation of regulatory genes downstream, increasing the frequency of opacity phase variation compared to a strain that lacks this element. In *S. pneumoniae*, these BOX elements have been shown to be located in the vicinity of genes required for genetic competence and virulence genes such as *nanA*, *ply* and *lytA* (Martin *et al.*, 1992).

It is clear from the above that pneumococcal phase variation may be multi-factorial and underpinned by distinct mechanisms from those operating in other pathogens. Nevertheless, the massive technological advances in genomics, transcriptomics and proteomics that have occurred in the last decade provide a new opportunity to understand the fundamental mechanism of colony opacity phase variation in *S. pneumoniae*.

1.7 Hypotheses

Central hypotheses relevant to the work described in this thesis are as follows.

- The progression from carriage to invasive disease will require complex alterations in pneumococcal virulence gene expression.
- Such gene expression alterations enable a given *S. pneumoniae* strain to adapt to discrete host niches, and that some of these niche-specific adaptations can be attributed to phenotypic phase variation.

1.8 Aims

In order to better comprehend the transition of the pneumococcus from asymptomatic carriage to disease, the phenomenon of colony opacity of *S. pneumoniae* needs to be better understood. Studies reported to date indicate inconsistencies with regard to distinct expression patterns of proteins and genes between colony opacity variants. This study sets out to be the first comprehensive comparison of the differences between O and T variants at the proteomic, transcriptomic and genomic level. Analysis of O and T strain pairs from multiple *S. pneumoniae* strains should enable identification of molecular changes that are consistently associated with O or T phenotype. Further characterisation and examination of the functions of any identified genes or proteins using *in vitro* and *in vivo* models may provide important information on the mechanisms underlying phase variation in the pneumococcus and its role in pathogenesis. Accordingly, the aims of this thesis are:

1. To conduct a comprehensive characterisation of O and T *S. pneumoniae* strain pairs belonging to multiple serotypes at the proteomic, transcriptomic and genomic levels.
2. To characterise the role of key proteins/genes involved in phase variation identified in Aim 1 in pneumococcal pathogenesis using targeted mutagenesis.

Chapter 2: MATERIALS AND METHODS

2.1 Bacterial Strains

The bacterial strains used in this study are listed in Table 2.1.

Table 2.1 Bacterial strains used in the work of this thesis.

Strain	Description [a]	Reference/Source
D39WT	Serotype 2	(Dochez & Avery, 1917)
D39O	Serotype 2, opaque phenotype	(Mahdi *et al.*, 2008)
D39T	Serotype 2, transparent phenotype	(Mahdi *et al.*, 2008)
WCH16O	Serotype 6A, opaque phenotype	(Mahdi *et al.*, 2008)
WCH16T	Serotype 6A, transparent phenotype	(Mahdi *et al.*, 2008)
WCH43O	Serotype 4, opaque phenotype	(Mahdi *et al.*, 2008)
WCH43T	Serotype 4, transparent phenotype	(Mahdi *et al.*, 2008)
D39OΔ*spxB*	D39O with *spxB* deleted and replaced with *erm* (Erythromycin resistant)	This study
D39TΔ*spxB*	D39T with *spxB* deleted and replaced with *erm*	This study
XL10-Gold	*Escherichia coli*, competent cells	Stratagene, CA, USA
XL10-pAL3:*spxB*	*E. coli* containing pAL3 vector with *spxB* insert	This study
D39O-pAL3::*spxB*	D39O containing pAL3 vector with *spxB* insert	This study
$D39O^{D39O spxB}\Delta spxB$	D39OΔ*spxB* back-transformed with *spxB* from D39O	This study
$D39T^{D39O spxB}\Delta spxB$	D39TΔ*spxB* back-transformed with *spxB* from D39O	This study
$D39O^{D39T spxB}\Delta spxB$	D39OΔ*spxB* back-transformed with *spxB* from D39T	This study
$D39T^{D39T spxB}\Delta spxB$	D39TΔ*spxB* back-transformed with *spxB* from D39T	This study
SpnD39IIIA	D39 expressing SpnIII HsdS “A” allele	(Manso *et al.*, 2014)
SpnD39IIIB	D39 expressing SpnIII HsdS “B” allele	(Manso *et al.*, 2014)
SpnD39IIIC	D39 expressing SpnIII HsdS “C” allele	(Manso *et al.*, 2014)

SpnD39IIID	D39 expressing SpnIII HsdS "D" allele	(Manso *et al.*, 2014)
SpnD39IIIE	D39 expressing SpnIII HsdS "E" allele	(Manso *et al.*, 2014)
SpnD39IIIF	D39 expressing SpnIII HsdS "F" allele	(Manso *et al.*, 2014)
D39OΔ*comD*	D39O with *comD* deleted and replaced with *erm*	This study
D39TΔ*comD*	D39T with *comD* deleted and replaced with *erm*	This study
WCH16OΔ*comD*	WCH16O with *comD* deleted and replaced with *erm*	This study
WCH16TΔ*comD*	WCH16T with *comD* deleted and replaced with *erm*	This study
WCH43OΔ*comD*	WCH43O with *comD* deleted and replaced with *erm*	This study
WCH43TΔ*comD*	WCH43T with *comD* deleted and replaced with *erm*	This study

[a] Bacterial strains are *S. pneumoniae* unless indicated.

2.2 Growth and Phenotypic Analysis of Bacteria

2.2.1 Liquid culture

S. pneumoniae strains were routinely grown in THY broth (Todd-Hewitt broth [THB]) (Oxoid, Hampshire, England] with 1% Bacto yeast extract (Becton, Dickinson and Company (BD), New Jersey, USA], in a chemically-defined medium (C+Y (Lacks & Hotchkiss, 1960)) or TSB (30 g/l Bacto™ Tryptic Soy Broth, soybean-casein digest medium [Becton, Dickinson and Company, Maryland, USA]). Strains used for mouse challenge studies were grown in serum broth (SB) (10 g/l peptone [Oxoid], 10 g/l Lab Lemco powder [Oxoid], 5 g/l NaCl and 10% (v/v) heat-inactivated horse serum [Gibco®, Auckland, New Zealand]). *E. coli* strains were routinely grown in Luria broth (LB) (10 g/l tryptone, 5 g/l yeast extract and 5 g/l NaCl).

2.2.2 Solid culture

S. pneumoniae were routinely grown on blood agar (BA) plates (39 g/l Columbia base agar [Oxoid], supplemented with 5% (v/v) defribinated horse blood) and incubated at 37°C in 5% CO_2 overnight before inoculation into broth. *E. coli* strains for transformation were cultured on LB agar (10 g/l peptone, 5 g/l yeast extract, 5 g NaCl, 15 g/l agar).

To identify colonies of either opaque or transparent phenotypes, *S. pneumoniae* were grown on THY+catalase or C+Y+catalase agar plates (THY broth or C+Y broth supplemented with 1.5% [w/v] agar-agar [Merck, Darmstadt, Germany] supplemented with 200 U/ml catalase [Sigma, NSW, Australia]). After incubation at 37°C in 5% CO_2 in air overnight, the colony phenotype was determined under oblique transmitted light, as described by Weiser *et al*. (1994).

2.2.3 Freezer stocks/storage

S. pneumoniae strains were grown overnight on BA, inoculated into THY and incubated at 37°C in 5% CO_2 for approximately 4 h. Thereafter, glycerol was added to 15% (v/v) and 0.5 ml to 1 ml aliquots dispensed into cryogenic tubes (CryoTube™ Vials [Nunc, Denmark]) which were then stored at -80°C.

2.2.4 Antibiotics

The antibiotics used to select for *S. pneumoniae* in this study were erythromycin (Ery) (Roche) at 0.2 µg/ml and Gentamicin (Gent) (Sigma) (to inhibit background flora in *in vivo* samples) at 5 µg/ml. For growth of *E. coli* for cloning, 500 µg/ml Ery was added to the medium. Diagnostic optochin disks were purchased from Oxoid.

2.2.5 Bacterial growth curves

To measure pneumococcal growth, bacteria were inoculated into sterile C+Y from an overnight BA culture and incubated until A_{600} 0.20 in 5% CO_2 in air at 37°C. This was then diluted 1:10 in C+Y. Thereafter, 200 µl of the diluted culture was transferred into a Costar® 96-well Flat-bottomed Cell Culture Plate (Corning Incorporated, New York, USA) in triplicate and A_{600} was measured every 20 min for 18 h, at 37°C on a SpectraMax M2 Microplate Reader (Molecular Devices, California, USA). Data obtained were analysed using SoftmaxPro® software v 4.6 (Molecular Devices).

2.2.6 Pneumococcal capsular serotyping

Serotype-specific capsule production for D39O, D39T, WCH16O, WCH16T, WCH43O and WCH43T was confirmed by Quellung reaction, as follows: Each strain was grown on BA at 37°C in 95% air, 5% CO_2 overnight and then inoculated into 1 ml SB (Section 2.2) and incubated for a further 3 h. Ten microlitres of culture was mixed with 2 µl type-specific serum (types 2, 6A and 4 [Staten Serum Institut, Denmark]) and the culture was observed by light microscopy after 2 min at room temperature. A positive reaction is indicated by "swelling" of refractive material around the bacterial cells.

2.2.7 Optochin sensitivity

To distinguish *S. pneumoniae* from phenotypically similar colonies, colonies were tested for their sensitivity to optochin. The bacterium is plated onto BA followed by embedding a 5 µg optochin disc (Oxoid) and incubating at 37°C in 5% CO_2 in air overnight. A zone of growth inhibition around the disc confirms optochin sensitivity and identity as *S. pneumoniae*.

2.2.8 Colony opacity analysis

To check for colony opacity phenotype, pneumococcal strains were plated for single colonies on THY+catalase or C+Y+catalase agar plates and incubated at 37°C in 5% CO_2 in air for 18-36 h. Colony phenotypes were observed under oblique transmitted light to determine the degree of "transparency". Colonies were also observed under a dissecting microscope (Nikon SMZ1000) and visualised through cellSens software.

2.3 Preparation of cell pellets

Strains D39O, WCH16O and WCH43O and their transparent counterparts were cultured in 200 ml C+Y at 37°C in 5% CO_2 in air to an absorbance reading of A_{600} 0.50. To harvest the cell pellets, the cultures were centrifuged at 3,600 × *g* at 4°C for 20 min (3K-18 Centrifuge, Sigma), and the supernatant discarded. For genomic and transcriptomic analyses, pellets were harvested from 50 ml cultures, whereas 100 ml cultures were

harvested for proteomic analyses. This procedure was repeated on four separate occasions to obtain four biological replicates. Cell pellets were stored at -80°C until required.

2.4 Oligodeoxynucleotides

The oligodeoxynucleotides (primers) used in this study were purchased from Sigma and are listed in Table 2.2.

Table 2.2 Primers used in the work of this thesis.

Primer name [a]	Primer sequence (5' → 3') [b]	Primer description	Primer source
SpxB pAL3 F	TCCAATTCTATGTAATCGAATTCTCC AAG [cd]	Amplification of *spxB* for pAL3 cloning	This study
SpxB pAL3 R	GAAAATCAAAGAATGAATTCTACAA GTTTC [c]	Amplification of *spxB* for pAL3 cloning	This study
J214	GAAGGAGTGATTACATGAACAA	*erm* cassette 5'	Our laboratory
J215	CTCATAGAATTATTTCCTCCCG	*erm* cassette 3'	Our laboratory
J257	TCTGTTTGAAGAAGAAGGTATCTTG	Sequencing *erm* cassette insert	Our laboratory
J258	TCTGTTTGAAGAAGAAGGTATCTTG	Sequencing *erm* cassette insert	Our laboratory
SpxB seqF	TGACTAGATTTCTTTGTTATAAAAC AG	Sequencing of *spxB* – upstream flank	This study
SpxB seqR	TCCAAGATATGCTCCAAGTCAGC	Sequencing of *spxB* – downstream flank	This study
SpxB F Int	ACCCAATGTACAACGGTATCGCTG	Sequence within *spxB*	This study
SpxB Flank F	GTTGCAGGTAAGCCATATATCCAG	Mutagensis of *spxB* – upstream flank	This study
SpxB Flank R	AACCCCGTCTTTGTAAATGGCATC	Mutagenesis of *spxB* – downstream flank	This study
SpxB UpSeq F	TCTGTTTGAAGAAGAAGGTATCTTG	Sequencing of *spxB* – upstream flank	This study
SpxB EryY	CGGGAGGAAATAATTCTATGAGCCG AAAATCAAATATGAAACTTGTA[e]	Mutagensis of *spxB*, with *erm* cassette 3'-end	This study

SpxB EryX	TTGTTCATGTAATCACTCCTTCTTCA ATTTTTTTAAACTTGGAGAATA[e]	Mutagensis of *spxB*, with *erm* cassette 5'-end	This study
ComD Flank F	CCCATCTGACAATCGAATATCTA	Mutagensis of *comD* – upstream flank	This study
ComD Flank R	CCCATATGGATCCCACATTGATG	Mutagensis of *comD* – downstream flank	This study
J215/ComD F	CGGGAGGAAATAATTCTATGAGAGC AAGAAATTGATATAATGGTTATA[e]	Mutagensis of *comD*, with *erm* cassette 3'-end	This study
J214/ComD R	TTGTTCATGAATCACTCCTTCATTTC ATTACTTTTTTCGTAGGAAAA[e]	Mutagensis of WCH43 *comD*, with *erm* cassette 5'-end	This study
J214/ComD D39 R	TTGTTCATGAATCACTCCTTCATTTC ATTACTTTTTTCTTTGTAAAATAAA[e]	Mutagensis of D39 and WCH16 *comD*, with EryR cassette tail-end	This study
16S rRNA F	CGTGAGTAACGCGTAGGTAA	Q-PCR (Sprr01) for 16S rRNA	This study
16S rRNA R	ACGATCCGAAAACCTTCTTC	Q-PCR (Sprr01) for 16S rRNA	This study
cbpD F	ACCGACGATTGGTTCCATTA	Q-PCR for *cbpD*	(Trappetti *et al.*, 2011c)
cbpD R	CCAACACTGCCACTATCCAA	Q-PCR for *cbpD*	(Trappetti *et al.*, 2011c)
cglA F	TCAGTTGCAGTTGAACGAAG	Q-PCR for *cglA*	(Harvey, 2010)
cglA R	CTGTCGCACCTGTCAAACTA	Q-PCR for *cglA*	(Harvey, 2010)
comX1 F	GTCCAAGGGACTGTGTATAAGTGT	Q-PCR for *comX1*	(Ogunniyi *et al.*, 2012)
comX1 R	CTATAATCTCTTAGTGTTTCATGAAA G	Q-PCR for *comX1*	(Ogunniyi *et al.*, 2012)
fusA F	ACATCATCGACACACCAGGA	Q-PCR for *fusA* (Spd0253)	This study
fusA R	AGTCAGCACCGATTTTGTCC	Q-PCR for *fusA* (Spd0253)	This study
groEL F	CAGATGCCCGTTCAGCCATGGT	Q-PCR for *groEL*	(Ogunniyi *et al.*, 2012)

groEL R	CAATCCCACGACGAATACCGATTG	Q-PCR for *groEL*	(Ogunniyi *et al.*, 2012)
lctO F	TGGTGTGCATGAGTTTGGTT	Q-PCR for *lctO*	(Ogunniyi *et al.*, 2012)
lctO R	CAGCACCTTCTGGCAGGTAT	Q-PCR for *lctO*	(Ogunniyi *et al.*, 2012)
ldh F	AGTGGGACCTGGTGATGAAG	Q-PCR for *ldh* (SP1220)	This study
ldh R	AACAATCCCTGCGAGCTCTA	Q-PCR for *ldh* (SP1220)	This study
luxS F	CCCTATGTTCGCTTGATTGGGG	Q-PCR for *luxS*	(Trappetti *et al.*, 2011c)
luxS R	AGTCAATCATGCCGTCAATGCG	Q-PCR for *luxS*	(Trappetti *et al.*, 2011c)
murC F	AAGAACCATTGCCTTGTTGG	Q-PCR for *murC* (SP1521)	This study
murC R	GGATGTCTCCTGCTCCCATA	Q-PCR for *murC* (SP1521)	This study
phtD F	GTATTAGACAAAATGCTGTGGAG	Q-PCR for *phtD*	(Plumptre *et al.*, 2014b)
phtD R	CTGTATAGGAGTCGGTTGACTTTC	Q-PCR for *phtD*	(Plumptre *et al.*, 2014b)
ppc F	CGACTCACGCAGAAAAATCA	Q-PCR for *ppc* (SPD0953)	This study
ppc R	AGGGGAACAATCTGAACACG	Q-PCR for *ppc* (SPD0953)	This study
psaA F	GGTACATTACTCGTTCTCTTTCTTTCT	Q-PCR for *psaA*	(Ogunniyi *et al.*, 2012)
psaA R	GTTTTCAGTTTTCTTGGCATTTTCTAC	Q-PCR for *psaA*	(Ogunniyi *et al.*, 2012)
purA F	AAACCCTGTAGCTGGTGGTG	Q-PCR for *purA* (SP0019)	This study
purA R	CACGGATACGTTCTCCCACT	Q-PCR for *purA* (SP0019)	This study
pyk F	CAATCGACAAGAACGCTCAA	Q-PCR for *pyk* (SPD0790)	This study
pyk R	CGTCAGTTGAAGATGGAGCA	Q-PCR for *pyk* (SPD0790)	This study
SP1837 F	AGTGGGACCTGGTGATGAAG	Q-PCR for SP1837	This study
SP1837 R	AACAATCCCTGCGAGCTCTA	Q-PCR for SP1837	This study

SP1837 seq F	CATGTTTCAAACCGTGTTTGAGGTAC	Partial sequencing of SP1837	This study
SP1837 seq R	CTCAATTAGACTTTTTAGGGGCAGGA	Partial sequencing of SP1837	This study
SP1837 16R	AGAAACTGTTTTGAAAGTCTCAATGA	Partial sequencing of SP1837	This study
spxB F	CAACATGTGCTACCCAGACG	Q-PCR for *spxB*	This study
spxB R	CGAGCATCGATGACAACAGT	Q-PCR for *spxB*	This study
SPD0323 F	GTCGGCATTGTATTCTTTATATCG	Partial sequencing of SPD0323	This study
SPD0323 R	CACTCGGTTCTTATATGGGATAAC	Partial sequencing of SPD0323	This study
SPD1128 F	GAGGTAGTCGAATCTTTAACTTCTTT	Partial sequencing of SPD1128	This study
SPD1128 R	GAGAGAGTATCGAACAGTATAGCTAAAG	Partial sequencing of SPD1128	This study
SPD0466 F	ATTGGATTCTCCTCCTAAGA	Partial sequencing of SPD0466	This study
SPD0466 R	ATCCACTGTTTCAGCCTTGGCTAG	Partial sequencing of SPD0466	This study
SP0347 F	CCCAAACTTTTTGGCG	Partial sequencing of SP0347	This study
SP0347 R	CCTTGATTGCGATTCACTAC	Partial sequencing of SP0347	This study
SP1090 F	GGGGCAAAAAAGTTTATCAG	Partial sequencing of SP1090	This study
SP1090 R	CCAACCAGCATGACATTG	Partial sequencing of SP1090	This study
SP201 F	GTGCTAGAAGAAGCCGAAGAG	Partial sequencing of SP0201	This study
SP201 R	CTGAGTAATCAGTCTTTACTTGTTGG	Partial sequencing of SP0201	This study
16O-C5-10150 F	GAATCGGTATTTCGACAGAAG	Partial sequencing of WCH16O contig C5	This study

16O-C5-10150 R	GATGCCAATCTCAAGGAAATC	Partial sequencing of WCH16O contig C5	This study

[a] 'R' denotes the reverse complementary to target
[b] Primers were derived from the *S. pneumoniae* TIGR-4 genome as deposited in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database
[c] Underline on primer sequence indicates *Eco*RI site
[d] Nucleotide changed from published NCBI sequence is indicated in red font
[e] *erm* cassette sequence is underlined

2.5 Identification of Differentially Expressed Proteins by 2D-DIGE

2.5.1 Preparation of membrane and cytosolic fractions

Cell pellets for two-dimensional differential gel electrophoresis (2D-DIGE) were resuspended in 10 ml PBS containing one protease inhibitor cocktail tablet (cOmplete, EDTA-free, Roche Diagnostic), 1 µg/ml DNase I (Roche), 1 µg/ml RNase A (Roche) and 1000 U mutanolysin (Sigma). Cells were disrupted using an Aminco French Pressure Cell at 12,000 p.s.i. and the lysates separated into cytosolic and membrane fractions by ultracentrifugation (Optima^{TM} L-100 XP Ultracentrifuge [Beckman Coulter, California, USA]) at 250,000 × *g* (Ti70 Rotor [Beckman Coulter]) for 1 h at 4°C. The cytosolic (supernatant) and membrane (pellet resuspended in 1 ml TUC buffer (7 M Urea, 2 M Thio-urea, 4% [w/v] CHAPS, 30 mM TRIS and 50% [w/v] Acetonitrile) fractions were stored at -20°C until required.

2.5.2 2D-DIGE

2D-DIGE and associated statistical analyses were performed by The Adelaide Proteomics Centre (APC), University of Adelaide, SA, Australia. Briefly, protein fractions were quantified using an EZQ^{TM} Protein Quantitation Kit (Life Technologies, Carlsbad, USA), according to manufacturer's instructions and equal amounts of both cytosolic and membrane protein fractions were labelled with Cy2, Cy3, or Cy5 using CyDye^{TM} DIGE fluors (GE Healthcare, Little Chalfont, UK), using Cy2 as the internal pool standard (IPS), according to the manufacturer's instructions. Samples were firstly subjected to isoelectric focusing (IEF) to separate proteins according to isoelectric point (pI) on an IPGphor II apparatus (GE Healthcare), followed by SDS-PAGE, which separates the proteins

according to size (Ettan DALT12 electrophoresis unit [GE Healthcare]). After electrophoresis, gels were scanned on an Ettan DIGE Imager (GE Healthcare) and images were captured using ImageQuant software v 7.0 (GE Healthcare). The images were then analysed using DeCyder 2D software v 7.0 (GE Healthcare) and the spot maps of the respective O and T variants of each strain were compared. Intensities of differentially expressed protein spots were compared using an unpaired two-tailed Student's *t*-test. Proteins of interest were excised from the SDS-PAGE gels using an Ettan Spot Picker (GE Healthcare) and identified using liquid chromatography electrospray ionisation ion-trap mass spectrometry (LC-ESI-IT MS) and in the case of the identification of more than one protein on a spot, emPAI (exponentially modified protein abundance index) scores reported by MASCOT (v 2.2, Matrix Science) were taken into account during analysis of data.

2.6 Transformation of bacteria

2.6.1 Transformation of *S. pneumoniae*

For transformation experiments where the gene of interest is replaced using an antibiotic resistant cassette, pneumococci were grown in THY or C+Y to A_{600} 0.5. This was then diluted 1/10 in THY or C+Y competence medium (THY or C+Y supplemented with 0.2% [w/v] bovine serum albumin [BSA], 0.2% [w/v] glucose and 0.02% [w/v] $CaCl_2$), respectively and 50 ng of competence-stimulating peptide 1 (CSP-1) (Chriontech, VIC, Australia) for strains D39 and WCH16, or CSP-2 (Chriontech) for strain WCH43, was added and incubated for 15 min at 37°C. Next, 500 ng of the PCR product or plasmid with the gene of interest was added and incubated for 3 h before plating onto BA supplemented with the appropriate antibiotic, and incubated at 37°C in 5% CO_2 in air, overnight. Transformants were analysed by PCR (Section 2.8.2) and DNA sequencing (Section 2.8.8).

2.6.2 Preparation of pneumococcal competent cells and back-transformation

For generation of back-transformations, which requires a higher transformation efficiency than that obtained using the method described in Section 2.6.1, competent *S. pneumoniae*

cells were initially prepared. Firstly, the pneumococci were grown in complete-CAT medium, (cCAT) (10 g/l Bacto Casamino acids [BD], 5 g/l Bacto tryptone, 5 g/l Bacto yeast extract, 4% [v/v] 0.4 M K2HPO4, 0.002% [w/v] glucose, 150 mg/l glutamine) to A_{600} 0.50-0.60. Cells were then diluted 1 in 10 in CTM (cCATsupplemented with 0.2% [w/v] BSA, 1% [v/v] 0.01 M $CaCl_2$). Following this, cells were further incubated until A_{600} 0.2 and then centrifuged at 3,600 × *g* for 15 min. The pellet was then resuspended in $1/10^{th}$ volume CTM adjusted to pH 7.8 and 15% (v/v) glycerol. Aliquots of 50 µl were stored at -80°C until required.

For transformation, 50 ng of CSP-1 and 500 µl CTM was added to an aliquot of the competent pneumococci and incubated for 15 min at 37°C. Next, 10-100 ng/ml donor DNA was added to the transformation mixture and incubated at 32°C for 30 min, followed by 2 h at 37°C. Following 2 h incubation at 37°C, enrichment for erythromycin-sensitive transformants was performed by adding erythromycin to the transformation mixture to a final concentration of 90 µg/ml and then incubated at 37°C for 20 min. Ampicillin was added to the transformation mixture to a final concentration of 0.9 mg/ml and incubated at 37°C for 4 h to selectively kill erythromycin-resistant bacteria. Bacteria were plated onto BA without antibiotics and grown overnight at 37°C in 5% CO_2 in air. To screen for potential back-transformants (those that had lost antibiotic resistance), colonies were patched onto duplicate BA plates with and without the relevant antibiotics.

2.6.3 Preparation of *E. coli* competent cells and transformation

Competent *E. coli* XL10-Gold (XL10) cells were prepared by growing the cells in LB medium to A_{600} 0.60 at 37°C, followed by incubation at 4°C for 30 min. Cells were centrifuged and the pellet resuspended in ice cold 100 mM $MgCl_2$, followed by another centrifugation and resuspension in 100 mM $CaCl_2$, and incubation for a further 30 min at 4°C. 30% (v/v) glycerol was added to the competent cells and aliquots of 100 µl were stored at -80°C until required. For transformation of plasmid DNA into the competent *E. coli* XL-10, equal volumes of KCM (100 mM KCl, 30 mM $CaCl_2$, 50 mM $MgCl_2$) and competent cells were mixed, together with the ligation mixture containing the plasmid. This was placed on ice for 30 min, followed by heat-shock at 42°C for 90 s. After this, the

cells were placed into 1 ml LB and incubated at 37°C for one h with shaking. Since the plasmid contained an erythromycin-resistance cassette, *E. coli* XL10 transformants were selected on LB agar supplemented with 500 µg/ml erythromycin and incubated at 37°C in air, overnight. Presence of the appropriate insert was confirmed by PCR analysis of recombinant plasmid DNA (Section 2.8.2) and DNA sequencing (Section 2.8.8).

2.6.4 Over-expression of *spxB* in strain D39O

The overexpression of *spxB* in strain D39O was achieved by cloning the *spxB* gene from strain D39T into a derivative of the *Streptococcus-Escherichia coli* shuttle vector pVA838. The vector constitutively expressed *lux* genes under the *S. pneumoniae* promoter aminopterin resistance operon (*ami*) (Beard *et al.*, 2002). The pAL3 vector described in this thesis is a derivative of pVA838 where the *lux* genes have been deleted and the *Eco*RI ends religated (Trappetti *et al.*, 2011c). To generate a *spxB*-expressing plasmid (pAL3:*spxB*), *spxB*-specific primer sequences carrying *Eco*RI sites (Table 2.2) were used to PCR-amplify *spxB* (Section 2.8.2) from a D39T DNA template, and product size was checked by gel electrophoresis (Section 2.8.5). The 1.8 kb product was then purified (Section 2.8.5) and digested with 20 U *Eco*RI-HF (New England BioLabs® Inc. [NEB], Massachusetts, USA), according to manufacturer's instructions. Purified pAL3 DNA was also digested using *Eco*RI-HF. 5 U Antarctic Phosphatase (NEB) was added to the vector reaction to prevent the ends re-ligating (dephosphorylation). This reaction was incubated at 37°C for 1 h, while the DNA digest was kept at 4°C until required. The ligation step involved the mixture of the two reactions at a 10:1 molar ratio (insert : vector) to a maximum of 200 ng in total using T4 DNA Ligase (NEB) at 16°C, according to the manufacturer's instructions. This plasmid was then transformed into *E. coli* as described in Section 2.6.3.

Prior to transforming the pAL3:*spxB* clone into *S. pneumoniae* strain D39O, plasmid was extracted using a QIAprep Spin Miniprep Kit (Qiagen, VIC, Australia), according to the manufacturer's instructions, and eluted in sterile ultrapure water. Transformation of *S. pneumoniae* was as described in Section 2.6.1.

2.7 Biochemical Assays

2.7.1 Protein quantification

The total protein in lysates for western blot analyses and GAPDH assays was determined using the Pierce™ BCA Protein Assay kit according to the manufacturer's instructions (Thermo Scientific, Illinois, USA).

2.7.2 SDS-PAGE and Western Immunoblotting

Relative protein expression was determined by quantitative western blotting of whole cell lysates. Bacteria were grown to A_{600} 0.19 in SB or A_{600} 0.50 in C+Y, concentrated 40-fold and lysed by treatment with 0.1% (w/v) DOC. The total protein in the lysates was determined using the BCA Protein Assay kit (Section 2.7.1). Samples containing approximately 15 µg of the protein solubilised in LUG buffer (2.5% [v/v] β-mercaptoethanol, 31.25 mM Tris-HCl, 1% [w/v] SDS, 5% [v/v] glycerol, and 0.025% [w/v] bromophenol blue; pH 6.8) and 10 µl Novex® Sharp Pre-Stained Protein Standard (Life Technologies) were loaded onto Bolt® 4-12% Bis-Tris Plus Gels (Life Technologies) and subjected to electrophoresis at 180 V for 35 min using a Blot® Mini Gel Tank (Life Technologies). Separated proteins were electroblotted onto nitrocellulose membrane (Invitrogen) using an iBlot Dry Blotting System (Invitrogen) according to the manufacturer's instructions, followed by blocking of the membrane using 1% (w/v) skim milk powder dissolved in Tris-buffered saline containing 0.02% Tween 20 (TTBS), for 1 h at room temperature. The membrane was then probed with the appropriate mouse-polyclonal antiserum at a dilution of 1:2000 to 1:5000 in TTBS supplemented with 0.05% (w/v) skim milk powder, and then reacted with donkey anti-mouse IRDye 800CW secondary antibody (LI-COR Biosciences, Nebraska, USA) at a dilution of 1:50000. The filter was scanned using the Odyssey Infrared Imaging System (LI-COR Biosciences). Relative quantification of protein expression between the strains was performed using the Odyssey Infrared Imaging System application software v 3.0.21 by comparing units of fluorescence (LI-COR Biosciences).

Alternatively, colourmetric detection was performed using a goat anti-mouse 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 four times for 5 min with TTBS. After washing, the membrane was equilibrated by rinsing with 15 ml DIG buffer 3 (100mM Tris, 100 mM NaCl, 50 mM $MgCl_2$, pH 9.5) for 2 min 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-indolyphosphate toluidinium salt in dimethylformadmide) 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.

2.7.3 GAPDH Assay

Pneumococci were grown in 10 ml C+Y to A_{600} 0.50 and the cell pellet was harvested as described in Section 2.3. The pellet was resuspended in 1 ml PBS and lysed by ultrasonication (Branson sonifier cell distruptor 200, Sonicator Heat Systems-Ultrasonics Inc. NY) on ice, applying four cycles of 10 pulses (duty cycle, 60%; output, 6) with one minute intervals between cycles. The cell lysate was centrifuged at 13,000 × *g* for 10 min at 4°C and supernatant transferred to another microfuge tube. Samples were kept on ice until required. The GAPDH assay involved measurement of the kinetics of the reduction of NAD (β-Nicotinamide adenine dinucleotide hydrate) according to Fillinger *et al*. (2000), with modifications. Using a 1 ml disposable cuvette, 900 µl triethanolamine/sodium arsenate buffer (125 mM triethanolamine [Sigma], 5 mM L-cysteine [Sigma], 20 mM sodium arsenate [Sigma], 50 mM disodium hydrogen phosphate (Na_2HPO_4) [Merck], pH 9.2), 2 mM NAD (Sigma) and 60 µl cell extracts were mixed and A_{340} was measured at 25°C. Initially, the reaction mix was allowed to react for about 10 s to ensure that there was no baseline NAD reduction. Following this, 4 mM D-glyceraldehyde-3-phosphate (G-3-P) substrate (Sigma), was added and the reaction was recorded for 3-5 min at 25°C. One unit of GAPDH reduces one micromole of NAD per minute and is calculated as follows, where the ΔA_{340}/minute at 25°C is derived from the linear portion of the graph:

$$Units\ per\ mg = \frac{\Delta A340 \text{ per minute}}{6.22 \times \text{mg enzyme per ml reaction mixture}}$$

2.7.4 Uronic acid assay

This assay was performed essentially as described by Morona *et al.* (2006). CPS samples were prepared by growing the pneumococci in 10 ml TSB to A_{600} 0.2 in 5% CO_2 in air, at 37°C. Samples were pelleted at 3,600 × *g* and resuspended in 500 ml of 150 mM Tris-HCl (pH 7.0), 1 mM $MgSO_4$. The samples were then treated with 0.1% (w/v) DOC and incubated at 37°C for 30 min. Further digestion of the samples was carried out by addition of 100 U of mutanolysin, 50 U DNase I and 50 µg RNaseA and incubation overnight at 37°C. Prior to determining the relative quantities of uronic acid, 100 µg proteinase K was added to each sample and incubated at 56°C for 4 h. Samples were diluted two-fold, in duplicate, to a final volume of 100 µl in ultrapure water, including 100 µl of ultrapure water as blank. To each sample, 600 µl of concentrated H_2SO_4/0.0125 M $Na_2B_4O_7$ solution were added, vortexed and then heated at 100°C for 5 min before cooling the samples on ice. This was followed by the addition of 10 µl 3-phenylphenol solution (0.15% [w/v] 3-phenylphenol in 0.5% [w/v] NaOH) and 10 µl 0.5% (w/v) NaOH to the duplicate tube as a control. After mixing the samples, 200 µl of each was transferred to a microtitre tray and A_{520} was measured using a SpectraMax M2 Microplate Reader. The difference in the amount of uronic acid was calculated by subtracting the NaOH-only control from the respective test sample reading (3-phenylphenol-added sample).

2.8 DNA Isolation, Manipulation and Analysis

2.8.1 DNA extraction for PCR

DNA extraction for PCR was carried out using the Wizard™ Genomic DNA purification Kit (Promega Corporation, Madison, WI). Cell pellets were treated with 0.1% (w/v) DOC and incubated for 30 min at 37°C before preparing DNA according to the manufacturer's instructions.

2.8.2 DNA extraction for genome sequencing

Whole chromosomal DNA extraction for Ion Torrent^{TM} sequencing (Life Technologies) was carried out using the Qiagen Blood and Cell Culture DNA Kit with Qiagen Maxi tips (tip 500/G) (Qiagen), according to the manufacturer's instructions, except that cells were lysed in 0.1% (w/v) DOC for 30 min at 37°C prior to the addition of 100 U mutanolysin, lysozyme and Proteinase K treatments, as described in Section 2.3. To check the integrity of the chromosomal DNA preparations, the samples were analysed using agarose gel electrophoresis (Section 2.8.5), followed by preliminary quantification using a NanoDrop 1000 Spectrophotometer (ThermoFisherScientific) with NanoDrop 1000 v 3.7.1 software.

2.8.3 DNA extraction for allele quantification

To generate genomic DNA for allele quantification, colonies from BA were placed into 1 ml TSB and incubated at 37°C in 5% CO_2 in air for 3 h, followed by pelleting the cells before DNA extractions. DNA samples for methylome analysis were extracted from *in vivo* and *in vitro* samples using a DNeasy® Blood and Tissue kit (QIAGEN) according to the manufacturer's protocol, with some modifications. In brief, cells were harvested by centrifugation at 18,000 × *g* on a benchtop centrifuge and the pellet was resuspended in 200 µl lysis buffer containing 0.1% (w/v) DOC and 20 mg/ml lysozyme. Pretreatment for Gram-positive bacteria was carried out, followed by "Purification of total DNA from animal tissues" (Spin-column protocol), as per the manufacturer's instructions and samples were eluted in 30 µl elution buffer from the kit.

2.8.3.1 SpnD39III Allele quantification

SpnD39III allele quantification was carried out at the Institute for Glycomics, Griffith University, QLD, Australia, as described by Manso *et al.* (2014).

2.8.4 Polymerase chain reaction

PCR reactions were performed using a Mastercycler Flexilid thermal cycler (Eppendorf, NSW, Australia) in a final volume of either 25 µl or 50 µl. Standard reactions were carried out using 2 × Phusion Flash PCR Master Mix (Thermo Fisher Scientific, VIC, Australia),

according to the manufacturer's instructions. Typical reaction conditions comprised of 25 cycles of denaturation at 95°C for 30 s, annealing at 55°C for 30 s and extension at 68°C for 15 s per kb of expected PCR product.

2.8.5 Overlap-extension polymerase chain reaction

Site-directed mutagenesis using overlap-extension polymerase chain reaction (OE-PCR) (Ho *et al.*, 1989) was used to amplify the product required for *S. pneumoniae* transformations. Primers used to amplify the respective fragments are listed on Table 2.2.

2.8.6 PCR product purification

PCR products were purified using a MinElute PCR Purification Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions with the exception that DNA was eluted in ultrapure water.

2.8.7 Agarose gel electrophoresis

To analyse PCR products or chromosomal DNA preparations, gel electrophoresis was performed using 0.8-2.0% (w/v) agarose (Agarose low EEO [AppliChem, Germany]) in TBE buffer (44.5 mM Tris, 44.5 mM boric acid, 1.25 mM EDTA, pH 8.4). Prior to loading, DNA was mixed with $1/10^{th}$ volume of loading buffer (15% [w/v] Ficoll, 0.1% [w/v] bromophenol blue, 100 ng/ml RNAse A). Gels were electrophoresed in 0.5 × TBE buffer at 180 V and were then stained using GelRed™ (Biotium, California, USA) according to the manufacturer's instructions. DNA bands were visualised by transillumination with short wavelength ultraviolet light using a Gel/Chemi Doc XR system (Bio-Rad, NSW, Australia) and analysed using Quantity One v 4.6.9 software.

2.8.8 DNA sequencing

DNA sequencing of PCR products was performed by Australian Genome Research Facility Ltd (AGRF). For sequencing using Sanger Sequencing, reactions were prepared according AGRF specifications. The resultant sequences were visualised and analysed using Chromas Lite v 2.0 (Technelysium) and DNAMAN v 4.0 (Lynnon Biosoft).

2.9 RNA Isolation, Manipulation and Analysis

2.9.1 RNA extraction

RNA was isolated from bacterial cell pellets using extraction with acid-phenol:chloroform:isoamly alcohol (APIA) (25:24:1; pH 4.5) (Ambion, Austin, TX., USA) as described previously (LeMessurier *et al.*, 2006; Ogunniyi *et al.*, 2002). Cells were harvested as described in Section 2.3 and the pellets were resuspended in 300 µl pre-warmed APIA. This was followed by incubation at 65°C for 5 min before adding 400 µl NAES (50 mM sodium acetate, 10 mM EDTA, 1% [w/v] SDS, pH 5.1, treated with diethyl pyrocarbonate [DEPC]), mixing and incubating for a further 5 min at 65°C. The mixture was cooled for 1 min on ice and centrifuged at 18,000 × *g* for 2 min at 4°C to separate the clear upper aqueous phase (nucleic acid) from the lower organic (mostly protein) phase. The aqueous phase was carefully withdrawn into a separate tube and added to 400 µl of pre-warmed APIA for a second round of extraction. The extractions were performed 2 additional times until there was no white material at interface between the aqueous and organic phases. Following the extractions, RNA was precipitated from the aqueous phase by adding 2.5 volumes of 100% ethanol and 1/10th volume 0.05% (w/v) DEPC-treated 3 M sodium acetate and stored at -80°C overnight. The precipitated RNA was pelleted by centrifugation at 13,000 × *g* for 30 min at 4°C, washed in 70% (v/v) ethanol and repelleted by centrifugation under the same conditions for 10 min. The washing procedure was repeated twice, after which the supernatant was removed and any residual ethanol was removed from the pellet by drying in a SpeedVac for 5 min. The dried pellet was then resuspended in 50 µl nuclease-free water (Roche). Prior to DNase treatment, RNA was heated to 95°C for 3 min to denature any RNA : DNA complexes that could otherwise compromise the DNase treatment. Contaminating DNA was then removed by treatment with 5 U RQ1 RNase-free DNase (Roche) at 37°C for 1 h in the presence of 1 U/µl RNasin Ribonuclease Inhibitor (Promega Life Sciences, WI., USA). Samples were then stored at -80°C.

2.9.2 Reverse transcription polymerase chain reaction (RT-PCR)

DNA contamination of RNA preparations was assessed by one-step RT-PCR using the Access RT-PCR system kit (Promega) according to the manufacturer's instructions, using primers 16S F and 16S R (Table 2.2). A standard 12 µl RT-PCR reaction contained DEPC-treated water, 0.2 mM of each dNTP, 1 µM of each primer, 1 mM MgSO4, AMV/*Tfl* reaction buffer, 0.1 U/µl AMV Reverse Transcriptase, 0.1 U/µl *Tfl* DNA polymerase, 0.25 µl RQ1 DNase stop solution and 0.25 µl of the relevant template.

RT-PCR was carried out initially with a reverse transcription cycle at 48°C for 45 min, followed by 25 cycles of amplification comprising denaturation at 95°C for 30 s, annealing at 55°C for 30 s and extension at 68°C for 20 s. Successful DNAse treatment was confirmed by the absence of a 200 bp RT-PCR product in reactions lacking the reverse transcriptase when analysed by agarose gel electrophoresis.

2.9.3 Comparative transcriptomic hybridization

2.9.3.1 Preparation of microarray slides

Microarray slides were obtained from the Bacterial Microarray Group at St. George's Hospital, University of London, and consisted of PCR products for each of the identified ORFs of the *S. pneumoniae* TIGR4 genome, supplemented with additional PCR products of ORFs identified in the genome of *S. pneumoniae* R6. The microarray slide was blocked with blocking solution (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. Following blocking, the slide was washed twice with ultrapure water followed by washing twice in 100% isopropanol. The slides were then dried by centrifugation at low speed (100 × *g* [Eppendorf Centrifuge 5403]) for 2 min.

2.9.3.2 Generation of probes

Approximately 3 µg RNA was labelled using the Genisphere 3DNA Array 900 MPX^{TM} kit (Genisphere®, P.A., USA) essentially as described in the manufacturer's instructions, but with the following modifications. Firstly, Superscript^{TM} III Reverse Transcriptase (Invitrogen) was used for preparation of the RNA-RT reaction mix to generate cDNA.

Secondly, the first set of array washes were performed at 37°C, and after hybridisation of cDNA to the microarray slides, they were washed for 20 min in 2 × SSC supplemented with 0.02% (w/v) SDS and 0.075% (w/v) dithiothreitol (DTT) at 65°C, then for 15 min in 2 × SSC and 0.075% (w/v) DTT at 48°C and finally for 15 min in 0.2 × SSC at 37°C. The slides were then dried by centrifugation at 1,000 × *g* for 2 min.

Slides were scanned at 10 µm resolution using a Genepix 4000B Scanner (Molecular Devices, USA). Detector photomultiplier (PMT) voltages were adjusted individually for each slide so that the total red (Alexa Fluor 647) and green (Alexa Fluor 546) fluorescence signals for each channel were approximately equal, while minimising the total number of features with signal above the maximum detectable. Foreground and background mean pixel intensity values were extracted from the scanned images for both channels (Alexa Fluor 546, Alexa Fluor 647) using the Spot plugin (CSIRO, Australia) within the R statistical software package (http://www.R-project.org). The Limma plugin for R (Smyth, 2005) was used for data processing and statistical analysis. After background subtraction, the foreground intensities were $\log_2$ transformed and a single ratio (Alexa Fluor 647/Alexa Fluor 546) value was obtained for each probe. Ratio values were normalized using the print-tip Loess normalization routine (Smyth & Speed, 2003). The replicate arrays were normalized to each other to give similar ranges of mRNA expression values. For each probe across the arrays, a linear model was fitted to determine a final fluorescence expression value for each mRNA probed (Smyth, 2004). These statistics were used to rank the mRNAs from those most likely to be differentially expressed to the least likely, using false-discovery rate values of $p < 0.05$. Microarray analysis comparing RNA between D39O and D39T, WCH16O and WCH16T, and WCH43O and WCH43T was performed on at least four independent hybridisations for each pair-wise comparison, including at least one dye reversal per comparison for each strain.

2.9.4 Real-time RT-PCR (qRT-PCR)

qRT-PCR was performed using the Superscript™ III Platinum SYBR® Green One-Step qRT-PCR kit (Invitrogen), with specifically designed primers (Table 2.2) on a LightCycler® 480 II [Roche]. The cycling program used for real-time RT-PCR was 50°C

for 15 min, 95°C for 5 min, followed by 40 cycles of 95°C for 15 s, 55°C for 30 s and 72°C for 15 s. The acquisition of fluorescence was undertaken at the end of the 72°C step of each cycle. The cycling phase was followed by a 40°C step for 1 min. Melt curves were generated by cycles comprising 95°C for 5 s, 65°C for 1 s, which was increased at a rate of 0.11 °C/s to 97°C. The acquisition of fluorescence was carried out continuously during the gradual temperature increase phase.

2.9.4.1 qRT-PCR data analysis

The relative amount of target mRNA present in different RNA samples was calculated using the comparative cycle threshold ($2^{\Delta\Delta Ct}$) method (Livak & Schmittgen, 2001). The amount of target mRNA in one sample was compared with the amount of the same target mRNA in another sample, normalised to the internal control 16S rRNA concentration. Alternatively, the amount of target mRNA relative to the amount of 16S rRNA was also calculated in some instances.

The standard deviations (SD) were determined as $\sqrt{((SDsample)^2 + (SD\ 16S)^2)}$, and this was applied to the formulas: SD+ = $2^{\Delta\Delta Ct\text{-}SD}$ -$2^{\Delta\Delta Ct}$ and SD- = $2^{\Delta\Delta Ct+SD}$.

2.10 Preparation and analysis of samples for genome sequencing

2.10.1 DNA library preparation

In order to sequence genomic DNA using Ion Torrent™ technology, the DNA was fragmented to approximately 100-200 bp, and adapters were ligated to the DNA, which was followed by nick-translation and amplification of the DNA library.

2.10.2 DNA fragmentation, ligation of adapters and size-selection

DNA fragmentation, ligation of adapters and size selection was carried-out using the Ion Xpress™ Fragment Library Kit (Invitrogen) according to the manufacturer's instructions, with some modifications.

When purifying the 100 bp adapter-ligated DNA, 1.55 × sample volume of Agencourt^{R} AMPure^{R} beads (Beckman Coulter) were added instead of 1.8 × sample volume described in the manufacturer's instructions. DNA fragments were selected by adding 1.20 × sample volume of beads, followed by incubation for 5 min at room temperature. The DNA fragments attached to the beads were separated by the placing the tubes in a magnetic rack until the solution was clear. The supernatant was then removed and placed into a new DNA Lo-Bind tube (Eppendorf, Hamburg, Germany) and DNA samples were purified using the Qiagen Min-Elute Kit (Section 2.8.5), except that samples were eluted in 23 µl low TE buffer (from Ion Xpress^{TM} Fragment Library Kit). Six cycles were used during the library amplification step.

For the preparation of 200 bp DNA fragments, size-selection was carried out by firstly adding 0.9 × sample volume of beads and vigorous vortexing. After a 5 min incubation, the solution containing adapter-ligated DNA was placed on the magnetic stand until clear and the supernatant transferred to a new tube. This was followed by the addition of 0.2 × sample volume of beads and then allowing the reaction to occur at room temperature. DNA-attached beads were washed with 70% ethanol twice and samples air-dried before eluting in low TE buffer. Eight cycles were used during the library amplification step.

Initial DNA yield and quality for both 100 and 200 bp fragments were analysed using a Nanodrop spectrophotometer. This was then confirmed using an Agilent 2100 Bioanalyzer (Agilent Technologies Inc., California, USA) using the High Sensitivity DNA Assay. Samples were stored at -20°C until required.

2.10.2.1 Enrichment and whole genomic DNA sequencing

Depending on which kit was used to generate the DNA library in Section 2.10.2, enrichment and amplification of the DNA library was performed using the Ion OneTouch^{TM} 100 Template Kit or the Ion OneTouch^{TM} 200 Template Kit v2 (Life Technologies) on the Ion OneTouch^{TM} System, according to manufacturer's instructions. To ensure that the ratios of enriched : unenriched Ion Sphere^{TM} Particles (ISPs) were suitable for sequencing, quality control was performed using the Ion Sphere^{TM} Quality Control Kit (Life Technologies) on the Qubit Fluorometer 2.0 (Life Technologies)

according to the manufacturer's instructions. If the DNA library was at an acceptable ISP ratio, dilutions of the DNA libraries were prepared using the Ion PGM™ OneTouch Calculator v 2.0 and then sent to the ACGF Cancer Genomics Facility (Adelaide, Australia) for sequencing on the Ion Personal Genome Machine® (PGM™) System (Life Technologies).

2.10.2.2 Identification of INDELs, SNPs and homopolymeric tracts

Identification of insertions/deletions (INDELs) and single nucleotide polymorphisms (SNPs) between the opaque and transparent variants was carried by firstly aligning the sequenced genome with the appropriate reference genomes using Bowtie2 (2.0.0.0-beta6) (Langmead & Salzberg, 2012). Strains D39 and WCH43 (clonally related to TIGR4) had publicly available genome sequences (NC_008533.1 (Lanie *et al.*, 2007) and NC_003028 (Tettelin *et al.*, 2001) respectively), while *de novo* genome assemblies using MIRA v 3.4.1.1 (Chevreux *et al.*, 1999) were used for WCH16. SAMtools (v 0.1.18) (Li *et al.*, 2009) was used to identify SNPs and indels from the alignments, whereas BEDTools (v 2.24.0) (Quinlan & Hall, 2010) was used to compare the lists generated by SAMtools. Homopolymeric tracts were specifically filtered out of these data by using an in-house custom algorithm (S. Bent [University of Adelaide], personal communication). Artemis (v 14.0.0) (Rutherford *et al.*, 2000) was used to visualise variants (SNP or INDEL) in the alignments. The Basic local alignment search tool v 2.2.24+ (BLAST) (Altschul *et al.*, 1990) was used as appropriate, and indels and SNPs of interest were confirmed by Sanger sequencing (Section 2.8.8).

2.11 *In vivo* murine model

2.11.1 Mice

In vivo models of pneumococcal disease in this project used 5- to 6-week old outbred female CD1 (Swiss) mice obtained from the Laboratory Animal Services breeding facility at the University of Adelaide. Experiments were carried out in compliance with the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes (7th Edition, 2004) and the South Australian Animal Welfare Act 1985. Ethics approval for all

experiments in this project was granted by the Animal Ethics Committee of the University of Adelaide.

2.11.2 Growth of challenge strain

Strains of *S. pnuemoniae* to be used for challenge studies were inoculated from an overnight BA plate into sterile SB and grown to A_{600} 0.17 or 0.20, corresponding to 5×10^7 or 1×10^8 cfu/ml, respectively.

2.11.3 Intranasal challenge model

Mice were anaesthetised by intraperitoneal (i.p.) injection of pentobarbital sodium (Nembutal) (Rhone-Merieux, QLD, Australia) at a dose of 66 µg/g body weight, after which 1×10^7 CFU of bacteria (Section 2.11.2) suspended in 50 µl serum broth was administered into the nares. The challenge dose was confirmed retrospectively by plating on BA. For pathogenesis experiments, mice were euthanased 48 h after the challenge by CO_2 asphyxiation and samples collected from the nares (nasal wash) and posterior vena cava (blood). All samples were serially diluted and plated on BA supplemented with 5 µg/ml gentamicin (to inhibit contaminating flora) and incubated overnight at 37°C in 5% CO_2 in air. Following this, *S. pneumoniae* colonies were enumerated to determine the viable count. Statistical significance of differences was calculated on log-transformed data using the unpaired (one-tailed) *t*-test.

2.11.4 Sepsis model

Female outbred 5- to 6-weekold CD-1 (Swiss) mice were inoculated intravenously with 1×10^5 CFU of *S. pneumoniae* (Section 2.11.2) (confirmed retrospectively by viable count) in a volume of 100 µl in SB. Groups of six mice were inoculated for each strain and blood was collected by submandibular bleeding at 4 and 24 h post-infection. At 30 h post challenge, mice infected with each strain were euthanased by CO_2 asphyxiation and blood collected from the vena cava. After perfusion with PBS, brain, spleen and liver tissues were excised and homogenised in 1 ml PBS using a Precellys® 24 tissue homogeniser (Bertin Technologies, France). All tissue samples were plated after serial dilution onto BA

supplemented with gent plates for enumeration of viable pneumococci and on catalase-supplemented TSB agar plates to determine colony opacity phenotype. Statistical significance of differences was calculated on log-transformed data using the unpaired (two-tailed) *t*-test.

Chapter 3: PROTEOMIC ANALYSIS OF OPACITY PHASE VARIABLE PNEUMOCOCCI

3.1 Introduction

The propensity of *S. pneumoniae* to cause disease lies in its ability to translocate from the nasopharynx where it colonises asymptomatically, to invasion of deeper host tissues. This transition would require it to adapt to different microenvironments and thus, regulate the expression of its proteins and other factors to best allow it to colonise and resist host immune responses in each niche. One apparent adaptive mechanism of the pneumococcus involves the reversible phase variation in colony phenotype from "T" to "O" and vice versa.

Pneumococcal colony opacity has been demonstrated to have a role in pathogenicity in mouse sepsis models, where all mice challenged with the O form of a virulent pneumococcal strain died, whereas, only a few mice succumbed to infection with the T variant (Weiser, 1998). Furthermore, in the same study, it was shown that only organisms of the O phenotype were recovered from the spleen in a mouse model of sepsis. Variation in the amount of capsule has been reported to be associated with colony opacity variation, with a higher cell wall teichoic acid to capsule ratio in the T form and vice versa in the O form (Kim & Weiser, 1998). This could be of importance as pneumococcal teichoic acid contains ChoP, which is important as it mediates binding to the platelet-activating factor receptor (PAFr) on respiratory epithelium, hence its importance in colonisation of the nasopharynx (Gosink *et al.*, 2000; Rosenow *et al.*, 1997). However, Weiser *et al.* (1994) showed that colony opacity variation still occurs in unencapsulated mutants, suggesting that the variation in the amount of capsule is not solely responsible for the difference in colony opacity phenotype, but could account for the difference in degree of systemic virulence.

To date, there has been no simultaneous comparative studies of the protein profiles of opacity variants in different pneumococcal strains. The inconsistencies in the literature

with regard to distinct expression patterns of protein between pneumococcal colony opacity variants (Overweg *et al.*, 2000; Weiser *et al.*, 1994) could have arisen because the pneumococcal strains and analytical techniques in each of those studies were different. In order to more accurately identify the commonly regulated proteins between O and T phenotypes, strains belonging to three serotypes (2, 4 and 6A) with distinct pathogenicity characteristics were analysed in this study. In our mouse intranasal challenge model, strain D39 (serotype 2) causes severe pneumonia and high-grade bacteraemia, WCH43 (serotype 4) demonstrates the "classical" disease progression from the nasopharynx to the lungs and dissemination to the blood and then to the brain, while WCH16 (serotype 6A) appears to progress directly to the brain with minimal lung and blood involvement (Mahdi *et al.*, 2008; Mahdi *et al.*, 2013; Orihuela *et al.*, 2003). This chapter describes the expression profiles of proteins of the O and T variants of the three strains, using 2D-DIGE and identification of protein spots of interest using mass spectrometry, as part of a three-way approach to elucidate the mechanism of colony opacity phase variation in pneumococci. Proteins of interest were then subjected to mutational and biochemical analyses. Expressional levels of selected proteins were also evaluated using quantitative Western blot analysis. To ensure consistency and reproducibility, four biological replicate cultures of the various O and T variants of each of the three strains were prepared, as described in Section 2.3 and extracts from the same cultures were used for transcriptomic and genomic analyses (described in Chapters 4 and 5, respectively), as well as for proteomic analysis, as described below.

3.2 Results

3.2.1 Identification of O and T variants

Mixed populations of O and T variants of the three strains were streaked-out for single colonies on THY or C+Y agar plates supplemented with catalase to distinguish between the phenotype under oblique, transmitted light, as described in Materials and Methods (Section 2.2.8). Catalase was added to the media to decompose the H_2O_2 produced by pneumococci during aerobic growth. O colonies are typically of convex elevation, compared with those of T phenotype, which have an umbonate elevation, giving a

doughnut-shaped appearance as shown in Figure 3.1. Optimum incubation times for viewing colony opacity were markedly different from strain to strain: 18 h for D39, 24 h for WCH16, and 36 h for WCH43. To minimise agar-depth and plate batch bias, variants of the same strain were grown on the same agar plate. This also allowed the opacity of the colonies to be distinguished more easily. Pure colonies of each variant were then selected, from which frozen aliquots were prepared (Section 2.2.3). These were used as working stocks for subsequent experiments.

3.2.2 Differentially Expressed Proteins in Phenotypic Variants of *S. pneumoniae*

The relative protein expression levels of four biological replicates for each of the O and T variants of strains D39, WCH16 and WCH43 were compared using 2D-DIGE. Each variant was labelled with either Cy3 or Cy5 dye, and using Cy2 dye as an internal standard. For each of the four biological replicates of each strain, a technical replicate of the O variant labelled with either Cy3 or Cy5 was prepared to eliminate any bias in dye labelling. After standardisation to Cy2, the fluorescence intensities of proteins spots labelled with Cy3 and Cy5 on each gel were compared. Differential expression of specific spots between O and T variants of each strain was determined by statistical analysis using the student's two-tailed *t*-test by the Adelaide Proteomics Centre (Section 2.5.2). The spots that returned a *p*-value of <0.05 were considered significantly differentially expressed. Such spots were manually inspected (Figure 3.2) and those deemed potentially artefactual (for example, those that were due to horizontal or vertical streaking and low-level noise) were eliminated from further analysis. Protein spots were picked robotically from the 2D-DIGE gels or from replica Coomassie-stained gels for those with low protein abundance (Figure 3.3). As a result of the analyses, the top 15 protein spots of strains D39 and WCH43 and the top 13 protein spots of strain WCH16 displaying a significant difference in expression between O and T variants were chosen for robotic picking, followed by identification using mass spectrometry.

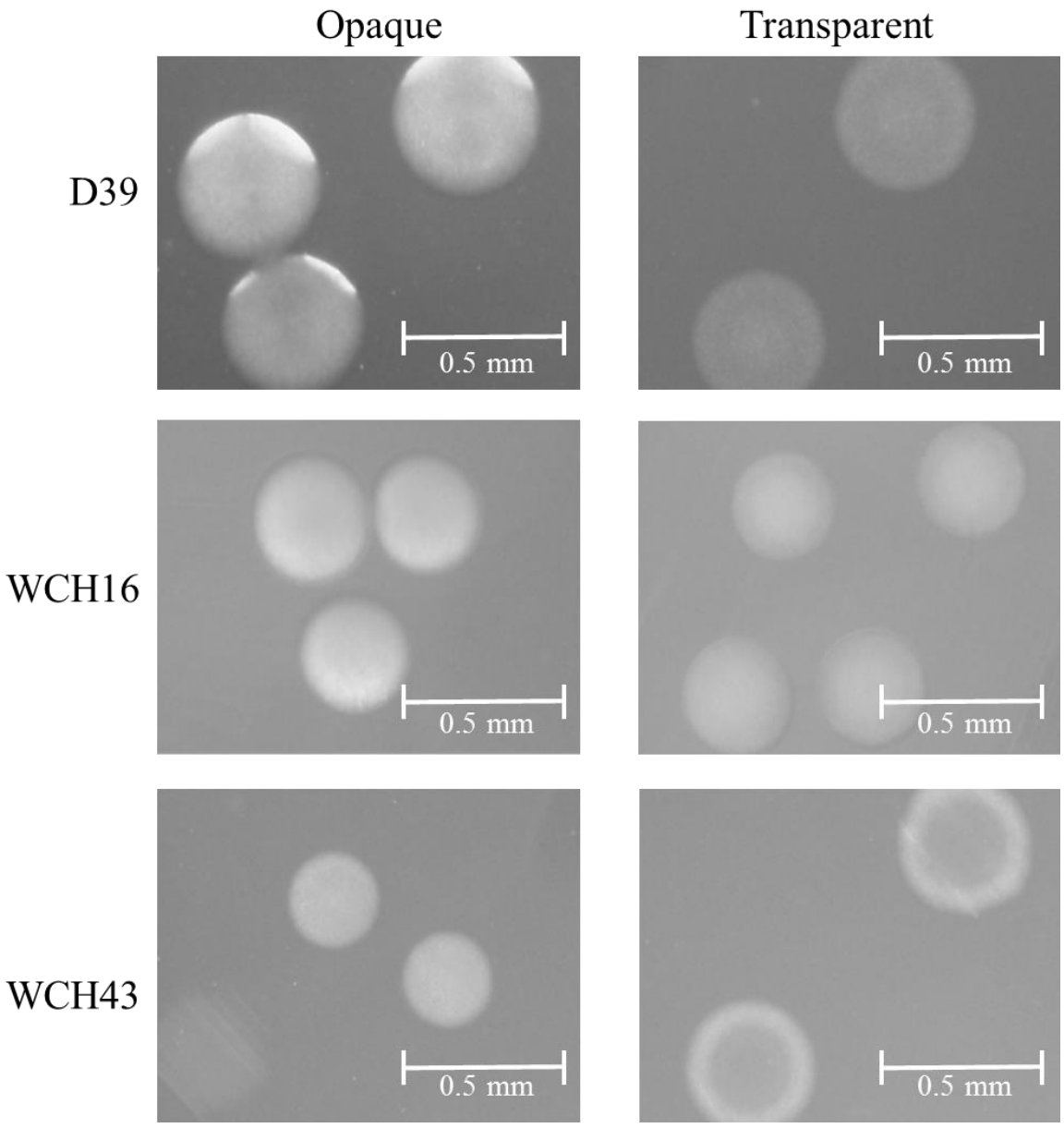


Figure 3.1. Colony morphology of O and T variants of *S. pneumoniae* strains.

Colony morphologies of O and T variants of *S. pneumoniae* strains D39, WCH16 and WCH43 grown on clear media (THY + catalase). Images were captured using a Nikon SMZ1000 dissecting microscope and cellSens software.

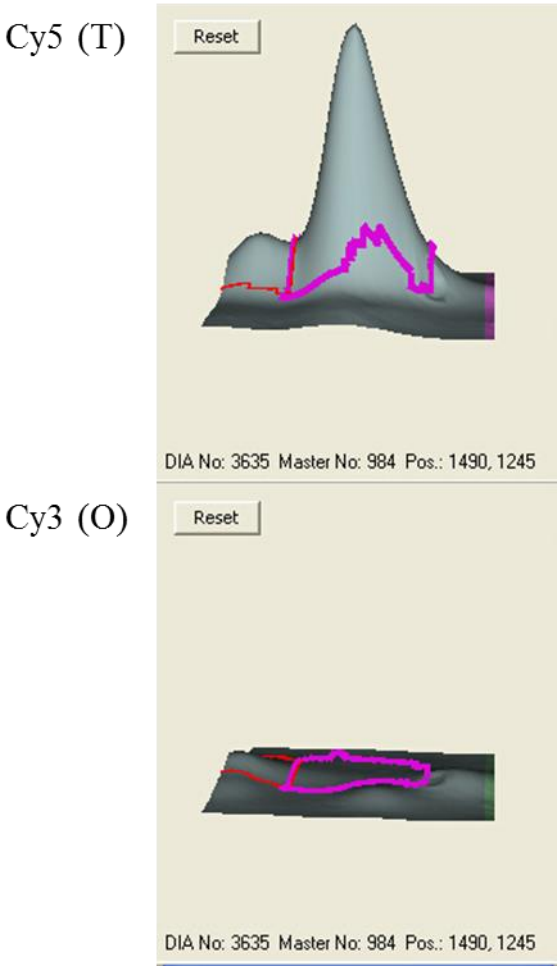
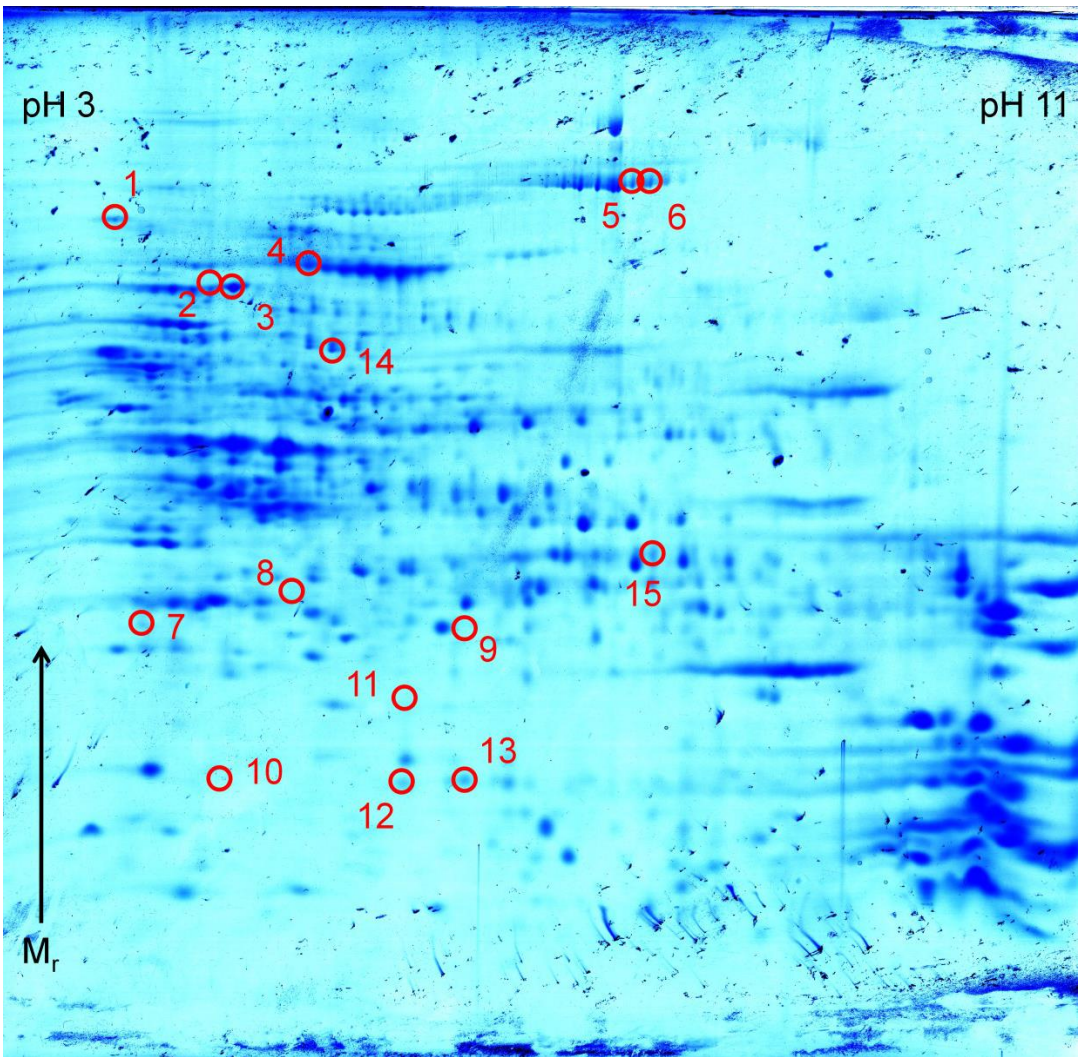


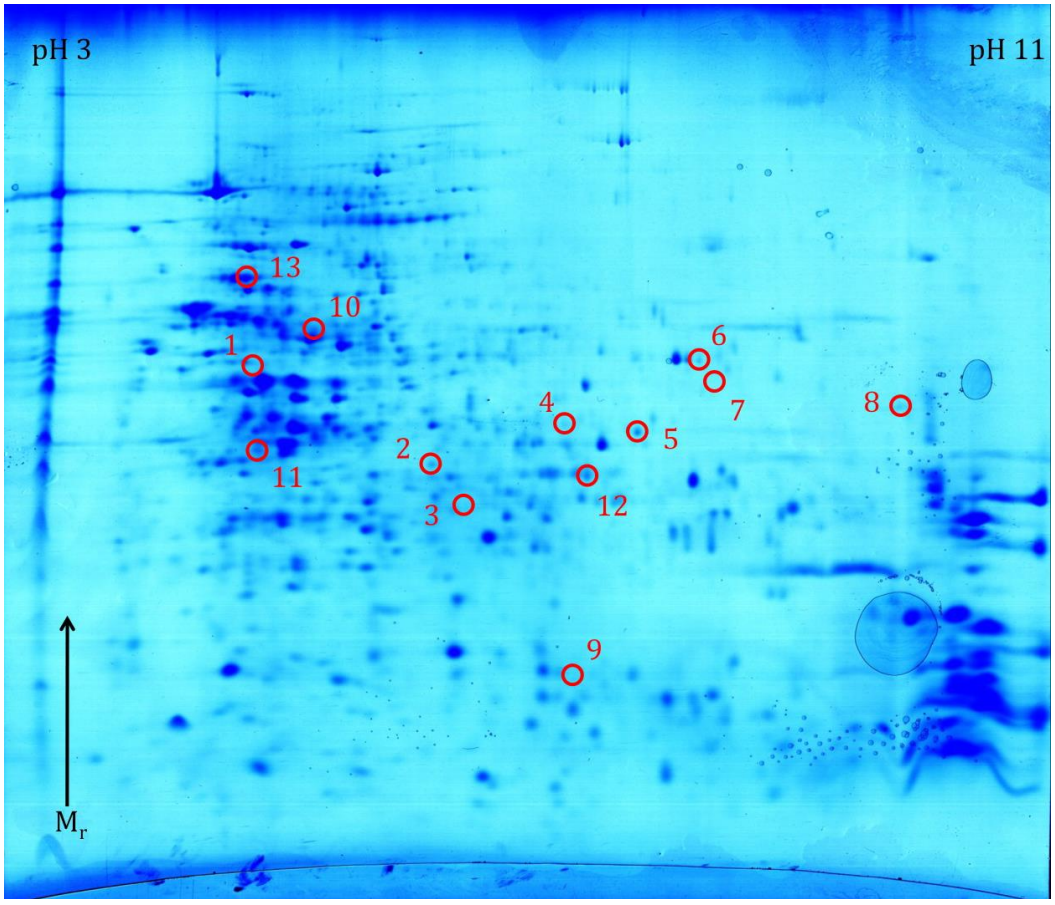
Figure 3.2. 3D view of a protein spot

An example of 3D view of a spot (master spot 984 correlates with spot 10 on Figure 3.3, WCH43) showing differential levels in the Cy5 vs. Cy3 channels as viewed on DeCyder 2D software. The peak in Cy5 shows that there was a significant upregulation of this protein in WCH43T vs. WCH43O.

D39



WCH16



WCH43

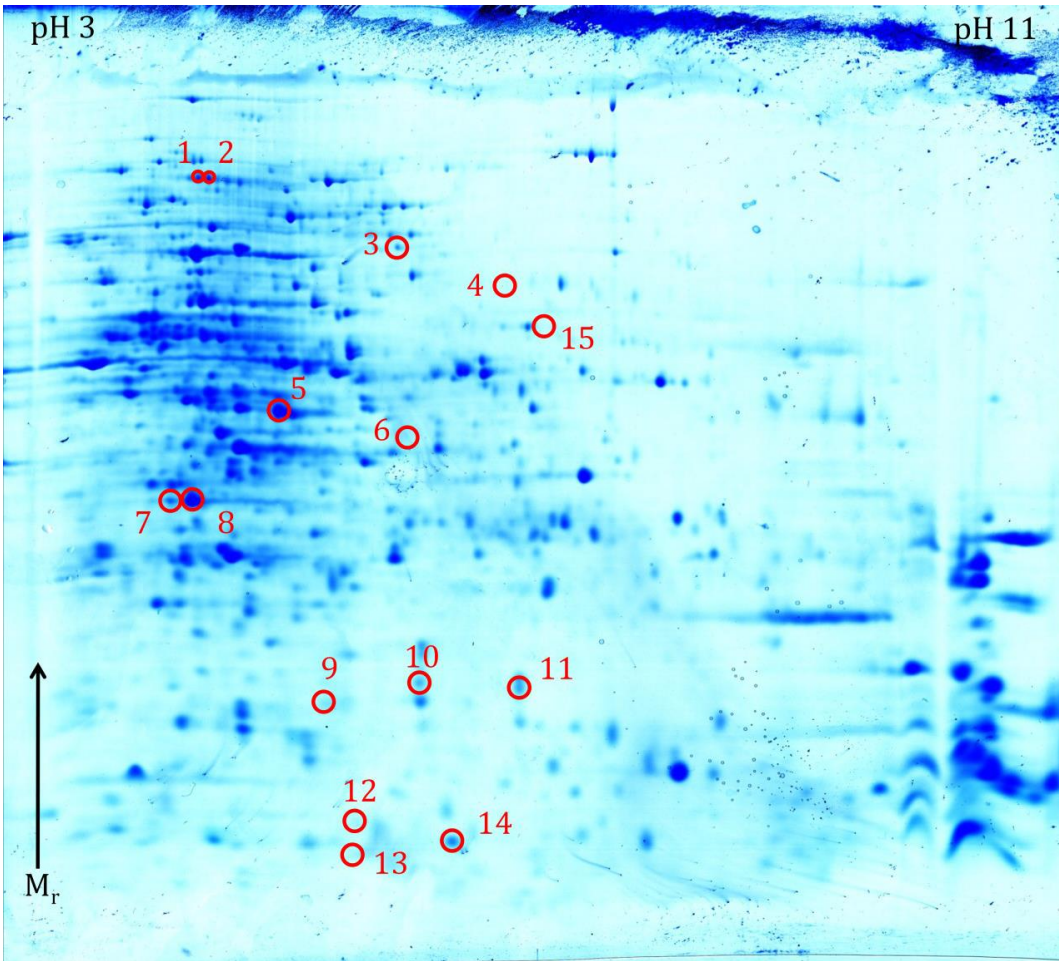


Figure 3.3. Coomassie-stained 2D gels of strains D39, WCH16 and WCH43.

Coomassie-stained 2D gels of 2D-DIGE gels of strains D39, WCH16 and WCH43. Protein spots picked and identified using LC-ESI-IT MS are indicated. These numbers correspond to spot numbers on Tables 3.1, 3.2 and 3.3.

3.2.2.1 Analysis of differential protein expression profiles of D39O vs. D39T

For D39, a total of 982 distinct protein spots were detected on the master gel and 602 of these were present in 75% of the spot maps (i.e. 9 out of 12 gels). Univariate statistical testing (*t*-test) detected 65 differentially expressed spots with a *p*-value <0.05. At α=0.05, the q-value extended false discovery rate (FDR) was calculated to be 13% (that is, 13% of these spots are predicted to be false positives). Of the 65 spots, 37 spots were upregulated in D39O and 28 were upregulated in the D39T. The top 15 spots that were picked and identified using mass spectrometry (Table 3.1) had fold-changes >2.0 and *p*-value <0.05 (Figure 3.3). The proteins upregulated in D39O were bifunctional acetaldehyde-CoA/alcohol dehydrogenase (AdhE; multiple protein spots; 10.9- and 4.4—folds), elongation factor G (EF-G; 4.2-fold), a hypothetical protein (SpneT_02001699; 3.5-fold), single-stranded DNA-binding protein (SsbB; 3.4-fold), SpxB (2.9-fold) and a response regulator (ComE; 2.5-fold). The proteins upregulated in D39T were GAPDH (multiple protein spots; 15.8-, 2.5- and 2.1—folds), EF-Tu (5.4-fold), lactate dehydrogenase (Ldh; 3.0-fold), a sugar ABC transporter ATP-binding protein (2.5-fold) and ribosome recycling factor (2.3-fold).

3.2.2.2 Analysis of differential protein expression profiles of WCH16O vs. WCH16T

For WCH16, a total of 1168 protein spots were detected on the master gel and 674 of these could be detected in 75% of the spot maps. Statistical analysis detected 73 differentially expressed protein spots (α=0.05). Of the 73 spots, 46 spots were upregulated in WCH16O and 27 were upregulated in WCH16T. The corresponding q-value extended FDR was calculated to be 66%. The top 12 spots were picked (Figure 3.3) and identified using mass spectrometry (Table 3.2). The proteins upregulated in WCH16O were guanosine 5'-monophosphate oxidoreductase (GMP reductase GuaC; 3.7-fold), maltose operon transcriptional repressor (MalR; 2.2-fold), chaperone protein DnaJ (2.1-fold) and GAPDH (2.0-fold). The proteins upregulated in WCH16T were fructose-bisphosphate aldolase (Fba; 4.2-fold), GTP-binding protein Era (2.9-fold) and a branched chain amino acid transport ABC transporter substrate-binding protein (2.7-fold).

Table 3.1. Table showing the identities of protein spots from Figure 3.3 (D39).

Spot number[a]	Protein	GI accession no.	Predicted molecular mass (kDa) / pI	Combined IonScore[b]	Sequence coverage (%)	emPAI value[c]	Fold-change (O/T) [d]	*p*-value[e]
1	**elongation factor G**	**15900207**	**76.8 / 4.9**	**1572**	**56**	**2.7**	**4.2**	**0.0059**
2	**pyruvate oxidase**	**15900627**	**65.3 / 5.1**	**975**	**42**	**1.2**	**2.9**	**0.0007**
2	PTS system, mannose-specific IIAB components	15900218	35.8 / 5.2	451	49	1	2.9	0.0007
3	**hypothetical protein SpneT_02001699**	**111657073**	**35.4 / 5.2**	**1353**	**68**	**8.4**	**3.5**	**0.0004**
3	pyruvate oxidase	15900627	65.3 / 5.1	1087	39	1.1	3.5	0.0004
4	**GTP-binding protein LepA**	**15901063**	**67.8 / 5.1**	**364**	**20**	**0.4**	**2.9**	**0.0341**
4	glyceraldehyde-3-phosphate dehydrogenase	12619268	33.0 / 5.0	80	14	0.2	2.9	0.0341
4	recP peptide	153792	72.0 / 6.0	303	15	0.2	2.9	0.0341
4	threonyl-tRNA synthetase	148984063	74.8 / 5.2	355	22	0.2	2.9	0.0341
4	translation initiation factor IF-2	15900468	105.7 / 8.8	66	6	0.1	2.9	0.0341
5	**bifunctional acetaldehyde-CoA/alcohol dehydrogenase**	**15903879**	**98.7 / 6.1**	**1192**	**33**	**1**	**4.4**	**0.0184**
6	**bifunctional acetaldehyde-CoA/alcohol dehydrogenase**	**15903879**	**98.7 / 6.1**	**1955**	**39**	**2.2**	**10.9**	**0.0141**
7	**glyceraldehyde-3-phosphate dehydrogenase**	**15901835**	**35.9 / 5.3**	**392**	**38**	**1.2**	**-2.5**	**0.0435**
7	elongation factor Tu	15901337	43.9 / 4.9	436	25	1.1	-2.5	0.0435
7	deoxyribose-phosphate aldolase	15900730	23.2 / 5.1	318	52	1	-2.5	0.0435
7	30S ribosomal protein S2	15902019	28.8 / 5.1	106	20	0.2	-2.5	0.0435
8	**lactate dehydrogenase**	**4138534**	**35.1 / 5.0**	**348**	**26**	**0.7**	**-3.0**	**0.0068**
8	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	15901131	26.3 / 5.2	203	15	0.6	-3.0	0.0068
8	6-phosphofructokinase	15902840	35.2 / 5.3	250	26	0.6	-3.0	0.0068
8	uridylate kinase	15900823	26.7 / 5.4	142	19	0.4	-3.0	0.0068
8	phosphoglyceromutase	15901490	26.0 / 5.1	55	15	0.3	-3.0	0.0068
8	pyruvate oxidase	15900627	65.3 / 5.1	67	9	0.1	-3.0	0.0068

Continues on following page

Table 3.1 continued

Spot number[a]	Protein	GI accession no.	Predicted molecular mass (kDa) / pI	Combined IonScore[b]	Sequence coverage (%)	emPAI value[c]	Fold-change (O/T) [d]	*p*-value[e]
9	**ribosome recycling factor**	**15900824**	**20.6 / 6.0**	**257**	**38**	**1.1**	**-2.3**	**0.0001**
9	glyceraldehyde-3-phosphate dehydrogenase	15901835	35.9 / 5.3	301	29	0.8	-2.3	0.0001
9	tRNA (guanine-N(7)-)-methyltransferase	15900463	24.4 / 5.9	192	26	0.5	-2.3	0.0001
9	uracil phosphoribosyltransferase	225856421	23.7 / 5.6	56	11	0.1	-2.3	0.0001
10	**elongation factor Tu**	**15901337**	**43.9 / 4.9**	**139**	**13**	**0.2**	**-5.4**	**0.0221**
10	glyceraldehyde-3-phosphate dehydrogenase	12619268	33.0 / 5.0	106	8	0.2	-5.4	0.0221
11	**glyceraldehyde-3-phosphate dehydrogenase**	**12619268**	**33.0 / 5.0**	**116**	**8**	**0.3**	**-15.8**	**0.0061**
12	**glyceraldehyde-3-phosphate dehydrogenase**	**15901835**	**35.9 / 5.3**	**406**	**20**	**1.2**	**-2.1**	**0.0412**
12	single-stranded DNA-binding protein	15903766	14.9 / 5.9	241	49	0.8	-2.1	0.0412
12	arginine repressor	15902837	18.1 / 5.8	136	23	0.4	-2.1	0.0412
12	hypothetical protein SP_1465	15901315	17.0 / 6.1	112	20	0.4	-2.1	0.0412
13	**single-stranded DNA-binding protein**	**15903766**	**14.9 / 5.9**	**370**	**49**	**3.2**	**3.4**	**0.0086**
13	50S ribosomal protein L22	15900150	12.2 / 10.8	68	25	0.6	3.4	0.0086
13	glyceraldehyde-3-phosphate dehydrogenase	12619268	33.0 / 5.0	187	13	0.6	3.4	0.0086
14	**sugar ABC transporter, ATP-binding protein**	**15902792**	**55.0 / 5.5**	**1708**	**60**	**3.3**	**-2.5**	**0.0032**
14	bifunctional N-acetylglucosamine-1-phosphate uridyltransferase	15900863	49.4 / 5.4	171	10	0.2	-2.5	0.0032
14	glucose-6-phosphate 1-dehydrogenase	15901104	56.8 / 5.4	126	6	0.1	-2.5	0.0032
15	**response regulator (ComE)**	**7920143**	**29.9 / 6.0**	**1022**	**61**	**5.7**	**2.5**	**0.0207**
15	phosphate transporter ATP-binding protein	15901251	30.4 / 6.7	233	29	0.4	2.5	0.0207

[a]As depicted in Figure 3.3 (D39) and identified using LC-ESI-IT MS. The dominant ID, based on combined IonScore, sequence coverage and emPAI value is indicated in bold

[b]The sum of ion scores (scores based on the probability that ion fragmentation matches are a non-random event)

[c]Exponentially Modified Protein Abundance Index, the relative quantitation of the proteins in a mixture based on protein coverage by the peptide matches in the database search results (Ishihama *et al.*, 2005)

[d]A positive value indicates upregulation in the O variant

Table 3.2. Table showing the identities of protein spots from Figure 3.3 (WCH16).

Spot number[a]	Protein	GI accession no.	Predicted molecular mass (kDa) / pI	Combined IonScore[b]	Sequence coverage (%)	emPAI value[c]	Fold-change (O/T) [d]	*p*-value[e]
1	**glyceraldehyde-3-phosphate dehydrogenase, type I**	**149003848**	**35.9 / 5.3**	**821**	**48**	**2.2**	**2.0**	**0.0119**
1	DNA polymerase III subunit beta	15899951	42.0 / 5.0	177	15	0.4	2.0	0.0119
1	30S ribosomal protein S1	15900746	43.9 / 5.0	188	16	0.3	2.0	0.0119
1	asparagine synthetase AsnA	15901793	37.7 / 5.0	178	29	0.3	2.0	0.0119
1	lactate dehydrogenase	4138534	35.1 / 5.0	136	11	0.3	2.0	0.0119
1	elongation factor Tu	15901337	43.9 / 4.9	173	19	0.2	2.0	0.0119
1	phosphoglycerate kinase	15900413	41.9 / 4.9	141	14	0.2	2.0	0.0119
2	**ABC transporter substrate-binding protein - branched chain amino acid transport**	**237650408**	**40.5 / 5.1**	**407**	**30**	**1**	**-2.7**	**0.0194**
3	**GTP-binding protein Era**	**15900846**	**34.0 / 6.4**	**75**	**13**	**0.4**	**-2.9**	**0.0486**
5	**guanosine 5'-monophosphate oxidoreductase**	**169833329**	**36.1 / 6.3**	**1113**	**70**	**6.6**	**3.7**	**0.0010**
5	30S ribosomal protein S2	15902019	28.8 / 5.1	394	42	1.2	3.7	0.0010
6	**chaperone protein DnaJ**	**15900433**	**40.9 / 6.8**	**440**	**32**	**0.7**	**2.1**	**0.0209**
8	**maltose operon transcriptional repressor**	**15901927**	**37.2 / 8.7**	**721**	**47**	**3**	**2.2**	**0.0402**
8	rhamnosyl transferase Cps6aS	20331049	38.9 / 8.8	210	20	0.4	2.2	0.0402
8	putative glycosyl transferase, family 8	148988108	36.7 / 9.2	89	19	0.2	2.2	0.0402
9	**fructose-bisphosphate aldolase**	**15900513**	**31.5 / 5.0**	**272**	**44**	**1**	**-4.2**	**0.0250**
9	hypothetical protein SP_1465	15901315	17.0 / 6.1	78	14	0.4	-4.2	0.0250

[a]As depicted in Figure 3.3 (WCH16) and identified using LC-ESI-IT MS. The dominant ID, based on combined IonScore, sequence coverage and emPAI value is indicated in bold

[b]The sum of ion scores (scores based on the probability that ion fragmentation matches are a non-random event)

[c]Exponentially Modified Protein Abundance Index, the relative quantitation of the proteins in a mixture based on protein coverage by the peptide matches in the database search results (Ishihama *et al*., 2005)

[d]A positive value indicates upregulation in the O variant

3.2.2.3 Analysis of differential protein expression profiles of WCH43O vs. WCH43T

For WCH43, a total of 1285 protein spots were detected on the master gel and 879 of these were present in 75% of the spot maps. Statistical analysis identified 183 spots that exhibited a significant change in protein abundance (α=0.05). As with the other *S. pneumoniae* strains investigated in this study, the number of upregulated protein spots in the O variant (n=118) was greater than the number upregulated in the T variant (n=65). The q-value extended FDR for α=0.05 was 11%. The top 15 differentially regulated spots were picked (Figure 3.3) and identified using mass spectrometry (Table 3.3). Proteins that were identified as upregulated in WCH43O include anaerobic ribonucleoside triphosphate reductase (Nrd; 6.6-fold), formate acetyltransferase (6.0-fold) and Fba (5.4-fold). In WCH43T, the proteins identified as upregulated include GAPDH (18.2-fold), Fba (8.7-fold) and ribose-phosphate pyrophosphokinase (7.3-fold).

3.2.3 Contribution of SpxB to Colony Opacity in Strain D39

Spot 2 on the Coomassie-stained gel (Figure 3.3) of D39 was analysed using mass spectrometry and identified putatively as SpxB. The molecular size of this protein is predicted to be 65.3 kDa, with a pI of 5.1. A previous study by Overweg *et al.* (2000) of *S. pneumoniae* strains belonging to serotypes 6B and 9V showed an increase in SpxB in the T variants, which is in contrast to the findings of this study. Western blot analysis was carried out to compare the amount of SpxB produced by the O/T pairs (Figure 3.4). Although the intensity of the SpxB signal for D39O appeared slightly stronger than for D39T, consistent with the proteomics findings, quantitative analysis of the image in Figure 3.4 found no statistically significant difference in SpxB expression between the O/T pairs (data not shown). These discrepancies, coupled with the previous implication of SpxB in colonisation (Regev-Yochay *et al.*, 2007; Spellerberg *et al.*, 1996), justified further investigation of the association of this protein (if any) with O/T phase variation.

Table 3.3. Table showing the identities of protein spots from Figure 3.3 (WCH43).

Spot number[a]	Protein	GI accession no.	Predicted molecular mass (kDa) / pI	Combined IonScore[b]	Sequence coverage (%)	emPAI value[c]	Fold-change (O/T) [d]	*p*-value[e]
1	**formate acetyltransferase**	**15900375**	**88.2 / 5.1**	**2257**	**60**	**4.2**	**6.0**	**0.0001**
1	pyruvate kinase	15900780	54.8 / 5.0	389	26	0.7	6.0	0.0001
1	elongation factor Tu	15901337	43.9 / 4.9	96	12	0.2	6.0	0.0001
1	glyceraldehyde-3-phosphate dehydrogenase	15901835	35.9 / 5.3	169	22	0.2	6.0	0.0001
2	**formate acetyltransferase**	**15900375**	**88.2 / 5.1**	**3168**	**70**	**8.3**	**5.2**	**0.0029**
2	DNA mismatch repair protein MutS	15903929	96.6 / 5.2	173	11	0.1	5.2	0.0029
2	elongation factor G	15900207	76.8 / 4.9	109	6	0.1	5.2	0.0029
2	pneumolysin	15901747	52.9 / 5.2	85	10	0.1	5.2	0.0029
2	signal recognition particle-docking protein FtsY	15901105	47.8 / 4.4	64	8	0.1	5.2	0.0029
3	**anaerobic ribonucleoside triphosphate reductase**	**15900138**	**84.5 / 5.6**	**1360**	**38**	**2.1**	**6.6**	**0.0081**
4	**glutathione reductase**	**15902736**	**49.0 / 5.8**	**269**	**18**	**0.7**	**-5.1**	**0.0037**
4	elongation factor Tu	15901337	43.9 / 4.9	85	11	0.2	-5.1	0.0037
4	translation initiation factor IF-2	303255249	102.9 / 8.8	125	6	0.1	-5.1	0.0037
5	**glyceraldehyde-3-phosphate dehydrogenase**	**15901835**	**35.9 / 5.3**	**1482**	**70**	**20.9**	**7.5**	**0.0001**
5	capsular polysaccharide biosynthesis protein Cps4J	15900287	39.4 / 5.3	361	20	0.5	7.5	0.0001
5	asparagine synthetase AsnA	15901793	37.7 / 5.0	93	11	0.2	7.5	0.0001
5	pyruvate kinase	15900780	54.8 / 5.0	125	6	0.2	7.5	0.0001
5	pyruvate oxidase	1161270	65.2 / 5.1	132	6	0.2	7.5	0.0001
5	UDP-N-acetylglucosamine-2-epimerase	15900286	41.5 / 5.2	114	11	0.2	7.5	0.0001
6	**ribose-phosphate pyrophosphokinase**	**15899974**	**35.6 / 5.7**	**535**	**42**	**1.4**	**-7.3**	**0.0008**
6	aminotransferase AlaT	15901817	45.9 / 5.5	384	27	1	-7.3	0.0008
6	hypothetical protein SP_1565	15901408	36.1 / 5.5	490	38	1	-7.3	0.0008
6	hypothetical protein CGSSp3BS71_09756	148984912	59.2 / 6.3	132	13	0.2	-7.3	0.0008
6	PhoH family protein	15900827	36.3 / 5.7	99	12	0.2	-7.3	0.0008

Continues on following page

Table 3.3 continued

Spot number[a]	Protein	GI accession no.	Predicted molecular mass (kDa) / pI	Combined IonScore[b]	Sequence coverage (%)	emPAI value[c]	Fold-change (O/T) [d]	*p*-value[e]
7	**fructose-bisphosphate aldolase**	**15900513**	**31.5 / 5.0**	**609**	**47**	**4**	**5.4**	**0.0003**
7	dihydroorotate dehydrogenase electron transfer subunit	15900840	27.4 / 4.7	205	21	0.8	5.4	0.0003
7	glyceraldehyde-3-phosphate dehydrogenase	15901835	35.9 / 5.3	268	28	0.7	5.4	0.0003
7	elongation factor G	15900207	76.8 / 4.9	308	18	0.3	5.4	0.0003
7	lactate dehydrogenase	4138534	35.1 / 5.0	183	17	0.3	5.4	0.0003
7	ABC transporter, substrate-binding protein	15900086	30.6 / 5.1	75	12	0.2	5.4	0.0003
8	**fructose-bisphosphate aldolase**	**15900513**	**31.5 / 5.0**	**1411**	**70**	**10.1**	**5.4**	**0.0101**
8	glyceraldehyde-3-phosphate dehydrogenase	12619268	33.0 / 5.0	232	22	0.5	5.4	0.0101
8	threonyl-tRNA synthetase	149001904	25.3 / 5.0	96	14	0.3	5.4	0.0101
8	elongation factor G	15900207	76.8 / 4.9	120	4	0.1	5.4	0.0101
9	**glyceraldehyde-3-phosphate dehydrogenase**	**15901835**	**35.9 / 5.3**	**437**	**49**	**1.6**	**-8.5**	**0.0002**
9	50S ribosomal protein L13	15900228	16.1 / 10.0	114	16	0.5	-8.5	0.0002
9	50S ribosomal protein L15	15900165	15.4 / 10.3	89	19	0.5	-8.5	0.0002
9	50S ribosomal protein L9	15902011	16.5 / 9.4	135	47	0.4	-8.5	0.0002
9	6-phosphogluconate dehydrogenase	15900298	52.7 / 4.9	121	12	0.1	-8.5	0.0002
9	NADH oxidase	15901319	50.5 / 5.0	93	3	0.1	-8.5	0.0002
10	**glyceraldehyde-3-phosphate dehydrogenase**	**15901835**	**35.9 / 5.3**	**552**	**35**	**3.9**	**-9.0**	**0.0001**
10	hypothetical protein SP_0022	15899970	20.0 / 5.6	185	56	0.9	-9.0	0.0001
10	uracil phosphoribosyltransferase	225856421	23.7 / 5.6	61	10	0.3	-9.0	0.0001
10	ribonucleotide-diphosphate reductase subunit alpha	15901044	81.8 / 5.3	112	4	0.2	-9.0	0.0001
11	**glyceraldehyde-3-phosphate dehydrogenase**	**15901835**	**35.9 / 5.3**	**718**	**35**	**3.5**	**-6.9**	**0.0001**
12	**fructose-bisphosphate aldolase**	**15900513**	**31.5 / 5.0**	**195**	**30**	**0.4**	**-8.7**	**0.0001**
12	glyceraldehyde-3-phosphate dehydrogenase	12619268	33.0 / 5.0	115	8	0.2	-8.7	0.0001
13	**glyceraldehyde-3-phosphate dehydrogenase**	**12619268**	**33.0 / 5.0**	**134**	**18**	**0.3**	**-5.4**	**0.0006**
14	**glyceraldehyde-3-phosphate dehydrogenase**	**15901835**	**35.9 / 5.3**	**511**	**30**	**1.9**	**-18.2**	**0.0002**
14	30S ribosomal protein S5	15900163	17.0 / 9.5	73	37	0.4	-18.2	0.0002

[a]As depicted in Figure 3.3 (WCH43) and identified using LC-ESI-IT MS. The dominant ID, based on combined IonScore, sequence coverage and emPAI value is indicated in bold

[b]The sum of ion scores (scores based on the probability that ion fragmentation matches are a non-random event)

[c]Exponentially Modified Protein Abundance Index, the relative quantitation of the proteins in a mixture based on protein coverage by the peptide matches in the database search results (Ishihama *et al.*, 2005)

[d]A positive value indicates upregulation in the O variant

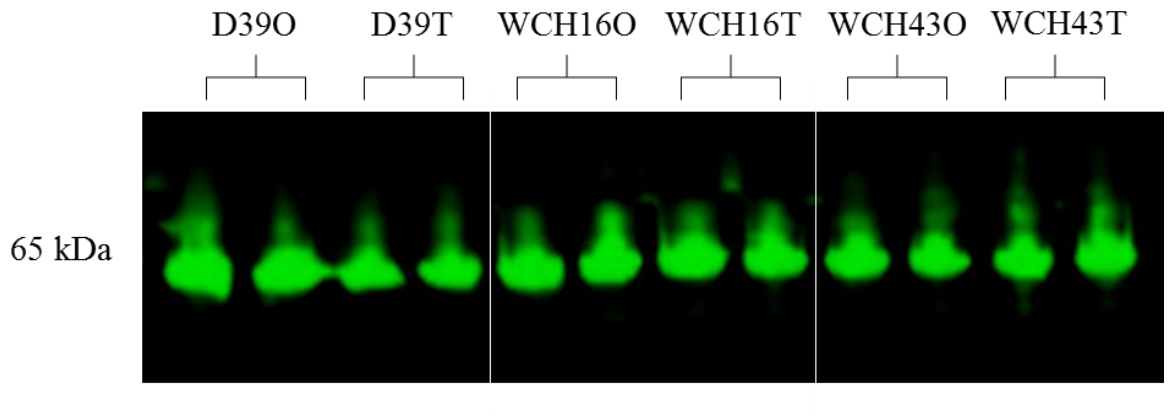


Figure 3.4. Expression of SpxB in O and T variants of D39, WCH16 and WCH43.

Lysates of the indicated strain were subjected to quantitative Western blot analysis using polyclonal anti-SpxB, as described in Section 2.7.2. This figure presents two out of four biological replicates for each strain.

3.2.3.1 Deletion, over-expression and back-transformation of D39 *spx*B

In order to assess whether *spxB* in D39 directly influenced colony phenotype, the *spxB* gene was deleted from D39O and D39T by targeted deletion replacement mutagenesis with an *erm* cassette (Section 2.8.5), as summarised in Figure 3.5. This involved the extension of the necessary fragments using primers listed in Table 2.2; regions flanking the *spxB* gene were amplified using the primers 'SpxB Flank F' with 'SpxB EryX' and 'SpxB Flank R' with 'SpxB EryY', whereas the *erm* cassette was amplified using the primers 'J214' and 'J215' (Figure 3.5). Primers 'SpxB EryX' and 'SpxB EryY' contained sequences complementary to the *erm* cassette so that when the three products are mixed, denatured and reannealed under PCR conditions, the three fragments would fuse to form one recombinant product containing the regions flanking the gene with the *erm* cassette in place of *spxB*. This was then used to transform D39O or D39T. Transformants that had incorporated the erythromycin resistance gene into their chromosome were selected on erythromycin-supplemented medium (Section 2.6.1). Erythromycin-resistant colonies were then checked for deletion of *spxB* and its replacement with *erm* by PCR and Sanger sequencing. Additionally, PCRs using primer sequences in the flanking regions of SpxB ('SpxB Flank R' and 'SpxB UpSeq F') and within the *erm* cassette ('J258' and 'J257') were carried out to check uptake of the *erm* cassette (Figure 3.6A) (expected sizes of PCR products were approximately 2 kb and 2.5 kb, respectively). Lack of SpxB expression was also confirmed by Western Blot analysis (Section 2.7.2) (Figure 3.6B).

Two independent D39 mutants lacking *spxB* were then plated for single colonies on THY-catalase plates to observe any impact on opacity phenotype. D39OΔ*spxB* colonies appeared larger, more mucoid and even more opaque than wild-type D39O (D39O WT) (Figure 3.7). Interestingly, the D39TΔ*spxB* mutant also showed the same hyper-opaque phenotype as D39OΔ*spxB* (Figure 3.7). This observation was unexpected given that proteomic analysis suggested SpxB levels were higher in D39O than D39T (Table 3.1). However, the resultant phenotype was in concordance with that observed by other groups (Allegrucci & Sauer, 2008; Carvalho *et al.*, 2013). Since the deletion of *spxB* in either O or T backgrounds was shown to produce markedly more opaque colonies, the *spxB* gene from D39O was cloned into plasmid pAL3 and transformed into D39O for ectopic expression.

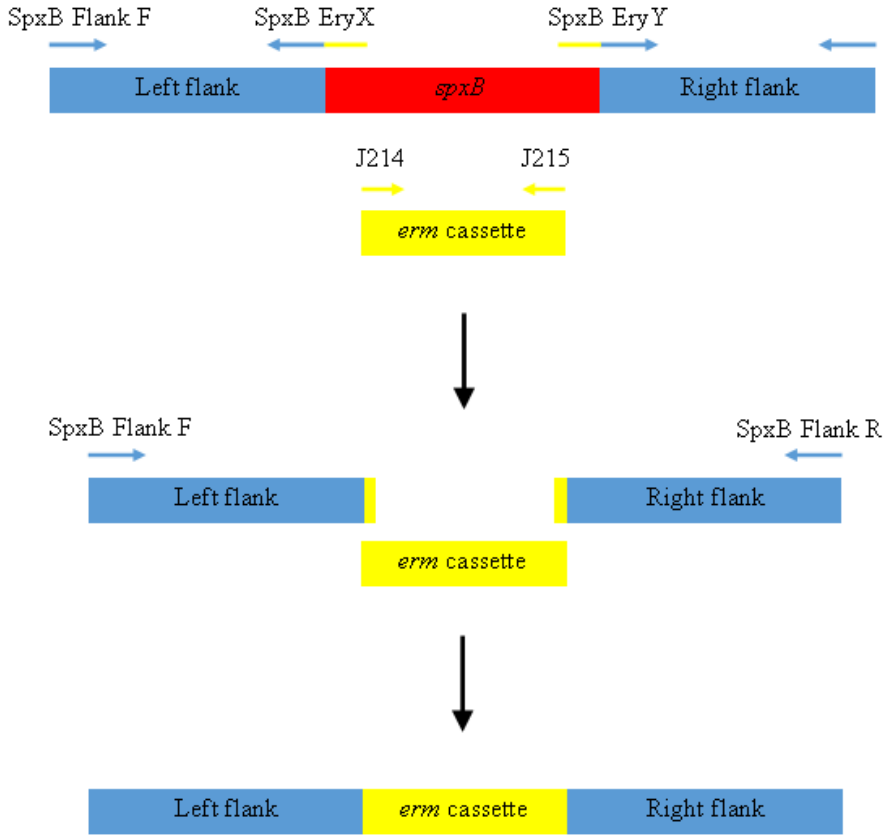


Figure 3.5. Schematic representation of *spxB* mutagenesis using OE-PCR.

Regions flanking either side of *spxB* were amplified using primers 'Flank F' with 'SpxB Ery X' and 'Flank R' with 'SpxB EryY', where the primers 'SpxB EryX' and 'SpxB EryY' contain sequences complementary to the erythromycin resistance cassette (*erm*). The primers 'J214' and 'J215' were used to amplify the *erm* cassette from a strain containing this gene. The three products (left flank, right flank and *erm* cassette) were mixed together and OE-PCR was used to fuse the regions, resulting in replacement of the *spxB* gene with the *erm* gene. This linear DNA fragment was then used to transform D39O or D39T. Transformants were selected on BA supplemented with the appropriate concentration of erythromycin. Figure is not drawn to scale.

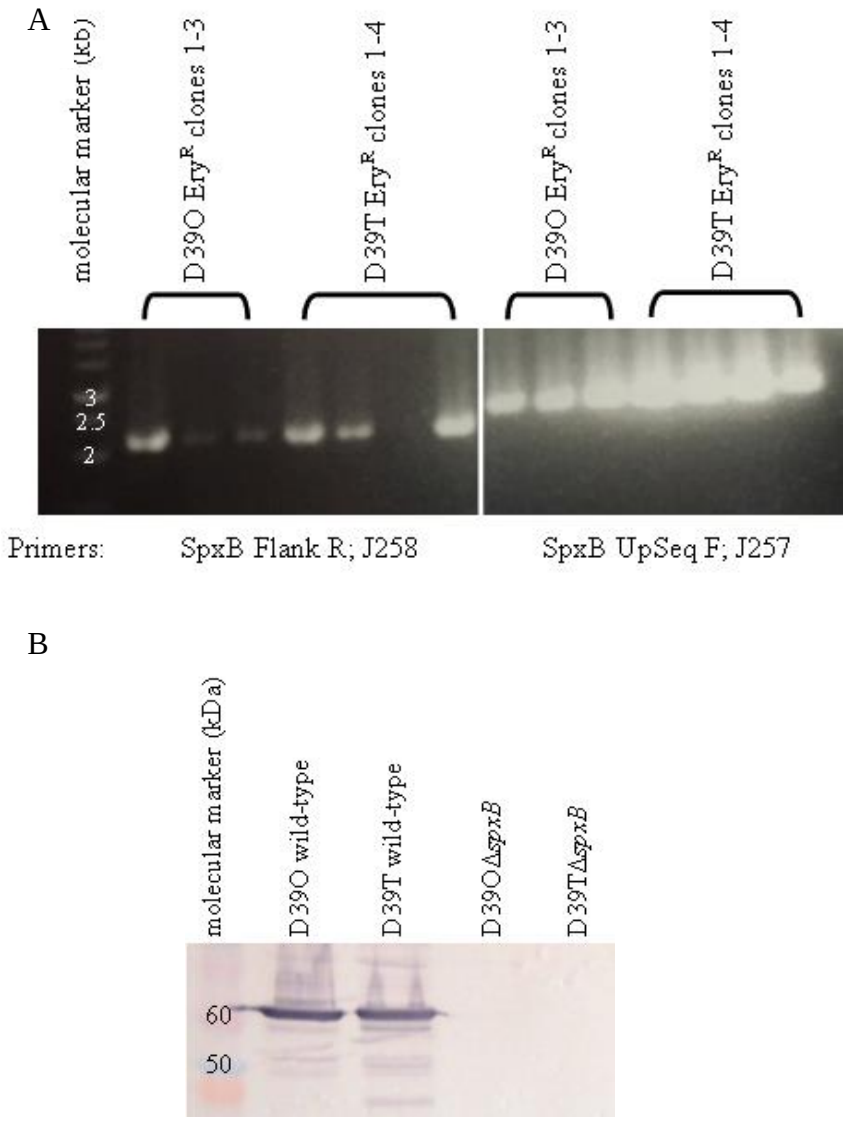


Figure 3.6. Confirmation of D39Δ*spxB* mutants using PCR and Western Blot analysis.

(A) Three and four putative *spxB*-negative clones of D39O and D39T, respectively, were selected from BA plates supplemented with Ery (Section 2.2.4). PCRs were carried out using primers flanking *spxB* with another within the *erm* cassette ('SpxB Flank R' with 'J258' and 'SpxB UpSeq F' with 'J257'). (B) One of each D39O and D39T clone that were positive in (A) were checked for the loss of SpxB expression using Western blot analysis (Section 2.7.2).

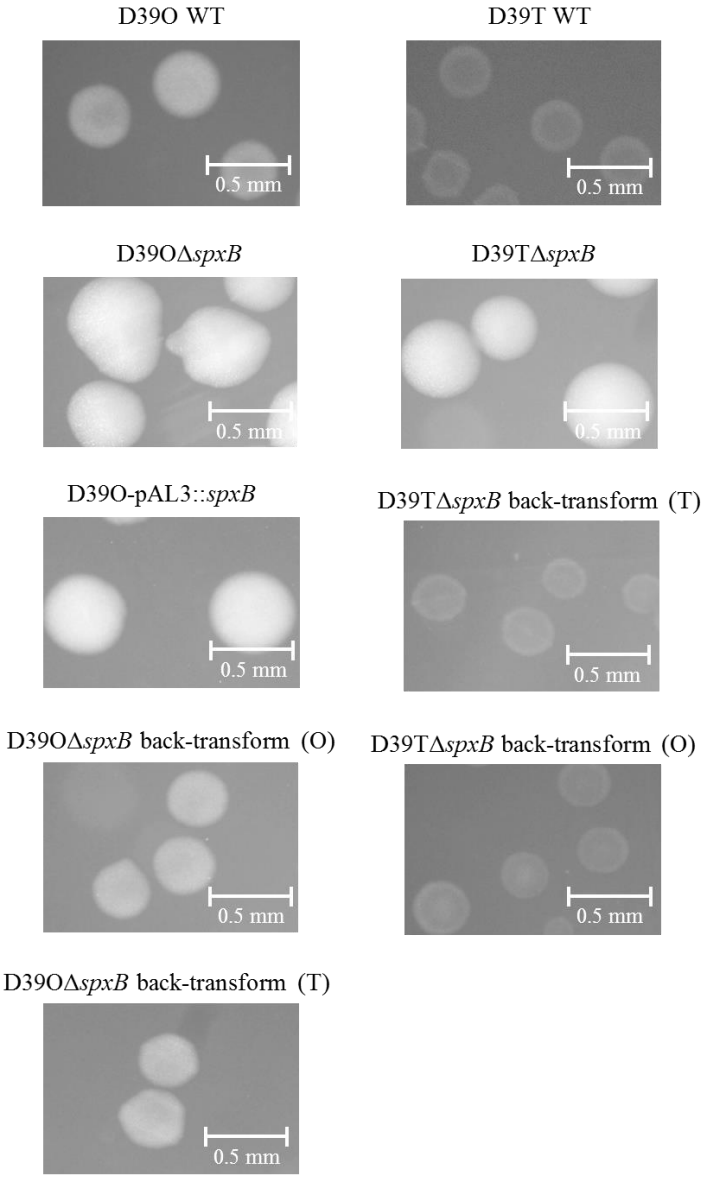


Figure 3.7. Colony morphology of various D39 *spxB* mutants.

Strains were streaked for single colonies on THY+catalase agar plates and images were captured using Nikon SMZ1000 on a dissecting microscope and viewed through cellSens software. D39O, strain D39 opaque; D39T, strain D39 transparent; D39OΔ*spxB*, D39O *spxB* mutant; D39TΔ*spxB*, D39T *spxB* mutant; D39O-pAL3::*spxB*, D39O SpxB over-expression SpxB from D39T using a plasmid with constitutively expressing promoter; D39TΔ*spxB* back-transform (T), D39TΔ*spxB* with *spxB* from strain D39T re-inserted back; D39OΔ*spxB* back-transform (O), D39OΔ*spxB* with *spxB* from strain D39O re-inserted back; D39TΔ*spxB* back-transform (O), D39TΔ*spxB* with *spxB* from strain D39O re-inserted back; D39OΔ*spxB* back-transform (T), D39OΔ*spxB* with *spxB* from strain D39T re-inserted back. The schematic representation of these contructs are on Figure 3.9.

The insert was sequenced to ensure that *spxB* had been inserted correctly (Figure 3.8) and since the vector contains an erythromycin resistance gene, transformants were selected on an erythromycin-supplemented agar plate. Surprisingly, the resultant mutant, D39O-pAL3::*spxB*, had a similar phenotype to the *spxB* mutants (Figure 3.7).

To determine the phenotypic impact of *spxB* from either D39O or D39T on different genome backgrounds, the *spxB* loci from both D39O and T were back-transformed into the mutants D39OΔ*spxB* and D39TΔ*spxB* (Section 2.6.2). This would generate clones with reconstituted *spxB* loci from either D39O or T on both D39O and D39T genetic backgrounds (Figure 3.9). Using the primers 'SpxB Flank F' and 'SpxB Flank R', the *spxB* region was amplified from both D39O and D39T. Each fragment was then directly transformed into D39OΔ*spxB* or D39TΔ*spxB* (Figure 3.9). Patch plating on BA with or without eryhtromycin was used to identify transformants that had lost erythromycin resistance (Section 2.6.2). Two clones of each putative back-transformant were subjected to PCR analysis using the primers 'SpxB Flank F' with 'SpxB Flank R' and the resulting fragment sizes were compared to that of the WT (Figure 3.10). All tested transformants produced PCR products of a similar size compared to the WTs, indicating successful uptake of *spxB* and loss of the *erm* cassette, which is about 700 bp (vs. 1.8 kb of *spxB*). Additionally, the presence of *spxB* in the back-transformants was confirmed by Sanger sequencing (Section 2.8.8). The opacity phenotypes were then observed on a THY-catalase plate (Figure 3.7). Strain D39OΔ*spxB* transformed with *spxB* obtained from either D39O or D39T variants had a similar phenotype to D39O WT. On the other hand, D39TΔ*spxB* transformed with *spxB* obtained from either D39O or D39T appeared to have reverted back to the phenotype observed for D39T WT (Figure 3.7).

3.2.4 Glyceraldehyde-3-phosphate Dehydrogenase (GAPDH)

Multiple protein spots identified as differentially regulated between O/T pairs were identified as the enzyme GAPDH (Table 3.1-3). These protein spots had distinct pI and molecular size, suggesting that they were different isoforms. A total of three and six isoforms of GAPDH were identified differentially regulated in D39 and WCH43, respectively. All but one of the protein spots were identified as upregulated in the T variant

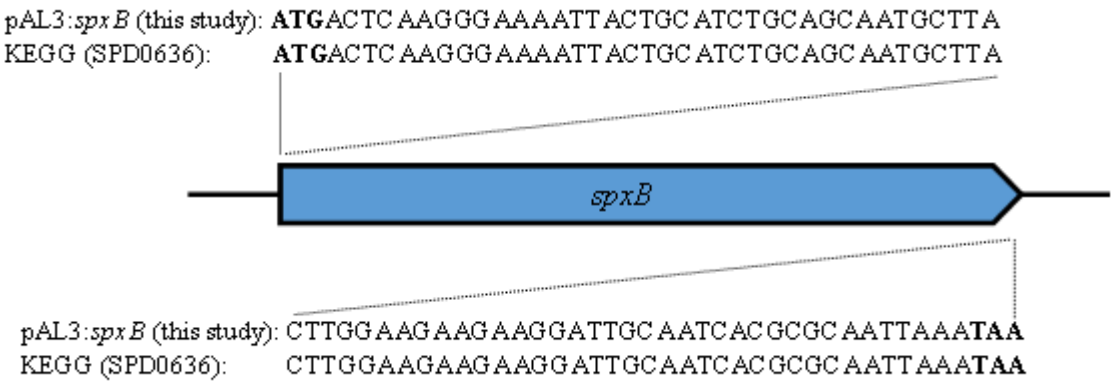


Figure 3.8. Sequencing of pAL3:*spxB*

Sanger sequencing of the vector used for over-expression of SpxB in D39O. The region containing *spxB* was sequenced and compared with the *spxB* gene sequence available on KEGG (SPD0636) to confirm the incorporation of the gene into the plasmid.

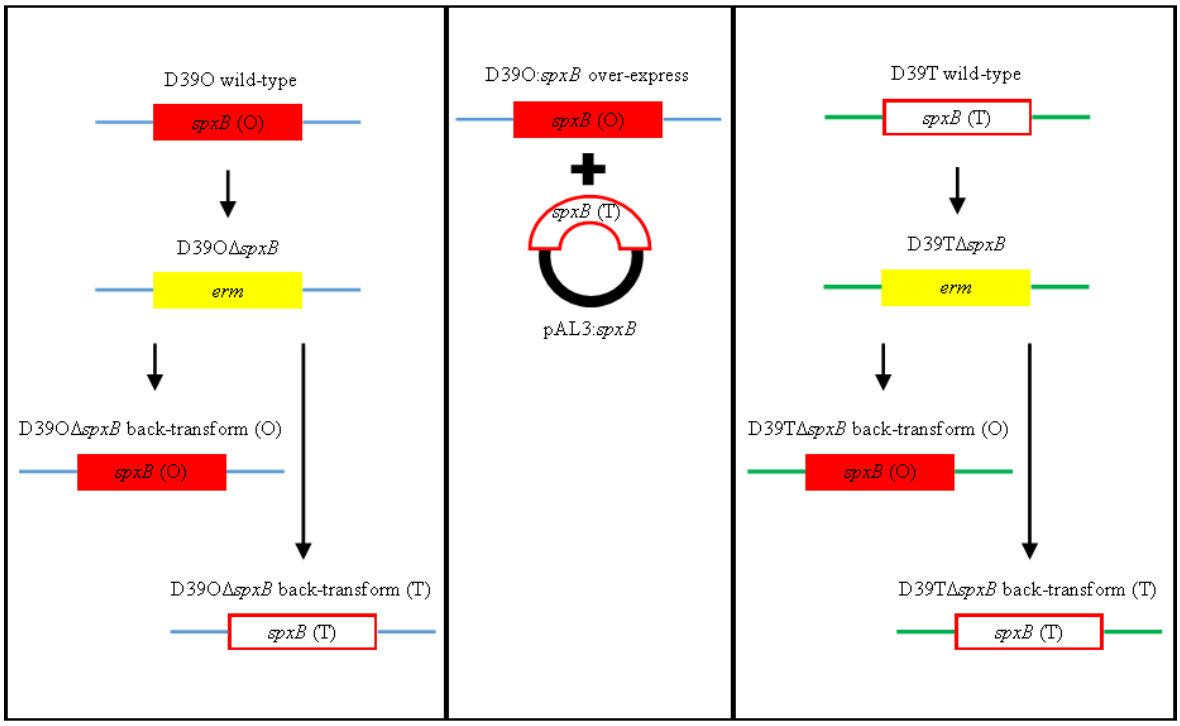


Figure 3.9. Schematic representation of *spxB* mutants used in this study.

D39O and D39T genetic backgrounds are represented by blue and green colours, respectively. D39O *spxB* is represented by a solid rectangular box and the wording *spxB* (O). D39T *spxB* is represented by a clear rectangular box and the wording *spxB* (T). SpxB mutants were created using overlap extension replacement PCR (Section 2.6) and the back-transformants were generated by replacing the *erm* cassette with the *spxB* gene from either D39O or D39T (Section 2.6.2). Selection of back-transformants were identified by the loss of Ery^R by patching. For the over-expression of SpxB in the pneumococcus, *spxB* was cloned into the plasmid vector pAL3 and transformed into competent D39O (Section 2.6.3-4). The plasmid contained an erythromycin resistant gene and successful uptake of the plasmid was indicated by growth on erythromycin supplemented BA. Figure is not drawn to scale.

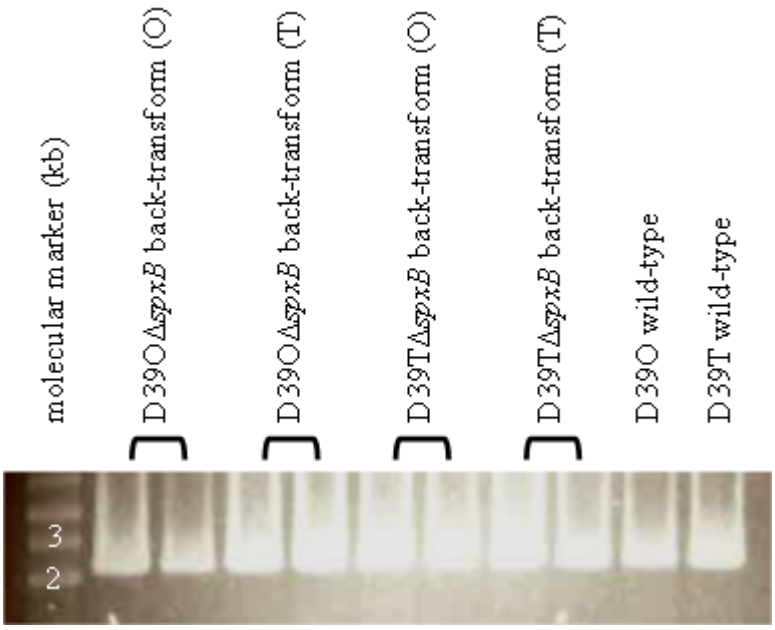


Figure 3.10. PCR to confirm presence of *spxB*.

Back-transformants that had lost resistance to erythryomycin were selected for sequencing. The regions flanking *spxB* were amplified and the size of the resulting PCR products were compared to that for the wild-types. Successful replacement of *erm* with *spxB* is indicated by the presence of a band of a similar size to that of the wild-type strains (~2.2 kb), A 1.1-kb band would indicate failed replacement of *erm* with *spxB*.

of D39 and WCH43. However, the presence of isoforms has made it difficult to assess and draw any conclusions as to which variant GAPDH is upregulated in.

3.2.4.1 GAPDH activity in O and T variants

In order to investigate the above differences, GAPDH activity was compared between the O and T variants of strains D39, WCH16 and WCH43. The GAPDH assay involved measuring the increase in A_{340} as a result of the reduction of NAD in the respective whole cell lysates in the presence of G-3-P substrate and the GAPDH activity in the T variant was calculated relative to the O variant of that strain. The GAPDH activity in lysates of strains D39 and WCH43 were about 1.5-fold higher in the T variant compared to their O counterpart (p <0.001 and <0.05, respectively), while there was no significant difference in GAPDH activity between the variants of strain WCH16 (Table 3.4). This was largely consistent with proteomic analysis (Tables 3.1-3), except that in WCH16, there was a two-fold upregulation in the O variant.

3.2.5 Quantitative Western Blot Analysis of Various Proteins

It was previously reported that O and T *S. pneumoniae* variants express varying levels of certain pneumococcal proteins involved in virulence and metabolism (Overweg *et al.*, 2000; Weiser *et al.*, 1994). The cell lysates of strains D39, WCH16 and WCH43 were therefore tested for differences in expression levels between the variants of various known pneumococcal virulence factors and cell surface proteins by quantitative Western blotting using polyclonal murine antisera (Section 2.7.2). Of the 13 pneumococcal proteins tested, four [NanA, Ply, pyruvate kinase (Pyk) and PhtD] showed significant differences in the expression level between O/T pairs in at least one of the three strains (Figure 3.11). Interestingly, none of the proteins tested was shown to be consistently upregulated in a particular variant across all three strains. NanA was upregulated in the O variant of both WCH16 and WCH43, but upregulated in the T variant of D39. D39 O/T pairs did not express significantly different levels of either Ply, Pyk or PhtD. Expression levels of both Ply and Pyk were significantly upregulated in WCH16O, but PhtD expression was upregulated in WCH16T. On the other hand, although Pyk expression was significantly

Table 3.4. GAPDH activity in pneumococcal phase variants.

GAPDH activity was assayed in extracts from three biological replicates of each strain (D39O, WCH16O and WCH43O and their relative T counterparts), as described in Section 2.7.3. Results are expressed for the T variant for each strain relative to the respective O variant. Data were analysed using Student's unpaired *t*-test.

Strain	**GAPDH activity T/O ± SEM**	***p*-value**
D39	1.50 ± 0.12	$p < 0.001$
WCH16	0.87 ± 0.17	ns
WCH43	1.49 ± 0.17	$p < 0.05$

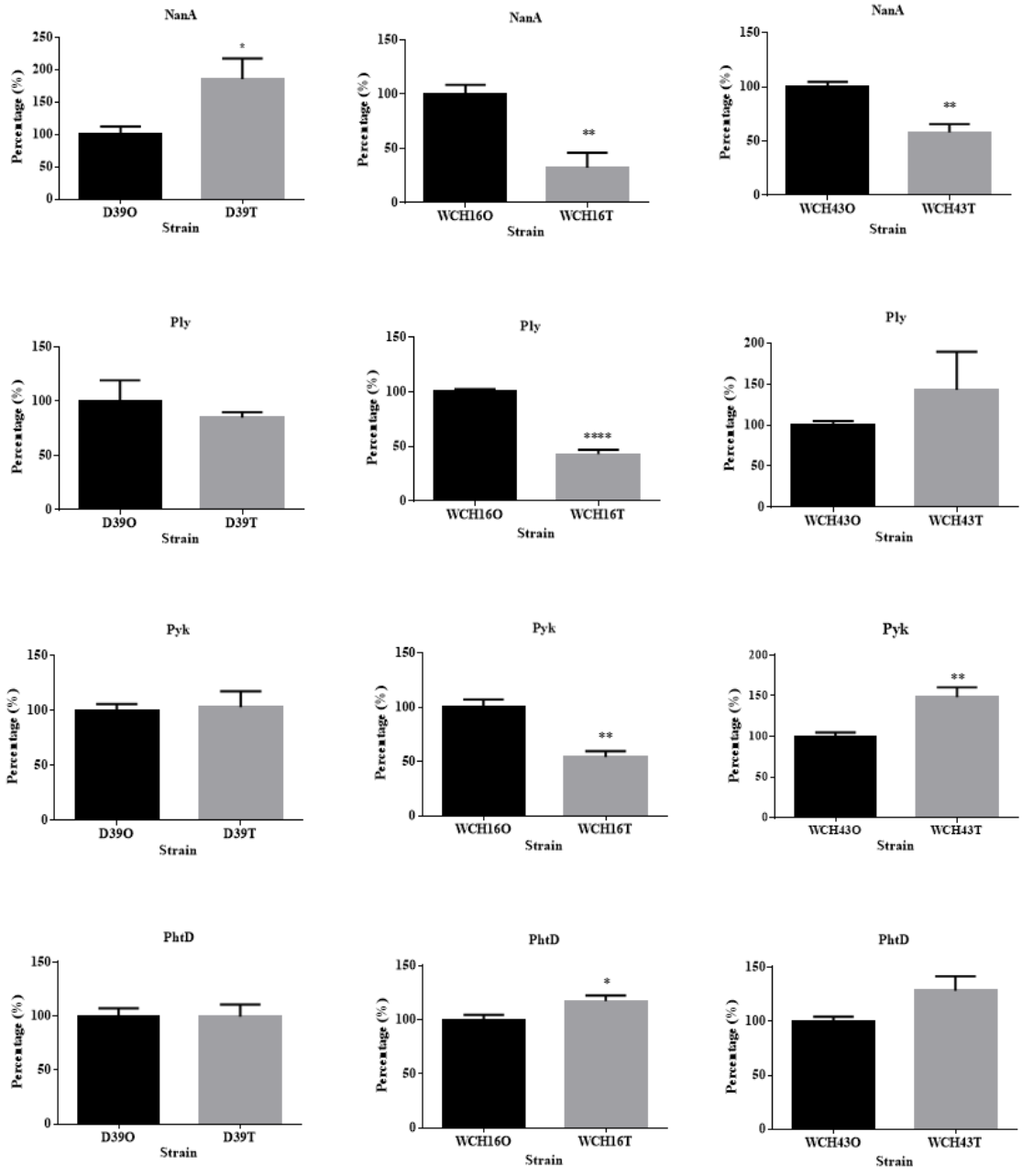


Figure 3.11. Quantitative Western blot analysis.

Data are expression of NanA, Ply, Pyk and PhtD from four biological replicates as expressed as a percentage of expression by the opaque variant of each strain ±SEM. Statistical significance was determined by two-tailed Student's *t*-test (*, p <0.05; **, p <0.01; ****, p <0.0001).

upregulated in WCH43T, there were no significant differences in levels of Ply and PhtD in the WCH43 background. The nine proteins that did not show differences in protein expression levels were LytA, AliA, PiuA, MalX, CbpA, GlpO, PsaA, PspA and SpxB.

3.3 Discussion

Pneumococcal colony opacity phase variation has been shown to correlate with differences in pathogenicity in various animal models (Weiser, 1998). This is also evident in other respiratory pathogens such as *H. influenzae* (Weiser, 1993) and *N. meningitidis* (Kupsch *et al.*, 1993; Makino *et al.*, 1991). The ability of these pathogens to disseminate from nasopharyngeal colonisation to cause invasive disease or thrive in their host may be dependent on their ability to reversibly adapt to changes in the microenvironment (Hammerschmidt *et al.*, 1996; Li-Korotky *et al.*, 2010; Weiser, 1993; Weiser *et al.*, 2001). For example, the phase variation of Opa was important in mediating the invasion of the *Neisseria* species into human epithelial cells, as non-invasive *N. meningitidis* and *N. gonorrhoeae* strains were shown to be devoid of this ability (de Vries *et al.*, 1996; Makino *et al.*, 1991). For *S. pneumoniae*, the phenomenon of colony opacity phase variation, first described nearly two decades ago (Weiser *et al.*, 1994), was suggested to play a role in pathogenicity and adaptation to different microenvironments (Cundell *et al.*, 1995b; Kim & Weiser, 1998; Weiser *et al.*, 1994). In a mouse challenge model, studies showed that when apparently identical *S. pneumoniae* strains are isolated from the nasopharynx and other tissues, those from the nasopharynx are largely in the T phase, whereas those from other tissues are in the O phase (Briles *et al.*, 2005; Kim & Weiser, 1998; Weiser *et al.*, 1994).

Although colony opacity phase variation has been clearly demonstrated in *S. pneumoniae*, there is discordance in the literature regarding the contribution of protein expression to this phenomenon, as discussed in the Introduction (Section 1.6.1). One of the main reasons for this could be due to the differences in pneumococcal strains and analytical techniques used. Therefore, in order to gain further insight on the types of proteins that might be central to the transition from nasopharyngeal colonisation to invasive disease, a detailed proteomic analysis of colony opacity in three *S. pneumoniae*

strains with distinct pathogenicity patterns were performed. Cross-comparisons of up to 15 differentially regulated protein spots of O/T pairs of the three strains showed that there was very little overlap between them. The proteins could be classified into five main groups – (i) those involved in glycolysis, (ii) pyruvate metabolism, (iii) cell growth/DNA and protein synthesis, (iv) maintenance of cellular health and (v) those involved sugar/amino acid transport (Table 3.5). Most of the proteins involved in pyruvate metabolism and transcription/translational proteins were found to be upregulated in the O variants, whereas the ABC transporters and those involved in glycolysis up to pyruvate were found in the T variants.

3.3.1 Differential expression levels of proteins involved in glycolysis and pyruvate metabolism

Glycolysis is the metabolic pathway that converts glucose into pyruvate, generating a small amount of energy (ATP) and reducing power (NADH). Pyruvate, the end product of glycolysis, can be converted back to carbohydrates (gluconeogenesis). Alternatively, pyruvate can be fermented to produce lactate, or take-up lactate from the environment to produce more pyruvate via Ldh. Furthermore, pyruvate can also be converted to ethanol and a variety of amino acids. Two of the proteins involved in glycolysis were identified as differentially expressed between O and T variants of the strains used in this study – Fba (upregulated in WCH16T and WCH43O vs. their respective phase counterparts) and GAPDH (predominantly upregulated in the T forms of D39 and WCH43 but the opposite for WCH16, when compared to their alternative phases). Both of the aforementioned proteins exist in several isoforms, resulting in multiple spot identification on the 2D-gels. On the other hand, four proteins involved in pyruvate metabolism were identified in this study as being differentially expressed between the variants – SpxB (D39O > D39T); formate acetyltransferase (WCH16O > WCH16T); AdhE (D39O > D39T); and Ldh (WCH16T > WCH16O). The differential expression level of SpxB between O and T will be discussed in detail later in this chapter (Section 3.3.6).

Fba catalyses the reversible reaction that splits fructose 1,6-bisphosphate into dihydroxyacetone phosphate and G-3-P. Pneumococcal Fba may also act as an adhesin, as

Table 3.5. Summary table of protein spot identifications (including isoforms) and fold-changes of each strain as identified from Figure 3.3.

Upregulated in the O variant		Upregulated in the T variant	
Protein	**Fold-change**	**Protein**	**Fold-change**
D39			
Pyruvate oxidase	2.9	Lactate dehydrogenase	3.0
Bifunctional acetaldehyde-CoA/alcohol dehydrogenase	4.4	Ribosome recycling factor	2.3
Bifunctional acetaldehyde-CoA/alcohol dehydrogenase	10.9	Elongation factor Tu	5.4
Elongation factor G	4.2	Glyceraldehyde-3-phosphate dehydrogenase	2.5
GTP-binding protein LepA	2.9	Glyceraldehyde-3-phosphate dehydrogenase	2.1
Single-stranded DNA-binding protein	3.4	Glyceraldehyde-3-phosphate dehydrogenase	15.8
Hypothetical protein Spne_T02001699	3.5	Sugar ABC transporter, ATP-binding protein	2.5
Response regulator (ComE)	2.5		
WCH16			
Guanosine 5'-monophosphate oxidoreductase	3.7	GTP-binding protein Era	2.9
Chaperone protein DnaJ	2.1	Fructose-bisphosphate aldolase	4.2
Glyceraldehyde-3-phosphate dehydrogenase	2.0	ABC transporter substrate binding protein-branched chain amino acid transporter	2.7
Maltose operon transcriptional repressor	2.2		
WCH43			
Formate acetyltransferase	6.0	Ribose-phosphate pyrophosphokinase	7.3
Formate acetyltransferase	5.2	Glyceraldehyde-3-phosphate dehydrogenase	8.5
Anaerobic ribonucleoside triphosphate reductase	6.6	Glyceraldehyde-3-phosphate dehydrogenase	9.0
Glyceraldehyde-3-phosphate dehydrogenase	7.5	Glyceraldehyde-3-phosphate dehydrogenase	6.9
Fructose-bisphosphate aldolase	5.4	Glyceraldehyde-3-phosphate dehydrogenase	18.2
Fructose-bisphosphate aldolase	5.4	Glyceraldehyde-3-phosphate dehydrogenase	5.4
		Fructose-bisphosphate aldolase	8.7
		Glutathione reductase	5.1

Continues on following page

Note:

Proteins involved in different cellular process are colour-coded as below:

Pyruvate metabolism
Transcriptional/translational proteins
Glycolysis
Maintaining cellular health
Sugar/amino acid transport
Competence

recombinant-Fba and anti-rFba antibodies inhibit the adherence of *S. pneumoniae* to A549 type II lung epithelial cells (Blau *et al.*, 2007). However, the identification of multiple isoforms of Fba in both WCH43 variants, made it difficult to interpret and draw any conclusions about Fba regulation in O/T variants. The other protein differentially expressed was GAPDH, an NAD-dependent protein with a role in glycolysis and gluconeogenesis. Although possessing many isoforms, most of the protein spots identified as GAPDH were significantly upregulated in the T variants of D39 and WCH43 (between approximately 2- to 18-fold), but were upregulated by approximately two-fold in WCH16O vs. WCH16T. Differences in protein abundance were also reflected in GAPDH enzymatic activity, which showed that there was a significant difference in GAPDH activity between O and T variants for strains D39 and WCH43. On the other hand, WCH16 did not show a significant difference in GAPDH activity between the phase variants. This would suggest that in the T variants, the increase in GAPDH activity results in carbon metabolic flux being directed to the production of pyruvate and generation of ATP during glycolysis. An interesting feature of several bacterial housekeeping metabolic enzymes is that they may have some additional function in pathogenesis (Pancholi & Chhatwal, 2003). Apart from being involved in glycolysis/gluconeogenesis, group B streptococcal GAPDH is also released from the cytoplasm and displayed on the cell surface, where it has a role in adherence to host cells (Pancholi & Fischetti, 1992), as well as inducing apoptosis in murine macrophages (Oliveira *et al.*, 2011). C1q, a major component of the classical complement pathway produced by macrophages and immature dendritic cells, has a role in response to microbial infection and opsonisation, facilitating the phagocytosis of pathogens (Euteneuer *et al.*, 1986; Rojas-Espinosa *et al.*, 1972). GAPDH is a C1q ligand and pneumococcal GAPDH-C1q interaction leads to complement activation (Terrasse *et al.*, 2012). Terrasse *et al*. (2012) demonstrated that strains deficient in surface-exposed GAPDH had decreased levels of C1q recognition compared to WT strains. In group A streptococcus (GAS), GAPDH is known as SDH (streptococcal surface dehydrogenase), a fibronectin-binding protein, which facilitates colonisation (Pancholi & Fischetti, 1992) and is important for GAS virulence (Jin *et al.*, 2011). The direct interaction of SDH with human pharyngeal cells has been shown to regulate intracellular signaling events and cause DNA condensation, the penultimate stage of apoptosis (Pancholi & Fischetti, 1997).

Formate acetyltransferase (also known as pyruvate formate-lyase, Pfl) is a key enzyme in fermentation and catalyses the production of formate from pyruvate. Mutations of the *pfl* gene in the pneumococcus obliterated formate production and resulted in altered fatty acid composition, reduced virulence, lower CFU in the nasopharynx and lungs of mice (after i.n. challenge) and lower CFU in blood after i.v. challenge), compared to the WT (Yesilkaya *et al.*, 2009). Furthermore *pfl*-deficient mutants exhibited increased production of GAPDH and LDH (Zhu & Shimizu, 2004). These findings are in agreement with this study, where Pfl was upregulated in the O form of strain WCH43, whereas GAPDH and Ldh were expressed at higher levels in the T variants of at least one of WCH43 and D39. Ldh catalyses the reversible conversion of pyruvate to lactate, converting NADH to NAD^+. In *N. gonorrhoeae*, L-lactate stimulates the consumption of oxygen to allow the organism to deplete this substrate and ultimately limit the formation of reactive oxygen intermediates by phagocytic cells (Atack *et al.*, 2014; Fu *et al.*, 1989). *S. pneumoniae* with an *adhE* mutation could not grow aerobically in minimal media, were sensitive to oxidative stress and showed division defects (Taniai *et al.*, 2008). The O form of the pneumococcus is usually associated with invasive disease where oxidative stress from host immune cells is more prevalent and hence, an upregulation of *adhE* may promote their survival and proliferation in challenging environments such as the blood over T variants.

3.3.2 O variants had increased expression in proteins involved in cell proliferation under stressful conditions

In this study, a total of nine proteins that were differentially expressed in O/T pairs were involved in either cell growth or DNA/protein synthesis. Six of these (EF-G, GTP-binding protein LepA, SsbB, GMP reductase, chaperone protein DnaJ and Nrd) were found to be upregulated in the O form of one or more strains, while in the T variant, ribosome recycling factor, GTP-binding protein Era and ribose-phosphate pyrophosphokinase were upregulated. However, none of the proteins were similarly expressed across all three strains tested.

The elongation phase of protein synthesis involves the delivery of aminoacyl tRNAs to the ribosome by the EF-Tu/EF-Ts complex (Douglass & Blumenthal, 1979; Jekowsky *et*

al., 1977) followed by translocation of the bacterial ribosome along the mRNA by EF-G (Rodnina *et al.*, 1997). These elongation factors are also able to act as folding templates for denatured polypeptides in *E. coli* under stressful conditions such as acid and thermal stress (Caldas *et al.*, 2000; Len *et al.*, 2004). Recently, LepA, a highly conserved GTP-binding protein that is present in all bacteria has been demonstrated to be another elongation factor required for accurate and efficient protein synthesis (Liu *et al.*, 2011; Qin *et al.*, 2006). LepA, now termed elongation factor 4 (EF4), recognises ribosomes after a defective translocation reaction and induces back-translocation (Liu *et al.*, 2011; Qin *et al.*, 2006). This reaction allows EF-G to translocate the tRNA correctly. The ribosome recycling factor, also known as ribosome releasing factor, along with EF-G is responsible for the dissociation of ribosomes from the mRNA after termination of translation and is essential for bacterial growth in *E. coli* (Ito *et al.*, 2002; Janosi *et al.*, 1994). Most of these proteins associated with elongation during protein synthesis were upregulated in D39O (vs. D39T) and consistent with this, Overweg *et al.* (2000) also demonstrated that the other elongation factor, EF-Ts, was upregulated in the O variant of the pneumococcus. It is unclear why the ribosome-recycling factor was upregulated in D39T. However, the fold-change was low (2.3-fold) and thus may not have a huge impact. Interestingly, there were more proteins detected as upregulated in the O form compared to the T form across all three strains (Section 3.3.2).

Another protein detected as upregulated in the O variant was NrdD. In *E. coli, nrdD* is essential for anaerobic growth (Garriga *et al.*, 1996). Additionally, NrdD also regulates the total rate of DNA synthesis so that the ratio of DNA to cell mass is maintained during cell division (Herrick & Sclavi, 2007). The O variant of the pneumococcus is typically isolated from tissues such as the blood (Weiser *et al.*, 1994), where oxygen is not readily available. Hence, upregulation of a protein that enables it to grow anaerobically would confer an advantage during dissemination of this pathogen through areas with low oxygen tension. Furthermore, during stressful conditions, bacterial cells require chaperones to ensure the correct folding, unfolding, translation, translocation and degradation of newly synthesised proteins to ensure cell viability (Mayer & Bukau, 2005). Chaperone protein DnaJ is expressed at a higher level in D39O (vs. D39T) in this study. DnaJ is important for the

survival of *E. coli* in temperatures over 30°C (Sell *et al.*, 1990). In *Salmonella enterica* serovar Typhimurium, DnaJ coupled with DnaK is important for the invasion of epithelial cells and survival within macrophages, facilitating systemic disease in a mouse model (Takaya *et al.*, 2004). Together, this implicates DnaJ as a heat shock protein but also, more importantly, a virulence factor and thus correlates with the increase of this protein in the predominantly blood associated O variant of D39 found in this study. Furthermore, an infection is usually associated with an elevated body temperature and a *S. enterica* serovar Typhimurium DnaK/J mutant has been found to be non-viable at temperatures over 39°C (Takaya *et al.*, 2004).

Although most of the proteins associated with DNA replication and repair were upregulated in the O variant, the protein associated with cell division (GTP-binding protein Era) was upregulated in the T variant. The expression of GTP-binding protein Era of *E. coli* is temperature dependent and expressed at 40°C but repressed at 27°C (Gollop & March, 1991) and the loss of this gene at over 40°C resulted in depressed synthesis of heat-shock proteins, resulting in inhibition of cell growth and viability (Britton *et al.*, 1998; Gollop & March, 1991; Lerner & Inouye, 1991). Furthermore, alterations in the *era* gene also caused altered carbon metabolism (Lerner & Inouye, 1991; Pillutla *et al.*, 1996).

In *E. coli*, the single-stranded DNA-binding protein (SSB) is involved in maintaining the health of the cell through DNA metabolism, including the replication, repair and recombination of DNA [for review, see Meyer & Laine (1990)]. Its roles in DNA replication include the enhancement of helix destabilisation by DNA helicase, prevention of reannealing of single strands and protection from nuclease digestion. In the pneumococcus, a SSB paralog, coined SsbB is expressed during late genetic competence (Peterson *et al.*, 2004), creating a reservoir of internalised ssDNA allowing for successive recombination cycles and thus more efficient genetic transformation (Attaiech *et al.*, 2011). However, SsbB was upregulated in the O variant, a phase commonly associated with sterile sites such as the blood, and thus would have limited or no extracellular DNA. Perhaps it confers an advantage over its T counterpart by repairing DNA due to external stressors. Other differentially regulated proteins were involved in nucleotide metabolism, but one was upregulated in the O form (GMP reductase) while the other was upregulated in

the T form (ribose-phosphate pyrophosphokinase). GMP reductase catalyses the reversible and NADPH-dependent reductive deamination of GMP into inosine phosphate (IMP), maintaining intracellular balance of A and G nucleotides (Andrews & Guest, 1988; Martinelli *et al.*, 2011). Another enzyme involved in purine and pyrimidine synthesis is ribose-phosphate pyrophosphokinase, where ribose-5 phosphate is converted into phosphoribosyl pyrophosphate and ultimately, IMP (Hove-Jensen, 1988).

3.3.3 Proteins involved in protection from extracellular substances

The hypothetical protein Spne_T02001699 identified as upregulated in D39O has sequence homology to PTS system mannose-specific IIAB component. The gene encoding this protein was found to be repressed by CSP (Peterson *et al.*, 2004). The expression of this protein in *Listeria monocytogenes* (Ramnath *et al.*, 2000) and *L. lactis* (Kjos *et al.*, 2011) makes the bacteria sensitive to bacteriocins (leucocin A and lactococcin A, respectively) produced by their species. Bacteriocins can be used to inhibit the growth of other species or even members of the same species in a particular niche, such as in the nasopharynx, where a multitude of microbial species are in competition with each other. Since the T form is predominantly isolated from the nasopharynx, a downregulation of this protein and thus, reduction in sensitivity to bacteriocins would facilitate colonisation.

On the other hand, glutathione reductase (Gor) was found to be upregulated in WCH43T. Gor is important protein for maintenance of cellular processes by reducing glutathione disulfide to the sulfhydryl form glutathione (GSH), an important molecule in resistance to oxidative stress (Potter *et al.*, 2013) and maintenance of a reduced environment in the cell (Ritz & Beckwith, 2001). GSH provides protection from physiological stress, including reactive oxygen and nitrogen species, toxic concentrations of metal ions, and osmotic and acidic stresses (Lushchak, 2012; Masip *et al.*, 2006). Growth in an environment, such as in the nasopharynx, which has a higher oxygen level than the blood would give rise to oxidative stress. Although a *gor* mutant was found to have impaired growth in the presence of superoxide dismutase, copper, cadmium and to a lesser extent, zinc, it did not appear to be important for nasopharyngeal colonisation nor

invasive disease in a mouse challenge model (Potter *et al.*, 2013). Hence, it is unclear why Gor was upregulated in the T variant.

3.3.4 Expression of sugar and amino acid transporters

If a bacterial cell is unable to produce enough substrates required for energy production or DNA/protein synthesis, it will require transporters to import them from the environment. Those involved in sugar transport or uptake identified in this study were sugar ABC transporter, ATP- and substrate-binding proteins and MalR. Both the substrate and ATP-binding components of ABC transporters were upregulated in the T variant, suggesting that these may be required for the transport of substances required for growth during colonisation. On the other hand, MalR was preferentially expressed in the O variant of WCH16. MalR belongs to the LacI family of transcriptional repressors and regulates the transcription of two operons, MalXCD (Puyet *et al.*, 1993) and MalMP (Lacks & Hotchkiss, 1960; Weinrauch & Lacks, 1981), which are involved in maltosaccharide uptake and utilisation, respectively. Whole transcriptomic analysis of WCH16 and WCH43 showed that the expression of *malR* was upregulated in the lungs compared to the nasopharynx in a mouse model of disease (Ogunniyi *et al.*, 2012), which was consistent with the findings of this study where expression of MalR was higher in WCH16O relative to WCH16T.

3.3.5 ComE was upregulated in D39O

Natural competence in *S. pneumoniae* is regulated by competence-stimulating peptide (CSP), and at a critical concentration, ComD kinase is activated (Ween *et al.*, 1999). The activated ComD phosphorylates the response regulator ComE which then binds to and activates the transcription of its own promoter (Pestova *et al.*, 1996; Ween *et al.*, 1999; Ween *et al.*, 2002). The deletion of *comE* increased colonisation in a rat challenge model (Kowalko & Sebert, 2008). This association of increased nasopharyngeal colonisation fitness in absence of ComE is consistent with the increase in ComE in D39O (compared to D39T, commonly isolated from the nasopharynx) found in this study.

3.3.6 Regulation of pneumococcal SpxB in phase variation

SpxB is an enzyme that catalyses the oxidation of pyruvate to acetyl phosphate, generating hydrogen peroxide and carbon dioxide. The involvement of this enzyme in aerobic metabolism in Gram-positive bacteria has been established (Carvalho *et al.*, 2013; Goffin *et al.*, 2006; Seki *et al.*, 2004), and pneumococcal SpxB has been implicated in virulence (Orihuela *et al.*, 2004; Pericone *et al.*, 2000; Ramos-Montanez *et al.*, 2008; Spellerberg *et al.*, 1996), fermentation (Taniai *et al.*, 2008), capsule production (Carvalho *et al.*, 2013), as well as phenotypic phase variation (Allegrucci & Sauer, 2008; Belanger *et al.*, 2004; Overweg *et al.*, 2000). A previous study showed that SpxB was upregulated in T phase (Overweg *et al.*, 2000). However, in this study, SpxB was shown to be upregulated in D39O but not between the other O/T pairs. One plausible explanation for the observed differential expression of SpxB in D39T vs D39O could be that more cell death occurs in T variants than in O variants during growth to A_{600} 0.5 (Regev-Yochay *et al.*, 2007). Cells at this density are in mid- to late-logarithmic growth phase, whereby some SpxB is released into the medium, which would not be quantified from cell lysates used in the proteomic analysis. Surprisingly, there was no detectable difference in SpxB expression between O and T variants of any of the three strains by quantitative Western blotting. SpxB is known to exist in multiple isoforms when run on a 2D-gel (Bae *et al.*, 2006), thus, if all the isoforms were quantified in all variants of the three strains, there may or may not be an overall significant change in SpxB expression. However, multiple isoforms were not found amongst the top 12-15 spots that was picked for MS identification. Other likely explanations for the discrepancies in SpxB expression between O/T pairs between studies could be differences in the pneumococcal strain, growth media and cell density. For example, Overweg *et al*. (2000) used a serotype 9V strain grown in THY (nutrient-rich medium) to A_{550} 0.3 (early to mid-log phase) compared to serotypes 2, 4 and 6A grown in C+Y (chemically-defined medium) to mid- to late-logarithmic phase in this study. Although the results obtained by Overweg *et al*. (2000) seem compatible with the fact that T variants are better at colonisation than their O counterparts, hydrogen peroxide is not the only factor necessary for pneumococcal colonisation – a strain deficient in *spxB* also exhibited decreased adherence to glycoconjugates (Spellerberg *et al.*, 1996). Furthermore,

they showed that a strain with complete knockout of the *spxB* gene was less efficient at nasopharyngeal colonisation, pneumonia and sepsis (compared to WT) in animal models. However, this is in contrast to the study conducted by Syk *et al*. (2014), who demonstrated that although an *spxB*-deficient serotype 4 strain appeared to colonise better, there was no difference in virulence between the WT and the mutant.

In order to further investigate the role of SpxB in pneumococcal phase variation, *spxB* was deleted in strain D39, which resulted in a larger and hyper-opaque phenotype as documented in other studies (Allegrucci & Sauer, 2008; Carvalho *et al.*, 2013). This in part, may be due to the reduced production of hydrogen peroxide and thus an increase in biomass production over time as there is decreased cell death or due to an increase in capsule production (Carvalho *et al.*, 2013). The same group also stipulated that SpxB activity was regulated in response to local oxygen tension, as D39 grown under anaerobic conditions also had a larger and more O appearance on blood agar plates and the capsule amount was similar to the *spx*B-deficient D39 on a chemically defined medium. This is consistent with the fact that the O form is found more often in blood (less readily available oxygen). However, it is unlikely that the reduction of SpxB is a switch that determines pneumococcal opacity phenotypic variation, as ectopic over-expression of SpxB in the O variant did not result in a T phenotype. Instead, it showed a phenotype similar to the complete knockout mutants. Furthermore, when the *spxB* genes from either O or T variants were transformed into both D39O and D39T *spxB* mutants, the transformants reverted back to their WT phenotypes, regardless of the origin of the *spxB* gene. This suggests that the *spxB* gene itself is not responsible for the change in pneumococcal colony opacity variation, but rather, it can act in concert with other proteins to generate the different colony opacity phenotypes. Due to time constraints, a number of additional assays to better characterise the impact of SpxB on opacity phenotype were not performed. For future studies, it would be important to assay SpxB activity to ensure that D39O-pAL3::*spxB* was actually over-expressing enzymatically active SpxB. Furthermore, it would be interesting to assess whether the observed differential regulation of SpxB in D39 could be replicated in the other two strains using isogenic *spxB* mutants.

3.3.7 Quantitative Western Blot Analysis of Other Proteins

Using quantitative Western blot analysis, the relative expression profiles of 13 well-known virulence and cell surface proteins were compared. Of these, only NanA, Ply, pyruvate kinase and PhtD showed differential expression between any of the O/T variants. It was not surprising to find that the proteins tested were not all differentially regulated in the same phenotype in all three strains. D39, WCH16 and WCH43 have distinct pathogenic profiles (Mahdi *et al.*, 2008), suggesting that there may be differences in the relative importance of individual virulence-related proteins such as those tested to the physiology of different pneumococcal strains.

The pneumococcus is known to produce at least three sialidases, the best characterised of which is NanA (Xu *et al.*, 2011), which has been shown to be constitutively transcribed at a high level and present in all strains (King *et al.*, 2005). NanA is implicated as an important adherence promoter and virulence protein for the pneumococcus (Brittan *et al.*, 2012) and has been shown to contribute to nasopharyngeal colonisation and the translocation into the lungs (Orihuela *et al.*, 2004), sepsis (Manco *et al.*, 2006), development of otitis media (Tong *et al.*, 2000), as well as crossing the BBB (Uchiyama *et al.*, 2009). Several studies have shown that at the transcriptomic level, there was higher expression of *nanA* in the T variant compared to the O variant of the same pneumococcal strain *in vitro* (King *et al.*, 2004; Li-Korotky *et al.*, 2010), but this was variable when analysed across various strains harvested *in vivo* (Mahdi *et al.*, 2008). Furthermore, *nanA* expression was found to be higher in the nasopharynx than other sites such as the blood (LeMessurier *et al.*, 2006; Manco *et al.*, 2006). Although NanA has been implicated as an important protein for colonisation and pathogenesis, King *et al.* (2004) did not detect a role for NanA in colonisation and a *nanA* mutation had no impact on the virulence of pneumococcal strain D39 in a mouse i.p. challenge model (Berry & Paton, 2000). Furthermore, the present data showed an upregulation of NanA in WCH16O (vs. WCH16T), a serotype 6A strain, using quantitative Western blot analysis. This was also in contrast to the findings of King *et al.* (2004), whereby the opposite was true for their serotype 6A strain. Additionally, in the current study, NanA expression was upregulated in D39T and WCH43O, when compared to their respective counterparts. It is uncertain why

the findings of King *et al.* (2004) differed to those for this and other studies, but could perhaps be attributable to the growth medium and growth phase at the time of cell harvest.

Another well-known virulence factor that showed a difference in expression between the phenotypic variants is Ply. Several groups have shown that the loss of the *ply* gene led to reduced virulence (Berry *et al.*, 1989b; Berry & Paton, 2000; Canvin *et al.*, 1995; Ng *et al.*, 2002; Wellmer *et al.*, 2002) and this has been shown to be partly due to delayed recruitment of inflammatory cells to the lungs (Canvin *et al.*, 1995). Ply also activates the classical complement pathway in an antibody-independent fashion causing a depletion of opsonic activity (Paton *et al.*, 1984). The ability of Ply to cause inflammation (Berry *et al.*, 1989a; Cockeran *et al.*, 2001; Houldsworth *et al.*, 1994) has been linked to its ability to bind to TLR-4 (Malley *et al.*, 2003). Ply can also induce apoptosis in murine DC's (Colino & Snapper, 2003), respiratory cells (Schmeck *et al.*, 2004; Srivastava *et al.*, 2005) and neuronal cells (Bermpohl *et al.*, 2005) at sublytic concentrations. The toxin's cytolytic activity also impairs mucous-mediated clearance by inhibiting the ciliary beating in the respiratory tract (Feldman *et al.*, 1990; Feldman *et al.*, 2007; Steinfort *et al.*, 1989). It has been suggested that inflammation, in part due to Ply, may be required for initial colonisation before progression to disease (Weiser, 2010). The initial influx of neutrophils does not clear all of the pneumococci and the bacterium persists long after the inflammatory cells are no longer detected (Matthias *et al.*, 2008). Furthermore, pneumococci deficient in Ply or strains expressing Ply point mutants deficient in pore-forming activity had diminished inflammatory responses (Matthias *et al.*, 2008; van Rossum *et al.*, 2005). WCH16, a strain that is not as effective at causing bacteraemia as D39 or WCH43 in mouse models (Mahdi *et al.*, 2008), had significantly lower Ply expression in the T variant. Therefore, perhaps the difference in Ply between WCH16 T and O variants reflects the reduced propensity of WCH16 to invade. However, in the present study, there were no significant differences in Ply expression between the phenotypic variants of strains D39 and WCH43. This observation also correlated with gene expression data reported by Mahdi *et al.* (2008).

Data presented in this study reported that there was a significant change in expression level of Pyk in WCH16 and WCH43 O/T pairs, albeit in an opposite manner (WCH16O >

WCH16T; WCH43T > WCH43O). There was no difference in Pyk expression in D39. Pyk encodes the enzyme responsible for the final step in the glycolytic pathway (conversion of phosphoenolpyruvate and adenosine diphosphate [ADP] to pyruvate and ATP). The absence of consistency between strains in the expression patterns of Pyk suggests this protein may have limited importance in determining the O/T phenotype.

The final protein described in this chapter as having differential expression is PhtD. Together with PhtB, PhtD has been shown to have a role in progression from colonisation to lung disease (Hava & Camilli, 2002; Kallio *et al.*, 2014). PhtD is able to recruit fH and restrict C3 opsonisation, facilitating the persistence of the pneumococcus in the nasopharynx (Li *et al.*, 2007; Lysenko *et al.*, 2005). PhtD has also been implicated in metal scavenging (Rioux *et al.*, 2011) and zinc homeostasis (Plumptre *et al.*, 2014b). In this study, the differential expression of PhtD only reached statistical significance for one of the three strains tested (WCH16 O/T pair). WCH16 is better at colonising the murine nasopharynx than D39 or WCH43 (Mahdi *et al.*, 2008), and had increased expression of PhtD in the T variant. The increased expression in a variant that is predominantly associated with epithelial adherence was in concordance with a study whereby *phtD* or *phtE* deletion mutants had roughly 50% reduced ability to adhere to Detroit 562 (nasopharyngeal) or 549 (type II pneumocytes) (Khan & Pichichero, 2012).

3.4 Conclusion

The findings of the proteomic analysis of the three strains examined in this study suggest that a combination of metabolic profiles and overall protein expression contribute to the O and T phenotypes, and that these opacity variations are likely to be strain dependent. There does not appear to be one single protein that is consistently expressed by all O variants compared to all T variants, or vice versa. However, for the most part, it appears that proteins belonging to the same functional group were similarly regulated in the same variant. For example, those involved in stress responses (DnaJ and SsBb) and pyruvate metabolism (SpxB, AdhE and Pfl) were generally upregulated in the O variant. On the other hand, proteins such as those involved in glycolysis, and with an adherence property such as GAPDH, were predominantly upregulated in the T variant. Furthermore, the

findings of proteomic analysis suggest that opacity phase variation is a complex and multifactorial event, involving a combination of factors that act together to produce a certain opacity phenotype, which also contributes to pathogenic characteristics. For example, both strains D39 and WCH43 are more likely to cause bacteraemia, while strain WCH16 is more adept at colonising the nasopharynx, where the T variant is more commonly isolated. This could be one reason why the protein expression profiles of D39 and WCH43 are more similar to each other than to WCH16.

Chapter 4: TRANSCRIPTOMIC ANALYSIS OF PNEUMOCOCCAL OPACITY PHASE VARIANTS

4.1 INTRODUCTION

In the pneumococcus, DNA microarray analysis has been employed to conduct large-scale gene expression studies, including investigating genes expressed during colonisation (Hendriksen *et al.*, 2008), invasive disease (LeMessurier *et al.*, 2006; Mahdi *et al.*, 2012; Ogunniyi *et al.*, 2012; Orihuela *et al.*, 2004) and biofilm formation (Trappetti *et al.*, 2011b), as well as the effects of mutation of certain genes on global gene expression (Mackinnon *et al.*, 1993; Molzen *et al.*, 2010; Ogunniyi *et al.*, 2010). However, there are no dedicated studies comprehensively comparing gene expression profiles between pneumococcal opacity phase variants. Previous studies employed qRT-PCR to compare expression of specific genes of interest between opacity phase variants (Li-Korotky *et al.*, 2009; Li-Korotky *et al.*, 2010; Mahdi *et al.*, 2008). For example, Mahdi *et al.* (2008) compared *in vitro* expression levels of well-known virulence factors between O/T pairs but were unable to find a common differentially expressed gene between serotypically and pathogenically distinct pneumococcal opacity phase variable strains. Furthermore, different groups produce different results. For example, Mahdi *et al.* (2008) reported that expression of *ply* was significantly upregulated in the O variant of a serotype 6A strain, but was not significantly different between the variants of two other serotypes that they tested, nor was it differentially regulated in a separate serotype 6A strain examined by Li-Korotky *et al.* (2009). Such discrepancies between studies have also been seen for other virulence genes such as *nanA*, *psaA* and *cbpA* and these have been summarised in Chapter 1 (Table 1.1). The lack of consistency between studies could be due to differences in growth medium and growth phase at which the cells were harvested. Hence, the work described in this chapter employed DNA microarray analysis to compare transcriptome profiles to identify commonly-regulated genes between different and T variants across the three distinct strains. Genes of interest that exhibited differential expression between O/T pairs were confirmed using qRT-PCR. The contribution of these genes to the O or T phenotype were assessed using various functional assays.

4.2 RESULTS

4.2.1 Gene expression profiles between O/T pairs

In order to examine the impact of colony opacity phase variation on *S. pneumoniae* gene expression, the transcriptomes obtained from DNA microarray data between O/T pairs of three phenotypically distinct serotypes (D39, WCH16 and WCH43) were compared (Section 2.9.3). Total cellular RNA extracts were treated with DNase, labelled and hybridised to the printed microarray slides comprising PCR products derived from the open reading frames (ORFs) found in the published *S. pneumoniae* TIGR-4 genome (Tettelin *et al.*, 2001), and additional ORFs present in the R6 genome (Hoskins *et al.*, 2001). Array slides were obtained from the Bacterial Microarray Group at St. George's, University of London. The array design is available in BμG@Sbase (accession no. A-BUGS-14; http://bugs.sgul.ac.uk/A-BUGS-14) and, also, ArrayExpress (accession no. A-BUGS-14). Four independent biological replicates for each strain were used, resulting in a total of eight hybridisations performed per O/T pair, as each pair was tested twice using dye reversal. For example, cDNA of D39O and D39T from the first biological replicate were labelled with Cy^3 and Cy^5 respectively, and hybridised to a microarray slide. To control for dye bias, this was then repeated using dye reversal; that is, D39O was labelled with Cy^5, while D39T was labelled with Cy^3 (Figure 4.1). This was carried out for all the biological replicates of the three strains tested. Pooled transcriptomic data were analysed as described in Chapter 2 (Section 2.9.3.2). Differences in gene expression between O and T variants were initially compared within each strain, before analysing the data across all three strains for commonality in differential gene expression.

4.2.2 Transcriptome analysis of D39O vs. D39T

Transcriptome analysis of strain D39 produced the largest number of differentially expressed genes between the variants, identifying 268 differentially regulated genes with an adjusted *p*-value <0.05, fold-change of 2.0 or more and a Bayesian value of greater than 7 [Table 4.1; Appendix A (differentially regulated hypothetical proteins); Appendix B (raw data)] using the software and methodology described in Section 2.9.3.2. Of these, 44% (117 genes) were upregulated in the O form, with a fold-change range from 2.01 to 25.79.

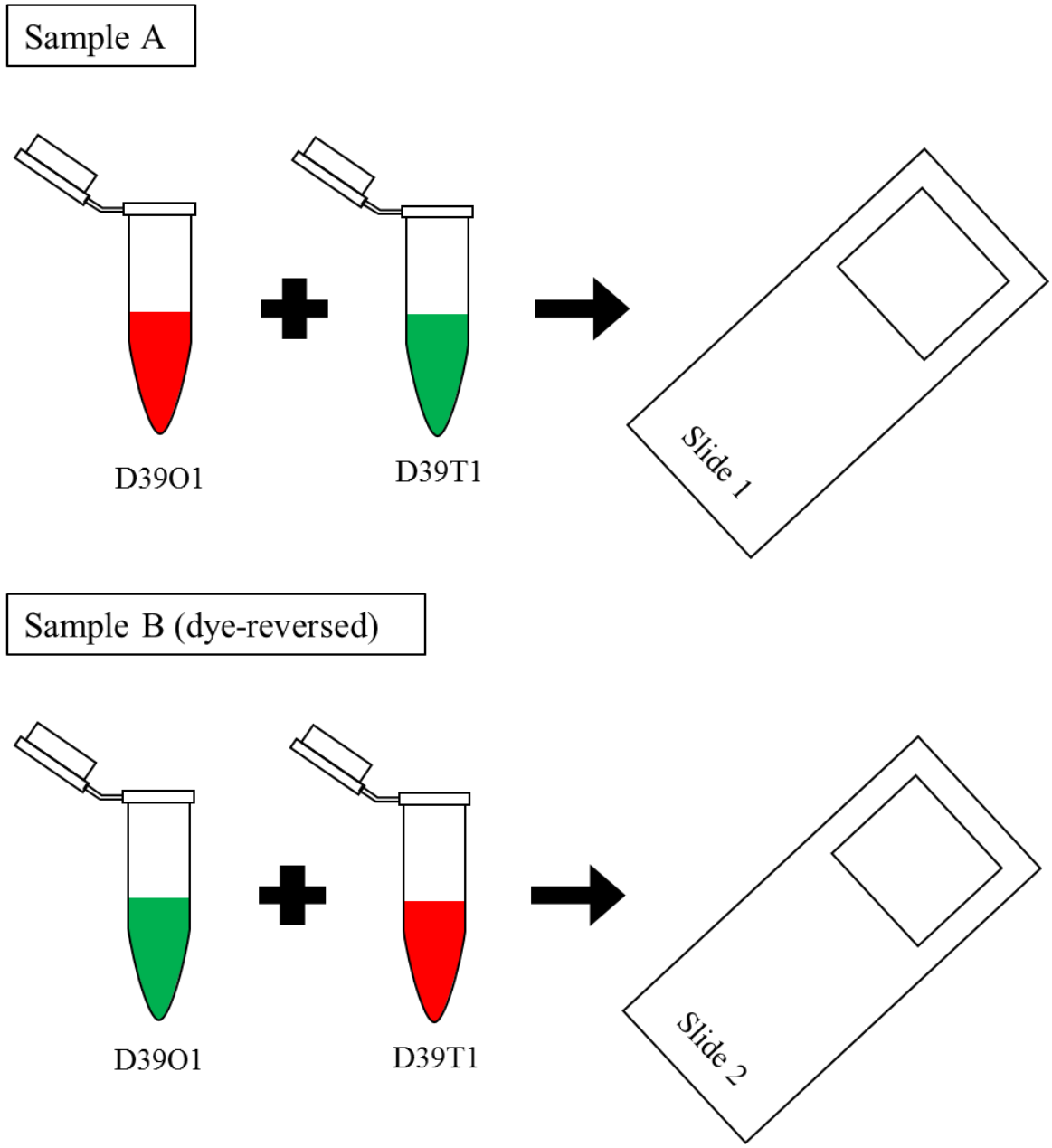


Figure 4.1. Dye-reversal strategy used in microarray analysis.

Figure showing a dye-reversal strategy used in this study, in which the dyes (labels) are swapped between the technical replicates. In sample A, D39O biological replicate 1 (D39O1) labelled with Cy5 (red) is mixed with D39T biological replicate 1 (D39T1), which was labelled with Cy3 (green) and then hybridised onto the microarray slides. To reduce any biases that may occur in hybridisation intensity due to the dye, a dye-reversed replicate was set-up (Sample B). In this case, D39O1 is labelled with Cy3 while D39T1 is labelled with Cy5 and hybridised onto another slide. This is repeated for the four biological replicates of D39, WCH16 and WCH43 O/T pair.

Table 4.1. Differential gene expression between *S. pneumoniae* O and T phase variants as determined by microarray analysis.

Gene name and/or annotation[a]	TIGR4 annotation[a]	Fold change[b] and *p*-value[c] in:					
		D39		WCH16		WCH43	
		Fold-change	*p*-value	Fold change	*p*-value	Fold-change	*p*-value
Putative Virulence Determinants							
bacteriocin BlpU	0041			-1.43	6.13E-04		
sensor histidine kinase	0084			-1.31	2.60E-03		
hypothetical protein (competence-induced bacteriocin A CibA)	0125	19.59	2.49E-10				
choline binding protein G	0390			1.69	3.41E-03		
two-component system, AgrA family, response regulator AgrA	0526	2.23	2.61E-07				
bacteriocin BlpJ	0532			-1.43	2.35E-04		
bacteriocin BlpO	0541			-1.47	1.06E-03		
superoxide dismutase, manganese-dependent	0766			-1.66	3.05E-04		
adherence and virulence protein A	0966			1.51	4.73E-04		
hypothetical protein (*phtD*)	1003	-3.73	9.57E-07	-1.69	1.12E-05		
hypothetical protein (*phtE*)	1004	-4.14	7.09E-07	1.80	1.55E-05		
hypothetical protein (*phtB*)	1174	-4.72	4.96E-11	-1.73	1.89E-03		
C3-degrading proteinase	1449	2.45	7.65E-07				
bacterocin transport accessory protein	1499	2.21	5.90E-08				
hemolysin	1466			1.36	1.02E-04		
metallo-beta-lactamase superfamily protein	1646	2.35	5.90E-08				
endopeptidase O	1647			1.51	1.19E-03		
TetR family transcriptional regulator	1858	2.93	1.83E-08				
pneumolysin	1923	-2.53	4.40E-07				
transcriptional regulator PlcR	1946	2.47	2.11E-08				
hypothetical protein (superoxide dismutase)	2054					-1.69	1.45E-02
transcriptional repressor, putative	2077	3.04	5.01E-08				
iga; immunoglobulin A1 protease	SpR6 1042	-3.12	1.40E-08				
phpA (phtA); histidine motif-containing protein	SpR6 1060	-3.97	7.01E-09	-1.71	2.86E-06		
pspC; choline binding protein A	SpR6 1995	-2.55	2.23E-07				
Genes involved in transcription							
DNA-binding response regulator	0083	2.08	4.80E-07				
PadR family transcriptional regulator, regulatory	0100	2.62	4.80E-07				
DNA-binding response regulator	0376	-2.25	2.17E-07				
DNA-binding response regulator VncR	0603	3.64	2.36E-10				
transcriptional regulator	0716			-1.72	3.26E-04		
MerR family transcriptional regulator	0739	2.02	6.50E-07				
TetR family transcriptional regulator	0743			1.51	1.40E-03		
KH domain-containing protein	0776			-1.31	6.89E-04		
PadR family transcriptional regulator, regulatory	0789	2.36	4.12E-07				
topology modulation protein	0837	7.89	3.61E-11				
transcriptional repressor, putative	0893			1.39	1.55E-05		
VicX protein	1225	-2.75	5.90E-08				
GntR family transcriptional regulator	1446			1.40	2.36E-04		
iron-dependent transcriptional regulator	1638	2.52	1.91E-09				
gpmA; phosphoglyceromutase	1655			1.70	5.94E-05		
type II restriction endonuclease DpnI	1850	2.57	1.83E-06				

Continues on following page

Table 4.1 continued

Gene name and/or description[a]	TIGR4 annotation[a]	Fold change[b] and *p*-value[c] in:					
		D39		WCH16		WCH43	
		Fold-change	*p*-value	Fold change	*p*-value	Fold-change	*p*-value
hypothetical protein (DNA-binding protein)	1915			2.75	6.29E-09		
transcriptional regulator	1942			1.34	4.71E-04		
transcriptional regulator CtsR	2195	2.28	4.71E-08				
transcriptional regulator MutR	1115	-2.12	1.26E-06				
xerS; site-specific tyrosine recombinase XerS	1159	2.10	5.01E-08				
GntR family transcriptional regulator	1714					-1.63	1.33E-02
maltose operon transcriptional repressor	2112			1.39	2.36E-05		
ribosomal subunit interface protein	2206					1.97	8.76E-04
hypothetical protein (site-specific DNA-methyltransferase (adenine-specific))	2045			1.71	7.79E-04		
SPSpoJ protein	2240			1.70	2.85E-05		
Genes involved in translation							
gidA; tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA	0120			-1.57	1.17E-05		
rplD; 50S ribosomal protein L4	0210			-1.38	1.30E-03		
rplW; 50S ribosomal protein L23	0211			-1.38	2.03E-03		
rplB; 50S ribosomal protein L2	0212			-1.30	2.53E-03		
DNA-directed RNA polymerase subunit alpha	0236	-2.06	2.66E-07				
rplQ; 50S ribosomal protein L17	0237			1.51	1.35E-05		
prolyl-tRNA synthetase	0264	-2.31	1.76E-08				
elongation factor P	0435			1.30	8.12E-04		
gatC; aspartyl/glutamyl-tRNA amidotransferase subunit C	0438			1.53	1.35E-05		
rpmB; 50S ribosomal protein L28	0441			-1.51	6.72E-05		
infB; translation initiation factor IF-2	0556	-2.53	5.68E-08	1.53	1.35E-05		
valS; valyl-tRNA synthetase	0568	-2.04	1.78E-06				
pheT; phenylalanyl-tRNA synthetase subunit beta	0581	-2.58	2.14E-07				
hypothetical protein (16S rRNA (guanine1207-N2)-methyltransferase)	0841			1.35	1.19E-03		
rpsA; 30S ribosomal protein S1	0862	-6.35	8.00E-10				
smpB; SsrA-binding protein	0976	-3.46	1.40E-08				
truB; tRNA pseudouridine synthase B	1212	-2.22	2.54E-08				
GTP-binding protein LepA	1200	-3.73	1.67E-10	-1.33	2.51E-04		
DNA processing protein DprA	1266	3.97	2.77E-08	1.41	3.05E-03		
rplS; 50S ribosomal protein L19	1293			-1.32	3.07E-03		
rplL; 50S ribosomal protein L7/L12	1354	-2.52	1.22E-07				
rplJ; 50S ribosomal protein L10	1355	-2.09	8.99E-08				
alaS; alanyl-tRNA synthetase	1383	-3.59	6.21E-08				
spoU rRNA methylase	1457			1.35	1.74E-04		
glyS; glycyl-tRNA synthetase subunit beta	1474	-2.56	2.16E-07				
glyQ; glycyl-tRNA synthetase subunit alpha	1475	-2.29	3.08E-08				
asnC; asparaginyl-tRNA synthetase	1542	-3.03	7.31E-10				
hypothetical protein (t-RNA (adenine22-NI-)-methyltransferase)	1610	2.49	1.92E-08				
ileS; isoleucyl-tRNA synthetase	1659	-6.98	2.36E-10				
engA, yfgK, yphC; GTP-binding protein EngA	1709	-2.87	2.49E-10				

Continues on following page

Table 4.1 continued

Gene name and/or description[a]	TIGR4 annotation[a]	Fold change[b] and *p*-value[c] in:					
		D39		WCH16		WCH43	
		Fold-change	*p*-value	Fold change	*p*-value	Fold-change	*p*-value
rRNA methyltransferase RsmB	1734			-1.30	2.35E-04		
iojap-related protein	1744	-2.06	8.50E-09				
hypothetical protein (tRNA threonylcarbamoyladenosine biosynthesis protein TsaE)	1944			1.56	1.09E-04		
DNA-directed RNA polymerase subunit beta	1960	-2.99	4.71E-08	-1.64	1.98E-07		
rpoB; DNA-directed RNA polymerase subunit beta	1961			-1.52	2.49E-06		
type II DNA modification methyltransferase	1969			1.57	1.97E-07		
rpoB; DNA-directed RNA polymerase subunit beta	1961			-1.52	2.49E-06		
type II DNA modification methyltransferase	1969			1.57	1.97E-07		
rRNA (guanine-N1-)-methyltransferase	2103			1.40	1.21E-03		
rpmF; 50S ribosomal protein L32	2134			-1.57	1.79E-04		
rpmG; 50S ribosomal protein L33				-1.67	5.94E-04		
NifR3 family TIM-barrel protein	2189	-2.40	1.76E-08				
rplI; 50S ribosomal protein L9	2204	-2.06	1.22E-07				
Post-translational protein alternations							
ATP-dependent Clp protease ATP-binding subunit	0338	3.35	2.70E-08				
type I restriction-modification system, S subunit	0508	-2.01	1.28E-06	-1.39	2.92E-03		
type I restriction-modification system, M subunit	0509	-2.48	5.51E-07				
16S rRNA (guanine1207-N2)-methyltransferase	0841	2.09	1.44E-06				
intein-containing protein	0871			-1.88	4.06E-05		
prsA; foldase protein PrsA	0981			1.57	5.43E-04		
bifunctional methionine sulfoxide reductase A/B protein	1359	-2.55	5.01E-07	-1.30	3.07E-03		
spxA, yjbD; transcriptional regulator Spx	1405	2.14	8.00E-10				
ATP-dependent RNA helicase, putative	1586	-3.34	2.07E-08	-1.68	4.78E-04		
OxaA-like protein precursor	1975			1.60	2.55E-03		
DNA repair, recombination, and modification							
mutL; DNA mismatch repair protein	0173			1.45	9.59E-04		
glutathione peroxidase	0313	2.36	1.59E-06				
recU; Holliday junction-specific endonuclease	0370	2.31	8.34E-08				
ribonuclease HIII	0403	2.89	1.73E-07				
hrcA; heat-inducible transcription repressor	0515	2.53	2.77E-08	1.74	2.60E-04		
uvrC; excinuclease ABC subunit C; K03703 excinuclease ABC subunit C	0618			1.32	2.05E-03		
peptidyl-prolyl cis-trans isomerase, cyclophilin-type	0771			1.69	4.00E-06		
Beta-galactosidase 3	0060					1.89	2.60E-02
DNA topoisomerase IV subunit B	0852			1.30	4.66E-04		
DNA topoisomerase IV subunit A	0855	-2.46	5.83E-10				

Continues on following page

Table 4.1 continued

Gene name and/or description[a]	TIGR4 annotation[a]	Fold change[b] and *p*-value[c] in:					
		D39		WCH16		WCH43	
		Fold-change	*p*-value	Fold change	*p*-value	Fold-change	*p*-value
DNA polymerase III subunits gamma and tau	0865	-2.66	1.13E-06				
dnaG; DNA primase	1072	2.27	5.90E-08	-1.48	4.00E-06		
radC; DNA repair protein RadC	1088	25.71	4.49E-10				
exonuclease RexB	1151	-2.16	2.14E-07				
uracil-DNA glycosylase	1169	2.19	1.77E-06				
DNA repair protein RecN	1202	-3.45	8.81E-09				
A/G-specific adenine glycosylase				-1.55	5.39E-05		
conserved hypothetical protein	1575	-2.20	5.90E-07				
ATP-dependent DNA helicase RecG	1697			-1.41	6.48E-05		
exodeoxyribonuclease	1845			-1.35	5.03E-04		
ATP-dependent DNA helicase RecG	1697			-1.41	6.48E-05		
exodeoxyribonuclease	1845			-1.35	5.03E-04		
single-stranded DNA-binding protein	1908	25.79	1.82E-11				
replicative DNA helicase	2203	-2.13	4.50E-08				
Nucleoside/nucleotide metabolism							
dut; deoxyuridine 5'-triphosphate nucleotidohydrolase	0021	2.37	5.79E-08				
phosphorylase Pnp/Udp family protein	0075			1.59	5.97E-04		
adk; adenylate kinase	0231	-4.12	2.62E-07				
thyA; thymidylate synthase	0669	2.80	2.61E-07				
SAP domain-containing protein	1292			1.39	1.06E-03		
glmM; phosphoglucosamine mutase	1559	-3.04	1.40E-08				
PhnA protein; phosphonoacetate hydrolase	1602	2.41	7.19E-10				
acetyltransferase (GNAT family)	1943			1.54	1.09E-04		
rpoB; DNA-directed RNA polymerase subunit beta	1961	-2.39	1.51E-07				
Competence/DNA transformation							
competence-induced protein Ccs16	0030	12.52	1.82E-11				
competence protein CelA	0954	11.36	8.01E-08				
adaptor protein MecA 1/2	1362	2.01	6.00E-07	1.36	6.24E-04		
competence damage-inducible protein A	1941	6.46	2.00E-09				
transcriptional regulator ComX2	2006	3.26	1.37E-06				
competence protein ComGF	2048	6.89	7.03E-07	1.94	1.47E-03		
competence protein CglD	2050	9.80	3.76E-08	2.70	8.02E-05		
competence protein CglC	2051	5.31	2.52E-07	2.54	2.66E-04		
competence protein CglB	2052			2.25	3.26E-04		
competence protein CglA	2053	9.19	3.37E-09	2.36	3.87E-04		
choline binding protein D	2201	3.65	2.10E-07	1.64	2.50E-04		
competence protein ComF	2207	4.01	1.55E-08	1.83	1.43E-03		
helicase; competence protein ComFA	2208			1.42	2.18E-03		
competence stimulating peptide 2	2237			-1.50	3.35E-03		
Energy metabolism							
hypothetical protein (CoA-binding protein)	0033	5.43	5.07E-08				
Beta-galactosidase 3	0060					1.89	2.60E-02
PTS system transporter subunit IIB	0061					1.54	3.25E-02
mevalonate kinase	0381	2.58	4.49E-10				
enoyl-(acyl-carrier-protein) reductase	0419	-2.91	2.60E-07				
fructose-bisphosphate aldolase	0605			1.40	1.95E-03		
lactate oxidase	0715					-15.98	3.86E-08
hypothetical protein (glyoxalase family protein)	0731			-1.77	1.11E-03		

Continues on following page

Table 4.1 continued

Gene name and/or description[a]	TIGR4 annotation[a]	Fold change[b] and *p*-value[c] in:					
		D39		WCH16		WCH43	
		Fold-change	*p*-value	Fold change	*p*-value	Fold-change	*p*-value
mannose-6-phosphate isomerase	0736					1.41	4.16E-02
HAD superfamily hydrolase	0805					-1.57	1.93E-02
bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase/5,10-methylene-tetrahydrofolate cyclohydrolase	0825	-3.00	3.08E-08				
pyruvate kinase	0897	-3.15	7.01E-09				
acetyltransferase	0953			1.34	6.72E-05		
phosphoenolpyruvate carboxylase	1068	-3.66	4.89E-11	1.33	7.48E-05		
acetyltransferase	1082	2.54	1.94E-07				
eutD; phosphotransacetylase	1100	-2.20	7.74E-07				
glycogen branching protein	1121					3.30	1.83E-04
eno; phosphopyruvate hydratase	1128	-2.10	1.71E-07				
acetoin dehydrogenase complex, E3 component, dihydrolipoamide dehydrogenase, putative	1161	-4.39	4.95E-07				
dihydrolipoamide acetyltransferase	1162	-4.01	1.08E-07				
HAD superfamily hydrolase	1171			1.32	1.84E-03		
galactose-6-phosphate isomerase subunit LacA	1193	3.64	1.00E-06				
hypothetical protein (PTS system, galactitol-specific IIB component)	1197					2.32	1.28E-02
F0F1 ATP synthase subunit alpha	1210	-4.16	2.49E-10				
glutamate dehydrogenase	1306	-2.42	1.10E-08				
glycosyl transferase family protein	1365	-2.61	1.26E-06				
NADH oxidase	1469	-2.81	2.49E-10				
atpC; F0F1 ATP synthase subunit epsilon	1507	-3.76	4.49E-10				
F0F1 ATP synthase subunit beta	1508	-4.78	5.79E-08				
F0F1 ATP synthase subunit gamma	1509	-3.76	7.49E-09				
F0F1 ATP synthase subunit delta	1511	-4.04	5.26E-09				
F0F1 ATP synthase subunit B	1512	-3.64	3.63E-07				
F0F1 ATP synthase subunit A	1513	-2.32	6.50E-07				
putative glutathione S-transferase YghU	1550	-2.09	4.91E-07				
oxalate:formate antiporter	1587			-1.57	4.71E-04		
acyltransferase family protein	1624	2.41	5.79E-08				
N-acetylmannosamine-6-phosphate 2-epimerase	1685					2.65	2.58E-06
isochorismatase family protein	1745	-2.37	3.93E-10				
acetate kinase	2044			1.57	2.66E-07		
zinc-containing alcohol dehydrogenase	2055					1.75	3.90E-02
ROK family protein	2142			-1.61	2.79E-03		
L-fuculose kinase fucK	2167					2.36	3.74E-02
Co-factor metabolism							
short chain dehydrogenase/reductase family oxidoreductase	0675	-3.33	9.03E-10				
hydroxyethylthiazole kinase	0717	-2.97	5.37E-09	-1.67	2.24E-05		
lipoate-protein ligase, putative	1160	-4.47	2.77E-08				
formate--tetrahydrofolate ligase	1229			-1.51	5.94E-04		
phosphopantothenate--cysteine ligase	1230			1.82	3.13E-04		
phosphopantothenoylcysteine decarboxylase	1231			1.92	5.97E-08		
acpS; 4'-phosphopantetheinyl transferase	1699			-1.38	5.50E-04		

Continues on following page

Table 4.1 continued

Gene name and/or description[a]	TIGR4 annotation[a]	Fold change[b] and *p*-value[c] in:					
		D39		WCH16		WCH43	
		Fold-change	*p*-value	Fold change	*p*-value	Fold-change	*p*-value
7-cyano-7-deazaguanine reductase	1777			-1.38	2.35E-04		
coaD; phosphopantetheine adenylyltransferase	1968			1.40	1.55E-05		
nicotinate-nucleotide pyrophosphorylase	2016	2.26	2.76E-07	-1.35	2.35E-04		
Capsule loci							
cps2H; hypothetical protein	SpR6 0315	-2.88	4.71E-08				
cps2I; hypothetical protein	SpR6 0316	-3.80	5.31E-09				
cps2J; hypothetical protein	SpR6 0317	-4.74	1.22E-07				
cps2K; hypothetical protein	SpR6 0318	-3.64	2.36E-10				
cps2P; hypothetical protein	SpR6 0319	-3.53	3.46E-09				
cps2L; hypothetical protein	SpR6 0320	-2.86	5.97E-07	-1.48	1.76E-03		
cps2M; hypothetical protein	SpR6 0321	-3.27	8.12E-09	-1.49	1.12E-04		
cpsN; dTDP-glucose-4,6-dehydratase	SpR6 0322	-3.61	1.28E-08	-1.40	2.46E-04		
Cell wall or cell membrane-related							
transporter, putative	0101			1.45	5.69E-05		
glycosyl transferase	0102	3.56	3.11E-08	1.62	7.76E-06		
LysM domain-containing protein	0107			2.69	8.93E-04		
hypothetical protein	0760			-1.60	1.33E-04		
lipoprotein	0845	-2.48	1.36E-09				
adhesion lipoprotein	1002	-3.15	8.70E-09	-1.66	2.51E-04		
phosphoenolpyruvate-protein phosphotransferase	1176	-2.71	2.59E-09				
sensory box sensor histidine kinase	1226	-2.04	6.01E-08				
licC protein	1267			1.42	2.30E-03		
licB protein	1268			1.51	2.07E-03		
choline kinase	1269			1.31	2.03E-03		
ispD; 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	1271			1.56	5.07E-04		
licD1 protein	1273	-3.12	1.30E-08				
licD2 protein	1274	-3.91	6.31E-10				
spermidine/putrescine ABC transporter, ATP-binding protein	1389	-2.08	2.73E-09				
Gfo/Idh/MocA family oxidoreductase	1482	2.02	8.34E-08				
acyltransferase	1624			1.39	7.81E-05		
aquaporin	1778	-3.68	1.92E-08	-1.46	1.40E-03		
oligopeptide ABC transporter, permease protein AmiD	1889	-2.42	1.85E-07				
oligopeptide ABC transporter, ATP-binding protein AmiF	1887	-2.31	8.55E-08				
oligopeptide ABC transporter, ATP-binding protein AmiE	1888	-3.09	1.35E-07				
oligopeptide ABC transporter, permease protein AmiC	1890	-2.74	7.41E-07				
cell wall surface anchor family protein	1992	3.37	3.84E-08				
gpsA; NAD(P)H-dependent glycerol-3-phosphate dehydrogenase	2091			1.79	1.11E-03		
hypothetical protein	2143			-2.09	2.13E-03		
DltD protein	2173			1.87	1.01E-08		
DltB protein	2175			1.88	8.02E-05		
Fatty acid metabolism							
acetyl-CoA carboxylase biotin carboxyl carrier protein subunit	0423			-1.57	1.43E-03		
acetyl-CoA carboxylase subunit beta	0426			-1.57	3.08E-03		

Continues on following page

Table 4.1 continued

Gene name and/or description[a]	TIGR4 annotation[a]	Fold change[b] and *p*-value[c] in:					
		D39		WCH16		WCH43	
		Fold-change	*p*-value	Fold change	*p*-value	Fold-change	*p*-value
Cell division							
cell division protein FtsH	0013	-2.45	3.08E-08				
cell division ABC transporter, permease protein FtsX	0757	-2.35	3.71E-07				
rod shape-determining protein RodA, putative	0803	-2.57	1.83E-07				
septation ring formation regulator EzrA	0807	-3.76	4.01E-08				
signal recognition particle-docking protein FtsY	1244	-2.81	2.90E-08				
psr protein	1368	-2.11	3.87E-07				
cell division protein DivIVA	1661	-3.46	1.33E-09			-1.62	2.48E-02
YlmF protein	1662	-2.32	3.17E-07				
YlmE protein	1665	-2.16	7.05E-07				
cell division protein FtsZ	1666	-2.82	8.00E-10				
YlmH protein	2169	-2.79	1.22E-07				
rod shape-determining protein MreC	2218	-3.27	3.08E-10				
Stress-related							
DNA-damage-inducible protein J	0275	3.35	9.94E-09				
DNA-binding response regulator	0387			1.80	1.66E-04		
hypothetical protein; Fe-S cluster assembly protein SufD	0868			-1.84	1.27E-07		
greA; transcription elongation factor GreA	1517	-2.48	2.19E-09				
groEL; chaperonin GroEL	1906	3.14	3.46E-09	1.55	2.50E-04		
groES; co-chaperonin GroES	1907	15.60	1.22E-07	1.62	1.52E-03		
universal stress protein	1996	2.56	9.66E-07			1.74	1.45E-02
Amino acid biosynthesis and acquisition							
adenylosuccinate synthetase	0019			1.43	1.51E-03		
beta-N-acetylhexosaminidase	0057	2.54	1.09E-07				
L-serine dehydratase, iron-sulfur-dependent, beta subunit	0106			1.46	7.42E-04		
MutT/nudix family protein	0119			-1.49	2.00E-03		
aminopeptidase C	0281			1.66	9.98E-06		
xanthine/uracil permease family protein	0287			-2.27	2.37E-06		
Cof family protein	0286			1.69	7.82E-05		
ilvH; acetolactate synthase 3 regulatory subunit	0446	-2.68	1.78E-06				
serine acetyltransferase	0589	-2.63	1.08E-07				
hypothetical protein (peptidase)	0617			1.39	1.46E-03		
amino acid ABC transporter ATP-binding protein	0709			-1.32	1.96E-03		
thiamine-phosphate pyrophosphorylase	0718	-2.89	6.43E-08	-1.52	2.00E-03		
glutathione reductase	0784			1.56	1.16E-03		
aminotransferase, class-V	0869	-4.10	4.49E-10	-2.23	1.46E-09		
lysine decarboxylase	0916			-2.08	2.49E-06		
spermidine synthase	0918			-2.68	1.17E-05		
hypothetical protein	0919			-	1.72E-05		
agmatine deiminase	0921			-2.72	7.48E-05		
peptidase T	1008			1.32	2.47E-03	1.34	3.75E-02
hypothetical protein (aromatic ring hydroxylating protein)	1074	2.89	5.64E-07	-1.45	8.47E-04		

Continues on following page

Table 4.1 continued

Gene name and/or description[a]	TIGR4 annotation[a]	Fold change[b] and *p*-value[c] in:					
		D39		WCH16		WCH43	
		Fold-change	*p*-value	Fold change	*p*-value	Fold-change	*p*-value
glutamine amidotransferase, class I	1089	6.29	1.36E-09				
amino acid permease family protein	1001	-2.21	2.17E-07				
aspartate-semialdehyde dehydrogenase	1013	-2.44	2.93E-07				
methionine aminopeptidase	1084	-3.24	2.49E-07				
serine/threonine protein phosphatase	1201	-4.06	8.00E-10				
amino acid ABC transporter, amino acid-binding protein/permease protein	1241	-2.42	4.04E-08				
amino acid ABC transporter, ATP-binding protein	1242	-2.09	1.29E-06				
Cof family protein	1245	-2.20	3.37E-10				
Cof family protein	1246	-2.60	2.10E-08				
guanosine 5'-monophosphate oxidoreductase	1249			3.78	2.63E-09		
hypothetical protein (chorismate mutase)	1296	2.87	3.29E-07				
homoserine kinase	1360			-1.35	5.30E-04		
amino acid ABC transporter, ATP-binding protein	1460	-2.02	1.20E-07	-1.34	2.61E-03		
proline dipeptidase	1591			1.33	2.79E-03		
UDP-glucose 4-epimerase	1607			1.51	2.44E-03		
carbon-nitrogen hydrolase family protein	1625			-2.15	5.63E-04		
hypothetical protein (serine/threonine protein phosphatase)	1637	2.71	2.21E-09				
alr; alanine racemase	1698			-1.40	1.36E-04		
phospho-2-dehydro-3-deoxyheptonate aldolase	1700	-2.43	9.79E-08				
phospho-2-dehydro-3-deoxyheptonate aldolase	1701	-2.44	3.95E-08				
trpC; indole-3-glycerol-phosphate synthase	1814			-1.92	1.11E-03		
trpD; anthranilate phosphoribosyltransferase	1815			-1.89	8.02E-04		
anthranilate synthase component II	1816			-1.63	1.06E-03		
hypothetical protein (acyl-coA thioesterase)	1851	2.03	4.71E-08				
choline transporter	1861			-1.43	1.29E-04		
glutamyl-aminopeptidase	1865			1.31	7.60E-04		
UDP-N-acetylglucosamine 1-carboxyvinyltransferase	1966	-2.56	5.64E-09				
L-asparaginase	1998	2.05	2.51E-07	1.30	8.47E-04		
transcriptional regulator of arginine metabolism	2077	3.04	5.01E-08				
UTP-glucose-1-phosphate uridylyltransferase	2092			1.37	3.42E-05		
dihydroxy-acid dehydratase	2126			1.53	2.46E-04		
arginine deiminase	2148					1.92	1.45E-02
D-alanine--poly(phosphoribitol) ligase subunit 2	2173			1.88	1.72E-05		
D-alanine--poly(phosphoribitol) ligase subunit 1	2174			1.69	5.28E-06		
hypothetical protein (amidase)	2216			1.75	1.96E-03		
serine protease	2239			1.49	2.26E-03		
amino acid ABC transporter ATP-binding protein	SpR6 0111			1.38	1.52E-03		

Continues on following page

Table 4.1 continued

Gene name and/or description[a]	TIGR4 annotation[a]	Fold change[b] and *p*-value[c] in:					
		D39		WCH16		WCH43	
		Fold-change	*p*-value	Fold change	*p*-value	Fold-change	*p*-value
appA; ABC transporter substrate-binding protein - oligopeptide transport	SpR6 1194	-2.40	2.75E-07				
Anion/Cation acquisition							
trk family potassium uptake protein	0079	-2.22	4.01E-08				
oligopeptide ABC transporter oligopeptide-binding protein AliA	0366			-1.49	3.07E-03		
sensor histidine kinase, putative	0386			1.90	2.24E-05		
iron-compound ABC transporter, permease protein	1033	-2.65	3.08E-08				
iron-compound ABC transporter, ATP-binding protein	1035	-6.88	2.73E-09	-1.38	3.81E-05		
phosphate transporter ATP-binding protein	1397	-2.26	1.93E-07				
cation transporter E1-E2 family ATPase	1551			-1.37	3.23E-04		
cadmium resistance transporter	1625			-1.49	3.25E-03		
cmp-binding-factor 1	1980	-3.16	1.28E-08				
zinc ABC transporter, zinc-binding lipoprotein	2169	-2.86	8.01E-08				
adcB; zinc ABC transporter permease	2170			-1.32	1.31E-04		
hypothetical protein (cobalt transport protein)	2219	-2.88	4.81E-07				
cbiO; cobalt transporter ATP-binding subunit	2221	-2.78	8.24E-08				
zmpB; zinc metalloprotease	SpR6 0581	-2.79	4.11E-09				
Sugar transport							
sugar ABC transporter, ATP-binding protein	0846	-2.94	2.71E-08				
sugar ABC transporter, permease protein, putative	0847	-2.71	8.00E-10				
sugar ABC transporter, permease protein, putative; K02057 simple sugar transport system permease protein	0848	-4.00	3.46E-09				
ABC transporter ATP-binding protein	0867			-1.59	5.52E-04		
sugar ABC transporter ATP-binding protein	1580					1.56	1.93E-02
maltose/maltodextrin ABC transporter maltose/maltodextrin-binding protein	2108					1.75	8.76E-04
MalA protein	2111			1.59	1.09E-04		

[a]Gene name and annotation according to KEGG (http://www.kegg.com/) website

[b]A positive fold change indicates upregulation in the opaque form and a negative fold-change indicates upregulation in the transparent form

[c]The *p-value* used here is the adjusted *p*-value of four biological replicates

The remaining 151 genes were upregulated in the T form, with a fold-change range from 2.01 to 6.98. The five genes with the highest fold changes in D39O were *ssbB* (SP1908; 25.79-fold), *radC* (SP1088; 25.71-fold), gene encoding a hypothetical protein SP0125 (competence-induced bacteriocin A *cibA S. pneumoniae* GA07643, 98% identity match using BLAST alignment; 19.59-fold), *groES* (SP1907; 15.60-fold) and *ccs16* (gene encoding competence-induced protein SP0030; 12.52-fold). On the other hand, those that were highly upregulated in D39T were *ileS* (SP1659; 6.98-fold), iron-compound ABC transporter (SP1035; 6.88-fold), *rpsA* (SP0862; 6.34-fold), F_0F_1 ATP synthase subunit beta (SP1508, 4.78-fold) and *cps2J* (Spr0314; 4.74-fold).

4.2.3 Transcriptome analysis of WCH16O vs. WCH16T

When the same statistical cut-offs used to compare mRNA levels between D39O and D39T (Section 4.2.2) were applied to strain WCH16, it yielded ten genes (genes encoding GMP reductase, hypothetical proteins, spermidine synthase, xanthine/uracil permease, class-V aminotransferase and lysine decarboxylase) that were differentially regulated between WCH16O and WCH16T (as highlighted in Appendix C – pseudogenes have not been excluded from the table), and only one of these was similarly regulated between the O/T variants of both strains (SP0869; upregulated in the T forms of both D39 and WCH16; 4.10 and 2.23 fold-change, respectively). On the other hand, if the fold-change was dropped to ≥1.3, with an unaltered Bayesian value to that for D39, it would only yield 38 genes. Hence, the differentially regulated genes between WCH16 O and T phase variants were compared using an adjusted *p*-value <0.05, fold-change ≥1.3 and a positive Bayesian value [the larger the value, the greater the support for the hypothesis that there is a difference between the O/T pairs (Kass & Raftery, 1995)] (Table 4.1 and Appendix A). This generated 206 differentially regulated genes between WCH16O and WCH16T. Of these, 113 (55%) were upregulated in WCH16O and 93 (45%) were upregulated in WCH16T. The fold-change ranges of the differentially regulated genes were 1.30 to 3.78 in the O variant and 1.30 to 2.72 in the T variant. The top five differentially expressed genes in WCH16O were GMP oxidoreductase (SP1249; 3.75-fold), genes encoding hypothetical proteins (SP0097 and SP1914; 3.11- and 3.09-fold, respectively), *lysM* domain-containing gene (SP2063; 2.93-fold) and a gene encoding hypothetical protein SP1915 (DNA-binding protein *S. pneumoniae* SPA026; 2.75-fold). Those that were

upregulated in the T variants were agmatine deiminase (SP0921; 2.72-fold), spermidine synthase (SP0918; 2.68-fold), a gene encoding hypothetical protein SP0742 (orthologous to DegV family protein, *S. pneumoniae* TCH8431/19A; 2.28-fold), xanthine/uracil permease family (SP0287; 2.27-fold) and aminotransferase class-V (SP0869; 2.23-fold).

4.2.4 Transcriptome analysis of WCH43O vs. WCH43T

Using the same cut-offs as described in Section 4.2.2, three genes (genes encoding lactate oxidase, a hypothetical protein and N-acetylmannosamine-6-phosphate 2-epimerase) were found to be differentially regulated between WCH43O and WCH43T (as highlighted in Appendix D – pseudogenes have not been excluded from the table) but none of these were differentially regulated in the other two strains tested. If the fold-change was reduced to ≥1.3, with an unaltered Bayesian value to that for D39, it would yield the same three genes. Hence, the same statistical cut-offs as those used to analyse differences in WCH16 O/T pairs (Section 4.2.3) were used to analyse differences in gene expression between WCH43 O/T pairs. This generated 27 differentially regulated genes between WCH43O and WCH43T (Table 4.1 and Appendix A). The majority (20 genes; 74%) of the mRNA levels were upregulated in the O variant, with a fold-change range from 1.34 to 3.30. Although there were only seven genes detected as differentially expressed in the T variant, the fold-change range was much bigger (1.59-15.98). The five highly regulated genes in the O form were glycogen branching protein (SP1121; 3.30-fold), a gene encoding hypothetical protein SP2182 (2.91-fold), N-acetylmannosamine-6-phosphate 2-epimerase (SP1685; 2.65-fold), L-fuculose kinase *fucK* (SP2167; 2.36-fold) and a gene encoding hypothetical protein SP1197 (PTS system, galactitol-specific IIB component; 2.32-fold). In WCH43T, the five most upregulated genes were *lctO* (lactate oxidase; SP0715; 15.98-fold), a gene encoding hypothetical protein SP0800 (1.72-fold), a gene encoding hypothetical protein SP2054 (orthologous to superoxide dismutase in *S. pneumoniae* TCH8431/19A; 1.69-fold), a gene encoding a GntR family transcriptional regulator (SP1714; 1.63-fold), and *divIVA* (cell division initiation SP1661; 1.63-fold).

4.2.5 Comparative transcriptome analysis of D39, WCH16 and WCH43 O and T variants

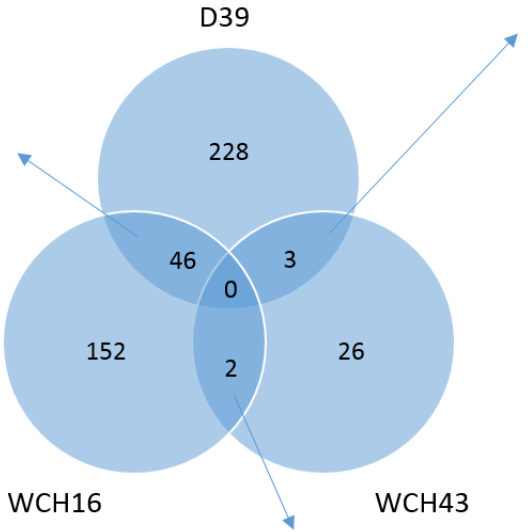
The aim of this study was to identify differently expressed gene(s) that are common between the O/T variants across the three strains. Since the changes in gene expression between the O/T variants of WCH16 and WCH43 were less pronounced and affected fewer genes compared to D39, less stringent statistical analyses were applied to these strains. There were 46 genes that were commonly regulated in D39 and WCH16, three genes between D39 and WCH43 and two genes between WCH16 and WCH43, as shown in a Venn diagram on Figure 4.2 (hypothetical proteins are not included). Surprisingly however, there were no genes similarly regulated using the specified statistical cut-offs (Sections 4.2.2 to 4.2.4) between all three strains. Genes encoding hypothetical proteins or those with unknown functions are listed in Appendix A.

Although there was not one single gene that was differentially expressed in the same pattern across the three strains, there were genes with similar functions that were similarly expressed. For example, the genes involved in competence/DNA transformation and stress-related genes were upregulated in the O forms of all three strains. On the other hand, those genes consistently upregulated in the T variants appear to be involved in translation, energy metabolism, cell wall or capsule synthesis, membrane-related, fatty-acid metabolism, anion/cation acquisition and amino acid biosynthesis and acquisition. Genes involved in other cellular processes were not considered to associated with either O or T phases, as these were not consistently upregulated in one phase over the other.

There were fold-change trends in some genes that were consistent with significantly differentially regulated genes in the other strains. To explain this feature, a "heat-map" of the fold-change was used to show apparent upregulation of certain genes. The heat-map gives an overall picture of the top 50 differentially regulated genes between all three strains (Figure 4.3). Since both strains WCH16 and WCH43 had fewer differentially expressed genes that reached statistical significance as described for D39 (Section 4.2.2), genes with significant values from combined analyses (combination of all values from O and T variant, regardless of strain) were used. That is, the data from all the experiments were pooled together and the top 50 genes that had a pooled *p*-value <0.01 and a positive Bayesian value (Appendix E) were chosen to generate the heat-map. However, due to the

Gene name and/or description	TIGR annotation	Fold change D39	Fold change WCH16
glycosyl transferase	0102	3.56	1.62
type I restriction-modification system, S subunit	0508	-2.01	-1.39
hrcA; heat-inducible transcription repressor	0515	2.53	1.74
infB; translation initiation factor IF-2	0556	-2.53	1.53
hydroxyethylthiazole kinase	0717	-2.97	-1.67
thiamine-phosphate pyrophosphorylase	0718	-2.89	-1.52
aminotransferase, class-V	0869	-4.10	-2.23
NifU family protein	0870	-2.60	-2.10
adhesion lipoprotein	1002	-3.15	-1.66
iron-compound ABC transporter, ATP-binding protein	1035	-6.88	-1.38
phosphoenolpyruvate carboxylase	1068	-3.66	1.33
ABC transporter, ATP-binding protein	1071	2.03	-1.69
dnaG; DNA primase	1072	2.27	-1.48
GTP-binding protein LepA	1200	-3.73	-1.33
DNA processing protein DprA	1266	3.97	1.41
bifunctional methionine sulfoxide reductase A/B protein	1359	-2.55	-1.30
adaptor protein MecA 1/2	1362	2.01	1.36
amino acid ABC transporter, ATP-binding protein	1460	-2.02	-1.34
ATP-dependent RNA helicase, putative	1586	-3.34	-1.68
aquaporin	1778	-3.68	-1.46
groEL; chaperonin GroEL	1906	3.14	1.55
groES; co-chaperonin GroES	1907	15.60	1.62
DNA-directed RNA polymerase subunit beta	1960	-2.99	-1.64
L-asparaginase	1998	2.05	1.30
nicotinate-nucleotide pyrophosphorylase	2016	2.26	-1.35
competence protein ComGF	2048	6.89	1.94
competence protein CglD	2050	9.80	2.70
competence protein CglC	2051	5.31	2.54
competence protein CglA	2053	9.19	2.36
choline binding protein D	2201	3.65	1.64
competence protein ComF	2207	4.01	1.83
cps2L; hypothetical protein	SpR6 0320	-2.86	-1.48
cps2M; hypothetical protein	SpR6 0321	-3.27	-1.49
cpsN; dTDP-glucose-4,6-dehydratase	SpR6 0322	-3.61	-1.40
phpA; histidine motif-containing protein	SpR6 1060	-3.97	-1.71

Gene name and/or description	TIGR annotation	Fold change D39	Fold change WCH43
cell division protein DivIVA	1661	-3.46	-1.62
universal stress protein	1996	2.56	1.74



Gene name and/or description	TIGR annotation	Fold change WCH16	Fold change WCH43
peptidase T	1008	1.32	1.34

Figure 4.2. Venn diagram of differentially expressed genes.

Venn diagram showing the relationship of the transcriptome data as determined by DNA microarray analysis between strains D39, WCH 16 and WCH43 O/T variants. Hypothetical proteins are not shown and a positive fold-change indicates upregulation in the O variant and a negative fold-change indicates upregulation in T variant.

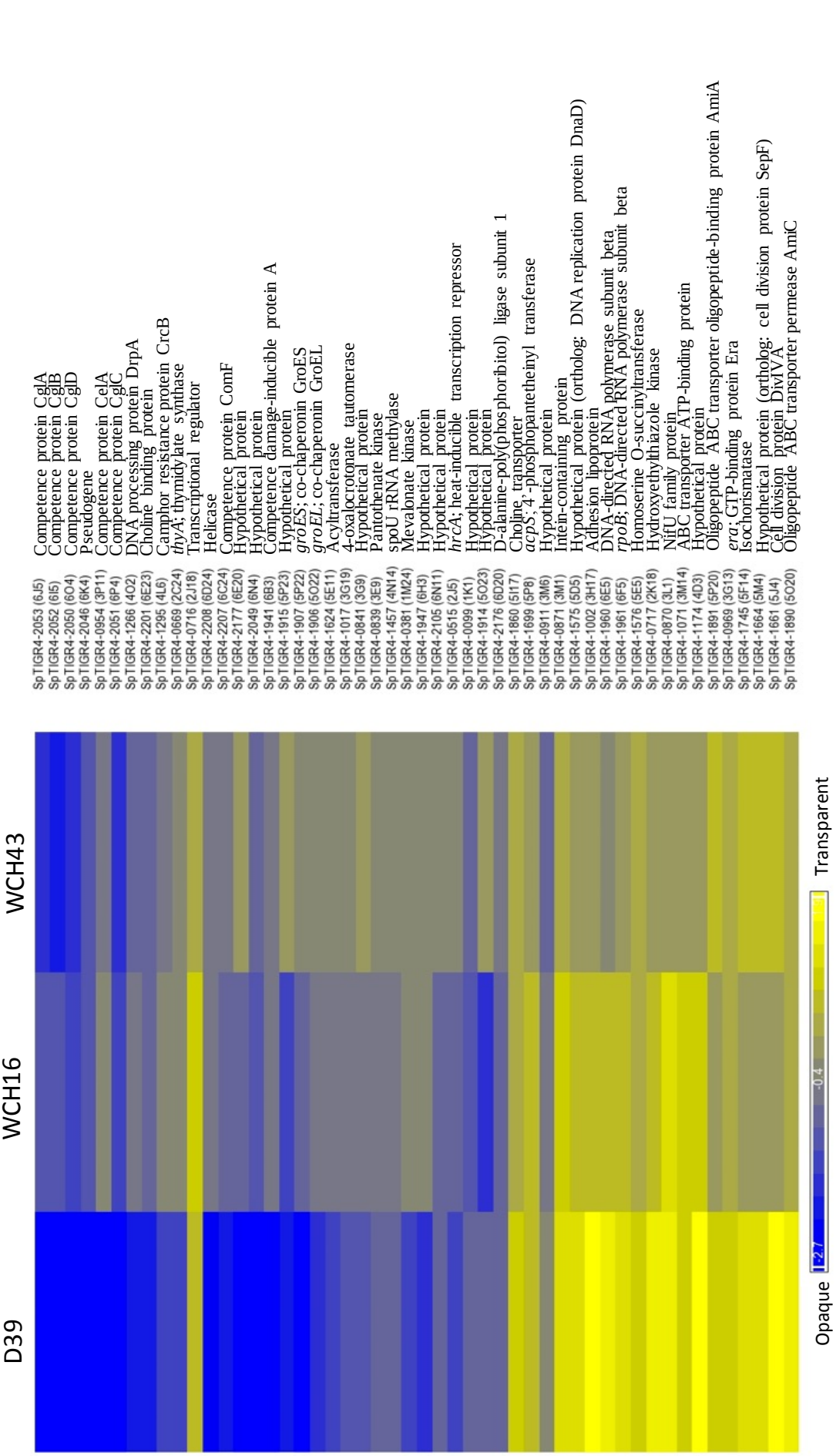

D39
WCH16
WCH43
SpTIGR4-2053 (6J5) Competence protein CglA
SpTIGR4-2052 (6I5) Competence protein CglB
SpTIGR4-2050 (6O4) Competence protein CglD
SpTIGR4-2046 (6K4) Pseudogene
SpTIGR4-0954 (3P11) Competence protein CelA
SpTIGR4-2051 (6P4) Competence protein CglC
SpTIGR4-1266 (4O2) DNA processing protein DrpA
SpTIGR4-2201 (6E23) Choline binding protein
SpTIGR4-1295 (4L6) Camphor resistance protein CrcB
SpTIGR4-0669 (2C24) thyA; thymidylate synthase
SpTIGR4-0716 (2J16) Transcriptional regulator
SpTIGR4-2208 (6D24) Helicase
SpTIGR4-2207 (6C24) Competence protein ComF
SpTIGR4-2177 (6E20) Hypothetical protein
SpTIGR4-2049 (6N4) Hypothetical protein
SpTIGR4-1941 (6B3) Competence damage-inducible protein A
SpTIGR4-1915 (5P23) Hypothetical protein
SpTIGR4-1907 (5P22) groES; co-chaperonin GroES
SpTIGR4-1906 (5O22) groEL; co-chaperonin GroEL
SpTIGR4-1624 (5E11) Acyltransferase
SpTIGR4-1017 (3G19) 4-oxalocrotonate tautomerase
SpTIGR4-0841 (3G9) Hypothetical protein
SpTIGR4-0839 (3E9) Pantothenate kinase
SpTIGR4-1457 (4N14) spoU rRNA methylase
SpTIGR4-0381 (1M24) Mevalonate kinase
SpTIGR4-1947 (6H3) Hypothetical protein
SpTIGR4-2105 (6N11) Hypothetical protein
SpTIGR4-0515 (2J5) hrcA; heat-inducible transcription repressor
SpTIGR4-0099 (1K1) Hypothetical protein
SpTIGR4-1914 (5O23) Hypothetical protein
SpTIGR4-2176 (6D20) D-alanine-poly(phosphoribitol) ligase subunit 1
SpTIGR4-1860 (5I17) Choline transporter
SpTIGR4-1699 (5P8) acpS; 4'-phosphopantetheinyl transferase
SpTIGR4-0911 (3M6) Hypothetical protein
SpTIGR4-0871 (3M1) Intein-containing protein
SpTIGR4-1575 (5D5) Hypothetical protein (ortholog: DNA replication protein DnaD)
SpTIGR4-1002 (3H17) Adhesion lipoprotein
SpTIGR4-1960 (6E5) DNA-directed RNA polymerase subunit beta
SpTIGR4-1961 (6F5) rpoB; DNA-directed RNA polymerase subunit beta
SpTIGR4-1576 (5E5) Homoserine O-succinyltransferase
SpTIGR4-0717 (2K18) Hydroxyethylthiazole kinase
SpTIGR4-0870 (3L1) NifU family protein
SpTIGR4-1071 (3M14) ABC transporter ATP-binding protein
SpTIGR4-1174 (4D3) Hypothetical protein
SpTIGR4-1891 (5P20) Oligopeptide ABC transporter oligopeptide-binding protein AmiA
SpTIGR4-0969 (3G13) era; GTP-binding protein Era
SpTIGR4-1745 (5F14) Isochorismatase
SpTIGR4-1664 (5M4) Hypothetical protein (ortholog: cell division protein SepF)
SpTIGR4-1661 (5J4) Cell division protein DivIVA
SpTIGR4-1890 (5O20) Oligopeptide ABC transporter permease AmiC
Opaque
2.7
-0.4
Transparent

Figure 4.3. Heat-map of the top 50 differentially regulated genes.

Heat-map showing the top 50 regulated genes between O and T across the three strains. Strains D39 and WCH16 used averaged duplicate spot values from each array (each comprising at least four biological replicates) and dye-reversal factored in. Since strain WCH16 and WCH43 had fewer differentially expressed genes with statistical significance described for D39 (Section 4.2.2) genes with significant values from combined analysis (pooled spot array data across all three strains) were used. Blue indicates upregulation in the O form and yellow indicates upregulation in the T form; colour intensity reflects fold-change in expression. The TIGR4 annotation is based on KEGG (www.kegg.jp) gene annotation for *S. pneumoniae* strain TIGR4.

large variability between the strains, these may not all be statistically significant according to the statistics described in Sections 4.2.2 to 4.2.4. On the heat-map, blue-coloured bars indicate upregulation in the O form, while the yellow-coloured bars indicate upregulation in the T form. Accordingly, there seems to be a consistent upregulation of genes involved in competence and/or transformation in the O form across the three strains. On the other hand, genes consistently upregulated in the T form appear to be related to cell division and transporters. The heat-map also gives an overview of the distinct patterns of gene regulation between the strains. For example, the distribution of blue (upregulated in O) and yellow (upregulated in T) coloured genes reflects the fact that a higher proportion of genes were upregulated in D39O vs D39T, compared with the O/T pairs of WCH16 and WCH43.Furthermore, the intensity of colour (both blue and yellow) is clearly greater for D39 than the other two strains reflecting the greater mean fold change in expression for differentially regulated genes.

4.2.6 Validation of differential gene expression by qRT-PCR analysis

The overall purpose of this study was to identify commonly-regulated proteins and genes between the O and T variants of the three *S. pneumoniae* strains. In order to validate key data derived from the microarray analysis above, the level of mRNA of selected genes of interest were compared using qRT-PCR (Section 2.9.4), using primer pairs listed in Table 2.1. The results of these analyses are presented in Table 4.2. The target genes were selected on the basis of having shown a significant upregulation in either the microarray data (Table 4.1) or proteomic analysis (Table 3.5) in either O or T variants of at least one of the strains. The mRNA extracts used for qRT-PCR analyses were the same as those used for the microarray analyses. The qRT-PCR data shown in (Table 4.2) were largely comparable with microarray data. For example, genes upregulated in the O variant, such as those involved in competence (*comXI*, *cglA* and *cbpD*) or expressed during stress (*ssbB*), were confirmed by qRT-PCR. On the other hand, genes shown to be upregulated in the T variant such as *phtD*, and *lctO* were also confirmed using qRT-PCR. However, there was negligible concordance between microarray and qRT-PCR data and that obtained by proteomic analysis in Chapter 3.

Table 4.2. Collective microarray, real-time qRT-PCR and proteomic data of interest.

Gene Annotation[a]	TIGR4[b]	Strain, method and fold change[c,d]								
		D39			WCH16			WCH43		
		M	Q	P	M	Q	P	M	Q	P
comX1	0014	3.59[f]	3.87	ND	-1.14[g]	1.79	ND	1.59[g]	1.45	ND
fusA	0273	-1.37[f]	-2.98	4.20	1.02[g]	1.06	ND	1.06[g]	-1.08	ND
fba	0605	1.02[f]	ND	ND	1.4	ND	-4.2	1.06[g]	ND	5.4
lctO	0715	-1.24[f]	-1.55	ND	1.08	-1.24	ND	-16.0	-47.2	ND
spxB	0730	-1.84[f]	-7.02	2.90	-1.00[g]	1.16	-1.33	-1.22[g]	-1.79	3.46
gor	0784	1.08[f]	ND	ND	1.56	ND	ND	1.15[g]	ND	-5.1
SP0846	0846	-2.71	ND	-2.5	1.12[g]	ND	ND	-1.08[g]	ND	ND
pyk	0897	-3.15	-7.91	2.12	1.08[g]	1.98	1.34	-1.10[g]	-1.67	2.28
phtD	1003	-3.73	-38.0	ND	-1.69	-2.11	ND	-1.02[g]	-2.49	ND
phtE	1004	-4.14	ND	ND	1.80	-2.53	ND	-1.08[g]	ND	ND
ppc	1068	-3.66	-7.75	ND	1.33	2.00	ND	-1.36[g]	-1.56	ND
lepA	1200	-3.73	ND	2.9	-1.33	ND	ND	1.04[g]	ND	ND
ldh	1220	-1.26[f]	-2.08	3.00	1.24[g]	-1.02	ND	1.06[g]	-1.30	ND
GroEL	1906	3.14	9.71	ND	1.55	2.76	ND	1.22[g]	-1.60	ND
ssbB	1908	25.79	ND	3.4	1.09[g]	ND	ND	2.57[g]	ND	ND
cglA	2053	9.19	9.04	ND	2.36	3.75	ND	3.12[g]	13.52	ND
malR	2112	1.47[f]	ND	ND	1.39	ND	2.2	1.16[g]	ND	ND
cbpD	2201	3.65	3.61	ND	1.64	3.32	ND	1.68[g]	-4.07	ND
purA[e]	0019	1.14[f]	-2.09	2.15	1.43	-1.04	1.58	1.39[g]	1.19	4.06
murC[e]	1521	-1.90[f]	-4.45	2.15	1.00[g]	-1.13	1.58	1.01[g]	-1.45	4.06
SP1837[e]	1837	-1.49[f]	-2.53	2.15	1.06[g]	-2.22	1.58	1.23[g]	-1.43	4.06

[a]Gene name according to KEGG (http://www.kegg.com/) website
[b]Gene number according to TIGR-4 annotation as per KEGG
[c]M denotes microarray data; Q denotes qRT-PCR data; P denotes proteomics data
[d]ND means no data available
[e]qRT-PCR was performed on these genes to identify the corresponding gene that encodes for the protein spot393 identified from proteomic analysis (Section 4.2.6)
[f]Fold change did not meet statistical significance criteria set in Section 4.2.2
[g]Fold change did not meet statistical significance criteria set in Section 4.2.3

There was only one protein spot (master spot 393) from proteomic analysis that was similarly upregulated in the O variant across all three strains. Using mass spectrometry, the protein spot produced three possible protein ID's (PurA, MurC and SP1837). When qRT-PCR was used to validate these results, the genes encoding all these proteins were shown to be upregulated in the T form, with the exception of *purA* in strain WCH43, which was upregulated slightly (1.19-fold) in the O form (Table 4.2). However, even for this gene, the fold-change was considered to be negligible

4.2.7 Competence-related genes are predominantly upregulated in O variants

S. pneumoniae is a naturally competent bacterium and currently, 23 CSP-inducible genes that are required for genetic transformation have been identified (Peterson *et al.*, 2004).These include early-induced genes such as the two-component sensor-regulator, *comDE* and the sigma factor *comX*, as well as late-induced genes such as the *cgl* family. ComCDE is regulated by CSP, which then activates the *comX* genes; *comX1* activates genes that are essential for transformation and *comX2* activates non-essential transformation genes. A range of competence-related genes were upregulated in the O variant of the three strains. Since *comX1* is turned on in response to CSP and activates downstream competence genes (Lee & Morrison, 1999; Peterson *et al.*, 2000), differences in *comX1* and *cglA* mRNA levels between O and T were quantitated by qRT-PCR to validate the microarray data.

In this study, ComE was identified as an upregulated protein on the comparative 2D-DIGE (Table 3.1) in strain D39O. Transcriptomic microarray analysis also identified several competence-related genes, including *cglA-D* and *cbpD*, which were significantly upregulated in strains D39O and WCH16O (Table 4.1). The upregulation of *cglA* and *comX1* in the O variant was verified for all three strains using qRT-PCR (Table 4.2). *comDE* has been shown to be essential for transformation and it has previously been shown that mutation of either *comD*, *comE* or *comDE* impaired the transformation ability of *S. pneumoniae* (Cheng *et al.*, 1997). Since the competence genes were shown to be differentially regulated both in the proteome and transcriptome data sets, *comD* deletion-replacement (with *erm* cassette) mutants were generated and plated for single colonies on THY+catalase agar plates. Construction of the *comD* mutations in O/T pairs for all three

strains was carried-out essentially as described for D39 *spxB* mutants (Section 3.2.3.1), using primer pairs described in Table 2.2. Since *comD* is ~1300 bases long, whilst the *erm* cassette is ~700 bp, amplification of the region flanking *comD* of erythromycin-resistant clones showing a smaller band size than their wild-type counterparts would indicate successful uptake of the *erm* OE-PCR product, and concomitant loss of *comD* (Figure 4.4). The clones were also confirmed for loss of *comD* by Sanger sequencing (Section 2.8.8).

All *comD* mutants were found to retain their parental O/T phenotypes. To further investigate transformability of *S. pneumoniae*, a chloramphenicol resistance (cml^R) cassette was inserted into an intergenic region (between SPD0024 and SPD0025), using primer pairs containing sequences from both *cml* flanking regions and the intergenic region (Section 2.6.1). Inserting the cml^R cassette into an intergenic region ensured that there would be no disruption to a gene, enabling assessment of whether non-specific recombination could on its own affect phase variation. This was first performed on strain D39, which resulted in no change in the phenotypes, and thus was not performed on the other two strains.

4.2.8 O and T variants have different growth rates

From the microarray analysis, a number of genes involved in energy metabolism, cell division and amino acid biosynthesis and acquisition were upregulated in the T variant. To investigate whether the differences observed in these genes could be reflected in a difference in growth rates between the O and T variants, growth curves were plotted for 18 hours in both C+Y and THY broths (Figure 4.5). Strains WCH16T and WCH43T appeared to grow better than their O counterparts in C+Y; by contrast, the D39O grew better in C+Y than its T counterpart. In D39, both variants appeared to initially grow at a similar rate, however, D39O grew to a much higher density. In WCH16, WCH16O had a greater log phase growth and reached a slightly higher density than WCH16. On the other hand, both WCH43O and WCH43T appeared to have a similar log phase rate-of-growth. However, WCH43O appeared to have a longer lag phase and reached a higher density compared to its T counterpart. The different growth abilities between the variants in different media may also explain to some extent, the phenotypic discrepancies between previously published studies (Table 1.1).

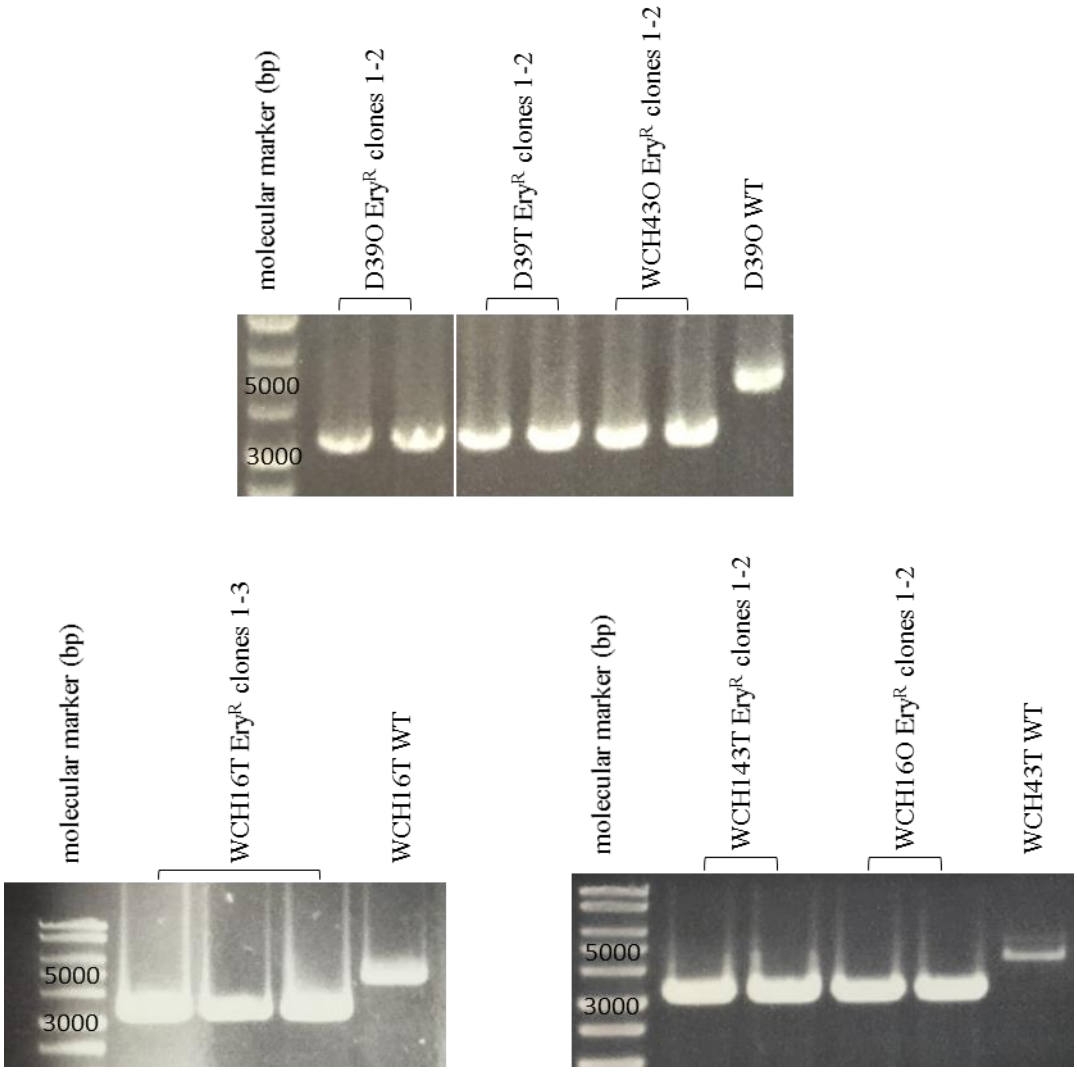


Figure 4.4. PCR products of putative *comD* clones.

PCR products of putative *comD* mutants of D39, WCH16 and WCH43 O/T pairs were analysed by agarose gel electrophorsis. Mutants were constructed using the OE method (Section 2.8.5). Regions containing *comD* were amplified (using primers "ComD Flank F" with "ComD Flank R"; Table 2.2) using WT as a control. A reduction in PCR product size is indicative of deletion/replacement of *comD* as *comD* is about 700 bp longer than the *erm* cassette. One clone from each strain was then confirmed for replacement of *comD* with *erm* cassette using Sanger sequencing (Section 2.8.8).

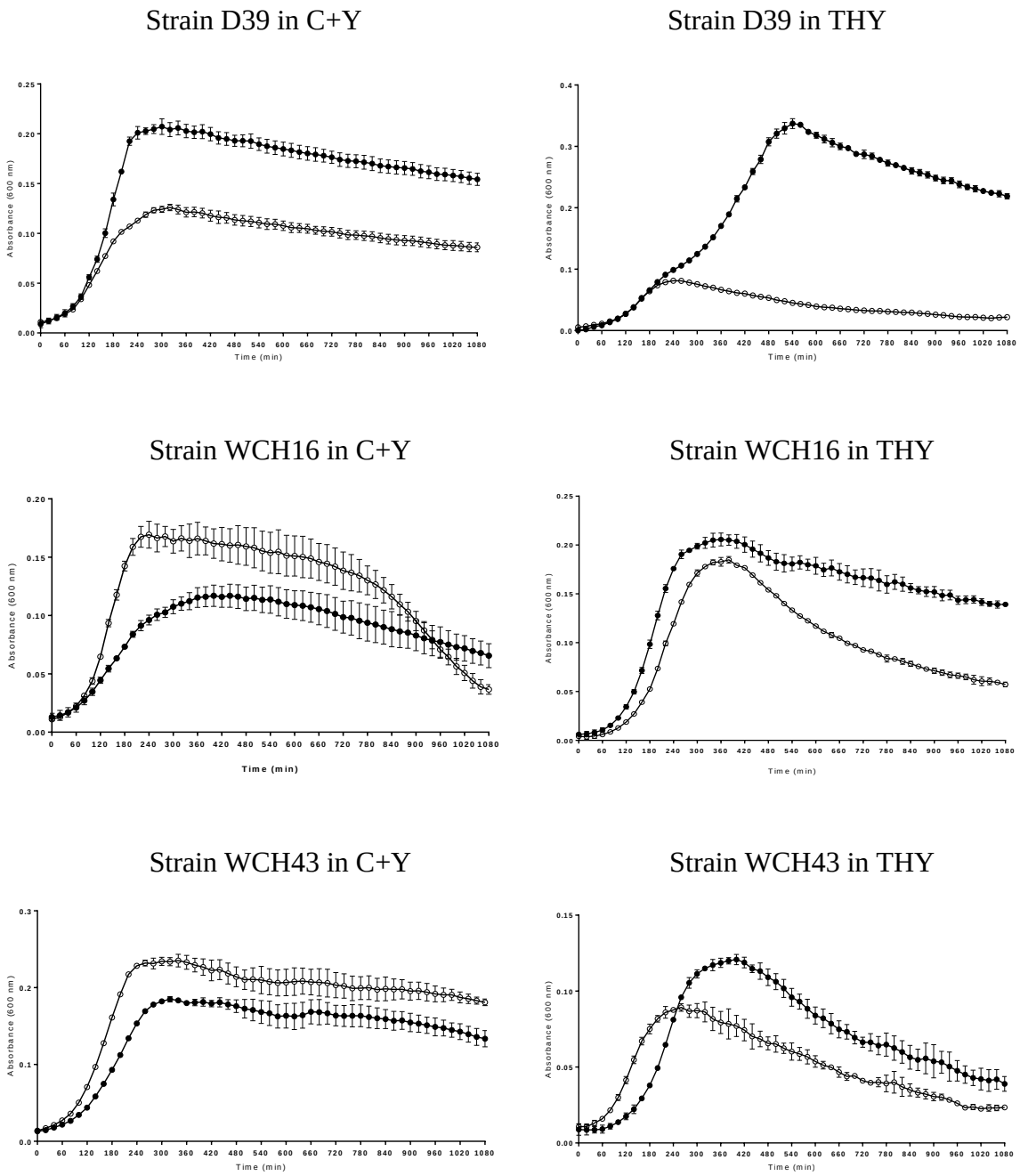


Figure 4.5. Growth curve of the O and T strains in different broths.

Bacterial growth at 37°C (defined by A_{600}) in the indicated medium was recorded over 18 hours (Section 2.2.5). Closed circles (●) indicate the O variant and open circles (○), the T counterpart for each strain. Data are A_{600} ± SEM for triplicate assays. These data are representative of two independent experiments.

4.3 DISCUSSION

Comparative whole transcriptome analysis used to investigate the differential expression of genes between pneumococcal opacity phase variants resulted in the identification of genes with similar functions associated with a certain phenotype. These genes are summarised and categorised according to their putative function in the pneumococcus (Table 4.1).

4.3.1 Differentially-regulated genes between O/T pairs

Genetic transformation by naturally competent pneumococci is mediated through a peptide pheromone signaling system. Firstly, a 41-residue precursor peptide known as competence stimulating peptide (CSP) is encoded by *comC* (Pestova *et al.*, 1996; Tomasz, 1965). The pre-peptide is then cleaved and exported by the ComAB transporter to generate a mature peptide (Hui *et al.*, 1995). The mature CSP accumulates extracellularly and then triggers the ComDE two component signal transduction system by firstly activating the histidine kinase ComD, which then phosphorylates the response regulator ComE resulting in induction of expression of early competence genes (Cheng *et al.*, 1997; Peterson *et al.*, 2004). The activation of the early competence genes provides a positive feedback loop, in which there is reactivation of *comABCDE*. Furthermore, ComE induces the expression of ComX, which encodes an alternative sigma factor, resulting in stimulation of a large group of late-phase competence genes. These include those that are required for DNA uptake and recombination, such as *ssbB* or other regulatory genes such as *cglA* (Lee & Morrison, 1999; Luo *et al.*, 2003; Peterson *et al.*, 2000).

In this study, although both proteome and transcriptome analyses did not identify exactly the same competence-related proteins or genes in a given strain (with the single exception of *ssbB* in strain D39), both analyses identified genes with similar functional properties. For example, in D39O, there was an upregulation of the *cgl* and *comX2* genes in transcriptome analysis but in proteome analysis, there was upregulation of ComE. Although different genes and proteins were upregulated, they belong to a set of genes that are essential for the pneumococci to achieve a competent state (Peterson *et al.*, 2004). Furthermore, the competence-related genes were upregulated in the O variant compared to T of both D39 and WCH16 (Table 4.1). Due to these findings, and since *comD* is a sensor,

comD mutants were constructed in all three strains and their O/T pairs to investigate whether deletion of this gene impacts on opacity phenotype. However, none of the mutants had altered opacity phenotypes. This suggests that abrogating the ability of the pneumococcus to respond to CSP does not affect its ability to switch phenotypes. Nevertheless, even though the *comD* mutation did not affect the opacity phenotype of the pneumococci, the expression of competence-related genes and proteins could have a different impact on pathogenesis of O versus T strains. Studies have also shown that the inhibition of the competent state is associated with a decrease in virulence in animal models, showing the importance of competence regulon in pathogenesis (Oggioni *et al.*, 2006; Zhu & Lau, 2011). For example, it is known that *comD* mutants are attenuated in terms of their capacity to cause bacteraemia and pneumonia in mouse infection models (Lau *et al.*, 2001). Since the O variant is usually associated with increased systemic virulence while the T variant is associated with enhanced colonisation (Weiser *et al.*, 1994), and competence has been associated with pneumococcal pathogenesis, an increase in level of competence in the O variants might support the notion that competence is causally linked with virulence. To further support this, ComE was shown to be upregulated in the O form in proteomic analysis, and this regulator has been negatively associated with colonisation fitness. A *S. pneumoniae comE* mutant has been shown to colonise better than its WT parent strain in an infant rat model of asymptomatic carriage, and an increase in *comE* transcription impaired colonisation (Kowalko & Sebert, 2008). Thus, increased ComE expression in O is compatible with the notion that O is usually associated with systemic virulence, whereas the decreased ComE expression in T is consistent with the association of the T variant with carriage.

The murein hydrolase, choline-binding protein D (CbpD), which is expressed during competence was also found to be highly expressed in the O variants of D39 and WCH16, and the T variant of WCH43 in this study. CbpD is believed to mediate cell lysis of a sub-population of non-competent pneumococci in a process called fratricide (Kausmally *et al.*, 2005; Steinmoen *et al.*, 2003; Trappetti *et al.*, 2011c). Lysis of non-competent pneumococci results in greater abundance of extracellular DNA that could be taken up by competent pneumococci to provide a selective advantage.

Pneumococcal colony opacity and competence have also been linked to ability to form biofilms. During the induction of competence via exogenous CSP, a *comD* receptor mutant was unable to form a biofilm and also had attenuated virulence in a pneumonia model, when compared to its parent strain (Oggioni *et al.*, 2006). Furthermore, only the O forms of *S. pneumoniae* are able to form biofilms, due to their enhanced ability to produce an extracellular matrix (Trappetti *et al.*, 2011b), a property dependent on the expression of *cbpD* (Trappetti *et al*., 2011c). Similar to the findings in this study, Trappetti *et al.* (2011a) found that there was a significant increase in the expression of competence genes, namely *comE* and *comX*, in a biofilm model, while there was no detectable upregulation of capsule-related genes. In contrast, Weiser and Kapoor (1999) reported that T variants had higher transformation efficiencies than their O counterparts, although expression of competence genes was not investigated. O variants have a higher CPS to TA ratio compared to their T counterparts, and an unencapsulated mutant of D39O had a higher transformation efficiency than its parent strain (Kim & Weiser, 1998). Thus, the thicker capsule may physically impede transformation of O variants, notwithstanding their higher expression of competence-related genes.

CSP-induced genes can be classified into three distinct groups – early, late and delayed. After the activation of the early competence genes that are essential for transformation such as *comABCDE* and *comX* genes, another 124 genes are activated, but only a small number of these (for example, *ssbB* and *cgl*) are actually required for transformation (Peterson *et al.*, 2004). The *ssbB, groES, groEL* and *cgl* family genes have been shown to have increased expression in the O variants in this study, as judged by microarray analysis (Table 4.1), some of which were subsequently confirmed using qRT-PCR (Table 4.2). These genes could be important for the pneumococcus during invasion of deeper tissues, as competent pneumococci are better at adapting to DNA-damaging stress than non-competent cells (Engelmoer & Rozen, 2011). The same group also found that the activation of the competent state protected against antibiotics that affect protein synthesis. SsbB also contributes to the genetic plasticity of the pneumococci by creating a reservoir of single-stranded DNA (ssDNA) by stabilising and internalising ssDNA to a high concentration (Attaiech *et al.*, 2011). Thus, having an increase expression of stress-related

genes may confer a survival advantage on the O forms over the T forms when in deeper tissues.

Differences in relative amounts of CPS and TA between O and T variants and the importance of the capsule to the pneumococcus have been described in the introductory chapter of this thesis (Section 1.4.1). In this study, the genes within the capsule loci were shown to be upregulated in the T variants of strains D39 and WCH16. This was unexpected, given the reported a higher CPS to TA ratio on O variants (Kim & Weiser, 1998). A previous study in our laboratory comparing D39O and D39T did not detect any difference in capsule gene expression by either microarray or qRT-PCR analyses, despite the O form having five-fold more capsular polysaccharide (Stroeher U. H., unpublished data). Moreover, in *S. pneumonaie*, two genes involved in sugar metabolism (*pgm* [phosphoglucomutase] and *galU* [UTP-Glc-1-phosphate uridylyltransferase]) located distant from the capsule locus have been shown to be involved in CPS production, as mutagenesis of either gene eliminated or reduced any detectable CPS (Hardy *et al.*, 2000; Mollerach *et al.*, 1998). Both Pgm and GalU are required for the synthesis of UDP-Glc, a precursor for CPS production. Hence, the increased production of CPS is not be dependent entirely on the genes within the *cps* loci, but rather may be affected by availability of precursors such as UMP and UDP-Glc.

4.3.2 Growth rates of O vs. T vary between media

Different groups attempting to elucidate the mechanism behind pneumococcal opacity phase variation have employed a range of growth media and culture conditions, complicating phenotypic comparisons between studies (Kim & Weiser, 1998; LeMessurier *et al.*, 2006; Li-Korotky *et al.*, 2009; Li-Korotky *et al.*, 2010; Mahdi *et al.*, 2008; Overweg *et al.*, 2000; Weiser *et al.*, 1996). In this study, all strains were grown in C+Y prior to RNA extraction. Numerous genes involved in amino acid acquisition and biosynthesis were upregulated in the T variant. In C+Y, the T variants showed better growth performance than O for strains WCH16 and WCH43, reaching a higher cell density. However, strain D39O grew better than D39T in C+Y. On the other hand, the O forms of all three strains exhibited better growth than their respective T counterparts in THY, as judged by maximum cell density, although WCH43O exhibited a more prolonged lag phase. It is not

certain why D39 behaved differently to the other two strains, as the microarray results indicated that in addition to upregulation of the amino acid acquisition and biosynthesis genes, genes involved in cell division were also shown to be upregulated in D39T (Table 4.1). The growth of the bacterium was assessed by measuring their optical density. However, the ability of O or T variants to grow to a certain optical density may be influenced by the light-scattering properties of the bacterial cells. Hence, A_{600} and CFU counts could potentially yield different "growth curves". It would also be interesting to compare the gene expression profiles of O and T variants in cells grown in different media (e.g. THY rather than C+Y), although this was beyond the scope of the current study.

4.3.3 Proteomic and transcriptomic analyses

Initially, this study aimed to identify proteins that were consistently differentially regulated between O and T variants in all three strains and compare these with transcriptomic profiles. However, there were very few direct correlations between protein expression (Chapter 3) and transcription described in this chapter. Nevertheless, there was some correlation in terms of the functional class of differentially regulated proteins/genes in O vs. T (e.g. competence, stress response and amino acid biosynthesis). On the other hand, there were examples of marked discordance between protein and mRNA levels, such as in SpxB, Ldh and Pyk (Table 4.2).

The lack of correlation between the data obtained using the two analytical approaches is not uncommon. In a review by Maier *et al.* (2009), it was shown that studies on *E. coli* had about 50% correlation between mRNA and protein levels. Isoforms may in part contribute to the discrepancies, as seen in the proteomic analysis of this study, where GAPDH appeared in multiple protein spots, particularly in strain WCH43. In the case of GAPDH, the activity of the enzyme in cell lysates was assayed to validate the proteomic results. The co-migration of two or more proteins on the same protein spot in the 2D gel could also affect the proteomic results, as it is usually assumed that each protein spot represents one protein. Hence, alternative confirmation methods such as quantitative Western blot analysis or biochemical assays are usually employed to confirm proteins of interest. Furthermore, there were differences in the number of protein spots selected for identification in proteomic analysis compared to the number of genes assessed in

transcriptomic analysis. It was not possible to pick and identify all protein spots from the 2D-DIGE gel. Thus, a selection of differentially regulated protein spots were identified using a pre-determined statistic. As a result, only about 15 protein spots per strain were identified, compared to the analysis of over 2000 genes detected from the microarray slide. The differences in number of identified protein spots compared to genes could be another reason for the lack of correlation between proteomic and transcriptomic analyses, as certain protein spots may have been missed. However, as mentioned previously, certain proteins and genes belonging to the same functional group were detected and these were further investigated through quantitative Western blot analysis, biochemical assays, qRT-PCR or mutational studies. In recent years, there has been rapid advancement in gene expression profiling, such as the use of RNA-sequencing (RNA-seq). While this study utilised microarray slides containing cDNA of ORFs from TIGR-4 genome (Tettelin *et al.*, 2001) and additional ORFs present in the R6 genome (Hoskins *et al.*, 2001), there may be genes that were missed from D39 and/or WCH16 as they are not present in either TIGR4 or R6. RNA-seq would be able to detect all changes and can also be used to identify small non-coding RNAs (sRNA) which have been recently identified as having a role in pneumococcus virulence (Mann *et al.*, 2012). sRNAs are short sequences in the genome which are present in the intergenic regions and do not get translated into a protein, but can control the expression of certain virulence genes. Hence, future studies involving whole gene expression profiling could be carried out using RNA-seq for a more comprehensive comparison between O/T pairs.

In order to optimise the potential for correlation between proteomic and transcriptomic data in this study, cell extracts used to these analyses (and subsequently genomic DNA) were from the same culture batch and were aliquoted into appropriate volumes and the lysates immediately stored at -80°C until required. Moreover, four biological replicate samples were prepared and analysed. Thus, the discrepancies between proteomic and transcriptomic data sets are unlikely to be due to lack of experimental precision or aberrant behavior of *S. pneumoniae* in one culture versus another.

4.4 CONCLUSION

This chapter demonstrated that there were differentially regulated genes between *S. pneumoniae* O and T variants. Although no single gene was consistently differentially expressed between O and T variants across all three strains, there were a number of genes belonging to particular functional classes that were upregulated in one variant over the other. Such examples include the competence and stress-related genes, which were also reflected in the proteomic analyses. The competence and stress-related genes appear to be important for the O variant, but do not appear to be responsible for the variation in actual colony morphology. Furthermore, the type of culture medium impacted their growth rates. Similarly, the different growth media could impact on the expression of specific genes that could account for the lack of consistency between different studies. When comparing proteome and transcriptome profiles, the limitations of such comparisons must be taken into account, including recognition that the data from the two approaches may not necessarily correlate.

Chapter 5: GENOMIC ANALYSIS OF OPACITY PHASE VARIABLE PNEUMOCOCCI

5.1 INTRODUCTION

The first *S. pneumoniae* strain to have its complete genome sequenced was TIGR4 (serotype 4) (Tettelin *et al.*, 2001), followed by the readily transformable, non-encapsulated, avirulent D39 derivative, R6 (Hoskins *et al.*, 2001). These works were performed using conventional Sanger dideoxynucleotide sequencing platforms (Sanger *et al.*, 1977; Sanger & Coulson, 1978). Recent technological advances ("next-generation sequencing") have made genome sequencing more readily accessible (Quail *et al.*, 2012).

In this study, Ion Torrent™ Personal Genome Sequencer® (Ion PGM™) technology was used. Ion Torrent is a next-generation sequencing technology that utilises DNA library preparations containing adapters, enabling attachment to and amplification on IonSphere™ Particles, which are then captured in microwells on a sequencing chip (Rothberg *et al.*, 2011) (Figure 5.1). Each well on the Ion sequencing chip contains a single template-bearing bead and each nucleotide is flushed over all wells in a sequential manner. Incorporation of a complementary nucleotide leads to a pH change that is detected by the chip and indicates the base at a specific position. Homopolymers are detected by a higher pH change proportional to the number of bases. The turn-around for sample processing can be as little as a few days.

In other bacteria, such as *N. meningitidis, H. influenzae* and *B. pertussis,* it has been shown that INDELs and SNPs can contribute to phase variation (Bart *et al.*, 2010; Hammerschmidt *et al.*, 1996; Jennings *et al.*, 1999; van Ham *et al.*, 1993; Weiser *et al.*, 1990). INDELs are more frequent in homopolymeric tracts and non-trimeric tandem DNA repeats as a result of slipped strand mispairing during replication. For example, in *N. meningitidis*, an INDEL in a cytosine homopolymeric tract results in a frameshift in the capsule polymerase gene, leading to premature termination of translation and thus a loss in capsule production (Hammerschmidt *et al.*, 1996). INDELs in homopolymeric tracts of LPS outer core transferase genes is also responsible for variation in LPS structures

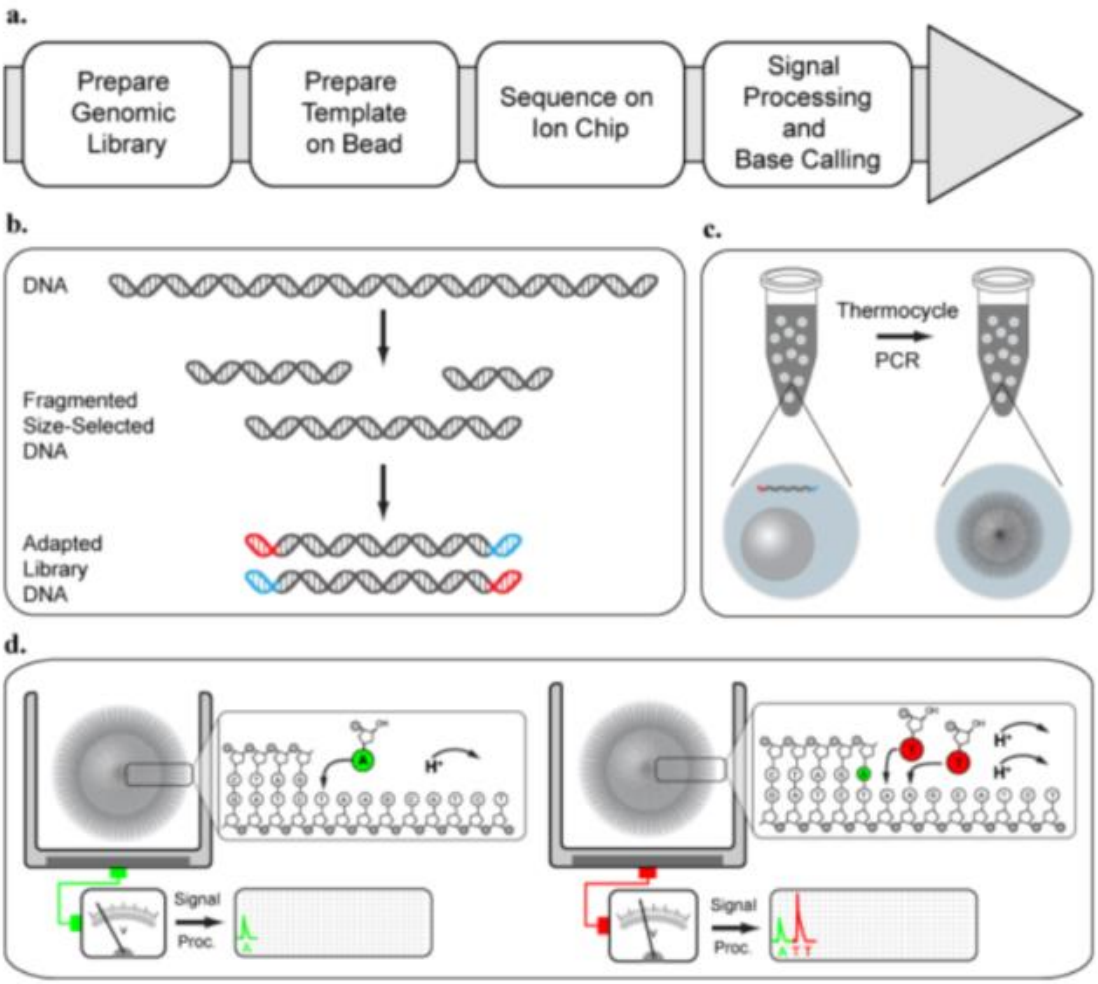


Figure 5.1. IonTorrent™ PGM workflow.

(A) The IonTorrent™ PGM workflow is a four step process beginning with library preparation from genomic DNA. (B) The DNA is fragmented, size-selected and ligated to sequence-specific adapters. (C) Template preparation involves the clonal amplification of the library fragments on beads using emulsion PCR. (D) Sequencing involves the sequential flow of nucleotides over the template-containing wells on a sequencing chip and bases are detected as a change in voltage. Figure is reproduced from Walther *et al*. (2013).

between *N. meningitidis* immunotypes (Jennings *et al.*, 1995; Jennings *et al.*, 1999). An example of SSM at non-trimeric tandem repeats is the CAAT repeat sequence locatedwithin the *lic1* locus of *H. influenzae*. Shifting the initiation of downstream genes in or out of frame, depending on the number of tandem repeats present leads to phase variation of LPS epitopes (Weiser *et al.*, 1990). On the other hand, single amino acid replacements cause by SNPs in structural genes can also have a phenotypic impact on the pathogenecity of an organism. For example, the sequencing of the *E. coli* FimH fimbrial adhesin gene revealed five amino acid polymorphisms that contribute to the ability of the strains to bind to collagen (Pouttu *et al.*, 1999), which is important for the initiation of systemic infection by *E. coli*.

Previously, in *S. pneumoniae*, McKessar (2003) sequenced 14 potentially phase variable genes in strain D39 that were selected as they had non-trimeric repeat units or homopolymeric tracts (>6 nt for A and T; and >5 nt for C and G in their ORFs). The genes examined encoded various putative surface proteins, regulatory proteins, glycosyl transferases and a methyl transferase. However, the study did not find any evidence of SSM between O and T forms in these genes.

While McKessar (2003) examined individual target genes in D39, in the present study, whole genome analysis was employed in an attempt to identify nucleotide sequence differences (SNPs or INDELs) that distinguish O and T variants of a given strain. Hence, the aim of this Chapter was to sequence the whole genomes of the phase variants of three different strains (serotypes 2, 4 and 6A) and compare these to assess whether there are any SNPs or INDELs that are consistently different between O and T variants regardless of strain background. These results will then be compared to the findings from the proteomic and transcriptomic analyses described in Chapters 3 and 4.

5.2 RESULTS

5.2.1 DNA Extraction

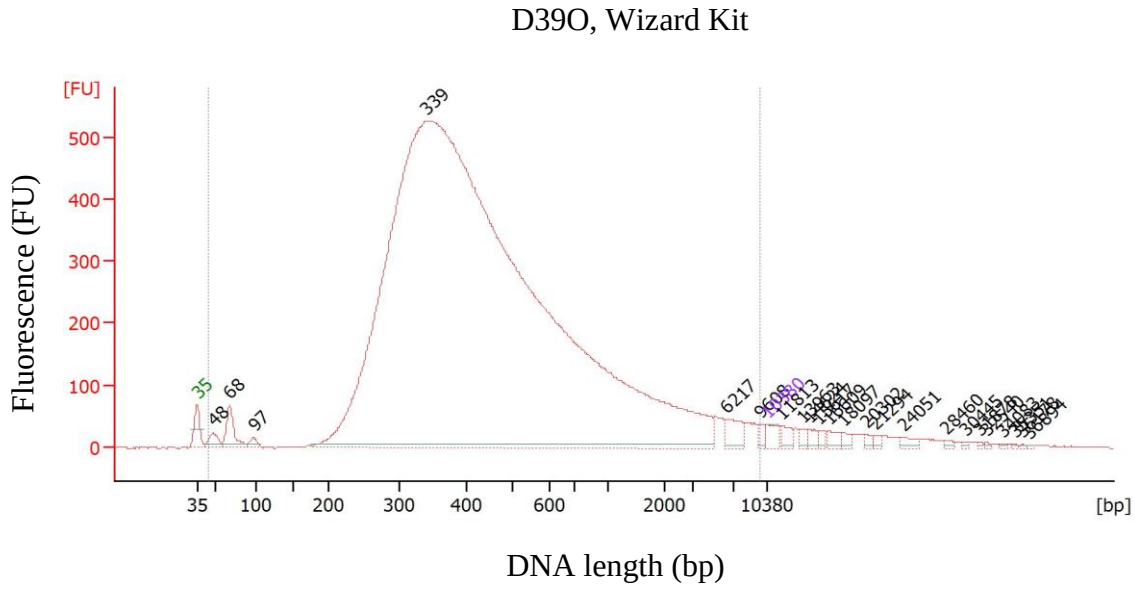
Library preparation begins with extraction and shearing of the DNA, followed by size selection of the fragments, as described in the Materials and Methods (Section 2.8.2). Pneumococcal DNA initially extracted using the Wizard™ Genomic DNA purification Kit

(Section 2.8.1) did not produce fragments of an appropriate size range and yield following these procedures, as confirmed by BioAnalyzer analysis (Figure 5.2A and B). For IonTorrent sequencing, it was recommended that the DNA fragments were between 100-200 bp, but the aforementioned DNA extraction kit yielded larger DNA fragment sizes and of a wider range. This was believed to be due to contaminating protein that remained bound to the DNA after the extraction process. Therefore, some samples were subjected to proteinase K treatment, but this did not improve the quality, size range and yield of the DNA fragments (Figure 5.2C and D). Subsequently, strains were re-cultured under the same conditions (Section 2.3) and DNA was extracted using the Qiagen Blood and Cell Culture DNA Kit, which uses an anion-exchange tip to help remove contaminants (Section 2.8.2). This method of DNA extraction produced a narrower range of fragment sizes as well as more consistent yields following the size-selection steps (Figure 5.2E and F). Hence, DNA samples extracted using the Wizard Kit were used as templates for confirmation of SNPs and INDELs, whereas DNA extracted using the Qiagen Kit was used for whole genome sequencing using the Ion PGM™.

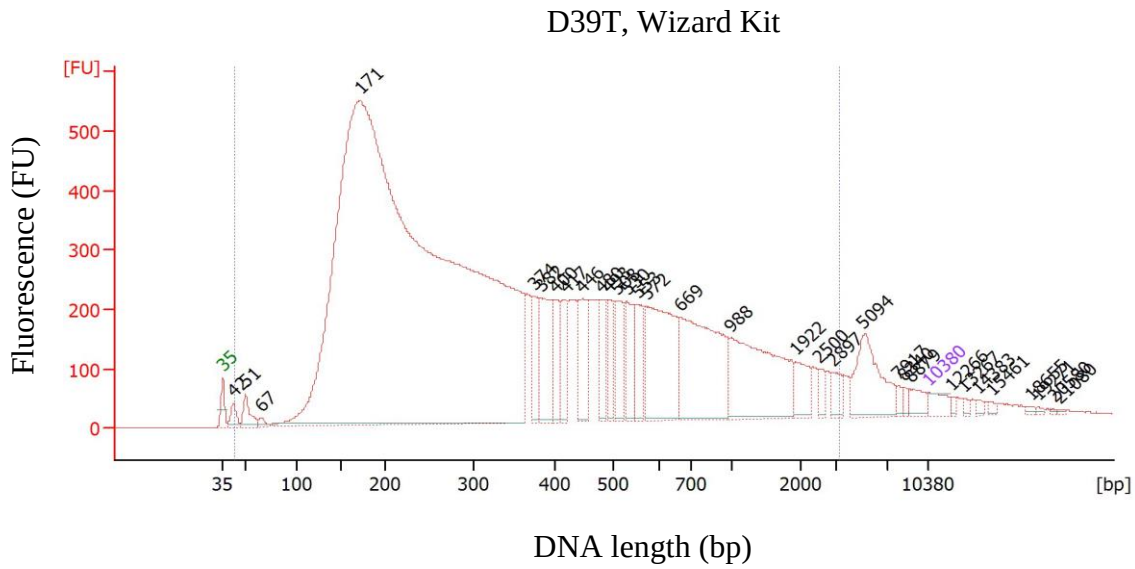
5.2.2 Optimisation of DNA library size selection

The protocol for DNA fragmentation and library preparation recommended by the manufacturer was optimised for *E. coli*. Hence, it was important to optimise the two size selection steps for pneumococcal DNA. To do this, several ratios of Agencourt® AMPure® bead volume : sample volume were tested in order to obtain suitably sized DNA fragments for sequencing on the IonTorrent™ PGM. The ratio of bead volume to sample volume determines the size of fragments that attach to the beads – the higher the ratio, the shorter the fragments that are attached. Initial size selection optimisation was performed on chromosomal DNA extracted from a serotype 1 *S. pneumoniae* strain (strain 4, provided by R. Harvey, University of Adelaide) using the Qiagen Blood and Cell Culture DNA Kit. Figure 5.3 summarises the conditions tested to optimise the DNA library size selection step for 100 bp fragments. Two AMPure® bead steps using different bead:sample volume ratios were used to obtain the correct sized DNA fragments. The aim of the first step was to remove smaller DNA fragments, whereas the second step was to remove larger DNA fragments. Apart from the ratio combinations shown in Figure 5.3 (1.80:1 followed by

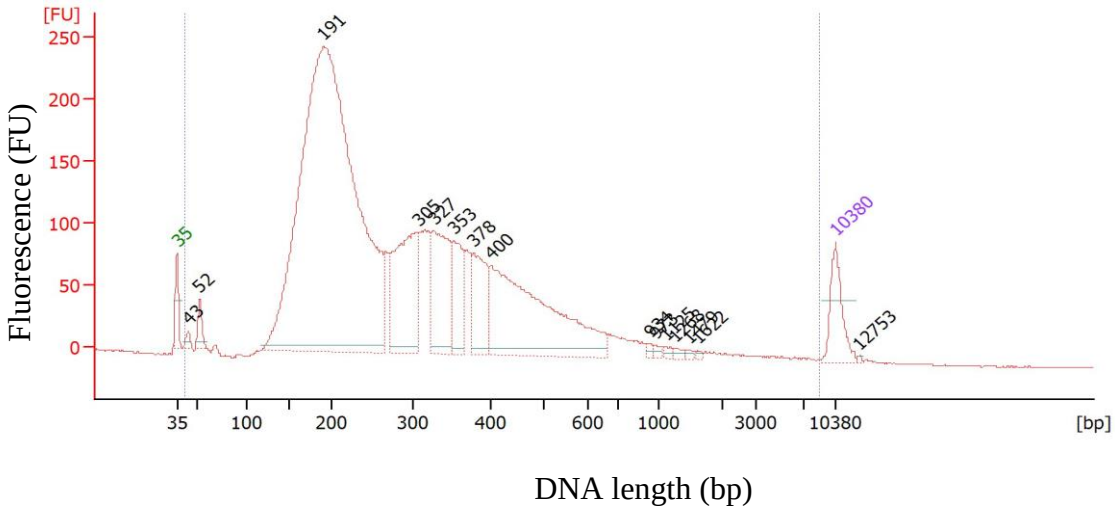
A


D39O, Wizard Kit
[FU]
Fluorescence (FU)
DNA length (bp)
[bp]

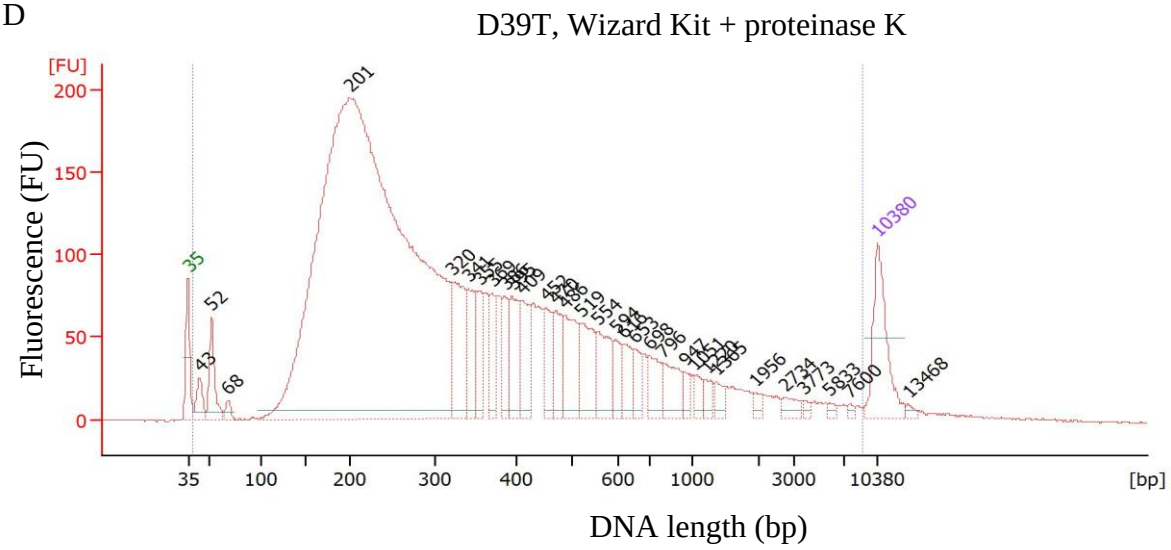
B


D39T, Wizard Kit
[FU]
Fluorescence (FU)
DNA length (bp)
[bp]

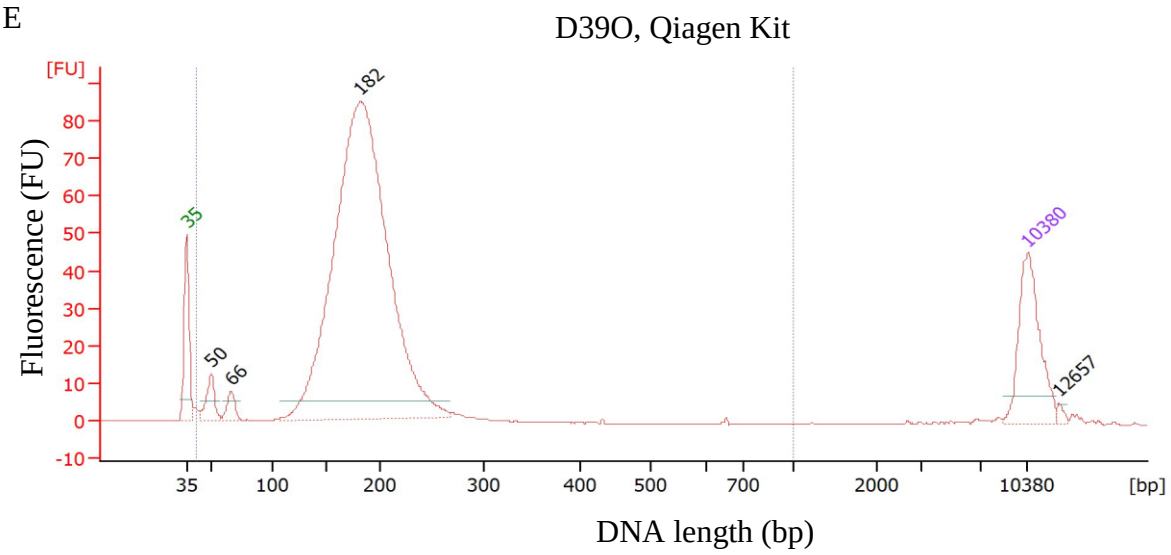
C


D39O, Wizard Kit + proteinase K
[FU]
Fluorescence (FU)
DNA length (bp)
[bp]

D


D39T, Wizard Kit + proteinase K
[FU]
Fluorescence (FU)
201
10380
35
52
43
68
13468
35
100
200
300
400
600
1000
3000
10380
[bp]
DNA length (bp)

E


D39O, Qiagen Kit
[FU]
Fluorescence (FU)
182
35
10380
50
66
12657
35
100
200
300
400
500
700
2000
10380
[bp]
DNA length (bp)

F

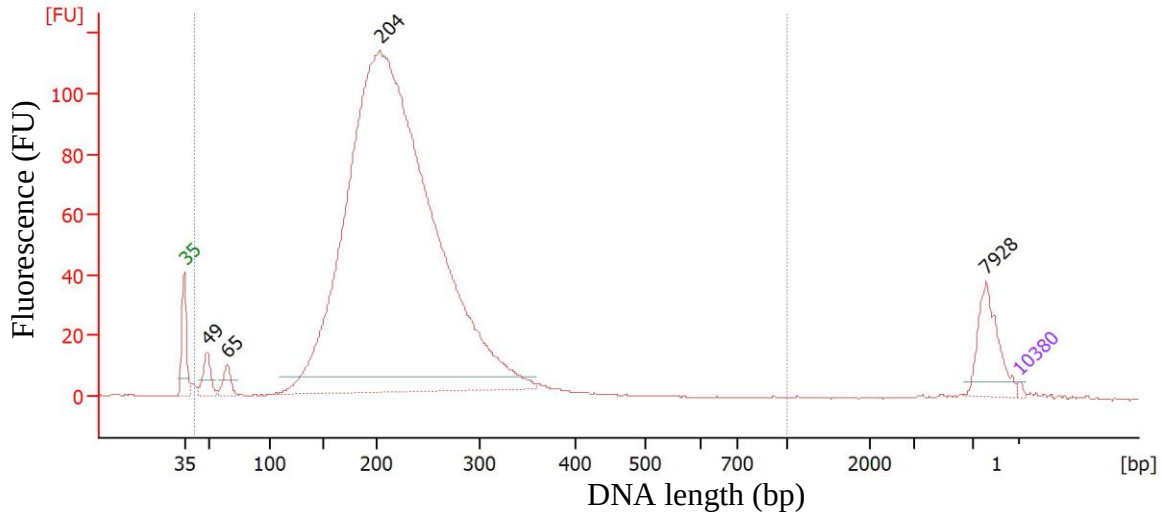
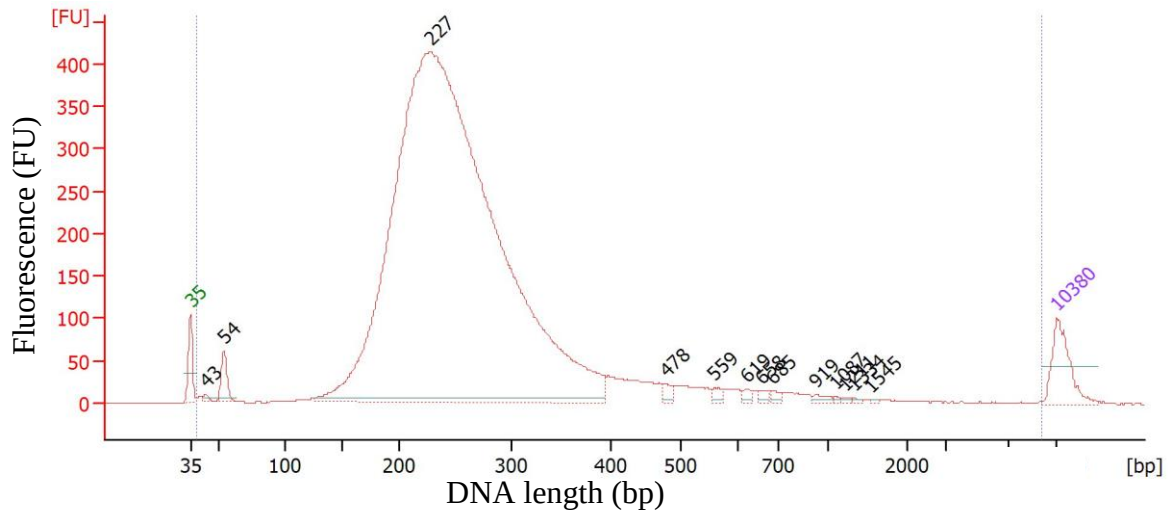

D39T, Qiagen Kit
[FU]
Fluorescence (FU)
204
35
7928
49
65
10380
35
100
200
300
400
500
700
2000
1
[bp]
DNA length (bp)

Figure 5.2. BioAnalyzer analysis of fragmented and size-selected DNA derived from various DNA extraction methods

A-D is DNA extracted using the Wizard™ Genomic DNA Purification Kit and E-F is DNA extracted using the Qiagen Blood and Cell Culture DNA Kit as described in the Materials and Methods (Sections 2.8.1). A, C and E were DNA extracts from strain D39O, whereas B, D and F were extracts from strain D39T. DNA depicted in C and D was treated with Proteinase K prior to fragmenting the DNA

A



B

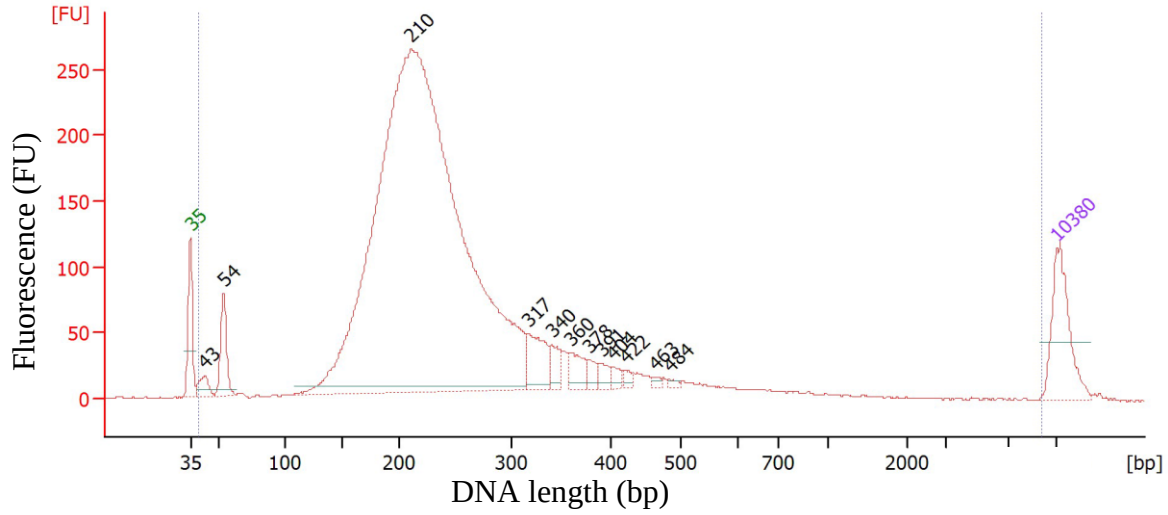


Figure 5.3. BioAnalyzer analysis of DNA fragments size-selected using different bead : sample volume ratios.

BioAnalyzer analysis of pneumococcal DNA fragmented using different bead : sample volume ratios. A, 1.80 × followed by 1.35 ×; B, 1.55 × followed by 1.20 ×. The first size selection step is after fragmentation of the DNA and the second step is after ligation of the adapters, as described in Section 2.10.

1.35:1; and 1.55:1 followed by 1.20:1), another set of ratios (1.60:1 followed by 1.4:1) was also tested, but the DNA yield dropped dramatically (under 5 ng/µl as measured by a NanoDrop spectrophotometer) and was thus rejected. Of the ratio combinations tested, 1.55:1 followed by 1.20:1 was deemed optimal. All analyses were carried out using the Ion 316^{TM} chip for whole genome sequencing, with an output of at least 100 Mb.

5.2.3 Identification of SNPs and INDELs in D39O versus D39T

The Ion Torrent read outputs for D39O and D39T were 358 Mb and 239 Mb, respectively (equivalent to approximately 160× and 110× coverage depth). Alignments of the reads to the published D39 genome [NC_008533.1 (Lanie *et al.*, 2007)] was produced by the Ion Torrent PGM server. SNPs were then filtered from these two alignments using SAMtools (Li *et al.*, 2009), which uses the statistical framework described by Li H. (2011).(Section 2.10.2.2). SAMtools takes into account the complexities of analysing SNPs, which requires looking at both read-depth and basecall quality, and assigns SNPs a quality score indicating the confidence in the alternative base call. SNPs with a low quality score (<50) were eliminated to minimise the analysis of false positives (Li *et al.*, 2009). Following this, BEDTools (Quinlan & Hall, 2010) was used to compare the lists of SNPs between D39O and D39T generated using SAMtools. This identified 11 potential variants in D39O and seven potential variants in D39T when compared to the reference D39 genome (Table 5.1A). In D39O, three SNPs (nt 423210, SPD0881 and SPD1834) were in a non-coding region, encoded a pseudogene, or did not result in an amino acid change, respectively. While it is possible that a SNP in non-coding sequence could potentially alter the expression of a nearby gene, the gene in the immediate vicinity of this SNP were not found to have phase-specific expression in the transcriptional analysis (Chapter 4). When D39T was compared to NC_008533.1, one of the SNPs (SPD0767) did not result in an amino acid change. Furthermore, four SNPs (SPD0306, SPD0616, SPD1617 and SPD1137) in D39T were found to have the same change as D39O and were thus disregarded. This left six potential SNPs between D39O and D39T, as highlighted in Table 5.1A. Of these, two were of potential interest as one was in the gene encoding SpxB (SPD0636) and the other was in a gene annotated as encoding a CPS biosynthesis protein (SPD1619/SP1837). SpxB was shown in this study to be differentially expressed between D39O and D39T on 2D-

Table 5.1. Putative SNPs (A) and INDELs (B) generated in this study comparing D39O and D39T to the *S. pneumoniae* D39 reference genome (NC_008533.1) (Lanie *et al.*, 2007)[a].

A

Reference position	Gene (SPD)	Gene annotation	NC_008533.1		D39O		D39T	
			Base	Codon (amino acid)	Base	Codon (amino acid)	Base	Codon (amino acid)
153282	0148	transporter major facilitator family protein	C	**C**GT (R)	A	**A**GT (S)	C	CGT (R)
303240	0306	*pbpX*; penicillin-binding protein 2X	A	**A**AT (N)	G	**G**AT (D)	G	**G**AT (D)
423210		non-coding region	C		G		C	
637923	0616	amino acid ABC transporter ATP-binding protein	A	**A**TC (I)	G	**G**TC (V)	G	**G**TC (V)
657517	0636	*spxb*; pyruvate oxidase	C	T**C**A (S)	T	T**T**A (L)	C	TCA (S)
752930	0740	sugar ABC transporter ATP-binding protein	G	G**G**T (G)	A	G**A**T (D)	G	GGT (G)
778495	0767	*dacc*; D-alanyl-D-alanine carboxypeptidase	C	AT**C** (I)	C	ATC (I)	T	AT**T** (I)
894389	0881	pseudogene	T		C		T	
1108236	1079	Type II restriction endonuclease	C	**C**TG (L)	C	CTG (L)	A	**A**TG (M)
1171761	1137	ABC transporter ATP-binding protein	C	CA**C** (H)	G	CA**G** (Q)	C	CAC (H)
1632063	1617	sensor histidine kinase	A	**A**CA (T)	G	**G**CA (A)	G	**G**CA (A)
1634539	1619	capsular polysaccharide biosynthesis protein	C	T**C**C (S)	T	T**T**C (F)	C	TCC (S)
1821503	1834	bifunctional acetaldehyde-CoA/alcohol dehydrogenase	C	AA**C** (N)	T	AA**T** (N)	C	AAC (N)
2032325	2055	*guaB*; inosine 5'-monophosphate dehydrogenase	G	**G**CA (A)	G	GCA (A)	A	**A**CA (T)

B

Reference position	Gene (SPD)	Gene annotation	Base		
			NC_008533.1	D39O	D39T
114145	0113	pseudogene	G--	G--	G**TA**
322245	0323	*csp2H*; polysaccharide polymerase	G-	G-	G**T**
399783		non-coding region	T-	T-	T**A**
810029		non-coding region	A**G**	AG	A-
817303	0802	S1 RNA-binding domain-containing protein	T-	T-	T**A**
885718	0871	regulatory protein	G**C**	GC	G-
1061710		non-coding region	C-	C-	C**T**
1159960	1128	polysaccharide biosynthesis protein	G-	G-	G**A**
1189845		non-coding region	T-	T-	T**A**
1507541		non-coding region	A**G**C	AGC	A-C
1686605		non-coding region	A**T**	AT	A-
1721184		non-coding region	C-	C-	C**T**
1750794		non-coding region	A-G	A-G	A**T**G
1948524		non-coding region	A-	A-	A**C**
1973329		non-coding region	G**C**	GC	G-
1994870		non-coding region	T-C	T-C	T**A**C
2038670	2059	pseudogene	G-	G-	G**A**

[a]Bases in bold indicate a change between the reference genome versus D39O and/or D39T. Highlighted genes indicate a putative SNP difference between D39O and D39T that contributes to a change in amino acid

DIGE (Table 3.1), while CPS production has been reported to differ between O/T pairs (Kim & Weiser, 1998). Accordingly, the regions containing the SNPs were analysed by Sanger sequencing (Section 2.8.8) in four biological replicates each for strains D39O, WCH16O and WCH43O and their respective T counterparts. The first SNP sequenced using primers "SpxB Flank F" with "SpxB Flank R" (Table 2.2) (cytosine/thymine) was at nt 657517 of NC_008533.1 in the *spxB* gene (nt 542 in SPD0636), changing the amino acid from serine to lysine, but the change was only present in D39 and not the other two strains (Figure 5.4). For the second SNP, primers "SP1837 seq F" and "SP1837 seq R" were used for D39 and WCH43 while the alternative reverse primer "SP1837 16R", was used for WCH16 as the sequence flanking this region is slightly different to the other two strains (Table 2.2). This gene had a base change (cytosine/thymine) at nt 1634539 (nt 652 in SPD1619), changing the amino acid from serine to phenylalanine. Further *in silico* investigation of this protein using BLAST and literature searches found that the annotation was misleading and the gene may actually be involved in teichoic acid rather than CPS biosynthesis (Denapaite *et al.*, 2012). However, Sanger sequencing also showed that this SNP only occurred between D39O and D39T and not the other strain pairs (WCH16O vs WCH16T and WCH43O vs WCH43T) in all four biological replicates (Figure 5.5).

In addition to SNPs, SAMtools also generate a list of INDEL sites. Stephen Bent (University of Adelaide, Australia) used a custom algorithm to filter out INDELs outside of homopolymeric tracts. The list of INDEL sites generated were compared between D39T and D39O and those that were present in both O and T phenotypes were discarded along with those with quality scores of <50. This reduced the list of INDEL sites between D39O and NC_008533.1 to 12 and these sites from MIRA-assembled D39O were viewed with Artemis (Rutherford *et al.*, 2000) to select DNA regions of about 90 bp spanning the INDEL. These sequences were then compared against MIRA-assembled D39T (Section 2.10.2.2). Using this method to screen for false call-outs, ten INDEL call-outs with low scores (51-92) were found to be false as the INDELs listed were not present in the reads when compared to the reference genome. The two remaining sites were found to be in intergenic regions (Table 5.1B), which were not close to any proteins or genes identified in proteomic or transcriptomic analyses (Tables 3.5 and 5.1). On the other hand, using the same method but comparing D39T to the reference genome, there were 66 putative INDEL

Figure 5.4. Sanger sequencing results of a region in SPD0636 (*spxB*) for opacity variants of D39, WCH16 and WCH43.

A 60 bp region in SPD0636 that contains a SNP, which is present at position 657517 (NC_008533.1) in D39O is shown and the changed base highlighted. The numbers following "O" or "T" indicate biological replicates.

Figure 5.5. Sanger sequencing results of a region in SPD1619 for opacity variants of D39, WCH16 and WCH43.

A 60 bp region in SPD1619 that contains a SNP, which is present at position 1634539 (NC_008533.1) in D39O is shown and the changed base highlighted. The numbers following "O" or "T" indicate biological replicates.

sites. Of these, 51 were found to be false call-outs, nine were in non-coding regions and two were present in pseudogenes, thus leaving four putative INDEL sites of interest (Table 5.1B). Two of these (SPD0323 and SPD1128) were of interest as they were located in genes annotated as being associated with CPS synthesis. These were Sanger sequenced using primers "SPD0323 F"/"SPD0323 R" or "SPD1128 F"/"SPD1128 R" (Table 2.2), but no sequence differences were found in this region between D39O and D39T.

5.2.4 Identification of SNPs and INDELs in WCH43O versus WCH43T

The read outputs for WCH43O and WCH43T were 294 Mb and 403 Mb, respectively (equivalent to approximately 130× and 180× coverage depth). These data were analysed essentially as described in Section 0 with the exception that TIGR4 (NC_003028.3) was used as the reference genome [both WCH43 and TIGR4 belong to the same clonal group, ST205 (Trappetti *et al.*, 2011b)]. Lower quality scores (<50) were also accepted as, unlike the comparison between D39 O/T pairs, there were not very many SNP call-outs (11 in D39 with quality scores of >50, but only two in WCH43 that had a quality score >50). As with the examination of D39 O/T SNPs, the WCH43 SNPs were checked on Artemis to ensure that the SNPs generated from the reads were genuine. In WCH43O versus TIGR4 there were six SNPs found, representing five regions 5.2A). However, three were disregarded because they were either within a transposase (SP1582), the SNP did not result in an amino acid change (SP1772) or the quality score and read depths were low (SP0491), making it difficult to conclusively determine whether the SNP was genuine. Furthermore, SP0491 was annotated as a hypothetical protein, but was only 51 amino acids long and appeared to have an INDEL rather than a SNP. When comparing the genome of WCH43T to TIGR4, there were 13 SNPs in six regions. Of these, when inspected with Artemis, only the five SNPs found in region SP1772 could be determined to be potentially variable. However, only one of them resulted in a change in amino acid (position 1686543; threonine in D39O versus isoleucine in D39T). The SNP present in SP0288 did not result in an amino acid change. Furthermore, the putative changes in the other genes were found to be false positives when inspected with Artemis.

Using the same method for the identification of INDELs as that used for strains D39O vs. D39T (Section 5.2.3), 92 putative sites containing INDELs were identified. However,

Table 5.2. Putative SNPs (A) and INDELs (B) generated in this study comparing WCH43O and WCH43T to the TIGR4 reference genome (NC_003028.3) (Tettelin *et al.*, 2001)[a].

A

Reference position	Gene (SP)	Gene annotation	NC_003028.3		WCH43O		WCH43T	
			base	Codon (amino acid)	base	Codon (amino acid)	base	Codon (amino acid)
115647	0114	hypothetical protein	T	CG**T** (R)	C	CG**C** (R)	T	CGT (R)
115665	0114	hypothetical protein (low score)	T	CT**T** (L)	G	CT**G** (L)	T	CTT (L)
192435	0206	hypothetical protein	G	**G**CT (A)	C	**C**CT (P)	G	GCT (A)
267083	0288	hypothetical protein	C	AA**C** (N)	C	AAC (N)	T	AA**T** (N)
1487744	1582	*IS*1167, transposase (low coverage)	A		G		A	
1682817	1772	cell wall surface anchor family protein	T	AC**T** (T)	C	AC**C** (T)	T	ACT (T)
1684614	1772	cell wall surface anchor family protein	G	TC**G** (S)	G	TCG (S)	A	TC**A** (S)
1686537	1772	cell wall surface anchor family protein	C	GC**C** (A)	C	GCC (A)	G	GC**G** (A)
1686543	1772	cell wall surface anchor family protein	T	AC**T** (T)	T	ACT (T)	C	AT**C** (I)
1690269	1772	cell wall surface anchor family protein	A	GC**C** (A)	A	GCC (A)	G	GC**G** (A)
1690272	1772	cell wall surface anchor family protein	C	AG**C** (S)	C	AGC (S)	T	AG**T** (S)

B

Reference position	Gene (SP)	Gene annotation	Base		
			NC_003028.3	WCH43O	WCH43T
40218		non-coding region	G-	G**A**	G-
119683	0117	surface protein A	G-C	G-C	G**AC**
137141	0138	hypothetical protein	C-T	C-T	C**AT**
188292	200/201	competence-induced protein Ccs4/hypothetical protein	T**CA**	T-A	TCA
322235	0347	capsular polysaccharide biosynthesis protein Cps4B	G-T	G-T	G**AT**
412334		non-coding region	T-	T-	T**A**
736070		non-coding region	A**G**	A**-**	AG
743804		non-coding region	G**C**	G-	GC
787651		non-coding region	G-	G**A**	A-
927894		non-coding region	A-	A-	AG
1026862	1090	redox-sensing transcriptional repressor rex	T-C	T**AC**	T-C
1039089		non-coding region	G-T	G-T	G**AT**
1291034		non-coding region	A**C**	AC	A-
1301236	1378	hypothetical protein	GC--	GC--	GC**AC**
1443291		non-coding region	A**G**	A-	AG
1471147		non-coding region	T-	T**A**	T-
1676885	1771	glycosyl transferase family protein	T-	T-	T**C**
1777223		non-coding region	A**C**	A-	AC
1849386	1946	transcriptional regulator PlcR	T-	T-	A**T**
2036829		non-coding region	A**C**	A-	AC
2069756		non-coding region	A**G**	A-	AG
2086555		non-coding region	G**C**	G-	GC

[a]Base in bold indicates a change between the reference genome versus WCH43O and/or WCH43T. Grey highlighted genes indicate genes that show putative SNPs between WCH43O and WCH43T that contribute to a change in amino acid and not in a non-coding region or a transposase. Blue highlighted genes indicate regions that are found in putative virulence-associated pneumococcal sRNA listed in Mann *et al.* (2012).

only two of these found between WCH43O and the reference TIGR4 strain (NC_003028.3) were of potential interest; 79 were false call-outs and 11 were in non-coding regions. The two significant INDELs were present in the genes SP1090 (redox-sensing transcriptional repressor Rex) and in sequence present in the overlapping portion of open-reading frames SP0200 and SP0201 (competence-induced protein Ccs4 and hypothetical protein), respectively (Table 5.2B). When WCH43T was compared to the TIGR4 reference strain, 41 INDEL sites were detected, but only five of these were considered to be significant, as the others were false call-outs (33) or in non-coding regions (3). One of the insertions of interest was found in SP0347 (Cps4B, insertion of an adenine (A) base at nt 711). However, when this region of WCH43T was Sanger sequenced using primers SP0347 F and SP0347 R (Table 2.2), the INDEL was not detected. As most INDELs identified were in non-coding regions, the positions of these INDELs were compared to a list of known virulence-associated sRNA in *S. pneumoniae* (Mann *et al.*, 2012). From this list, two INDELs were found in putative sRNA regions, as highlighted in Table 5.2B. These INDELs were both insertions of a base in WCH43T compared to WCH43O but these were not investigated further in this study due to time contraints.

5.2.5 Identification of SNPs and INDELs in WCH16O versus WCH16T

The read outputs for WCH16O and WCH16T were 584 Mb and 643 Mb, respectively (equivalent to approximately 270× and 290× coverage depth). Unlike the other strains, there was no published whole genome sequence for a *S. pneumoniae* strain belonging to its serotype or ST (serotype 6A, ST4966 [(Trappetti *et al.*, 2011b)]). Hence, the first step was to *de novo* assemble both WCH16O and WCH16T using MIRA. Using the *de novo*-assembled WCH16T genome sequence as a reference, WCH16T and WCH16O reads were aligned and SNPs and INDELs obtained as described in Section 5.2.3. These steps are depicted in a flowchart in Figure 5.6. Any SNPs and INDELs that were common to the two phase variants were excluded. This generated a total of 100 potential SNPs and INDELs from the WCH16O alignment, 50 of which were common to the WCH16T alignment. The remaining 50 putative SNPs and INDELs were manually checked, as described for D39O vs D39T INDELs (Section 2.10.22). Of these, 12 were found to be genuine SNPs, while the rest were false call-outs or were in intergenic regions. BLAST

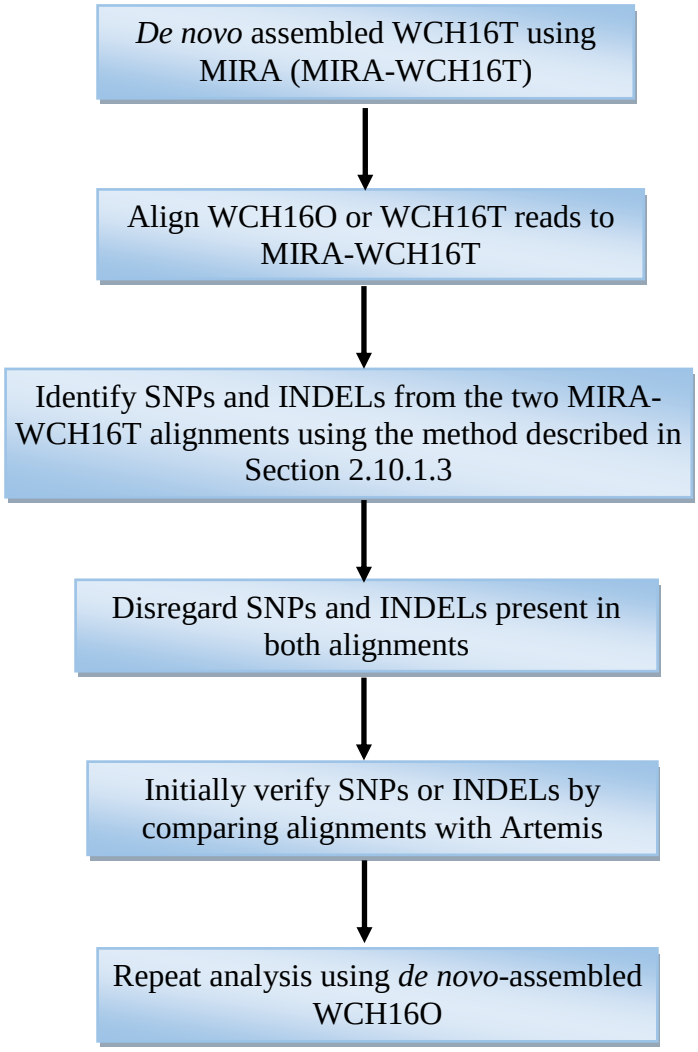


Figure 5.6. Flowchart of the generation of a list of INDELs from WCH16O and WCH16T reads.

Firstly, WCH16O reads obtained from Ion Torrent™ PGM were assembled using MIRA (Section 2.10.1.3) and then WCH16O or WCH16T reads were aligned to the *de novo* (MIRA)-assembled WCH16O in two separate alignments. This was followed by the identification of SNPs and INDELs, and those present in both WCH16O and WCH16T were disregarded as it meant that there was no difference between the variants. The other SNPs and INDELs were initially verified by visualising the alignments with Artemis. This analysis was repeated using MIRA-assembled WCH16T as the reference genome. SNPs and INDELs present in both analyses were discarded

(http://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul *et al.*, 1990) was used to assign putative functions. There were no INDELs found using this comparison. Using the same methodology, but this time using *de novo*-assembled WCH16O as the reference, a list of 78 possible SNPs and INDELs were detected and 11 SNPs and 2 INDELs were found to align to the reference used. From the two analyses, four SNPs were consistently different between the two strains, as highlighted in Table 5.3A and B. However, these were within genes that had not been previously associated with phase variation nor detected in the proteomic or transcriptomic analyses in Chapters 3 or 4. Moreover, these SNPs were not detected between O and T variants of D39 or WCH43. However, a deletion of a base of interest was found in the comparison of IonTorrent WCH16T to *de novo*-assembled WCH16O. This was of particular interest even though the deletion was not present when *de novo*-assembled WCH16T was used as a reference strain, as the deletion was found to be present in the sensor protein CiaH. The *ciaRH* two component system has been shown to repress the *comCDE* locus (Mascher *et al.*, 2003; Zahner *et al.*, 1996) and since competence genes and proteins showed a difference between the O/T phenotypes in thetranscriptome and proteome analyses, this region was sequenced to verify that it was not a false-negative. However, the Sanger-sequenced region confirmed that there was no addition of an adenine base in WCH16O.

5.2.6 Examination of Homopolymeric Tracts

Changes in the lengths of homopolymeric tracts through replicative slippage have been shown in other species to cause phase variation (Jennings *et al.*, 1999; Josenhans *et al.*, 2000). Stephen Bent (University of Adelaide, Australia) used a customised algorithm to generate a list of the positions of homopolymeric tracts within the publicly available D39 and TIGR4 genomes, as they are of a higher quality than the IonTorrent-sequenced D39 and WCH43 genomes. The frequency of SSM is known to increase with the length of the homopolymeric tract (Jennings *et al.*, 1999; Shinde *et al.*, 2003; Streisinger & Owen, 1985). Hence, this study investigated sites that contained ≥9 base repeats, because previous studies have shown that the variation in the longer homopolymeric tracts leads to phase variation (Jennings *et al.*, 1999; Kearns *et al.*, 2004; Willems *et al.*, 1990). Although phase variation has been reported to occur in shorter tracts, albeit with lower probability

Table 5.3. Putative SNP and INDEL differences between WCH16O and WCH16T identified in this study using (A) MIRA-assembled WCH16O and (B) MIRA-assembled WCH16T genomes[a].

A

Contig in WCH16O	Position	Reference (WCH16O)	WCH16T	Putative ID using BLAST
5	10150	TA	T	sensor protein CiaH
8	24031	T	C	aspartyl/glutamyl-tRNA amidotransferase subunit A
8	24048	G	A	aspartyl/glutamyl-tRNA amidotransferase subunit A
20	29707	T	G	efflux transporter, RND family protein, MFP subunit
35	5231	C	A	tRNA synthetases class II core domain family protein
57	15320	T	C	guanylate kinase
60	18545	A	G	chain A, active site restructuring regulates ligand recognition in class A penicillin-binding proteins (PBPs)
71	2336	G	T	GTP-binding protein LepA domain protein
94	9	AC	A	prophage lambdaSa2, site-specific recombinase phage integrase family
97	9246	T	G	dihydrofolate folylpolyglutamate synthetase
106	13209	G	T	amidohydrolase family protein
111	9358	G	T	flavodoxin
128	895	T	C	non-coding region (17 bp from the 5' side of CSP-1)
128	898	G	A	non-coding region (20 bp from the 5' side of CSP-1)

B

Contig in WCH16T	Position	Reference (WCH16T)	WCH16O	Putative ID using BLAST
1	3129	A	G	DNA-directed RNA polymerase, beta subunit
15	6229	C	T	guanylate kinase
29	14143	T	A	alcohol phosphatidyltransferase family protein
43	3298	C	A	FAD synthetase family protein
52	198	A	G	hypothetical protein
70	5788	A	G	putative tRNA binding domain protein
109	580	A	C	flavodoxin
113	12229	C	A	FolC bifunctional family protein
130	2652	G	T	efflux transporter, RND family protein, MFP subunit
146	1134	A	C	tRNA synthetases class II core domain family protein
150	5060	A	C	amidohydrolase family protein
164	3336	A	C	GTP-binding protein LepA domain

[a]Highlighted IDs indicates common differences between the Tables 3A and B

(Hammerschmidt *et al.*, 1996; Jennings *et al.*, 1999), it was not feasible to examine all such tracts as part of the present study due to time constraints.

In strain D39, there were 17 sites with a homopolymeric tract of ≥9 nt (Table 5.4) and 20 sites with a homopolymeric tract of ≥9 nt in strain TIGR4 (Table 5.5). It was not surprising that there was some overlap (five homopolymeric tracts) between the two strains, as highlighted in Tables 5.4 and 5.5. Although most of the homopolymeric tracts were in non-coding regions, there was one that could potentially be in the promoter region of a gene that was shown to be differentially regulated in D39O in the microarray data (Table 4.1). Using a promoter prediction tool (http://www.fruitfly.org/seq_tools/promoter.html), the homopolymeric tract was found to be in a putative promoter region for either SPD0466 (BlpT protein, protein of unknown function) or SPDO467 (BlpS). Ion Torrent sequencing detects the incorporation of each base as a change in pH that is proportional to the number of that base at that position. Thus, the inherent error in the detection of a single base accumulates as the homopolymer increases in length making the precise length of long homopolymers difficult to determine. Hence, Sanger sequencing was used to confirm the length of specific homopolymer tracts. Since a difference in gene expression was shown in the microarray data for strain D39 for SPD0466 (Table 4.1), the region containing the nine thymine bases was Sanger sequenced in D39O/T pairs but the results showed that there was no difference in the number of T bases between the phase variants. The other homopolymeric tracts were in non-coding regions or within genes that had not been associated with differences between O/T pairs by either transcriptomic or proteomic analyses, or virulence-associated pneumococcal sRNA listed by Mann *et al.* (2012). Hence, these were not further examined.

Using the *de novo*-assembled WCH16O and WCH16T genomes generated from this study, a list of regions containing ≥10 nt homopolymeric tracts was investigated, as it was more manageable than a list of homopolymeric tracts containing nine or more residues (25 vs. 114 putative differences in WCH16O and 31 vs. 120 putative differences in WCH16T). The regions containing the homopolymeric tracts were searched against the BLAST database it was found that most of the homopolymeric tracts were located within non-coding regions (Table 5.6). The 23 homopolyermic tracts that were identified in coding

Table 5.4. Homopolymeric tracts containing nine or more residues as found in the D39 genome (NC_008533.1)[a].

Position in NC_008533.1	Number of bases	Base	Gene (SPD)	Gene annotation
1985627	13	T		non-coding region
1522042	11	T		non-coding region
531108	10	T		non-coding region
1476574	10	T		non-coding region
1739198	10	A		non-coding region
1740053	10	T		non-coding region
476755	9	T		non-coding region[b]
541774	9	T	0530	amino acid ABC transporter amino acid binding protein
576900	9	A		non-coding region
816158	9	G	800/801	capsular biosynthesis protein/hypothetical protein
952572	9	T	0943	hypothetical protein
1314386	9	A	1294	hypothetical protein
1446997	9	T		non-coding region
1536755	9	C	1513	pseudogene
1686617	9	T		non-coding region
1747194	9	A	1753	subtilase family serine protease
1864576	9	T		non-coding region

[a]Highlighted homopolymeric tracts indicate that the same was found in TIGR4 (Table 5).
[b]In a putative promoter region of SPD0466 or SPD0467.

Table 5.5. Homopolymeric tracts containing nine or more residues as found in the TIGR4 (NC_003028.3)[a] genome.

Position in NC_003028.3	Number of bases	Base	Gene (SP)	Gene annotation
2100492	14	T		non-coding region
1256184	13	T		non-coding region
1593440	12	T		non-coding region
1692822	11	A		non-coding region
769503	10	A	0818	*IS*630-spni transposase
1547823	10	T		non-coding region
1850001	10	A		non-coding region
136952	9	A		non-coding region
160123	9	A		non-coding region
325015	9	A	0351	capsular polysaccharide biosynthesis protein Cps4F
505024	9	T		non-coding region[b]
565731	9	T		non-coding region
846781	9	A		non-coding region
994885	9	A		non-coding region
1378990	9	A	1465	nypothetical protein
1507165	9	T		non-coding region
1674303	9	C	1769	pseudogene
1850856	9	T		non-coding region
1867997	9	A	1954	serine protease subtilase
1976124	9	T		non-coding region

[a]Highlighted homopolymeric tracts indicate that the same was found in D39 (Table 4).
[b]In a putative promoter region of SPD0466 or SPD0467.

Table 5.6. Homopolymeric regions found in MIRA-assembled WCH16O (A) and WCH16T (B).

A

Contig	Position	Number of bases	Base	Putative ID
16ocut_c40	5225	11	T	non-coding region
16ocut_c50	6412	11	T	non-coding region
16ocut_c68	3788	11	T	non-coding region
16ocut_c118	4081	11	A	non-coding region
16ocut_c122	11412	11	A	non-coding region
16ocut_c140	375	11	T	non-coding region
16ocut_c5	8058	10	A	aminopeptidase N SP670_526
16ocut_c6	23969	10	T	conserved hypothetical protein (spr0461/sp0524)
16ocut_c7	638	10	A	non-coding region
16ocut_c34	18607	10	A	conserved hypothetical protein
16ocut_c47	11556	10	A	50S ribosomal protein L32
16ocut_c51	4	10	T	bifunctional riboflavin kinase/FMN adenyltransferase
16ocut_c66	7064	10	A	non-coding region
16ocut_c67	4254	10	T	amino acid ABC transporter, permease protein SPD0530
16ocut_c75	6211	10	A	non-coding region
16ocut_c78	6476	10	T	non-coding region
16ocut_c80	7828	10	T	GntR family regulatory protein SPN 994038_12130
16ocut_c86	13988	10	A	bifunctional riboflavin kinase/FMN adenyltransferase
16ocut_c119	3482	10	T	hypothetical protein SPAP_0194
16ocut_c129	7816	10	T	non-coding region
16ocut_c140	36	10	T	putative N-acetylmannosamine-6-phosphate 2-epimerase (*nane*)
16ocut_c146	149	10	T	non-coding region
16ocut_c152	6997	10	A	non-coding region
16ocut_c160	3527	10	T	hypothetical protein (SP1465/1294)
16ocut_c200	628	10	A	non-coding region

B

Contig	Position	Number of bases	Base	Putative ID
16tcut_c3	13137	11	T	non-coding region
16tcut_c14	149	11	A	non-coding region
16tcut_c20	14710	11	T	non-coding region
16tcut_c25	15930	11	T	non-coding region
16tcut_c37	21489	11	A	undecaprenylphosphate glucosephosphotransferase
16tcut_c57	33729	11	A	putative uncharacterised protein
16tcut_c75	45	11	A	non-coding region
16tcut_c75	10522	11	A	metal cation ABC transporter ATP-binding protein
16tcut_c91	6108	11	A	non-coding region
16tcut_c118	282	11	T	non-coding region
16tcut_c130	131	11	A	non-coding region
16tcut_c144	2317	11	A	*S. pneumoniae* partial integrative and conjugative elemental ICE6094, strain pn19
16tcut_c167	3241	11	T	non-coding region
16tcut_c171	11232	11	A	mevalonate diphosphate decarboxylase
16tcut_c172	4111	11	T	beta-glucosidase.6-phospho-beta-glucosidase/beta-galactosidase
16tcut_c218	395	11	A	non-coding region
16tcut_c228	822	11	A	*S. pneumoniae* strain jnr. 7/87 truncated putative phosphoenolypyruvate protein phosphotransferase (PTS1), BVH-11-3 and BVH-11 genes, complete CDS
16tcut_c243	38	11	T	non-coding region
16tcut_c6	9228	10	T	non-coding region
16tcut_c6	12156	10	A	arginine repressor
16tcut_c8	1479	10	A	conserved hypothetical protein
16tcut_c22	6572	10	T	non-coding region
16tcut_c23	8724	10	T	non-coding region
16tcut_c28	22960	10	A	non-coding region
16tcut_c34	10171	10	A	conserved hypothetical protein
16tcut_c62	10168	10	T	non-coding region
16tcut_c70	4617	10	A	non-coding region
16tcut_c105	9933	10	A	non-coding region
16tcut_c112	5501	10	A	amino acid ABC transporter, ATP-binding protein
16tcut_c132	5501	10	T	hypothetical protein
16tcut_c137	3575	10	T	non-coding region

regions were not present in genes of particular interest. Furthermore, differences in homopolymeric tracts that are likely to be genuine rather than sequencing artefacts should have presented in both Tables A and B, but none of the 23 homopolymeric tracts were consistent between tables.

5.3 DISCUSSION

This chapter has described the sequencing and comprehensive genomic comparison between three serotypically different O/T pairs. Using Ion PGM^{TM} to sequence the strains produced a read output of at least 110× genome coverage. SNP and INDEL differences between O and T variants were then identified through bioinformatic analysis, with particular emphasis on regions correlating with the transcriptomic or proteomic analyses (Chapters 3 and 4). Sanger sequencing was used to confirm results where appropriate. However, the SNPs between O/T pairs identified were found to be strain-specific, while all the INDELs were found to be not genuine. All but one of the homopolymers investigated were distant from any of the genes that were found to be differentially expressed in the transcriptomic data (Table 4.1). Moreover, none were close to, or within genes encoding differentially expressed proteins identified using protemomic analysis (Table 3.5).

SpxB is a protein that has been demonstrated to be differentially expressed between the D39 O/T pairs both in this study (by proteomic analysis [Table 3.1] and confirmed using qRT-PCR [Table 4.2]) as well as in other studies (Mahdi *et al.*, 2008; Overweg *et al.*, 2000). In addition to differences in protein and gene expression levels, one of the SNPs of interest identified in this chapter was found in *spxB* in D39 (Table 5.1A). It is not known whether this amino acid substitution (leucine in O versus serine in T) would impact the activity of the protein itself. Furthermore, when this region was sequenced in WCH16 and WCH43, the change in nucleotide between O/T pairs seen in D39 was not present (Figure 5.4). SpxB is clearly important for the pathogenicity of the pneumococcus, especially during establishment of nasopharyngeal colonisation (LeMessurier *et al.*, 2006; Pericone *et al.*, 2000; Regev-Yochay *et al.*, 2007; Spellerberg *et al.*, 1996). The production of H_2O_2 was initially thought to be important as cell death at stationary phase leading to a decrease in cell biomass has been attributed to increased H_2O_2 accumulation (Regev-Yochay *et al.*, 2007), which could have explained the colony morphology differences observed on a clear agar plate between O/T pairs. However, the *spxB* mutagenesis study in the D39 O/T pair

undertaken in Chapter 3 (Section 3.2.3) demonstrated that the reintroduction of *spxB* derived from either D39O or D39T into D39OΔ*spxB* or D39TΔ*spxB* did not alter their colony phenotypes from their respective parent WT strains. Therefore, even though there is SNP in *spxB*, at least in D39, mutagenesis study indicate that SpxB does not appear to be responsible for pneumococcal phase variation. Therefore, no further experiments in relation to SpxB was carried-out, as it is beyond the scope of this thesis.

Another SNP of interest identified in strain D39 was in SPD1619, annotated as a capsular polysaccharide biosynthesis protein (Section 5.2.4). However, this annotation is misleading, as the gene is not part of the serotype 2 *cps* locus, and the genes upstream encode a glycosyl transferase family protein (SPD1620) and an ABC transporter ATP-binding protein/permease (SPD1621). The R6 ortholog Spr1654 has been previously identified as an aminotransferase with a possible role in TA biosynthesis (Denapaite *et al.*, 2012). Furthermore, when the mRNA levels of SPD1619 (SP1837) were compared using qRT-PCR, there was an upregulation of this gene in the T variants of all three strains ranging from 1.43-fold (WCH43T) to 2.53-fold (D39T) (Table 4.2). The significance of TA levels in phase variants was reported in pneumococcal strains in types 6A, 6B and 18C by Kim and Weiser (1998), who showed that the more virulent O form had a higher ratio of CPS to cell-associated TA compared to the less virulent T counterpart. However, MS analysis of the protein spot for SPD1619 (SP1837) also identified peptides from adenylosuccinate synthetase (SPD0024) and UDP-N-acetylmuramate-L-alanine ligase (SPD1349) (described in Chapter 4). This could be a co-migration issue whereby proteins of the same mass and pI migrate at the same rate and thus, when the protein spot is picked and identified, it contains more than one possible protein ID. However, since these SNPs were only found in strain D39, they are unlikely to be a major determinant of pneumococcal colony opacity phase variation, but rather, a strain-specific trait.

Since there was no publicly available genome sequence for strain WCH16 (serotype 6A), the reads obtained from Ion PGM^{TM} were *de novo*-assembled using the MIRA program for both opacity variants. Since the SNPs or INDELs would be present in different contigs in WCH16O and WCH16T, the sequences of WCH16T obtained from IonTorrent^{TM} PGM were firstly compared to MIRA-assembled WCH16O and then confirmed vice versa (that is, comparing IonTorrent WCH16O to MIRA-assembled

WCH16T). *De novo* assembled Ion Torrent data will inevitably contain more errors compared to a finished genome. This undoubtedly contributed to the many false-positive SNPs and INDELs. Thus, it is important to confirm putative SNPs and INDELs using an alternative method such as Sanger sequencing. There were only four putative differences that were common to both lists, all of which were SNPs (Table 5.3), but none of these were investigated further as they did not appear to be related to virulence or be part of the *cps* locus, competence or genes/proteins that were detected in proteomic or transcriptomic analyses.

In order to determine which INDELs and homopolymeric tracts would be confirmed with Sanger sequencing, the positions of these nucleotide changes in relation to the genes associated with the proteomics data, as well as the putative promoters of the genes identified during transcriptomic analysis, were assessed. This is because the presence of a SNP or INDEL within a gene, in contrast to a promoter, is unlikely to cause a difference in the amount of mRNA transcript, but it may affect translation and thus a difference in protein level would be observed in proteomic analysis. Using these criteria, only one homopolymeric tract identified in both D39 and WCH43, which was associated with a gene (SPD0466/SP0524) upregulated 3.51-fold in D39O (Table 4.1), but with no significant changes in the other two strains, was sequenced. However, there was no difference in the number of bases in this tract between the variants of strain D39 that were sequenced. Hence, this was not considered to be responsible for the change in colony opacity phenotype, and was not sequenced in the other strains. However, since the publicly available genomes were used to identify homopolymeric tract sites for D39 and WCH43, it is possible, but unlikely there may be unidentified homopolymeric tracts due to genetic differences between the reference genomes and the genomes of the strains used in this study.

A number of the INDELs and homopolymeric tracts identified in this study were present in non-coding regions. These regions can contain small non-coding RNAs (sRNAs) which have an important role in controlling gene expression and can affect virulence in prokaryotes (Papenfort & Vogel, 2010; Waters & Storz, 2009). Furthermore, they can also target virulence gene expression at both transcriptional and post-transcriptional levels (Mangold *et al.*, 2004). In the pneumococcus, nearly 90 putative sRNAs have been

identified, and most of these have important global and niche-specific roles (Mann *et al.*, 2012). Mann *et al*. (2012) identified these sRNAs in the reference genome TIGR4 and thus, the INDELs in the non-coding regions of a strain from the same ST, WCH43, were compared. This identified two potential INDELs in WCH43 that were located within the listed sRNAs that may have an impact in pneumococcal pathogenesis, with one of them affecting the lungs in particular. However, these were not investigated in this study due to time constraints, but can be investigated in future studies.

From this study, the SNP call-outs from the sequenced strains appeared to be genuine (confirmed by Sanger sequencing), but the INDELs were all false-positives. This high level of INDEL errors appears to be a limitation of the Ion PGM™ (Bragg *et al.*, 2013; Elliott *et al.*, 2012; Quail *et al.*, 2012; Yeo *et al.*, 2012) and is more prominent in, but not limited to, homopolymeric tracts (Elliott *et al.*, 2012; Ross *et al.*, 2013). Hence, in this study, call-outs of interest were sequenced using an alternative platform, Sanger sequencing, as an additional verification method. However, it is possible that there may have been INDELs that have not been identified using Ion Torrent sequencing. The error in calling-out INDELs in homopolymeric tracts is not surprising as the Ion PGM™ identifies the nucleotide present by measuring minute changes in pH which is converted to a voltage signal proportional to the number of molecules of the given nucleotide incorporated during each cycle. The longer the homopolymeric tract that is present, the greater the potential for error when converting voltage change to number of nucleotides incorporated. Bragg *et al.* (2013) reported that the INDEL false-positive calls are increased significantly (more than 30%) after about six homopolymeric residues and that deletions are the dominant type of error. Although Sanger sequencing is the pioneer in DNA sequencing and has been known to be reliable for short DNA sequences (up to about 1 kb), this is not ideal for whole genome sequencing due to the cost and time involved.

Two other commonly available next-generation sequencing platforms include Illumina sequencing by synthesis (Bentley *et al.*, 2008) and Pacific BioSciences (PacBio) single-molecule real-time sequencing (SMRT) (Eid *et al.*, 2009; Flusberg *et al.*, 2010). Illumina sequencing utilises fluorescently-labelled nucleotides which are added in a sequential manner into flow cell surfaces containing the DNA fragments. After each nucleotide incorporation, the emitted fluorescence is detected by a camera and the fluorescent-label

then cleaved, to allow the incorporation of the next nucleotide. Hence, the sequence of bases in the DNA fragment in each flow cell surface is detected one base at a time. On the other hand, PacBio employs SMRT sequencing in a zero-mode waveguide (ZMW) (Levene *et al.*, 2003), a structure used to detect minute amount of light emission during the incorporation of a single nucleotide incorporation. The DNA polymerase at the bottom of the ZMW incorporates a fluorescently-labelled nucleotide to the DNA template during which time the fluorescent label is excited and the light captured by a sensitive detector. After incorporation, the fluorescent dye is cleaved off and diffuses away. The whole process repeats, emitting a sequential burst of light representating the nucleotides, building the DNA sequence. When Ion PGM™, Illumina (MiSeq and HiSeq) and PacBio RS sequencing technologies were compared, Illumina was found to have the lowest error rate (<0.4%) while Ion PGM™ was second (1.78%) and PacBio came last (13%) (Quail *et al.*, 2012). Furthermore, the same group also found that Illumina had the highest number of error-free reads (76.45%), followed by Ion PGM™ (15.92%), while PacBio did not produce any error-free reads. The lack of error-free reads from PacBio is not unexpected, as PacBio produces the longest read length (up to 1500 bases at the time of publication) compared to Illumina and Ion PGM™, which produced about 150-200 base reads at the time of publication (Quail *et al.*, 2012). Thus, the long reads obtained by PacBio sequencing makes it more suitable for *de novo* sequencing than the other platforms. When it comes to SNP callings, Ion PGM™ performed best, while Illumina was best for INDEL callings (Laehnemann *et al.*, 2015). All platforms displayed issues when sequencing homopolymeric tracts (Laehnemann *et al.*, 2015; Quail *et al.*, 2012; Ross *et al.*, 2013). However, Illumina MiSeq had the highest limitation for the number of bases before an error is detected (up to 20 nt), whereas, Ion PGM™ could not sequence any homopolymeric tracts longer than 14 bases, and was unable to correctly predict the number of bases over 8 bases (Quail *et al.*, 2012). On the other hand, PacBio error rates (similar to the rate of detecting 8 or more bases using Ion PGM™) stayed the same when dealing with homopolymeric tracts, regardless of its length (Laehnemann *et al.*, 2015). This highlights the importance of using multiple sequencing methods as each has their advantages and limitations.

The results of the work carried out in this Chapter indicate that changes at DNA sequence level between O/T pairs are by-and-large unique to a given strain background.

There were no DNA changes detected that were consistently associated with a given phenotype in all strain lineages. This is consistent with the findings of both transcriptomic and proteomic levels where changes were individually validated, but were strain-specific.

Chapter 6: IMPACT OF PNEUMOCOCCAL EPIGENETIC DIVERSITY ON VIRULENCE

6.1 INTRODUCTION

The work described in Chapters 3, 4 and 5 of this thesis attempted to provide insights into the molecular basis for pneumococcal opacity phase variation using proteomic, transcriptomic and genomic analyses. However, these approaches failed to identify a single unifying molecular feature that was common to the respective O or T variants of the serotypically different strains tested. During the progress of these studies, an opportunity arose to participate in an international collaborative study of a pneumococcal Type I restriction-modification (R-M) system that underwent recombination impacting on target specificity. This enabled us to address the hypothesis that opacity phase variation could be influenced by epigenetic changes.

Epigenetics is the study of heritable alterations to the phenotype, but not to the genotype. An example of such alteration is DNA methylation, a heritable event that is characterised by the binding of a DNA methylase and DNA-binding protein to a DNA sequence with an overlapping target methylation site. This methylation can then regulate the expression of genes in that organism. An example of DNA methylase is the enzyme deoxyadenosine methlytransferase (Dam), which specifically targets the adenine base of 5'-GATC-3' sequences in the genome (Geier & Modrich, 1979; Marinus & Morris, 1973). In *E. coli*, Dam-dependent methylation is required for phase variation of the pyelonephritis-associated pilus (*pap*)-like family of fimbrial operons, an outer-membrane protein (Ag43) involved in autoaggregation and formation of biofilms, and also in the variation in expression of certain other *E. coli* virulence genes (Blyn *et al.*, 1990; Correnti *et al.*, 2002; Henderson & Owen, 1999; van der Woude & Baumler, 2004). Methylation of the adenine residue of the GATC sequence prevents the binding of the regulatory protein, thus resulting in "phase-ON" state, and vice versa for "phase-OFF". During DNA replication, competition between Dam and the regulatory protein of a given gene for the unmethylated GATC sites leads to phase variation (Low *et al.*, 2001). For example, in the extensively studied regulation of the *pap* operon, the leucine response regulatory protein

(Lrp) binds to one of the two binding sites in the *pap* regulatory region. When the GATC site distal to the *papBA* promoter ($GATC^{dist}$) was fully methylated, Lrp bound near this site inhibits the transcription of *pap* (Braaten *et al.*, 1994; Weyand & Low, 2000). Conversely, if the GATC site proximal to the *papBA* promoter ($GATC^{prox}$) was methylated, Lrp bound near this site would result in the transcription of *pap*, hence the "ON" phase. Another phase variable protein in *E. coli* is the outer membrane protein Agn43. Like *pap, agn* phase variation is Dam-dependent and requires the binding of the oxidative stress response regulatory protein (OxyR). However, unlike *pap* phase variation, which requires the binding of Lrp to one of the two GATC sites in the regulatory region of *pap*, the binding of OxyR to the three GATC sites in the *agn43* operator represses the transcription of *agn* (Haagmans & van der Woude, 2000; Henderson & Owen, 1999). In organisms such as *Salmonella enterica* serovar *Typhimurium* (Giacomodonato *et al.*, 2014) and *Haemophilus influenzae* (Watson et al., 2004), the lack of Dam-methylation increases the virulence of the strains, but this is not universal, as virulence of other bacteria such as *Shigella flexneri* is attenuated by the lack of Dam-methylation (Honma *et al.*, 2004).

6.1.1 Restriction Modification Systems

Restriction-modification (RM) systems are important components of prokaryote defences against incoming viral or other foreign DNA, and preliminary description of this phenomenon occurred in the 1950's (Bertani & Weigle, 1953; Luria & Human, 1952). RM systems have been classified into four (I-IV) main types, according to their subunit composition, sequence recognition, cleavage position, co-factor requirements and substrate specificity (Roberts *et al.*, 2003).

Type I systems are the most complex, containing three subunits encoded by host-specificity-determinant (*hsd*) genes; the *hsdR* gene is required for restriction activity, the *hsdM* gene encodes the methylase, and *hsdS* determines DNA sequence specificity (Boyer & Roulland-Dussoix, 1969; Hubacek & Glover, 1970). Type II RM systems on the other hand, comprise of a pair of enzymes (restriction endonuclease [RE] and methyltransferase [Mod]). The REs cleave DNA at a fixed position determined by their recognition sequence (Wilson, 1991) and hence are commonly used for recombinant DNA manipulations (Roberts, 1990). Type III RM enzymes comprise of two proteins (R and M) that are required for DNA recognition and cleavage (Hattman, 1964; Scott, 1970). Like Type I RM

enzymes, the Type III RM enzymes require ATPase activity for DNA cleavage (Szczelkun, 2011). Furthermore, Type III RM enzymes recognise short, non-palindromic sequences that may only be methylated on one strand, resulting in hemi-methylated progenies (Meisel *et al.*, 1992). A Type IV RM system was proposed by Janulaitis *et al*. (1992) as having similar characteristics to Type II enzymes, but differing in that the endonuclease is fused to a methyltransferase and the activity is stimulated by S-adenosyl methionine (AdoMet) (as opposed to a requirement for Mg^{2+} for the endonuclease but not methylase in Type II systems).

It has been recently shown that phase variation associated with type III RM enzymes, encoded by the *mod* gene, can regulate the expression of multiple genes. This genetic system has been coined a "phasevarion" (phase-variable regulon) (Srikhanta *et al.*, 2005). The first reported example of a phasevarion was identified by microarray expression analysis comparing a WT *H. influenzae* strain expressing *modA1* with a *modA1* knockout mutant (Srikhanta *et al.*, 2005). It was found that SSM of the tetranucleotide (5'-AGTC-3') repeat unit within the *mod* ORF determined the expression of this gene. In addition, Srikhanta *et al.* (2005) showed that the *modA1* knockout mutant expressed the same phenotype as a strain with an out-of-frame *modA1* gene. Microarray analysis comparing these mutants revealed a number of genes (including genes encoding two outer membrane proteins and heat-shock proteins) that were increased in *mod* mutants. Since then, phasevarions have also been described in *Neisseria* species (Srikhanta *et al.*, 2009) and *H. pylori* (Srikhanta *et al.*, 2011), suggesting a role for this mechanism in phase variable expression of multiple genes in several bacterial pathogens.

6.1.1.1 Pneumococcal RM Systems

In the pneumococcus, Type I, II and IV RM systems are present (Hoskins *et al.*, 2001; Tettelin *et al.*, 2001). At the time of the commencement of this project, only the DpnI/DpnII Type II RM system had been described in detail in the pneumococcus (Lacks *et al.*, 1986). In a population of *S. pneumoniae*, the cells would contain one of the two RE alleles, DpnI or DpnII (Muckerman *et al.*, 1982), which act on the methylated sequence 5'-G^{me}ATC-3' and the unmethylated sequence 5'-GATC-3', respectively (Lacks & Greenberg, 1975; Lacks & Greenberg, 1977). The purpose of the DpnI/DpnII system in the pneumococcus appears to be restriction of phage infection. Hence, if a *S. pneumoniae*

population is infected by an unmethylated dsDNA phage, isolates containing the DpnI system are not able to restrict the phage, resulting in lysis and release of the phage progenies. On the other hand, uptake of the unmethylated phage progenies by pneumococci with DpnII systems are able to survive due to their ability to restrict the unmethylated phage dsDNA (Lacks *et al.*, 1986). Conversely, if a population of pneumococci are attacked by methylated dsDNA phage, DpnI isolates would survive, whereas, DpnII-infected isolates would lyse. Having a population with mixed DpnI and DpnII systems has a protective effect on the pneumococcus ensuring its survival in a phage-containing environment.

This chapter describes investigation of a pneumococcal a Type I RM System, SpnD39III, which contains three co-transcribed genes: *hsdR, hsdM* and *hsdS*. The locus also contains a separately transcribed Cre tyrosine recombinase gene (*creX*) and two truncated/silent *hsdS* genes on the opposite strand downstream of the *hsd* operon (Loenen *et al.*, 2014; Roberts *et al.*, 2010). The actively transcribed *hsdS* gene contains two target recognition domains (TRDs) which share inverted repeats with the truncated *hsdS* genes. Recombination between these inverted repeats is thought to be enabled by the CreX recombinase and rearrangement of this locus leads to the generation of alternative *hsdS* variants which have different specificities (Figure 6.1). Thus, the methylation pattern throughout the genome will depend on which partial *hsdS* allele is being expressed. Polymorphisms in the *hsdS* genes, as described above, have also been reported in *Bacteriodes fragilis* and *Mycoplasmas pulmonis* (Cerdeno-Tarraga *et al.*, 2005; Dybvig *et al.*, 1998). Interestingly, the microarray data presented in this study (Table 4.1) showed that there is a 2-2.5-fold upregulation of *hsdS* (Spd0453) and *hsdM* (Spd0454), respectively, in strain D39T compared to D39O. This chapter describes analysis of the recombination between the *hsdS* genes, each encoding a different target specificity, and their influence on opacity phenotypes, virulence in a mouse model and the expression of certain virulence- and cell-associated proteins. Furthermore, the frequency of alternative SpnD39III alleles were compared between the opacity phase variable strains, D39O and D39T, before and after infection of mice.

6.2 RESULTS

6.2.1 Colony opacity phenotypes of D39WT and SpnD39IIIA-F variants

The D39 SpnD39III variants used in this study express one of the six possible *hsdS* variants (Figure 6.1) and are characterised by a 'locked' *spnD39III* allele as described in Manso *et al.* (2014). In brief, mutants were constructed by deleting the truncated *hsdS* genes and selecting one strain with a 'locked' *spnD39III* allele for each of the six possible variants. The absence of other mutations was confirmed by whole-genome sequencing of each mutant. Single-molecule real-time (SMRT) sequencing and methylome analysis (Fang *et al.*, 2012) were used to identify N6-adenine methylation targets for each of the locked variants, and methylome data confirmed the methylation (Manso *et al.*, 2014).

It was noted that certain *spnD39III* allele types had an impact on the colony opacity of the pneumococcus when assessed on catalase-supplemented THY agar plates (Section 2.2.8). SpnD39IIIA was characterised by its O-like colonies, whereas, SpnD39IIIB displayed predominantly T-type colonies (Figure 6.2). The other SpnD39III variants as well as the D39WT had mixed populations of O and T colonies, with SpnD39IIE and SpnD39IIIF yielding mostly O colonies (100% and 96%, respectively) , SpnD39IIID yielding 59% O colonies and SpnD39IIIC yielding 25% O colonies (Manso *et al.*, 2014).

6.2.2 Capsule expression by SpnD39III variants

Analysis of the gene expression of the SpnD39IIIA-D mutants by RNA-seq identified differences in expression of certain genes, including *luxS* and the capsular polysaccharide serotype 2 operon *cps2* (Manso *et al.*, 2014), both of which were downregulated in SpnD39IIIB. To validate this finding, the total CPS production by the variants was quantitated using a uronic acid assay (Section 2.7.4), since uronic acid is a component of the pneumococcal serotype 2 CPS repeat unit (Kenne *et al.*, 1975). Strain SpnD39IIIB had the lowest uronic acid content compared to the other strains tested (Figure 6.3). Strain SpnD39IIIC also showed a similar level of uronic acid content to SpnD39IIIB, but this was not significantly different to the WT. However, the other strains (SpnD39IIIA and SpnD39IIID-F) appear to have similar amounts of uronic acid to D39WT. This was

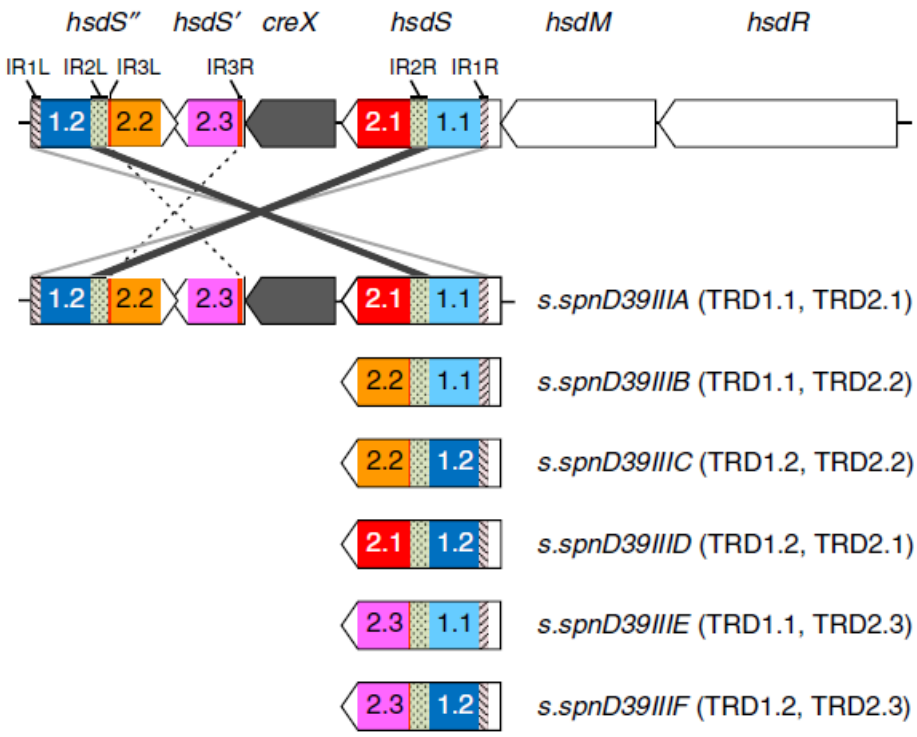


Figure 6.1. Schematic representation of *spnD39III* locus and the six alternative *hsdS* configurations.

The D39 type I RM system has three co-transcribed genes (*hsdR*, *hsdM* and *hsdS*), a separately transcribed Cre tyrosine DNA recombinase (*CreX*) and two truncated *hsdS* genes. The transcribed *hsdS* gene contains two target recognition domains (TRD) which share inverted repeats (IR) with the truncated *hsdS* genes. In this study, recombination of the three inverted repeats (IR1 to IR3) resulted in six possible allele types (*spnD39IIIA-F*).

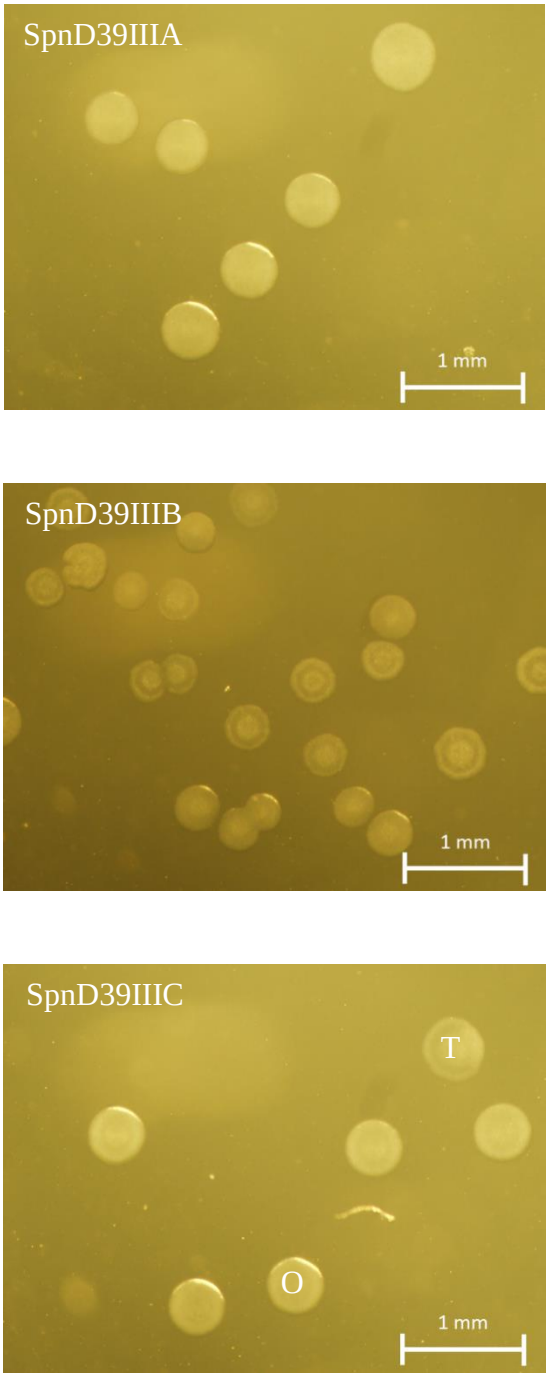


Figure 6.2. Colony opacity phenotypes conferred by SpnD39IIIA, B and C.

Colony morphology of SpnD39IIIA, SpnD39IIIB and SpnD39IIIC plated for single colonies on a THY+catalase plate (Section 2.2.8) and observed under oblique transmitted light using a Nickon SMZ1000 dissecting microscope. “O” and “T” on SpnD39IIIC represents an opaque and transparent colony, respectively. SpnD39IIID-F not shown.

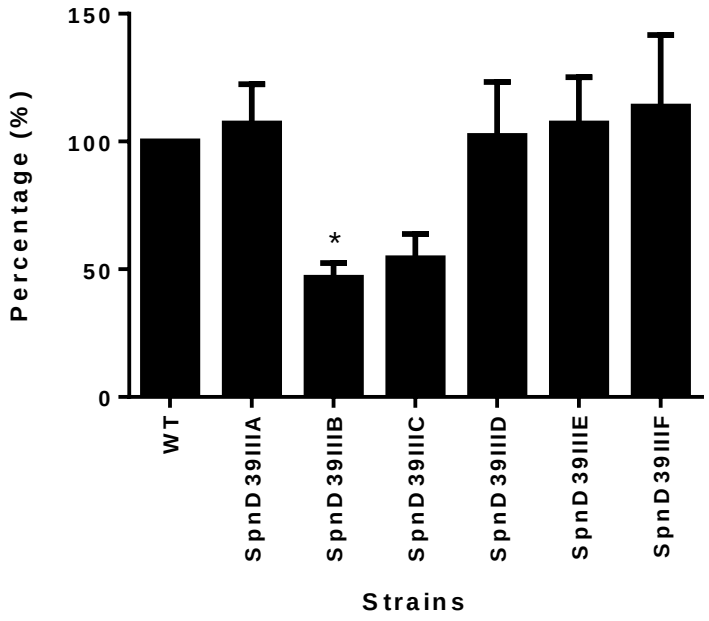


Figure 6.3. Uronic acid production by SpnD39IIIA-F mutants

The production of type 2 CPS by the SpnD39IIIA-F variant strains was quantified using a uronic acid assay (Section 2.7.4) and expressed as a percentage absorbance relative to strain D39WT. The mean of six samples are shown (in the case of SpnD39IIIE-F, including two technical replicates) and statistical analysis was performed using one-way analysis of variance. *p <0.05.

consistent with downregulation of capsule-associated genes in D39SpnIIIB reported by Manso *et al.* (2014).

6.2.3 Quantitative Western blot analyses of SpnD39IIIA-D locked strains

To investigate the impact of the *spnD39IIIA-D* locked mutants on the expression of certain virulence-related and cell surface proteins, relative protein expression was assessed using quantititave Western blot (Section 2.7.2). A selection of 12 anti-sera raised against Ply, CbpA, PspA, LytA, PsaA, MalX, GlpO, AliA, PiuA, PhtD, NanA and ClpP were used. However, there were no significant differences in expression of any of the proteins between any of the SpnD39III variants.

Manso *et al*. (2014) also reported that *luxS* was downregulated in SpnD39IIIB. Hence, LuxS production of the SpnD39III variants (SpnD39IIIA-F) was also assessed using quantitative Western blot analysis and its expression was calculated relative to D39WT (Figure 6.4). This showed that there is a significant reduction in the level of LuxS in SpnD39IIIB, as well as in SpnD39IIIC albeit to a lower extent than SpnD39IIIB. The SpnD39IIIA variant produced the highest level of LuxS, confirming the RNAseq findings of Manso *et al*. (2014).

6.2.4 Virulence phenotype of SpnD39III variants

To investigate the virulence phenotype of the SpnD39III variants, groups of six, six week old female CD1 mice were challenged intravenously with 1×10^5 pneumococci and blood samples were taken at 4 h and 30 h post-challenge. Bacterial loads were enumerated and opacity phenotypes examined on the appropriate agar plates. In addition, bacterial loads were also enumerated in the brain, spleen and liver at the end of the challenge period (30 h).

With the exception of SpnD39IIIF, there was a correlation between colony opacity and virulence. That is, the higher the percentage of O colonies of a particular strain, the more virulent the strain was (Figure 6.5). In the initial mouse challenge experiment, only four mutants (SpnD39IIIA-D) were available, hence these were assessed first. However, when SpnD39IIIE and SpnD39IIIF became available, the experiment was repeated using SpnD39IIIA-B and SpnD39IIIE-F. There were differences in survival rate for WT and

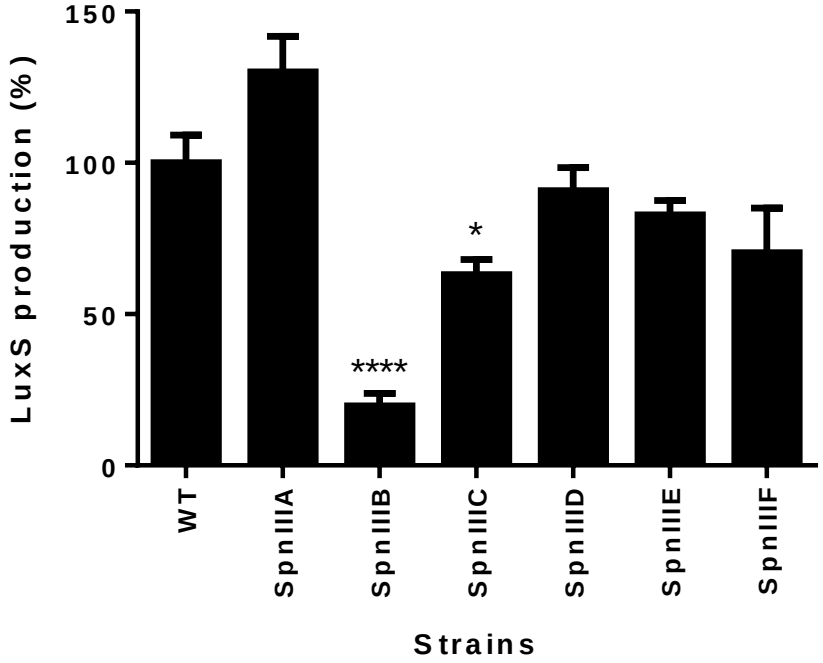
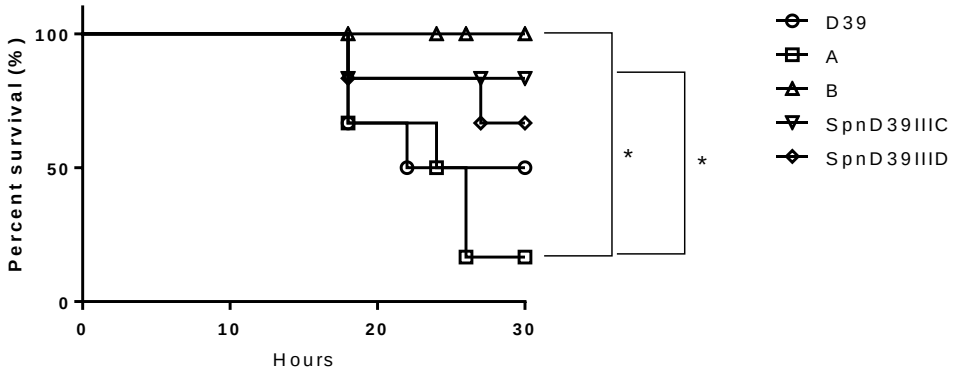


Figure 6.4. LuxS production by SpnD39IIIA-F mutants.

LuxS production by the SpnD39IIIA-F locked mutants was expressed as a percentage of that for D39 wild-type (WT) using quantitative Western blot analysis (Section 2.7.2). The mean of six samples are shown (in the case of SpnD39IIIE-F, including two technical replicates) and statistical analysis was performed using one-way analysis of variance. *, p <0.05; ****, p <0.0001.

A



B

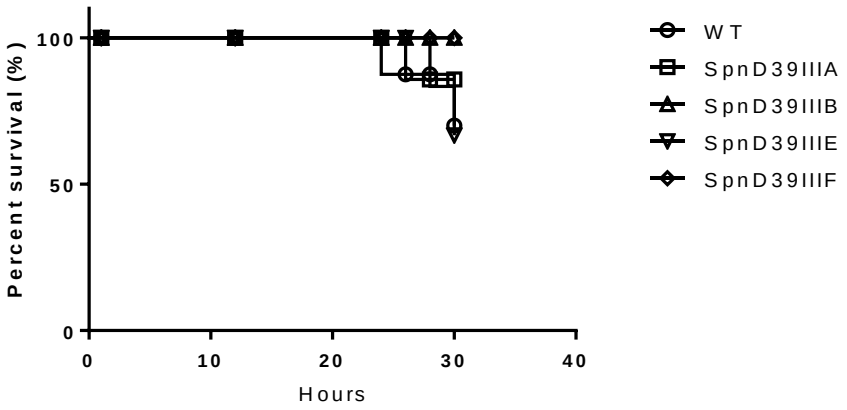


Figure 6.5. Survival of mice after intravenous challenge

The Kaplain-Meier curves showing the percent survival of mice over a 30 h period after intravenous (i.v.) challenge with D39WT or the indicated D39SpnIII locked mutants. Panels A and B represent two separate experiments, each containing groups of six CD1 mice. Statistical significance was determined using the log-rank test. *, p <0.05.

SpnD39IIIA-B and SpnD39IIIE-F. There was also a significant difference in the survival rate between SpnD39IIIA and SpnD39IIIB, and SpnD39IIIA and SpnD39IIIC. Differences in survival rate for WT and SpnD39IIIA- and SpnD39IIIB-challenged mice between the two challenge experiments were also observed. The reasons for these are uncertain, but may include slight variation in the age of mice, or other factors affecting disease kinetics. However, in both experiments, the most virulent strains were D39WT, SpnD39IIIA and SpnD39IIIE, while the least virulent was SpnD39IIIB. Furthermore, mice infected with either SpnD39IIIB or SpnD39IIIC had lower bacterial loads in their blood at 4 h compared to those infected with D39WT or the other SpnIIID39 variants (Figure 6.6A). At the later time point (30 h), mice challenged with the locked SpnD39IIIB, SpnD39IIIE and SpnD39IIIF strains exhibited significantly lower levels of bacteraemia compared to those infected with the other strains (Figure 6.6B). In the brain, spleen and liver, SpnD39IIIB-infected mice also exhibited significantly lower bacterial loads than those infected with the other strains (Figures 6.6C-E).

Blood samples collected at 30 h from all mice infected with the various strains were also plated on catalase-supplemented agar plates to determine O/T phenotype (Figure 6.7). All mice infected with SpnD39IIIE or SpnD39IIIF yielded 100% O colonies, as did all but two of the mice challenged with SpnD39IIIA and all but three of those challenged with SpnD39IIID. In contrast, all but one of the mice challenged with SpnD39IIIB yielded 100% T colonies. O/T phenotype ratios varied markedly, however, for mice challenged with D39WT or SpnD39IIIC, with some mice yielding 100% O colonies, while others exhibited mixtures of O and T phenotypes; the percentage of O colonies was zero for two of the SpnD39IIIC-infected mice.

6.2.5 Quantification of *spnD39III* subpopulations

To determine whether genetic switching between the *spnD39III* alleles was selected *in vivo* in mice infected with D39WT, the allele distribution of *spnD39III* was assessed from genomic DNA using a fluorescent GeneScan assay (fragment length analysis). This involved PCR amplification of the complete 4.2-kb *spnD39III* region with one of the PCR primers labelled with 6-fluorescein amidite (FAM). The PCR products were then digested with both *Dra*I and *Ple*I. This digestion was predicted to yield different sized FAM-labelled fragments for each of the variant forms. The pool of restriction fragments was then

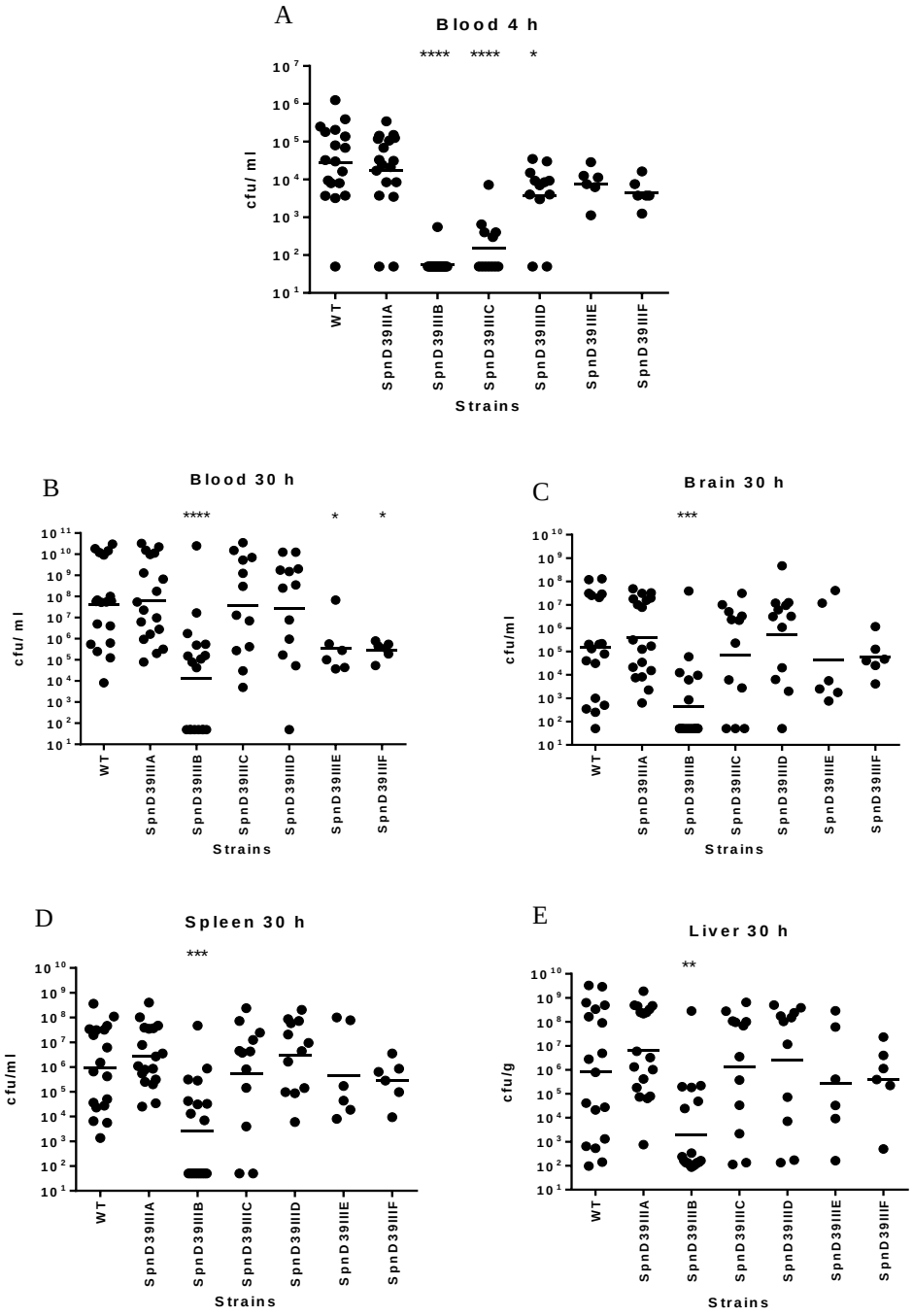


Figure 6.6. Bacterial loads at 4 and 30 h post intravenous challenge.

Numbers of pneumococci were enumerated in the blood at 4 h post challenge (A) and at 30 h post challenge in the blood (B), brain (C), spleen (D) and liver (E) from CD1 mice after intravenous challenge (Section 2.11.4). The geometric mean for each group is indicated by a horizontal bar. *, p <0.05, **, p <0.01, ***, p <0.001, ****, p <0.0001 relative to WT.

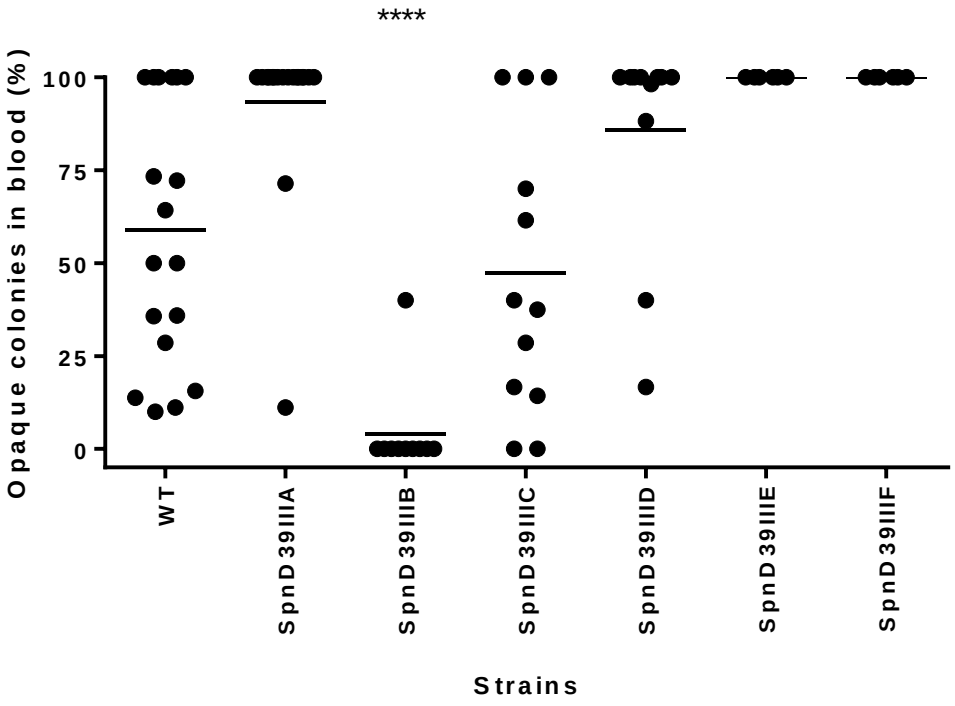


Figure 6.7. Opacity of SpnD39III strains in blood at 30 h post challenge

Groups of mice were challenged intravenously with the indicated strain and at 30 h post challenge, blood samples from each mouse were plated onto catalase agar plates and the percentage of O colonies was determined for each strain. ****, p <0.0001 relative to WT.

run on an ABI prism Gene Analyser and the area of the peak given by each labelled fragment, each corresponding to the prevalence of one of the variant forms, was quantified, as detailed in Manso *et al.* (2014).

In the first instance, pneumococcal genomic DNA harvested from *in vivo* experiments was checked for reproducibility of the DNA extraction and allele quantification methods when harvested directly from colonies on agar plates or grown in THY for 3 h before DNA extraction. The 3 h THY culture was required to increase bacterial counts as some *in vivo* samples contained very low bacterial CFU (under 10 colonies) when directly plated on agar plates. The *spnD39III* distribution was found to be reproducible when DNA was harvested from either solid or liquid culture (Table 6.1). To add to the robustness of the method, when independent PCR reactions were performed on diluted samples, the variations in allele frequencies were negligible (result not presented). This provides evidence for the reliability of the DNA extraction and allele quantification methods. To ensure consistency between the DNA samples, all samples were extracted after a 3 h incubation in liquid culture (THY broth).

After confirming the reliability of the allele quantification methods, pneumococcal DNA was extracted from the blood, liver, brain and spleen tissue of infected animals and *spnD39III* allele frequencies were compared with that in the initial inoculum. The D39WT inoculum used to challenge the mice had predominantly *spnD39IIIE* (22.4% *spnD39IIIA*, 7.9% *spnD39IIIB*, 0.5% *spnD39IIIC*, 3.4% *spnD39IIID*, 65.8% *spnD39IIIE* and 0.2% *spnD39IIIF*) (Figure 6.8A). A progressive change in the *spnD39III* allele distribution in the blood can be observed during the course of the infection and by 30 h post-challenge, *spnD39IIIA* was the dominant genotype. There was also a slight increase in *spnD39IIIB* (7.8% initially, 14.0% at 4h and 17.9% at 30 h). The same dominance of SpnD39IIIA was also seen at 30 h in the other tissues (spleen, liver and brain), and there was little difference in SpnD39III allele frequencies in bacteria isolated from the various tissues at 30 h post-challenge (Figure 6.8B). This shows that there is a selection for specific *spnD39III* allele types over the course of an infection, which is evident as early as 4 h in the blood (Figure 6.8A).

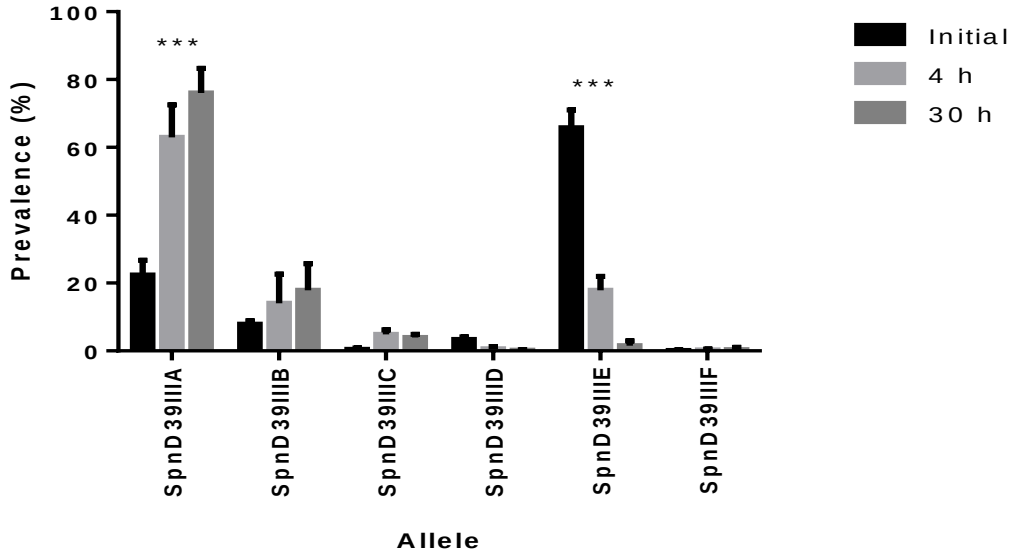
Table 6.1. Allele quantification reproducibility

Pneumococcal DNA were extracted (Section 2.8.3) from cells harvested from either colony from an agar plate or cell lysate from a broth culture. Following this, two independent reactions for each method were subjected to PCR, digest and GensScan analysis (Manso *et al*., 2014) to assess the reproducibility of the quantification method under both conditions.

Sample[a]	**SpnD39III allele type (%)**					
(Reaction number)	**A**	**B**	**C**	**D**	**E**	**F**
Plate (1)	5.3	63.2	2.0	28.3	1.12	-
Plate (2)	5.0	64.5	2.3	27.1	1.2	-
Broth (1)	5.79	64.3	3.4	25.3	1.2	-
Broth (2)	5.7	64.6	3.3	25.3	1.2	-

[a] "Plate" represents DNA extraction from colonies harvested from a plate, whereas, 'Broth' represents DNA extracted from the broth culture which was inoculated with colonies from the plate and incubated at 37°C in 5% CO_2 for 3 h. The numbers following the sample name represent the first (1) and second (2) independent PCR reactions to show the reproducibility of the data obtained using the GeneScan method to quantify allele distribution

A



B

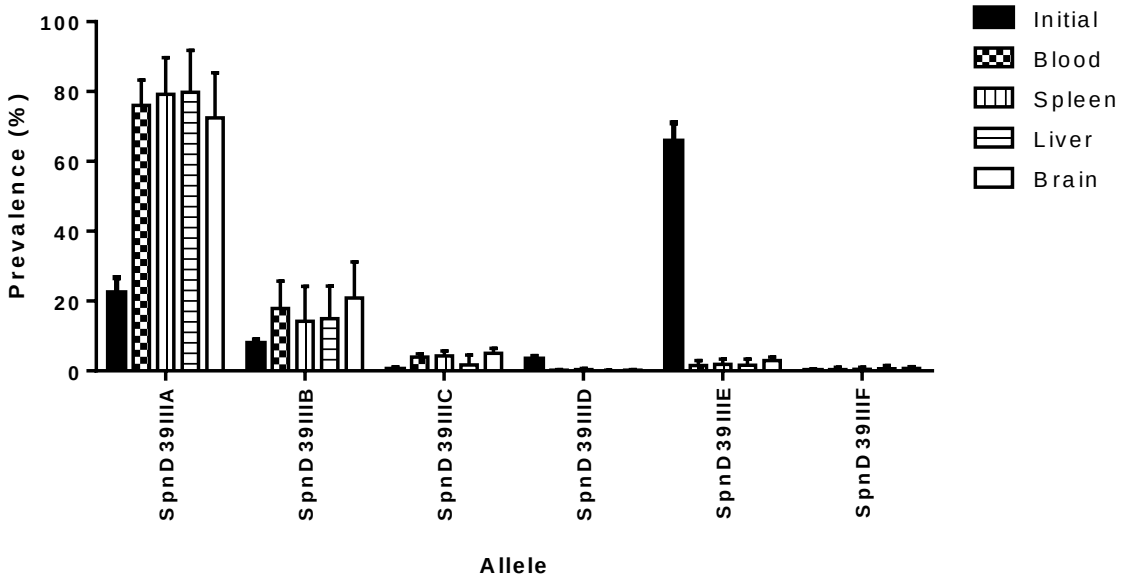


Figure 6.8. s*pnD39III* allele distribution after intravenous challenge of mice with D39WT

Mice were challenged intravenously with D39WT and *spnD39III* allele frequencies (mean % ± SD) were determined in DNA extracted from the blood, spleen, liver and brain from three to four mice (Manso *et al.*, 2014). A, variation in *spnD39III* allele distribution in the blood-borne bacteria over time. B, *spnD39III* allele distribution in blood, spleen, liver and brain at 30 h post challenge. ***, p <0.001.

6.2.6 Allele quantification of D39O and D39T DNA after intranasal challenge

To obtain DNA for allele quantification from strains D39O and D39T, two groups of 10 CD1 female, six-week old mice were challenged intranasally with 1×10^7 cfu of either D39O or D39T and pneumococcal DNA for allele quantification was harvested directly from the plates used to enumerate the bacteria and DNA extracted from these colonies (Section 2.8.3). The numbers of pneumococci were enumerated in nasal lavages and blood at 72 h post-challenge. Both groups showed similar numbers in the nasal wash, but animals challenged with D39O had significantly higher bacterial loads in the blood compared to those challenged with D39T (Figure 6.9). Furthermore, the opacity phenotypes of D39O and D39T did not change from the inoculum when plated for single colonies on THY+catalase agar plates at the end of the 72 h challenge period. That is, mice challenged with strains D39O and D39T still yielded exclusively O and T colonies, respectively.

To determine whether *spnD39III* allele frequencies switch *in vivo* in either D39O or D39T, allelic composition of pneumococcal DNA harvested from each of the animals were quantified, as described in Manso *et al*. (2014). The initial D39O inoculum used to challenge the mice comprised 65.7% *spnD39IIIB* and 34.3% *spnD39IIIA*, whereas, the D39T inoculum comprised 77.8% *spnD39IIIA* and 22.2% *spnD39IIIB*. For mice challenged with D39O, *spnD39IIIB* remained the dominant allele in all samples from the nasopharynx, although increases in *spnD39IIIE* were observed in six cases, and increases in *spnD39IIIC* and *spnD39IIID* alleles occurred in two mice each (Table 6.2). Allele switching was more pronounced in the blood. Pneumococcal DNA from two of the mice was 100% *spnD39IIIB*, while another mouse yielded 100% *spnD39IIIE*. *spnD39IIIE* also became the dominant allele in another mouse blood sample. However, another mouse blood sample had *spnD39IIIA* as the dominant allele (87.1%) (Table 6.2).

In mice challenged with D39T, changes in allele frequency were less dramatic. In nasopharyngeal samples, the population of *spnD39IIIA* and *spnD39IIIB* shifted slightly in favour of the *spnD39IIIB* allele in five of the six mice, with one mouse remaining essentially unchanged (Table 6.3). More marked alterations in allele frequency relative to that in the challenge inoculum, and divergence between mice, were observed in the blood, with one of the mice yielding 94.5% *spnD39IIIA*, while another exhibited 93.9%

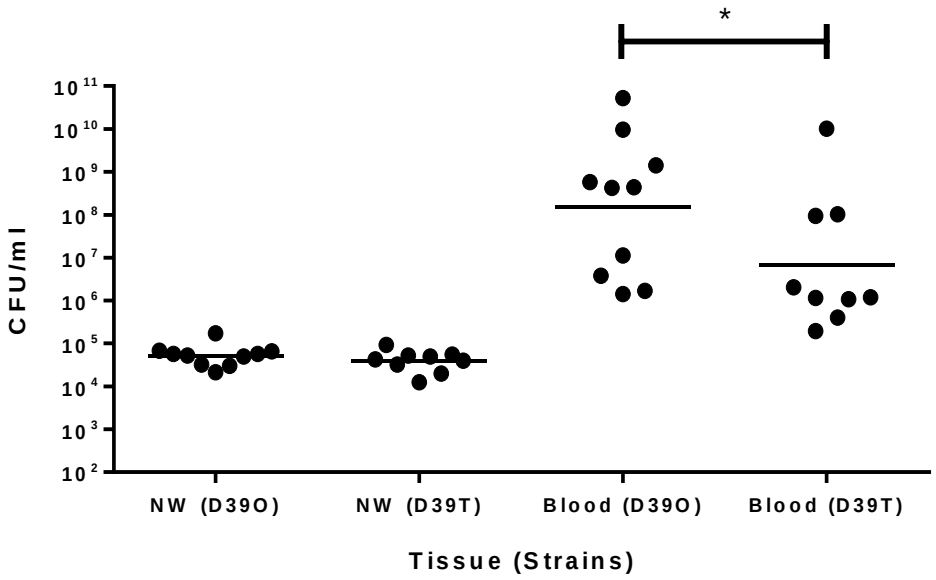


Figure 6.9. Numbers of pneumococci recovered from mice at 72 h post intranasal challenge.

Two groups of 10 six week old, female, CD1 mice were intranasally challenged (i.n.) with 1×10^7 CFU of either D39O or D39T under anaesthesia, as described in Section 2.11.3. The number of pneumococci were enumerated from the nasal wash (NW) and blood of each mouse at 72 h post challenge. The geometric mean CFU/ml for each group is indicated by the solid horizontal bar and statistical analysis was performed on log transformed data using a one-tailed *t*-test, *, $p > 0.05$.

Table 6.2. SpnD39IIIA-F allele quantification of *in vivo* D39O DNA.

Mice were challenged i.n. with D39O and the *spnD39III* allele frequencies (%) were determined (Manso *et al.*, 2014) in DNA extracted from pneumococcal colonies harvested from the agar plate that the bacteria were enumerated from. Isolates were enumerated from nasal wash (NO) and blood (BO) of 7-8 mice 72 h post challenge. D39O (initial) indicates the allele frequency of the inoculum used to challenge the mice.

Samples[a]	**Allele frequency (%)**					
	SpnD39IIIA	**SpnD39IIIB**	**SpnD39IIIC**	**SpnD39IIID**	**SpnD39IIIE**	**SpnD39IIIF**
D39O (initial)	34.3	65.7	-	-	-	-
NO1	5.4	77.7	3.6	-	13.3	-
NO2	-	100.0	-	-	-	-
NO3	36.6	54.7	-	6.0	2.7	-
NO4	37.8	46.2	3.0	1.7	11.6	-
NO5	30.1	67.1	-	-	2.8	-
NO7	22.4	67.2	-	-	10.4	-
NO8	24	66.3	-	-	9.7	-
NO9	21.3	78.7	-	-	-	-
NO10	28.7	71.3	-	-	-	-
BO1	-	100.0	-	-	-	-
BO3	1.9	87.1	-	9.3	1.7	-
BO4	87.1	12.9	-	-	-	-
BO5	-	-	-	-	100.0	-
BO8	17.9	18.1	7.3	-	56.7	-
BO9	-	100	-	-	-	-
BO10	8.5	91.5	-	-	-	-

[a]N indicates gDNA from nasal wash samples; B indicates gDNA from blood samples, O indicates mice challenged with D39O; the numbers following sample name corresponds to mouse identity number.

Table 6.3. SpnD39IIIA-F allele quantification of *in vivo* D39T DNA.

Mice were challenged i.n. with D39T and the *spnD39III* allele frequencies (%) were determined (Manso *et al.*, 2014) in DNA extracted from pneumococcal colonies harvested from the agar plate that the bacteria were enumerated from. Isolates were enumerated from nasal wash (NT) and blood (BT) of 4-6 mice 72 h post challenge. D39T (initial) indicates the allele frequency of the inoculum used to challenge the mice.

Samples[a]	**Allele frequency (%)**					
	SpnD39IIIA	**SpnD39IIIB**	**SpnD39IIIC**	**SpnD39IIID**	**SpnD39IIIE**	**SpnD39IIIF**
D39T (initial)	77.8	22.2	-	-	-	-
NT1	58.5	41.5	-	-	-	-
NT2	77.8	22.2	-	-	-	-
NT3	54.0	46.0	-	-	-	-
NT6	52.6	47.4	-	-	-	-
NT7	56.5	43.5	-	-	-	-
NT9	59.4	40.6	-	-	-	-
BT3	62.5	37.5	-	-	-	-
BT4	37.3	58.3	-	-	4.4	-
BT7	94.5	5.5	-	-	-	-
BT9	6.1	93.9	-	-	-	-

[a]N indicates gDNA from nasal wash samples; B indicates gDNA from blood samples, T indicates mice challenged with D39T; and the numbers following sample

spnD39IIIB. Only one mouse showed evidence of switching to one of the other alleles [in this case, *spnD39IIIE* (4.4%)] (Table 6.3)

6.3 DISCUSSION

This chapter has examined the involvement of six epigenetically different pneumococcal subpopulations which express distinct Type I RM target specificities in virulence. The existence of potentially switchable Type I systems has been reported previously for the pneumococcus and for other bacterial species, but no evidence was provided for differential methylation patterns or whether there was any phenotypic impact (Cerdeno-Tarraga *et al.,* 2005; Dybvig *et al.,* 1998; Tettelin *et al.,* 2001). Analysis of the locked SpnD39III mutants confirmed differences in genomic methylation patterns, as well as phenotypic differences in terms of colony opacity, and virulence protein and CPS expression that have an impact on bacterial virulence. Although the locked SpnD9IIIA and SpnD39IIIB mutants were almost exclusively O and T, respectively, there was a clear disconnect between these locked variants and the alleleic distribution of the D39O and D39T wild-type strains (D39T contained predominantly *spnD39IIIA*, whereas, D39O contained predominantly *spnD39IIIB*). In nature, there is a mixture of the six allele subpopulations and it is unknown what percentage of each of these is required for the observed phenotype.

Pneumococcal opacity phase variants have been reported to preferentially occupy certain host niches, with the T phenotype more likely to colonise the nasopharynx, whereas the O phenotype is more likely to invade into deeper tissues such as the blood (Briles *et al.*, 2005; Weiser *et al.*, 1994). In accordance with this premise, the locked SpnD39IIIA mutant (100% O phenotype) was unable to stably colonise the nasopharynx, whereas SpnD39IIIB (100% T phenotype) maintained similar nasopharyngeal bacterial cell counts for days 1-7 in a nasal colonization model (Manso *et al.*, 2014). Furthermore, in this study, SpnD39IIIA had higher numbers in the blood compared to SpnD39IIIB, at both 4 h and 30 h after intravenous challenge. The bacterial counts in the brain, spleen and liver were also very low for SpnD39IIIB compared to strains expressing the other SpnD39III variants. These results show that SpnD39III-locked mutants that yielded a higher proportion of O colonies (such as SpnD39IIIA) are more likely to cause systemic disease, whereas those that

express more T colonies (such as SpnD39IIIB) preferentially colonise the nasopharynx. Hence, the opacity phenotype of the locked SpnIIID39 mutants appears to reflect their virulence.

The differences between SpnD39IIIA and SpnD39IIIB were also evident at the transcriptional level, with the *dexB, luxS* and the SPD_0310 genes encoding type 2 CPS production downregulated in SpnD39IIIB (Manso *et al.*, 2014). In this study, quantitative Western blot analysis confirmed the downregulation of LuxS and a uronic acid assay confirmed the reduced amount of CPS in SpnD39IIIB compared to the other variants. LuxS, a protein associated with quorum sensing and biofilm formation (Trappetti *et al.*, 2011c) was significantly downregulated in SpnD39IIIB, compared to the other SpnD39III variants. This is consistent with the known role of LuxS in virulence and biofilm formation, and the finding that O variants are better able to form biofilms (Trappetti *et al.*, 2011b; Trappetti *et al.*, 2011c). Furthermore, there appears to be a link between LuxS expression and CPS production, through an unknown mechanism, with reduced amounts of both of these related to lower virulence, particularly in systemic challenge models. The increased CPS in the locked SpnD39III mutants that yielded a higher proportion of O colonies would be expected to facilitate survival in the blood by evading phagocytosis. Indeed, SpnD39IIIB, which produces the lowest amount of CPS, was significantly more readily phagocytosed by RAW 264.7 macrophages compared to the other variants (Manso *et al.*, 2014).

As mentioned previously, the SpnD39III locked mutants displayed differing percentages of O and T phenotypes. The O/T ratios of these strains were also reflected in their ability to cause disease. This was especially evident when comparing the virulence of SpnD39IIIA and SpnD39IIIB, where the exclusively O-yielding SpnD39IIIA was more virulent than the predominantly T-yielding SpnD39IIIB strain in an animal experiment (Figure 6.5). The virulence of the O phenotype was also shown in mice challenged with the wild-types D39O and D39T, where significantly more D39O compared to D39T were isolated from the blood of these mice (Figure 6.9). Hence, the more O colonies a strain produces, the more virulent the strain is. This has previously been documented in other studies (Weiser *et al.*, 1994; Weiser *et al.*, 1996).

Examination of *spnD39III* allele frequencies in mice challenged intranasally with WT D39 provided evidence for switching of allele types *in vivo,* and this could be seen as early as 4 h post challenge (Figure 6.8A). The initial inoculum used to challenge the animals contained a mixture of both colony opacity phenotypes and was predominantly the *spnD39IIIE* allele type. The shift from *spnD39IIIE* to a predominantly *spnD39IIIA* allele type in isolates from the blood is evidence that the re-isolated pneumococci had an epigenetic profile shift during infection and that selection for SpnD39IIIA is important for survival and proliferation of pneumococci in the blood. In addition, this epigenetic shift was also seen in bacteria isolated from the spleen, liver and brain, which had similar profiles to that of the blood at 30 h post challenge. On the other hand, this change was not seen in isolates from the nasopharynx during a carriage experiment (Manso *et al.*, 2014).

Changes in *spnD39III* allele frequency were also seen in animals that were challenged intranasally with D39O (Table 6.2). The original challenge inoculum was a mixture of 34.3% *spnD39IIIA* and 65.7% *spnD39IIIB* alleles, but *spnD39IIIC-E* alleles appeared in bacteria present in both the nasopharynx and the blood at 72 h. Presumably, expansion of allele types present in the bacterial population provided a survival advantage in both niches. However, there were examples of more extreme allele variations, with bacteria present in the nasopharynx of one of the mice and the blood of two others being exclusively *spnD39IIIB*. A third mouse had exclusively *spnD39IIIE*-expressing pneumococci in the blood. The dominance of the *spnD39IIIB* allele in the nasopharynx of one of the mice was compatible with the findings for the SpnD39IIIB-locked mutant, and it was unsurprising that bacteria were not detected in the blood in this particular animal. However, the dominance of the *spnD39IIIB* allele in bacteria in the blood of two of the mice (#1 and #9) is highly counterintuitive given the low systemic virulence of the SpnD39IIIB-locked mutant, particularly when these same mice had distinct allele frequencies the nasopharyngeal niche (Table 6.2). In contrast to the marked changes in allele frequency seen *in vivo* in mice challenged intranasally with D39O, only modest changes were seen in the majority of mice challenged with D39T (Table 6.3).

The D39T inoculum comprised 77.8% *spnD39IIIA* and 22.2% *spnD39IIIB*; in the nasopharyngeal niche, there was a modest increase in frequency of the *spnD39IIIB* allele, but no appearance of *spnD39IIIC-F*. In the four mice that yielded allele frequency data

from bacteria in the blood, the changes between *spnD39IIIA* and *spnD39IIIB* alleles were more divergent (*spnD39IIIA* frequency ranged from 6.1% to 94.5%), and a low level of *spnD39IIIE* was present in bacteria in the blood of one of the mice (Table 6.3). Hence, it appears that in the wild-type D39 strains, recombination at the *spnD39III* locus is much higher in D39O, whereas in D39T the locus is more stable. In the proteomics and transcriptomic analyses conducted in this study (Chapters 3 and 4), there was an increase in expression of competence-related proteins/genes in the O variants of all three strain backgrounds. Although this may be coincidental, there is a possibility that recombination of the IRs in the *hsdS* gene may be more efficient when the bacteria are in a highly competent state.

However, notwithstanding the above, the findings of this study show an apparent disconnect between the O/T phenotypes of SpnD39III-locked mutants and the *spnD39III* allele frequencies of WT D39O and D39T, where D39T contained predominantly *spnD39IIIA* and D39O was predominantly *spnD39IIIB.* It is important to note that 'locked' SpnD39III mutants do not occur in nature; populations of wild type pneumococci likely comprise subpopulations expressing any of the six possible *spnD39III* allele types. Each of these allele types will result in a distinct genomic methylation pattern with potentially diverse effects on gene expression. O/T phenotype is a phenomenon based on macroscopic appearance of an entire colony. Furthermore, not all *hsdS* alternatives were present in the other two strains (Oggioni, M., unpublished data). In spite of the analyses conducted in this study, the actual molecular basis for the difference in macroscopic appearance remains uncertain, but presumably reflects differential expression of surface structures (CPS, TA, surface proteins, etc.) and perhaps differences in the extent of autolysis of bacterial cells within the colony, collectively impacting on the opacity of the colony as a whole. This study has provided evidence for an impact of *spnD39III* allele on expression of factors (e.g. CPS) that would be expected to contribute to colony opacity. However, extrapolating phenotypic data from locked mutants to natural strains is problematic. We do not have data on the colonial phenotype of defined mixtures of bacterial cells expressing different *spnD39III* alleles. Furthermore, no data are available on the *in vivo* half-life of the various methylase enzyme complexes. Thus, it is possible that hybrid DNA methylation patterns may be present within a single bacterial cell if it has recently undergone *spnD39III* recombination and residual methylase activity with the previous target specificity remains

present in the cytoplasm. Such complexities make resolution of the apparent anomalies in O/T phenotype and *spnD39III* genotype a challenging task indeed.

Suffice it to say that notwithstanding the influence of *spnD39III* on opacity phenotype, it is not the sole determinant thereof. Moreover, it is important that previous and future studies on the regulation of pneumococcal virulence both *in vitro* and *in vivo* be interpreted in the context of the potential switching between epigenetically-different subpopulations during the course of any experimental.

Chapter 7: FINAL DISCUSSION

7.1 Introduction

The pneumococcus is a human pathogen that causes significant global mortality and morbidity (Section 1.1.1). The burden of disease is most prevalent in those under the age of five years and those over 65 years (Lynch & Zhanel, 2010). The treatment of pneumococcal infection is typically by antibiotics. However, this has been complicated due to the rise in antibiotic-resistant strains (Section 1.1.2). Hence, there is considerable research in preventative measures such as the development of vaccines (Section 1.2). Although the pneumococcus normally resides in the nasopharynx as part of the commensal flora, the bacterium is able to disseminate from there into deeper tissues to cause disease. Furthermore, the pneumococcus is known to exist in two colony opacity morphologies or phases (O and T), named according to their appearance when viewed under oblique, transmitted light on a clear solid medium (Figure 1.3). Pneumococci are able to spontaneously switch between phases, and in animal models, the T variants of the same strain appear to have enhanced ability to colonise the nasopharynx, whereas, the O variants are more likely to be isolated from deeper tissues such as the lungs, blood, or CSF (Briles *et al.*, 2005; Weiser *et al.*, 1994). The critical determinants that enable the pneumococcus to switch from one phenotype to another are poorly understood. Previous attempts to characterise this phenomenon at a molecular level have resulted in discordant reports. However, disagreement in the literature is likely to be a reflection of the fact that different strains and analytical techniques were used in each study (Table 1.1).

The primary objective of this study was to identify proteomic, transcriptomic and genomic changes that might be central to the transition of the pneumococcus from carriage (T phase) to invasive disease (O phase). This study used three different strains belonging to three different serotypes (D39, serotype 2; WCH16, serotype 6A; and WCH43, serotype 4), with distinct pathogenic profiles in animal models. In previous work, D39 and WCH43 have been found to frequently progress to bacteraemia following i.n. challenge, whereas, WCH16 frequently progresses directly to the brain with minimal lung and blood

involvement (Mahdi *et al.*, 2008). To further add to the robustness of this study, the samples used for each of the three analytical methods originated from the same starting culture, and repeated with four biological replicates. Additionally, during the progress of this project, an opportunity arose to collaborate with an international group to study a pneumococcal Type I RM system. This enabled the hypothesis that opacity phase variation could be influenced by epigenetic changes to be addressed. To date, this is the first study that has generated comprehensive proteomic, transcriptomic and genomic, profiles in an attempt to characterise the pneumococcal opacity phase variation phenomenon. Furthermore, this work is also the first to investigate the role of epigenetic changes on pneumococcal phase variation.

7.1.1 Proteomic analysis of opacity phase variable *S. pneumonaie*

Comparisons between proteome expression profiles of O/T pairs of each strain using 2D-DIGE showed very little consistency between strains in the expression of specific proteins in regards to opacity. The proteins identified as differentially expressed between O/T pairs were classified into five functional groups – those involved in (i) glycolysis, (ii) pyruvate metabolism, (iii) cell growth/DNA and protein synthesis, (iv) maintenance of cellular health and (v) sugar/amino acid transport (Table 3.5). Although no specific proteins were found to be consistently differentially expressed by one phase versus another in all three strains, there appeared to be some consistency between the putative functional groups of the proteins that were differentially expressed. For example, proteins predicted to be involved in stress responses (DnaJ and SsBb) and pyruvate metabolism (SpxB, AdhE and Pfl) were generally upregulated in the O variant versus the T variant (Table 3.5). On the other hand, a protein involved in glycolysis and adherence (GAPDH) was predominantly upregulated in the T variant compared to the O variant. However, this appears to be strain specific, as GAPDH was upregulated in the T variants of D39 and WCH43, but not in WCH16, and this was confirmed in an enzymatic assay (Table 3.4).

A previous study showed that SpxB was upregulated in the T form of a serotype 9V and 6B strain (Overweg *et al.*, 2000). In this study, SpxB was only differentially regulated in the D39 O/T variants (upregulation in O). However, quantitative Western blot analysis did not confirm the initial findings of the 2D-DIGE experiments. One possible reason for

the differences between studies is that SpxB is known to exist in a number of different isoforms (Bae *et al.*, 2006), but 2D-DIGE only detected one isoform as differentially expressed between D39O and D39T using the criteria described (Section 3.2.2.1). Thus, unless all SpxB protein spots were compared collectively, it is difficult to assess the actual change in SpxB expression between the O/T pairs of the strains tested. Furthermore, the growth media in which the cell lysates were harvested also differed between studies. In addition, cells were harvested at A_{600} 0.5 (mid- to late-logarithmic phase) in this study, whereas, Overweg *et al.* (2000), harvested cell lystates at A_{550} 0.3 (early to mid-logarithmic phase). To further explore the role of SpxB in opacity phase variation, targeted mutagenesis of *spxB* in D39O and D39T was performed. This resulted in a hyperopaque phenotype, while ectopic over-expression of SpxB in D39O did not result in a T phenotype (Figure 3.7). The hyperopaque phenotype in the *spxB*-deletion mutant has been documented in other studies (Allegrucci & Sauer, 2008; Carvalho *et al.*, 2013; Pericone *et al.*, 2002; Ramos-Montanez *et al.*, 2008). Although there were phenotypic changes as a result of *spxB* mutagenesis, it is unlikely that this protein is responsible for pneumococcal opacity phase variation. Rather, this was thought to be due to an increase in capsule production or as a result of increased cell biomass due to reduced SpxB-mediated production of H_2O_2 (Carvalho *et al.*, 2013).

7.1.2 Transcriptomic analysis of *S. pneumonaie* opacity phase variation

Whole transcriptome analysis between O/T pairs of strains D39, WCH16 and WCH43 by microarray was used to generate lists of genes that were consistently differentially regulated in one phenotype over another. These genes were subsequently grouped according to their putative functions. Similar to proteomic comparisons, specific genes were not consistently identified as having phase-specific expression in all three strains. However, there did appear to be some consistency between the putative functions of some genes in either the O or T phase. Genes that were upregulated in the O variant included those involved in competence and other stress-related functions (Table 4.1), and this was consistent with the proteomic studies (Table 3.5). Some of these (including *comX1*, *cglA*, *cbpD* and *groEL*) were confirmed using qRT-PCR (Table 4.2). On the other hand, genes that were predominantly upregulated in the T variant included those involved in cell

division, cell wall/membrane-related and amino acid biosynthesis and acquisition, but these were not found in proteomic analyses (Table 4.1).

In previous studies, inhibition of competence was responsible for a decrease in virulence (Oggioni *et al.*, 2006; Zhu & Lau, 2011). Furthermore, only O forms are able to form biofilms (Trappetti *et al.*, 2011b) and there was also a significant increase in the expression of competence genes, namely *comE* and *comX*, in these bacteria in a biofilm model (Trappetti *et al.*, 2011a). These correlate with the notion that the more invasive O form of the pneumococcus is associated with an increased expression of competence genes compared to T. However, mutagenesis of *comD*, which is the receptor of CSP, did not alter the opacity phenotype of the tested strains (Section 4.2.7). Hence, although the deletion of *comD* did not lead to a change in opacity phenotypes, it appears that competence is somehow associated with pneumococcal opacity phase variation, as shown in both proteomic (Table 3.1) and transcriptomic analyses (Table 4.1), where there was an increased expression of competence-related proteins and genes in the O variants (Table 4.2).

Many genes associated with cell division and amino acid acquisition and biosynthesis were shown to be upregulated in T variants in this study (Table 4.1). There were differences in growth rate between the O and T variants for individual strains, but this was not consistent between the strains (Figure 4.5).

7.1.3 Genomic analysis of opacity phase variable *S. pneumonaie*

Genomic comparisons were performed between D39, WCH43 and WCH16 O/T pairs in order to identify sequence differences that were present specifically in either the O or T phenotype, and may be responsible for the switch between the two. Previous work had investigated the role of 14 genes that contained non-trimeric repeat units or homopolymeric tracts >5-6 nt in length (McKessar, 2003). However, McKessar's work did not identify any evidence for a role for these sequences in the switching between D39O and D39T phenotypes. In this study, regions of interest selected for confirmation by Sanger sequencing were prioritised based on likely association with known virulence factors, such

as the capsule locus, or prediction resulting in a change in amino acid sequence due to SNPs.

While McKessar S. (2003) looked at individual genes, the present study sequenced and compared three sets of O/T pairs and identified putative SNPs and INDELs firstly between the O/T pairs and then those that were consistently different between O/T pairs of the three strains tested (Tables 5.1-3). This is the first known study to sequence three sets of O/T pairs and subsequently compare them for common SNPs and INDELs. Two of the SNPs found to be of potential interest in this study (SPD0636, *spxB*; and SPD1619, a gene with putative involvement in TA biosynthesis) were sequenced using Sanger sequencing, but this found that the SNP was only present in strain D39 (Figures 5.4-5). The SNP discovered in *spxB* between the D39O/T pair was of particular interest as SpxB was known to be differentially regulated in another study (Overweg *et al.*, 2000) and was found to be differentially regulated in the proteomic (Table 3.5) and transcriptomic (Table 4.1) analyses of this study, although differential-regulation was different between the methods (D39O > D39T in proteomics; D39T > D39O in transcriptomics). Furthermore, as described previously (Section 3.2.3.1), a *spxB* mutagenesis study produced hyper-opaque phenotypes in both D39O and D39T. Hence, *spxB* is not responsible for the O-T phenotypic changes, but rather, could be a secondary effect due to the production of H_2O_2, leading to a reduction in cell biomass, as described in Section 7.1.2. None of the INDELs or differences in homopolymeric tracts identified in genomic analyses could be subsequently confirmed by Sanger sequencing (Sections 5.2.3-6).

Hence, these results show that at a DNA sequence level, differences detected between O and T variants were largely strain-specific. Although there are limitations to the IonTorrent technology (Section 5.3), particularly in INDEL calling and the sequencing of homopolymeric tracts (Bragg *et al.*, 2013; Elliott *et al.*, 2012; Quail *et al.*, 2012; Yeo *et al.*, 2012), this work has largely supported previous work by McKessar (2003) that was not able to associate specific nucleotide changes with either O or T phenotype. Moreover, this study is the most comprehensive genomic analysis of opacity variants to date.

7.1.4 Pneumococcal Type I RM-system analysis

In addition to proteomic, transcriptomic and genomic analyses, this study investigated the involvement of six epigenetically different pneumococcal subpopulations that express distinct Type I RM target specificities. 'Locked' mutants displaying a single RM specificity displayed distinct phenotypes in terms of the proportion of O and T phase variants, as well as differences in expression of multiple genes (Manso *et al.*, 2014). In particular, two of these mutants, SpnD39IIIA and SpnD39IIIB, displayed almost exclusively O or T phenotypes, respectively (Manso *et al.*, 2014). Furthermore, they displayed pathogenic profiles similar to their dominant opacity phenotype (Figures 6.5 and 6.6), consistent with findings of earlier studies (Weiser *et al.*, 1994), as well as in this study for strain D39 (Figure 6.9). That is, the exclusively O phenotype SpnD39IIIA mutant was more likely to cause systemic disease, whereas the predominantly T variant-expressing SpnD39IIIB mutant preferentially colonised the nasopharynx. There was also an upregulation of LuxS and capsule expression in SpnD39IIIA versus SpnD39IIIB and these were consistent with the findings of Manso *et al.* (2014). LuxS has also previously been associated with O variants, which were better at forming biofilms (Trappetti *et al.*, 2011c) and had increased virulence (Trappetti *et al.*, 2011b). Additionally, in animal models, there was a shift in allele type distributions from a predominantly *spnD39IIIE* type to *spnD39IIIA* type (Figure 6.8A) during the progression of disease, further reinforcing the importance of the *spnD39IIIA* variant in the dissemination and proliferation of the pneumococci in the blood. However, there were no *spnD39III* allele changes detected in samples isolated from the nasopharynx (Manso *et al.*, 2014). Although the locked SpnD39IIIA and SpnD39IIIB mutants displayed phenotypes that were almost exclusively O or T, there was a clear disconnect with the allele distributions in D39O and D39T; it appeared that the initial inoculum of D39O was predominantly *spnD39IIIB,* whereas that of D39T was predominantly *spnD39IIIA*, and both D39O and D39T did not have any of the other *spnD39III* types in the initial inoculum. Meanwhile, after i.n. challenge in mice, there was an appearance of *spnD39IIIC-E* in both variants in the nasopharynx and blood at 30 h (Tables 6.2-3). This expansion of allele types presumably provides a selective advantage in their preferred niches. The dominance of the *spnD39IIIB* allele types in the blood of mice challenged with D39O was unexpected, given that the locked SpnD39IIIB

mutant displayed an almost exclusively T phenotype. However, it should be noted that these locked SpnD39III mutants do not occur in nature; rather, wild type *S. pneumoniae* populations may comprise mixtures of any of the six possible *spnD39III* allele types, at any frequency. Thus, although the actual molecular basis for pneumococcal opacity phase variation remains uncertain, perhaps a combination of these subpopulations may have a role in influencing the opacity phenotype. Study of this RM system did not result in identification of a "switch" that controls pneumococcal opacity phenotype. Nevertheless, it shows that epigenetic differences can contribute to the complex mechanism underlying pneumococcal opacity phase variation, which appears to be inherited but cannot be linked to specific nucleotide changes in the sequence. This could in part explain the lack of consistent differences between the O/T variants across the genome sequences of the three strains.

7.2 Concluding remarks

Pneumococcal opacity phase variation is poorly characterised at a molecular level. This study has attempted to understand the mechanism behind this phenomenon, by using proteomic, transcriptiomic and genomic approaches, as well as studies of epigenetic effects of a switchable Type I RM system. The criterion for opacity phase variation is based on the clarity of a colony growing on a transparent medium and the degree of opacity can be influenced by many factors including capsule : TA ratios, expression of cell surface molecules, differences in degree of autolysis and age of culture (Section 1.6.1). There was no single protein, gene or nucleotide differences that were found to be responsible for the O/T switching. However, there were some consistencies between the analytical methods in that some of the proteins or genes identified belonged to the same functional group, such those related to competence. Comparison of proteomic and transcriptomic data resulted in some inconsistencies between the two methods (Table 4.2) but was not unexpected. This is because detailed proteomics analysis was carried out on a selection of the protein spots that were picked and identified according to pre-determined criteria compared to whole gene expression profiling on over 2000 genes available on the the microarray slide. Furthermore, a difference in RNA level may not necessarily result in a difference in protein level, and this needs to be taken into account when comparing proteomic and

transcriptomic analyses. Specific nucleotide changes did not appear to be consistent between the strains. However, the Type I RM system appears to have an impact on O vs. T phenotype. The SpnD39III system investigated in this study also appeared to be a component of a central regulatory mechanism that has a role in pneumococcal fitness in distinct host niches.

The whole concept of what determines the O and T phase variants remains poorly defined and completely unrelated events could have a parallel effect on opacity phenotype. Thus, the colony opacity phenotype "switch" in *S. pneumoniae* is complex and may be a multi-factorial event with no single mechanism responsible for variation. Rather, it appears that a combination gene and protein expression patterns contributes to each phenotype, which may also contribute to their pathogenic characteristics. Gaining a better understanding of pneumococcal opacity phase variation is clearly important, as it has been shown to correlate with the transition from asymptomatic nasopharyngeal colonisation progression of the infection into deeper tissues thereby causing disease. The underlying mechanism is still unknown, but this study has provided an insight into the phenomic, transcriptomic, genomic and epigenetic differences between the O/T variants.

7.3 Future directions

This study was initially structured to identify a consistent mechanism across all three strains using a combination of analytical approaches. As with any study of whole proteome, transcriptome and genome profiling, large quantities of data were generated from each of the approaches in this study. Hence, further investigation of the proteins, genes or sequence differences identified were prioritised, based on the criteria described in each of the chapters, including commonality between the strains, the strength of the fold-changes and statistical analyses, whether or not these differences were in regions that were previously associated with O/T phase variants, and/or linked to differences indentified using the other analytical approaches used in this study. However, the results obtained using each of the approaches indicate that the differences could not be attributed to one single mechanism. Furthermore, each strain appears to achieve opacity switching differently, notwithstanding some commonality in terms of the functional classes of

differentially expressed genes or proteins. Hence, further work should involve the investigation of strain-specific differences in more detail to determine whether there might indeed be a common underlying mechanism or pathway.

Proteomic analysis (Chapter 3) identified several examples of spots with distinct pI that belonged to different isoforms of a given protein. Although the total activity of GAPDH between O/T pairs has been investigated in this study (Table 3.4), other potentially interesting enzymes such as Ldh, AdhE, PflB, Gor, and Fba were found to be differentially regulated between the O/T pairs in at least one of the strains tested (Table 3.5). Confirmation of these differences at the enzyme activity level is clearly warranted. Additionally, metabolomic profiles for substrates such as pyruvate, lactate or formate could assist in mapping out a biosynthesis pathway to ascertain whether there is a metabolic preference towards a particular substrate that differs between the phase variants, which may or may not affect their pathogenicity.

Opacity phase variants have been shown to vary in the relative amounts of capsule and TA they produce (Kim & Weiser, 1998), but there were no detectable changes in the expression of known capsule or TA synthesis genes/proteins in this study. This suggests that production of these macromolecules may be regulated by changes in the relative intracellular concentrations of precursors for capsule or TA biosynthesis, such as UMP or UDP-glucose. Pathways involved in synthesis of these precursors are clearly worthy of investigation at the enzyme activity level. One way this can be achieved is through metabolic profiling, or metabolomics. However, it may also be of interest to assess other metabolites, such as those involved in glycolysis or pyruvate, since these proteins and genes appear to be differentially regulated between the O/T pairs. This would enable mapping of metabolic pathways to further understand the differences between these two phenotypes. Furthermore, other protein spots that were identified as different between O and T variants of only one or two strains should be validated either biochemically or by quantitative Western blot, to provide a more comprehensive understanding of O/T protein expression in specific strains.

This study utilised microarray hybridisation analysis to investigate whole transcriptome profiles of O/T pairs of the three strains used in this study. The microarray slides used in

this study contained cDNA sequences covering ORFs in the published *S. pneumoniae* TIGR4 genome (Tettelin *et al.*, 2001), and additional ORFs present in the R6 genome (Hoskins *et al.*, 2001). Thus, ORFs not present in these two strains, as well as non-coding RNAs will be missed. Recent advances in sequencing technology mean that RNA sequencing (RNA-Seq) (Wang *et al.*, 2009) is now more reliable and economical, and would provide a wider and more precise measurement of transcripts for the study of whole transcriptome profiles. This is because, unlike hybridisation-based methods such as microarray analysis, RNA-seq is not confined to identified ORFs and does not rely on an existing genome sequence. RNAseq may also provide more robust data for low-abundance transcripts. Hence, it would be of interest to perform RNA-seq analysis on the phase variants of the three strains to determine if there are any differentially regulated genes that are not be present in the TIGR4/R6 genomes, or any other coding or non-coding regulatory RNAs that were not represented on the microarray slide. If indeed there are genes of interest that are detected, mutagenesis studies would be carried out to observe whether or not these are associated with a certain phenotype.

This study employed the IonTorrent^{TM} PGM genome sequencing, as it was the technology available at the time. Although the IonTorrent technology appears to be accurate for SNP calling, there were short-comings, namely with accurately identifying INDELs and sequencing homopolymeric tracts (Section 5.3). Furthermore, due to the short read lengths, *de novo* assembly using IonTorrent gives a fragmented draft genome. This shortcoming impacted the analysis of WCH16, as there is no publicly available genome for this strain or close relatives. Therefore, the WCH16 genome should be sequenced using SMRT sequencing (Brakmann, 2010; Eid *et al.*, 2009) available from Pacific Biosciences (PacBio). The longer read length is better suited for *de novo* assembly. Furthermore, it is able to detect epigenetic variations.

Finally, Chapter 6 described the study of six epigenetically different subpopulations of the pneumococcus in strain D39. Moreover, the distinct methylomes of these locked mutants had an impact on virulence and the expression of multiple genes. Preliminary data show that not all of the alternative HsdS alleles are present in WCH16 and WCH43 (Oggioni, M., unpublished data). Hence, it may be of interest to further investigate other

methyltransferase sites that may be present in all three strains. Furthermore, in strain D39, SpnD39III mutants were constructed in a D39 parent containing a mixture of both opacity types. It could be interesting to observe whether the same phenotypes can be achieved if the mutagenesis is carried out on homogeneous D39O and D39T cultures, and whether or not previously D39O strains would switch phenotype when locked into a particular methylation pattern, and vice versa. Collectively, these experiments would provide a better understanding of the molecular basis of pneumococcal opacity phase variation and pathogenesis.

References

ACIP (1997). Prevention of pneumococcal disease: recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Recomm Rep* **46**, 1-24.

Adamou, J. E., Heinrichs, J. H., Erwin, A. L., Walsh, W., Gayle, T., Dormitzer, M., Dagan, R., Brewah, Y. A., Barren, P., Lathigra, R., Langermann, S., Koenig, S. & Johnson, S. (2001). Identification and characterization of a novel family of pneumococcal proteins that are protective against sepsis. *Infect Immun* **69**, 949-958.

Ali, Y. M., Lynch, N. J., Haleem, K. S., Fujita, T., Endo, Y., Hansen, S., Holmskov, U., Takahashi, K., Stahl, G. L., Dudler, T., Girija, U. V., Wallis, R., Kadioglu, A., Stover, C. M., Andrew, P. W. & Schwaeble, W. J. (2012). The lectin pathway of complement activation is a critical component of the innate immune response to pneumococcal infection. *PLoS Pathog* **8**, e1002793.

Allegrucci, M. & 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.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-410.

Amin, Z., Harvey, R. M., Wang, H., Hughes, C. E., Paton, A. W., Paton, J. C. & Trappetti, C. (2015). Isolation site influences virulence phenotype of Serotype 14 *Streptococcus pneumoniae* strains belonging to Multilocus Sequence Type 15. *Infect Immun* **83**, 4781-4790.

Andersson, B., Dahmen, J., Frejd, T., Leffler, H., Magnusson, G., Noori, G. & Eden, C. S. (1983). Identification of an active disaccharide unit of a glycoconjugate receptor for pneumococci attaching to human pharyngeal epithelial cells. *J Exp Med* **158**, 559-570.

Andrews, S. C. & Guest, J. R. (1988). Nucleotide sequence of the gene encoding the GMP reductase of *Escherichia coli* K12. *Biochem J* **255**, 35-43.

Arguedas, A., Kvaerner, K., Liese, J., Schilder, A. G. & Pelton, S. I. (2010). Otitis media across nine countries: Disease burden and management. *Int J Pediatr Otorhinolaryngol* **74**, 1419-1424.

Arrecubieta, C., Garcia, E. & Lopez, R. (1995). Sequence and transcriptional analysis of a DNA region involved in the production of capsular polysaccharide in *Streptococcus pneumoniae* type 3. *Gene* **167**, 1-7.

Atack, J. M., Ibranovic, I., Ong, C. L., Djoko, K. Y., Chen, N. H., Vanden Hoven, R., Jennings, M. P., Edwards, J. L. & McEwan, A. G. (2014). A role for lactate

dehydrogenases in the survival of *Neisseria gonorrhoeae* in human polymorphonuclear leukocytes and cervical epithelial cells. *J Infect Dis* **2014**.

Attaiech, L., Olivier, A., Mortier-Barriere, I., Soulet, A. L., Granadel, C., Martin, B., Polard, P. & Claverys, J. P. (2011). Role of the single-stranded DNA-binding protein SsbB in pneumococcal transformation: maintenance of a reservoir for genetic plasticity. *PLoS Genet* **7**, e1002156.

Bae, S. M., Yeon, S. M., Kim, T. S. & Lee, K. J. (2006). The effect of protein expression of *Streptococcus pneumoniae* by blood. *J Biochem Mol Biol* **39**, 703-708.

Bakaletz, L. O. (2012). Bacterial biofilms in the upper airway - evidence for role in pathology and implications for treatment of otitis media. *Paediatr Respir Rev* **13**, 154-159.

Balachandran, P., Hollingshead, S. K., Paton, J. C. & Briles, D. E. (2001). The autolytic enzyme LytA of *Streptococcus pneumoniae* is not responsible for releasing pneumolysin. *J Bacteriol* **183**, 3108-3116.

Banerjee, A., Van Sorge, N. M., Sheen, T. R., Uchiyama, S., Mitchell, T. J. & Doran, K. S. (2010). Activation of brain endothelium by pneumococcal neuraminidase NanA promotes bacterial internalization. *Cell Microbiol* **12**, 1576-1588.

Bart, M. J., van Gent, M., van der Heide, H. G., Boekhorst, J., Hermans, P., Parkhill, J. & Mooi, F. R. (2010). Comparative genomics of prevaccination and modern *Bordetella pertussis* strains. *BMC Genomics* **11**, 627.

Basavanna, S., Khandavilli, S., Yuste, J., Cohen, J. M., Hosie, A. H., Webb, A. J., Thomas, G. H. & Brown, J. S. (2009). Screening of *Streptococcus pneumoniae* ABC transporter mutants demonstrates that LivJHMGF, a branched-chain amino acid ABC transporter, is necessary for disease pathogenesis. *Infect Immun* **77**, 3412-3423.

Bayle, L., Chimalapati, S., Schoehn, G., Brown, J., Vernet, T. & Durmort, C. (2011). Zinc uptake by *Streptococcus pneumoniae* depends on both AdcA and AdcAII and is essential for normal bacterial morphology and virulence. *Mol Microbiol* **82**, 904-916.

Beard, S. J., Salisbury, V., Lewis, R. J., Sharpe, J. A. & MacGowan, A. P. (2002). Expression of lux genes in a clinical isolate of *Streptococcus pneumoniae*: using bioluminescence to monitor gemifloxacin activity. *Antimicrob Agents Chemother* **46**, 538-542.

Belanger, A. E., Clague, M. J., Glass, J. I. & Leblanc, D. J. (2004). Pyruvate oxidase is a determinant of Avery's rough morphology. *J Bacteriol* **186**, 8164-8171.

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., Hoschler, K., Hurwitz, S., Ivanov, D. V., Johnson, M. Q., James, T., Huw Jones, T. A., Kang, G. D., Kerelska, T. H., Kersey, A. D., Khrebtukova, I., Kindwall, A. P., Kingsbury, Z., Kokko-Gonzales, P. I., Kumar, A., Laurent, M. A., Lawley, C. T., Lee, S. E., Lee, X., Liao, A. K., Loch, J. A., Lok, M., Luo, S., Mammen, R. M., Martin, J. W., McCauley, P. G., McNitt, P., Mehta, P., Moon, K. W., Mullens, J. W., Newington, T., Ning, Z., Ling Ng, B., Novo, S. M., O'Neill, M. J., Osborne, M. A., Osnowski, A., Ostadan, O., Paraschos, L. L., Pickering, L., Pike, A. C., Pike, A. C., Chris Pinkard, D., Pliskin, D. P., Podhasky, J., Quijano, V. J., Raczy, C., Rae, V. H., Rawlings, S. R., Chiva Rodriguez, A., Roe, P. M., Rogers, J., Rogert Bacigalupo, M. C., Romanov, N., Romieu, A., Roth, R. K., Rourke, N. J., Ruediger, S. T., Rusman, E., Sanches-Kuiper, R. M., Schenker, M. R., Seoane, J. M., Shaw, R. J., Shiver, M. K., Short, S. W., Sizto, N. L., Sluis, J. P., Smith, M. A., Ernest Sohna Sohna, J., Spence, E. J., Stevens, K., Sutton, N., Szajkowski, L., Tregidgo, C. L., Turcatti, G., Vandevondele, S., Verhovsky, Y., Virk, S. M., Wakelin, S., Walcott, G. C., Wang, J., Worsley, G. J., Yan, J., Yau, L., Zuerlein, M., Rogers, J., Mullikin, J. C., Hurles, M. E., McCooke, N. J., West, J. S., Oaks, F. L., Lundberg, P. L., Klenerman, D., Durbin, R. & Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53-59.

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

Benton, K. A., Everson, M. P. & Briles, D. E. (1995). A pneumolysin-negative mutant of *Streptococcus pneumoniae* causes chronic bacteremia rather than acute sepsis in mice. *Infect Immun* **63**, 448-455.

Benton, K. A., Paton, J. C. & Briles, D. E. (1997). Differences in virulence for mice among *Streptococcus pneumoniae* strains of capsular types 2, 3, 4, 5, and 6 are not attributable to differences in pneumolysin production. *Infect Immun* **65**, 1237-1244.

Bermpohl, D., Halle, A., Freyer, D., Dagand, E., Braun, J. S., Bechmann, I., Schroder, N. W. & Weber, J. R. (2005). Bacterial programmed cell death of cerebral endothelial cells involves dual death pathways. *J Clin Invest* **115**, 1607-1615.

Berry, A. M., Lock, R. A., Hansman, D. & Paton, J. C. (1989a). Contribution of autolysin to virulence of *Streptococcus pneumoniae*. *Infect Immun* **57**, 2324-2330.

Berry, A. M., Yother, J., Briles, D. E., Hansman, D. & Paton, J. C. (1989b). Reduced virulence of a defined pneumolysin-negative mutant of *Streptococcus pneumoniae*. *Infect Immun* **57**, 2037-2042.

Berry, A. M., Lock, R. A. & Paton, J. C. (1996). Cloning and characterization of *nanB*, a second *Streptococcus pneumoniae* neuraminidase gene, and purification of the NanB enzyme from recombinant *Escherichia coli*. *J Bacteriol* **178**, 4854-4860.

Berry, A. M. & Paton, J. C. (1996). Sequence heterogeneity of PsaA, a 37-kilodalton putative adhesin essential for virulence of *Streptococcus pneumoniae*. *Infect Immun* **64**, 5255-5262.

Berry, A. M. & Paton, J. C. (2000). Additive attenuation of virulence of *Streptococcus pneumoniae* by mutation of the genes encoding pneumolysin and other putative pneumococcal virulence proteins. *Infect Immun* **68**, 133-140.

Bertani, G. & Weigle, J. J. (1953). Host controlled variation in bacterial viruses. *J Bacteriol* **65**, 113-121.

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. & 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.

Black, S., Shinefield, H., Baxter, R., Austrian, R., Elvin, L., Hansen, J., Lewis, E. & Fireman, B. (2006). Impact of the use of heptavalent pneumococcal conjugate vaccine on disease epidemiology in children and adults. *Vaccine* **24 Suppl 2**, S2-79-80.

Blau, K., Portnoi, M., Shagan, M., Kaganovich, A., Rom, S., Kafka, D., Chalifa Caspi, V., Porgador, A., Givon-Lavi, N., Gershoni, J. M., Dagan, R. & Mizrachi Nebenzahl, Y. (2007). Flamingo cadherin: a putative host receptor for *Streptococcus pneumoniae*. *J Infect Dis* **195**, 1828-1837.

Blyn, L. B., Braaten, B. A. & Low, D. A. (1990). Regulation of pap pilin phase variation by a mechanism involving differential dam methylation states. *Embo J* **9**, 4045-4054.

Bogaert, D., De Groot, R. & Hermans, P. W. (2004). *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. *Lancet Infect Dis* **4**, 144-154.

Boyer, H. W. & Roulland-Dussoix, D. (1969). A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. *J Mol Biol* **41**, 459-472.

Braaten, B. A., Nou, X., Kaltenbach, L. S. & Low, D. A. (1994). Methylation patterns in pap regulatory DNA control pyelonephritis-associated pili phase variation in *E. coli*. *Cell* **76**, 577-588.

Bragg, L. M., Stone, G., Butler, M. K., Hugenholtz, P. & Tyson, G. W. (2013). Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. *PLoS Comput Biol* **9**, e1003031.

Brakmann, S. (2010). Single-molecule analysis: A ribosome in action. *Nature* **464**, 987-988.

Braun, J. S., Sublett, J. E., Freyer, D., Mitchell, T. J., Cleveland, J. L., Tuomanen, E. I. & Weber, J. R. (2002). Pneumococcal pneumolysin and H(2)O(2) mediate brain cell apoptosis during meningitis. *J Clin Invest* **109**, 19-27.

Briles, D. E., Yother, J. & McDaniel, L. S. (1988). Role of pneumococcal surface protein A in the virulence of *Streptococcus pneumoniae*. *Rev Infect Dis* **10 Suppl 2**, S372-374.

Briles, D. E., Hollingshead, S. K., Paton, J. C., Ades, E. W., Novak, L., van Ginkel, F. W. & Benjamin, W. H., Jr. (2003). Immunizations with pneumococcal surface protein A and pneumolysin are protective against pneumonia in a murine model of pulmonary infection with *Streptococcus pneumoniae*. *J Infect Dis* **188**, 339-348.

Briles, D. E., Novak, L., Hotomi, M., van Ginkel, F. W. & King, J. (2005). Nasal colonization with *Streptococcus pneumoniae* includes subpopulations of surface and invasive pneumococci. *Infect Immun* **73**, 6945-6951.

Brittan, J. L., Buckeridge, T. J., Finn, A., Kadioglu, A. & Jenkinson, H. F. (2012). Pneumococcal neuraminidase A: an essential upper airway colonization factor for *Streptococcus pneumoniae*. *Mol Oral Microbiol* **27**, 270-283.

Britton, R. A., Powell, B. S., Dasgupta, S., Sun, Q., Margolin, W., Lupski, J. R. & Court, D. L. (1998). Cell cycle arrest in Era GTPase mutants: a potential growth rate-regulated checkpoint in *Escherichia coli*. *Mol Microbiol* **27**, 739-750.

Broome, C. V. & Facklam, R. R. (1981). Epidemiology of clinically significant isolates of *Streptococcus pneumoniae* in the United States. *Rev Infect Dis* **3**, 277-281.

Brown, J. S., Gilliland, S. M. & Holden, D. W. (2001). A *Streptococcus pneumoniae* pathogenicity island encoding an ABC transporter involved in iron uptake and virulence. *Mol Microbiol* **40**, 572-585.

Brown, J. S., Gilliland, S. M., Ruiz-Albert, J. & Holden, D. W. (2002a). Characterization of pit, a *Streptococcus pneumoniae* iron uptake ABC transporter. *Infect Immun* **70**, 4389-4398.

Brown, J. S., Hussell, T., Gilliland, S. M., Holden, D. W., Paton, J. C., Ehrenstein, M. R., Walport, M. J. & Botto, M. (2002b). The classical pathway is the dominant complement pathway required for innate immunity to *Streptococcus pneumoniae* infection in mice. *Proc Natl Acad Sci U S A* **99**, 16969-16974.

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

Byrne, J. P., Morona, J. K., Paton, J. C. & Morona, R. (2011). Identification of *Streptococcus pneumoniae* Cps2C residues that affect capsular polysaccharide polymerization, cell wall ligation, and Cps2D phosphorylation. *J Bacteriol* **193**, 2341-2346.

Caballero, J. & Rello, J. (2011). Combination antibiotic therapy for community-acquired pneumonia. *Ann Intensive Care* **1**, 48.

Caldas, T., Laalami, S. & Richarme, G. (2000). Chaperone properties of bacterial elongation factor EF-G and initiation factor IF2. *J Biol Chem* **275**, 855-860.

Camara, M., Boulnois, G. J., Andrew, P. W. & Mitchell, T. J. (1994). A neuraminidase from *Streptococcus pneumoniae* has the features of a surface protein. *Infect Immun* **62**, 3688-3695.

Canvin, J. R., Marvin, A. P., Sivakumaran, M., Paton, J. C., Boulnois, G. J., Andrew, P. W. & Mitchell, T. J. (1995). The role of pneumolysin and autolysin in the pathology of pneumonia and septicemia in mice infected with a type 2 pneumococcus. *J Infect Dis* **172**, 119-123.

Carvalho, S. M., Farshchi Andisi, V., Gradstedt, H., Neef, J., Kuipers, O. P., Neves, A. R. & Bijlsma, J. J. (2013). Pyruvate oxidase influences the sugar utilization pattern and capsule production in *Streptococcus pneumoniae*. *PLoS One* **8**, e68277.

Cerdeno-Tarraga, A. M., Patrick, S., Crossman, L. C., Blakely, G., Abratt, V., Lennard, N., Poxton, I., Duerden, B., Harris, B., Quail, M. A., Barron, A., Clark, L., Corton, C., Doggett, J., Holden, M. T., Larke, N., Line, A., Lord, A., Norbertczak, H., Ormond, D., Price, C., Rabbinowitsch, E., Woodward, J., Barrell, B. & Parkhill, J.

(2005). Extensive DNA inversions in the *B. fragilis* genome control variable gene expression. *Science* **307**, 1463-1465.

Cheng, Q., Campbell, E. A., Naughton, A. M., Johnson, S. & Masure, H. R. (1997). The com locus controls genetic transformation in *Streptococcus pneumoniae*. *Mol Microbiol* **23**, 683-692.

Cheng, Q., Finkel, D. & Hostetter, M. K. (2000). Novel purification scheme and functions for a C3-binding protein from *Streptococcus pneumoniae*. *Biochemistry* **39**, 5450-5457.

Chevreux, B., Wetter, T. & Suhai, S. (1999). Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. In *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB)*, pp. pp. 45-56.

Choi, S. C., Parker, J., Richards, V. P., Ross, K., Jilly, B. & Chen, J. (2014). Draft Genome Sequence of an Atypical Strain of *Streptococcus pneumoniae* Isolated from a Respiratory Infection. *Genome Announc* **2**, 00822-00814.

Cilloniz, C., Ewig, S., Menendez, R., Ferrer, M., Polverino, E., Reyes, S., Gabarrus, A., Marcos, M. A., Cordoba, J., Mensa, J. & Torres, A. (2012). Bacterial co-infection with H1N1 infection in patients admitted with community acquired pneumonia. *J Infect* **65**, 223-230.

Claiborne, A. (1986). Studies on the structure and mechanism of *Streptococcus faecium* L-alpha-glycerophosphate oxidase. *J Biol Chem* **261**, 14398-14407.

Cockeran, R., Theron, A. J., Steel, H. C., Matlola, N. M., Mitchell, T. J., Feldman, C. & Anderson, R. (2001). Proinflammatory interactions of pneumolysin with human neutrophils. *J Infect Dis* **183**, 604-611.

Cohen, R., Varon, E., Doit, C., Schlemmer, C., Romain, O., Thollot, F., Bechet, S., Bonacorsi, S. & Levy, C. (2015). A 13-year survey of pneumococcal nasopharyngeal carriage in children with acute otitis media following PCV7 and PCV13 implementation. *Vaccine* **2015**, 010.

Colino, J. & Snapper, C. M. (2003). Two distinct mechanisms for induction of dendritic cell apoptosis in response to intact *Streptococcus pneumoniae*. *J Immunol* **171**, 2354-2365.

Correnti, J., Munster, V., Chan, T. & Woude, M. (2002). Dam-dependent phase variation of Ag43 in *Escherichia coli* is altered in a *seqA* mutant. *Mol Microbiol* **44**, 521-532.

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

Croucher, N. J., Harris, S. R., Fraser, C., Quail, M. A., Burton, J., van der Linden, M., McGee, L., von Gottberg, A., Song, J. H., Ko, K. S., Pichon, B., Baker, S., Parry, C. M., Lambertsen, L. M., Shahinas, D., Pillai, D. R., Mitchell, T. J., Dougan, G., Tomasz, A., Klugman, K. P., Parkhill, J., Hanage, W. P. & Bentley, S. D. (2011). Rapid pneumococcal evolution in response to clinical interventions. *Science* **331**, 430-434.

Cundell, D. R., Gerard, N. P., Gerard, C., Idanpaan-Heikkila, I. & Tuomanen, E. I. (1995a). *Streptococcus pneumoniae* anchor to activated human cells by the receptor for platelet-activating factor. *Nature* **377**, 435-438.

Cundell, D. R., Weiser, J. N., Shen, J., Young, A. & Tuomanen, E. I. (1995b). Relationship between colonial morphology and adherence of *Streptococcus pneumoniae*. *Infect Immun* **63**, 757-761.

Dave, S., Carmicle, S., Hammerschmidt, S., Pangburn, M. K. & 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.

de Vries, F. P., van Der Ende, A., van Putten, J. P. & Dankert, J. (1996). Invasion of primary nasopharyngeal epithelial cells by *Neisseria meningitidis* is controlled by phase variation of multiple surface antigens. *Infect Immun* **64**, 2998-3006.

Denapaite, D., Bruckner, R., Hakenbeck, R. & Vollmer, W. (2012). Biosynthesis of teichoic acids in *Streptococcus pneumoniae* and closely related species: lessons from genomes. *Microb Drug Resist* **18**, 344-358.

Dillard, J. P., Vandersea, M. W. & Yother, J. (1995). Characterization of the cassette containing genes for type 3 capsular polysaccharide biosynthesis in *Streptococcus pneumoniae*. *J Exp Med* **181**, 973-983.

Dintilhac, A., Alloing, G., Granadel, C. & Claverys, J. P. (1997). Competence and virulence of *Streptococcus pneumoniae*: Adc and PsaA mutants exhibit a requirement for Zn and Mn resulting from inactivation of putative ABC metal permeases. *Mol Microbiol* **25**, 727-739.

Dochez, A. R. & Avery, O. T. (1917). The elaboration of specific soluble substance by pneumococcus during growth. *J Exp Med* **26**, 477-493.

Douglas, R. M., Paton, J. C., Duncan, S. J. & Hansman, D. J. (1983). Antibody response to pneumococcal vaccination in children younger than five years of age. *J Infect Dis* **148**, 131-137.

Douglass, J. & Blumenthal, T. (1979). Conformational transition of protein synthesis elongation factor Tu induced by guanine nucleotides. Modulation by kirromycin and elongation factor Ts. *J Biol Chem* **254**, 5383-5387.

Dowson, C. G., Hutchison, A., Brannigan, J. A., George, R. C., Hansman, D., Linares, J., Tomasz, A., Smith, J. M. & Spratt, B. G. (1989). Horizontal transfer of penicillin-binding protein genes in penicillin-resistant clinical isolates of *Streptococcus pneumoniae*. *Proc Natl Acad Sci U S A* **86**, 8842-8846.

Duggan, S. T. (2010). Pneumococcal polysaccharide conjugate vaccine (13-valent, adsorbed) [prevenar 13(R)]. *Drugs* **70**, 1973-1986.

Dybvig, K., Sitaraman, R. & French, C. T. (1998). A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements. *Proc Natl Acad Sci U S A* **95**, 13923-13928.

Eberhardt, A., Hoyland, C. N., Vollmer, D., Bisle, S., Cleverley, R. M., Johnsborg, O., Havarstein, L. S., Lewis, R. J. & Vollmer, W. (2012). Attachment of capsular polysaccharide to the cell wall in *Streptococcus pneumoniae*. *Microb Drug Resist* **18**, 240-255.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., Dewinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korlach, J. & Turner, S. (2009). Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133-138.

Elliott, A. M., Radecki, J., Moghis, B., Li, X. & Kammesheidt, A. (2012). Rapid detection of the ACMG/ACOG-recommended 23 CFTR disease-causing mutations using ion torrent semiconductor sequencing. *J Biomol Tech* **23**, 24-30.

Engelmoer, D. J. & Rozen, D. E. (2011). Competence increases survival during stress in *Streptococcus pneumoniae*. *Evolution* **65**, 3475-3485.

Estabrook, M. M., Griffiss, J. M. & Jarvis, G. A. (1997). Sialylation of *Neisseria meningitidis* lipooligosaccharide inhibits serum bactericidal activity by masking lacto-*N*-neotetraose. *Infect Immun* **65**, 4436-4444.

Euteneuer, B., Storkel, S. & Loos, M. (1986). Contributions of C1q, bacterial lipopolysaccharide, and porins during attachment and ingestion phases of phagocytosis by murine macrophages. *Infect Immun* **51**, 807-815.

Faden, H., Stanievich, J., Brodsky, L., Bernstein, J. & Ogra, P. L. (1990). Changes in nasopharyngeal flora during otitis media of childhood. *Pediatr Infect Dis J* **9**, 623-626.

Faden, H., Duffy, L., Wasielewski, R., Wolf, J., Krystofik, D. & Tung, Y. (1997). Relationship between nasopharyngeal colonization and the development of otitis media in children. Tonawanda/Williamsville Pediatrics. *J Infect Dis* **175**, 1440-1445.

Fang, G., Munera, D., Friedman, D. I., Mandlik, A., Chao, M. C., Banerjee, O., Feng, Z., Losic, B., Mahajan, M. C., Jabado, O. J., Deikus, G., Clark, T. A., Luong, K., Murray, I. A., Davis, B. M., Keren-Paz, A., Chess, A., Roberts, R. J., Korlach, J., Turner, S. W., Kumar, V., Waldor, M. K. & Schadt, E. E. (2012). Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nat Biotechnol* **30**, 1232-1239.

Feldman, C., Mitchell, T. J., Andrew, P. W., Boulnois, G. J., Read, R. C., Todd, H. C., Cole, P. J. & Wilson, R. (1990). The effect of *Streptococcus pneumoniae* pneumolysin on human respiratory epithelium in vitro. *Microb Pathog* **9**, 275-284.

Feldman, C., Cockeran, R., Jedrzejas, M. J., Mitchell, T. J. & Anderson, R. (2007). Hyaluronidase augments pneumolysin-mediated injury to human ciliated epithelium. *Int J Infect Dis* **11**, 11-15.

Feldman, C. & Anderson, R. (2014). Review: current and new generation pneumococcal vaccines. *J Infect* **69**, 309-325.

Fernebro, J., Andersson, I., Sublett, J., Morfeldt, E., Novak, R., Tuomanen, E., Normark, S. & Normark, B. H. (2004). Capsular expression in *Streptococcus pneumoniae* negatively affects spontaneous and antibiotic-induced lysis and contributes to antibiotic tolerance. *J Infect Dis* **189**, 328-338.

Fillinger, S., Boschi-Muller, S., Azza, S., Dervyn, E., Branlant, G. & Aymerich, S. (2000). Two glyceraldehyde-3-phosphate dehydrogenases with opposite physiological roles in a nonphotosynthetic bacterium. *J Biol Chem* **275**, 14031-14037.

Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A., Korlach, J. & Turner, S. W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* **7**, 461-465.

Friedland, I. R., Paris, M. M., Hickey, S., Shelton, S., Olsen, K., Paton, J. C. & McCracken, G. H. (1995). The limited role of pneumolysin in the pathogenesis of pneumococcal meningitis. *J Infect Dis* **172**, 805-809.

Fu, H. S., Hassett, D. J. & Cohen, M. S. (1989). Oxidant stress in *Neisseria gonorrhoeae*: adaptation and effects on L-(+)-lactate dehydrogenase activity. *Infect Immun* **57**, 2173-2178.

Garcia, J. L., Sanchez-Beato, A. R., Medrano, F. J. & Lopez, R. (1998). Versatility of choline-binding domain. *Microb Drug Resist* **4**, 25-36.

Garriga, X., Eliasson, R., Torrents, E., Jordan, A., Barbe, J., Gibert, I. & Reichard, P. (1996). nrdD and nrdG genes are essential for strict anaerobic growth of *Escherichia coli*. *Biochem Biophys Res Commun* **229**, 189-192.

GAVI (2013). Pneumococcal vaccine support. Retrieved from http://www.gavi.org/support/nvs/pneumococcal/#.

Geier, G. E. & Modrich, P. (1979). Recognition sequence of the dam methylase of *Escherichia coli* K12 and mode of cleavage of Dpn I endonuclease. *J Biol Chem* **254**, 1408-1413.

Gertz, R. E., Jr., Li, Z., Pimenta, F. C., Jackson, D., Juni, B. A., Lynfield, R., Jorgensen, J. H., Carvalho Mda, G. & Beall, B. W. (2010). Increased penicillin nonsusceptibility of nonvaccine-serotype invasive pneumococci other than serotypes 19A and 6A in post-7-valent conjugate vaccine era. *J Infect Dis* **201**, 770-775.

Ghaffar, F., Friedland, I. R. & McCracken, G. H., Jr. (1999). Dynamics of nasopharyngeal colonization by *Streptococcus pneumoniae*. *Pediatr Infect Dis J* **18**, 638-646.

Giacomodonato, M. N., Llana, M. N., Castaneda Mdel, R., Buzzola, F., Garcia, M. D., Calderon, M. G., Sarnacki, S. H. & Cerquetti, M. C. (2014). Dam methylation regulates the expression of SPI-5-encoded *sopB* gene in *Salmonella enterica* serovar Typhimurium. *Microbes Infect* **16**, 615-622.

Giannakis, E., Male, D. A., Ormsby, R. J., Mold, C., Jokiranta, T. S., Ranganathan, S. & Gordon, D. L. (2001). Multiple ligand binding sites on domain seven of human complement factor H. *Int Immunopharmacol* **1**, 433-443.

Giffard, P. M. & Jacques, N. A. (1994). Definition of a fundamental repeating unit in streptococcal glucosyltransferase glucan-binding regions and related sequences. *J Dent Res* **73**, 1133-1141.

Goffin, P., Muscariello, L., Lorquet, F., Stukkens, A., Prozzi, D., Sacco, M., Kleerebezem, M. & Hols, P. (2006). Involvement of pyruvate oxidase activity and acetate production in the survival of *Lactobacillus plantarum* during the stationary phase of aerobic growth. *Appl Environ Microbiol* **72**, 7933-7940.

Gollop, N. & March, P. E. (1991). A GTP-binding protein (Era) has an essential role in growth rate and cell cycle control in *Escherichia coli*. *J Bacteriol* **173**, 2265-2270.

Goossens, H., Ferech, M., Coenen, S. & Stephens, P. (2007). Comparison of outpatient systemic antibacterial use in 2004 in the United States and 27 European countries. *Clin Infect Dis* **44**, 1091-1095.

Gosink, K. K., Mann, E. R., Guglielmo, C., Tuomanen, E. I. & Masure, H. R. (2000). Role of novel choline binding proteins in virulence of *Streptococcus pneumoniae*. *Infect Immun* **68**, 5690-5695.

Haagmans, W. & van der Woude, M. (2000). Phase variation of Ag43 in *Escherichia coli*: Dam-dependent methylation abrogates OxyR binding and OxyR-mediated repression of transcription. *Mol Microbiol* **35**, 877-887.

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. & 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. & Stoodley, P. (2009). Evolving concepts in biofilm infections. *Cell Microbiol* **11**, 1034-1043.

Hamel, J., Charland, N., Pineau, I., Ouellet, C., Rioux, S., Martin, D. & Brodeur, B. R. (2004). Prevention of pneumococcal disease in mice immunized with conserved surface-accessible proteins. *Infect Immun* **72**, 2659-2670.

Hammerschmidt, S., Muller, A., Sillmann, H., Muhlenhoff, M., Borrow, R., Fox, A., van Putten, J., Zollinger, W. D., Gerardy-Schahn, R. & Frosch, M. (1996). Capsule phase variation in *Neisseria meningitidis* serogroup B by slipped-strand mispairing in the polysialyltransferase gene (*siaD*): correlation with bacterial invasion and the outbreak of meningococcal disease. *Mol Microbiol* **20**, 1211-1220.

Hammerschmidt, S., Talay, S. R., Brandtzaeg, P. & Chhatwal, G. S. (1997). SpsA, a novel pneumococcal surface protein with specific binding to secretory immunoglobulin A and secretory component. *Mol Microbiol* **25**, 1113-1124.

Hammerschmidt, S., Wolff, S., Hocke, A., Rosseau, S., Muller, E. & Rohde, M. (2005). Illustration of pneumococcal polysaccharide capsule during adherence and invasion of epithelial cells. *Infect Immun* **73**, 4653-4667.

Harboe, Z. B., Thomsen, R. W., Riis, A., Valentiner-Branth, P., Christensen, J. J., Lambertsen, L., Krogfelt, K. A., Konradsen, H. B. & Benfield, T. L. (2009). Pneumococcal serotypes and mortality following invasive pneumococcal disease: a population-based cohort study. *PLoS Med* **6**, e1000081.

Hardy, G. G., Caimano, M. J. & Yother, J. (2000). Capsule biosynthesis and basic metabolism in *Streptococcus pneumoniae* are linked through the cellular phosphoglucomutase. *J Bacteriol* **182**, 1854-1863.

Hartel, T., Klein, M., Koedel, U., Rohde, M., Petruschka, L. & Hammerschmidt, S. (2011). Impact of glutamine transporters on pneumococcal fitness under infection-related conditions. *Infect Immun* **79**, 44-58.

Harvey, R. M. (2010). Genetic characterisation of *Streptococcus pneumoniae* serotype 1 isolates in relation to invasiveness. In *Microbiology and Immunology*, pp. 366. Adelaide: University of Adelaide.

Hattman, S. (1964). The Control of Host-Induced Modification by Phage P1. *Virology* **23**, 270-271.

Hava, D. L. & Camilli, A. (2002). Large-scale identification of serotype 4 *Streptococcus pneumoniae* virulence factors. *Mol Microbiol* **45**, 1389-1406.

Henderson, I. R. & Owen, P. (1999). The major phase-variable outer membrane protein of *Escherichia coli* structurally resembles the immunoglobulin A1 protease class of exported protein and is regulated by a novel mechanism involving Dam and *oxyR*. *J Bacteriol* **181**, 2132-2141.

Hendriksen, W. T., Bootsma, H. J., Estevao, S., Hoogenboezem, T., de Jong, A., de Groot, R., Kuipers, O. P. & Hermans, P. W. (2008). CodY of *Streptococcus pneumoniae*: link between nutritional gene regulation and colonization. *J Bacteriol* **190**, 590-601.

Herrick, J. & Sclavi, B. (2007). Ribonucleotide reductase and the regulation of DNA replication: an old story and an ancient heritage. *Mol Microbiol* **63**, 22-34.

Hicks, L. A., Harrison, L. H., Flannery, B., Hadler, J. L., Schaffner, W., Craig, A. S., Jackson, D., Thomas, A., Beall, B., Lynfield, R., Reingold, A., Farley, M. M. & Whitney, C. G. (2007). Incidence of pneumococcal disease due to non-pneumococcal conjugate vaccine (PCV7) serotypes in the United States during the era of widespread PCV7 vaccination, 1998-2004. *J Infect Dis* **196**, 1346-1354.

Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51-59.

Honma, Y., Fernandez, R. E. & Maurelli, A. T. (2004). A DNA adenine methylase mutant of *Shigella flexneri* shows no significant attenuation of virulence. *Microbiology* **150**, 1073-1078.

Hoskins, J., Alborn, W. E., Jr., Arnold, J., Blaszczak, L. C., Burgett, S., DeHoff, B. S., Estrem, S. T., Fritz, L., Fu, D. J., Fuller, W., Geringer, C., Gilmour, R., Glass, J. S., Khoja, H., Kraft, A. R., Lagace, R. E., LeBlanc, D. J., Lee, L. N., Lefkowitz, E. J., Lu, J., Matsushima, P., McAhren, S. M., McHenney, M., McLeaster, K., Mundy, C. W., Nicas, T. I., Norris, F. H., O'Gara, M., Peery, R. B., Robertson, G. T., Rockey, P., Sun, P. M., Winkler, M. E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C. A., Baltz, R. H., Jaskunas, S. R., Rosteck, P. R., Jr., Skatrud, P. L. & Glass, J. I. (2001). Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J Bacteriol* **183**, 5709-5717.

Houldsworth, S., Andrew, P. W. & Mitchell, T. J. (1994). Pneumolysin stimulates production of tumor necrosis factor alpha and interleukin-1 beta by human mononuclear phagocytes. *Infect Immun* **62**, 1501-1503.

Hove-Jensen, B. (1988). Mutation in the phosphoribosylpyrophosphate synthetase gene (prs) that results in simultaneous requirements for purine and pyrimidine nucleosides, nicotinamide nucleotide, histidine, and tryptophan in *Escherichia coli*. *J Bacteriol* **170**, 1148-1152.

Howard, L. V. & Gooder, H. (1974). Specificity of the autolysin of *Streptococcus* (*Diplococcus*) *pneumoniae*. *J Bacteriol* **117**, 796-804.

Hubacek, J. & Glover, S. W. (1970). Complementation analysis of temperature-sensitive host specificity mutations in *Escherichia coli*. *J Mol Biol* **50**, 111-127.

Hui, F. M., Zhou, L. & Morrison, D. A. (1995). Competence for genetic transformation in *Streptococcus pneumoniae*: organization of a regulatory locus with homology to two lactococcin A secretion genes. *Gene* **153**, 25-31.

Hyams, C., Camberlein, E., Cohen, J. M., Bax, K. & Brown, J. S. (2010a). The *Streptococcus pneumoniae* capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. *Infect Immun* **78**, 704-715.

Hyams, C., Yuste, J., Bax, K., Camberlein, E., Weiser, J. N. & Brown, J. S. (2010b). *Streptococcus pneumoniae* resistance to complement-mediated immunity is dependent on the capsular serotype. *Infect Immun* **78**, 716-725.

Hyams, C., Camberlein, E., Cohen, J. M., Bax, K. & Brown, J. S. (2012). The *Streptococcus pneumoniae* capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. *Infect Immun* **78**, 704-715.

Iannelli, F., Oggioni, M. R. & Pozzi, G. (2002). Allelic variation in the highly polymorphic locus *pspC* of *Streptococcus pneumoniae*. *Gene* **284**, 63-71.

Iovino, F., Molema, G. & Bijlsma, J. J. (2014). Platelet endothelial cell adhesion molecule-1, a putative receptor for the adhesion of *Streptococcus pneumoniae* to the vascular endothelium of the blood-brain barrier. *Infect Immun* **82**, 3555-3566.

Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J. & Mann, M. (2005). Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* **4**, 1265-1272.

Islam, S. T. & Lam, J. S. (2013). Wzx flippase-mediated membrane translocation of sugar polymer precursors in bacteria. *Environ Microbiol* **15**, 1001-1015.

Ito, K., Fujiwara, T., Toyoda, T. & Nakamura, Y. (2002). Elongation factor G participates in ribosome disassembly by interacting with ribosome recycling factor at their tRNA-mimicry domains. *Mol Cell* **9**, 1263-1272.

Iyer, R. & Camilli, A. (2007). Sucrose metabolism contributes to in vivo fitness of *Streptococcus pneumoniae*. *Mol Microbiol* **66**, 1-13.

Janosi, L., Shimizu, I. & Kaji, A. (1994). Ribosome recycling factor (ribosome releasing factor) is essential for bacterial growth. *Proc Natl Acad Sci U S A* **91**, 4249-4253.

Janulaitis, A., Petrusyte, M., Maneliene, Z., Klimasauskas, S. & Butkus, V. (1992). Purification and properties of the Eco57I restriction endonuclease and methylase--prototypes of a new class (type IV). *Nucleic Acids Res* **20**, 6043-6049.

Janulczyk, R., Iannelli, F., Sjoholm, A. G., Pozzi, G. & Bjorck, L. (2000). Hic, a novel surface protein of *Streptococcus pneumoniae* that interferes with complement function. *J Biol Chem* **275**, 37257-37263.

Jekowsky, E., Schimmel, P. R. & Miller, D. L. (1977). Isolation, characterization and structural implications of a nuclease-digested complex of aminoacyl transfer RNA and Escherichia coli elongation factor Tu. *J Mol Biol* **114**, 451-458.

Jennings, M. P., Hood, D. W., Peak, I. R., Virji, M. & Moxon, E. R. (1995). Molecular analysis of a locus for the biosynthesis and phase-variable expression of the lacto-*N*-neotetraose terminal lipopolysaccharide structure in *Neisseria meningitidis*. *Mol Microbiol* **18**, 729-740.

Jennings, M. P., Srikhanta, Y. N., Moxon, E. R., Kramer, M., Poolman, J. T., Kuipers, B. & van der Ley, P. (1999). The genetic basis of the phase variation repertoire of lipopolysaccharide immunotypes in *Neisseria meningitidis*. *Microbiology* **145 (Pt 11)**, 3013-3021.

Jin, H., Agarwal, S., Agarwal, S. & Pancholi, V. (2011). Surface export of GAPDH/SDH, a glycolytic enzyme, is essential for *Streptococcus pyogenes* virulence. *MBio* **2**, e00068-00011.

Johnsborg, O. & Havarstein, L. S. (2009). Regulation of natural genetic transformation and acquisition of transforming DNA in *Streptococcus pneumoniae*. *FEMS Microbiol Rev* **33**, 627-642.

Jonsson, S., Musher, D. M., Chapman, A., Goree, A. & Lawrence, E. C. (1985). Phagocytosis and killing of common bacterial pathogens of the lung by human alveolar macrophages. *J Infect Dis* **152**, 4-13.

Josenhans, C., Eaton, K. A., Thevenot, T. & Suerbaum, S. (2000). Switching of flagellar motility in *Helicobacter pylori* by reversible length variation of a short homopolymeric sequence repeat in *fliP*, a gene encoding a basal body protein. *Infect Immun* **68**, 4598-4603.

Kadioglu, A., Taylor, S., Iannelli, F., Pozzi, G., Mitchell, T. J. & Andrew, P. W. (2002). Upper and lower respiratory tract infection by *Streptococcus pneumoniae* is affected by pneumolysin deficiency and differences in capsule type. *Infect Immun* **70**, 2886-2890.

Kadioglu, A., Weiser, J. N., Paton, J. C. & Andrew, P. W. (2008). The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. *Nat Rev Microbiol* **6**, 288-301.

Kallio, A., Sepponen, K., Hermand, P., Denoel, P., Godfroid, F. & Melin, M. (2014). Role of Pht proteins in attachment of *Streptococcus pneumoniae* to respiratory epithelial cells. *Infect Immun* **82**, 1683-1691.

Kass, R. E. & Raftery, A. E. (1995). Bayes factors. *J Am Statist Assoc* **90**, 773-795.

Kausmally, L., Johnsborg, O., Lunde, M., Knutsen, E. & Havarstein, L. S. (2005). Choline-binding protein D (CbpD) in *Streptococcus pneumoniae* is essential for competence-induced cell lysis. *J Bacteriol* **187**, 4338-4345.

Kearns, D. B., Chu, F., Rudner, R. & Losick, R. (2004). Genes governing swarming in *Bacillus subtilis* and evidence for a phase variation mechanism controlling surface motility. *Mol Microbiol* **52**, 357-369.

Kelly, T., Dillard, J. P. & Yother, J. (1994). Effect of genetic switching of capsular type on virulence of *Streptococcus pneumoniae*. *Infect Immun* **62**, 1813-1819.

Kempf, M., Baraduc, R., Bonnabau, H., Brun, M., Chabanon, G., Chardon, H., Croize, J., Demachy, M. C., Donnio, P. Y., Dupont, P., Fosse, T., Gibel, L., Gravet, A., Grignon, B., Hadou, T., Hamdad, F., Joly-Guillou, M. L., Koeck, J. L., Maugein, J., Pechinot, A., Ploy, M. C., Raymond, J., Ros, A., Roussel-Delvallez, M., Segonds, C., Vergnaud, M., Vernet-Garnier, V., Lepoutre, A., Gutmann, L., Varon, E. & Lanotte, P. (2010). Epidemiology and antimicrobial resistance of *Streptococcus pneumoniae* in France in 2007: Data from the Pneumococcus Surveillance Network. *Microb Drug Resist*.

Kenne, L., Lindberg, B. & Svensson, S. (1975). The structure of capsular polysaccharide of the pneumococcus type II. *Carbohydr Res* **40**, 69-75.

Kerr, A. R., Adrian, P. V., Estevao, S., de Groot, R., Alloing, G., Claverys, J. P., Mitchell, T. J. & Hermans, P. W. (2004). The Ami-AliA/AliB permease of *Streptococcus pneumoniae* is involved in nasopharyngeal colonization but not in invasive disease. *Infect Immun* **72**, 3902-3906.

Khan, M. N. & Pichichero, M. E. (2012). Vaccine candidates PhtD and PhtE of *Streptococcus pneumoniae* are adhesins that elicit functional antibodies in humans. *Vaccine* **30**, 2900-2907.

Kim, J. O. & 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., Hippe, K. R., Gould, J. M., Bae, D., Peterson, S., Cline, R. T., Fasching, C., Janoff, E. N. & Weiser, J. N. (2004). Phase variable desialylation of host proteins that bind to *Streptococcus pneumoniae* in vivo and protect the airway. *Mol Microbiol* **54**, 159-171.

King, S. J., Whatmore, A. M. & 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.

King, S. J., Hippe, K. R. & Weiser, J. N. (2006). Deglycosylation of human glycoconjugates by the sequential activities of exoglycosidases expressed by *Streptococcus pneumoniae*. *Mol Microbiol* **59**, 961-974.

Kjos, M., Nes, I. F. & Diep, D. B. (2011). Mechanisms of resistance to bacteriocins targeting the mannose phosphotransferase system. *Appl Environ Microbiol* **77**, 3335-3342.

Klemm, P. (1986). Two regulatory fim genes, *fimB* and *fimE*, control the phase variation of type 1 fimbriae in *Escherichia coli*. *Embo J* **5**, 1389-1393.

Knecht, J. C., Schiffman, G. & Austrian, R. (1970). Some biological properties of Pneumococcus type 37 and the chemistry of its capsular polysaccharide. *J Exp Med* **132**, 475-487.

Koppe, U., Hogner, K., Doehn, J. M., Muller, H. C., Witzenrath, M., Gutbier, B., Bauer, S., Pribyl, T., Hammerschmidt, S., Lohmeyer, J., Suttorp, N., Herold, S. & Opitz, B. (2012). *Streptococcus pneumoniae* stimulates a STING- and IFN regulatory factor 3-dependent type I IFN production in macrophages, which regulates RANTES production in macrophages, cocultured alveolar epithelial cells, and mouse lungs. *J Immunol* **188**, 811-817.

Kowalko, J. E. & Sebert, M. E. (2008). The *Streptococcus pneumoniae* competence regulatory system influences respiratory tract colonization. *Infect Immun* **76**, 3131-3140.

Kung, E., Coward, W. R., Neill, D. R., Malak, H. A., Muhlemann, K., Kadioglu, A., Hilty, M. & Hathaway, L. J. (2014). The pneumococcal polysaccharide capsule and pneumolysin differentially affect CXCL8 and IL-6 release from cells of the upper and lower respiratory tract. *PLoS One* **9**, e92355.

Kupsch, E. M., Knepper, B., Kuroki, T., Heuer, I. & Meyer, T. F. (1993). Variable opacity (Opa) outer membrane proteins account for the cell tropisms displayed by *Neisseria gonorrhoeae* for human leukocytes and epithelial cells. *Embo J* **12**, 641-650.

Lacks, S. & Hotchkiss, R. D. (1960). A study of the genetic material determining an enzyme in pneumococcus. *Biochim Biophys Acta* **39**, 508-518.

Lacks, S. & Greenberg, B. (1975). A deoxyribonuclease of *Diplococcus pneumoniae* specific for methylated DNA. *J Biol Chem* **250**, 4060-4066.

Lacks, S. & Greenberg, B. (1977). Complementary specificity of restriction endonucleases of *Diplococcus pneumoniae* with respect to DNA methylation. *J Mol Biol* **114**, 153-168.

Lacks, S. A., Mannarelli, B. M., Springhorn, S. S. & Greenberg, B. (1986). Genetic basis of the complementary DpnI and DpnII restriction systems of *S. pneumoniae*: an intercellular cassette mechanism. *Cell* **46**, 993-1000.

Laehnemann, D., Borkhardt, A. & McHardy, A. C. (2015). Denoising DNA deep sequencing data-high-throughput sequencing errors and their correction. *Brief Bioinform* **2015**.

Langmead, B. & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359.

Lanie, J. A., Ng, W. L., Kazmierczak, K. M., Andrzejewski, T. M., Davidsen, T. M., Wayne, K. J., Tettelin, H., Glass, J. I. & Winkler, M. E. (2007). Genome sequence of Avery's virulent serotype 2 strain D39 of *Streptococcus pneumoniae* and comparison with that of unencapsulated laboratory strain R6. *J Bacteriol* **189**, 38-51.

Lau, G. W., Haataja, S., Lonetto, M., Kensit, S. E., Marra, A., Bryant, A. P., McDevitt, D., Morrison, D. A. & Holden, D. W. (2001). A functional genomic analysis of type 3 *Streptococcus pneumoniae* virulence. *Mol Microbiol* **40**, 555-571.

Lee, C. J., Banks, S. D. & Li, J. P. (1991). Virulence, immunity, and vaccine related to *Streptococcus pneumoniae*. *Crit Rev Microbiol* **18**, 89-114.

Lee, M. S. & Morrison, D. A. (1999). Identification of a new regulator in *Streptococcus pneumoniae* linking quorum sensing to competence for genetic transformation. *J Bacteriol* **181**, 5004-5016.

LeMessurier, K. S., Ogunniyi, A. D. & Paton, J. C. (2006). Differential expression of key pneumococcal virulence genes in vivo. *Microbiology* **152**, 305-311.

Len, A. C., Harty, D. W. & Jacques, N. A. (2004). Stress-responsive proteins are upregulated in *Streptococcus mutans* during acid tolerance. *Microbiology* **150**, 1339-1351.

Lerner, C. G. & Inouye, M. (1991). Pleiotropic changes resulting from depletion of Era, an essential GTP-binding protein in *Escherichia coli*. *Mol Microbiol* **5**, 951-957.

Levene, M. J., Korlach, J., Turner, S. W., Foquet, M., Craighead, H. G. & Webb, W. W. (2003). Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299**, 682-686.

Li-Korotky, H. S., Lo, C. Y., Zeng, F. R., Lo, D. & 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.

Li-Korotky, H. S., Lo, C. Y. & Banks, J. M. (2010). Interaction of pneumococcal phase variation, host and pressure/gas composition: virulence expression of NanA, HylA, PspA and CbpA in simulated otitis media. *Microb Pathog* **49**, 204-210.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-2993.

Li, J., Glover, D. T., Szalai, A. J., Hollingshead, S. K. & Briles, D. E. (2007). PspA and PspC minimize immune adherence and transfer of pneumococci from erythrocytes to macrophages through their effects on complement activation. *Infect Immun* **75**, 5877-5885.

Li, Y., Weinberger, D. M., Thompson, C. M., Trzcinski, K. & Lipsitch, M. (2013). Surface charge of *Streptococcus pneumoniae* predicts serotype distribution. *Infect Immun* **81**, 4519-4524.

Linares, J., Ardanuy, C., Pallares, R. & Fenoll, A. (2010). Changes in antimicrobial resistance, serotypes and genotypes in *Streptococcus pneumoniae* over a 30-year period. *Clin Microbiol Infect* **16**, 402-410.

Linder, T. E., Daniels, R. L., Lim, D. J. & DeMaria, T. F. (1994). Effect of intranasal inoculation of *Streptococcus pneumoniae* on the structure of the surface carbohydrates of the chinchilla eustachian tube and middle ear mucosa. *Microb Pathog* **16**, 435-441.

Liu, H., Chen, C., Zhang, H., Kaur, J., Goldman, Y. E. & Cooperman, B. S. (2011). The conserved protein EF4 (LepA) modulates the elongation cycle of protein synthesis. *Proc Natl Acad Sci U S A* **108**, 16223-16228.

Livak, K. J. & 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. & 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.

Loenen, W. A., Dryden, D. T., Raleigh, E. A. & Wilson, G. G. (2014). Type I restriction enzymes and their relatives. *Nucleic Acids Res* **42**, 20-44.

Long, J. P., Tong, H. H., Shannon, P. A. & 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.

Low, D. A., Weyand, N. J. & Mahan, M. J. (2001). Roles of DNA adenine methylation in regulating bacterial gene expression and virulence. *Infect Immun* **69**, 7197-7204.

Lu, Y. J., Gross, J., Bogaert, D., Finn, A., Bagrade, L., Zhang, Q., Kolls, J. K., Srivastava, A., Lundgren, A., Forte, S., Thompson, C. M., Harney, K. F., Anderson, P. W., Lipsitch, M. & Malley, R. (2008). Interleukin-17A mediates acquired immunity to pneumococcal colonization. *PLoS Pathog* **4**, e1000159.

Lu, Y. J., Leite, L., Goncalves, V. M., Dias Wde, O., Liberman, C., Fratelli, F., Alderson, M., Tate, A., Maisonneuve, J. F., Robertson, G., Graca, R., Sayeed, S., Thompson, C. M., Anderson, P. & Malley, R. (2010). GMP-grade pneumococcal whole-cell vaccine injected subcutaneously protects mice from nasopharyngeal colonization and fatal aspiration-sepsis. *Vaccine* **28**, 7468-7475.

Luo, P., Li, H. & Morrison, D. A. (2003). ComX is a unique link between multiple quorum sensing outputs and competence in *Streptococcus pneumoniae*. *Mol Microbiol* **50**, 623-633.

Luria, S. E. & Human, M. L. (1952). A nonhereditary, host-induced variation of bacterial viruses. *J Bacteriol* **64**, 557-569.

Lushchak, V. I. (2012). Glutathione homeostasis and functions: potential targets for medical interventions. *J Amino Acids* **2012**, 736837.

Lynch, J. P., 3rd & Zhanel, G. G. (2010). *Streptococcus pneumoniae*: epidemiology and risk factors, evolution of antimicrobial resistance, and impact of vaccines. *Curr Opin Pulm Med* **16**, 217-225.

Lysenko, E. S., Ratner, A. J., Nelson, A. L. & Weiser, J. N. (2005). The role of innate immune responses in the outcome of interspecies competition for colonization of mucosal surfaces. *PLoS Pathog* **1**, e1.

Mackinnon, F. G., Borrow, R., Gorringe, A. R., Fox, A. J., Jones, D. M. & Robinson, A. (1993). Demonstration of lipooligosaccharide immunotype and capsule as virulence factors for *Neisseria meningitidis* using an infant mouse intranasal infection model. *Microb Pathog* **15**, 359-366.

Madeddu, G., Laura Fiori, M. & Stella Mura, M. (2010). Bacterial community-acquired pneumonia in HIV-infected patients. *Curr Opin Pulm Med* **16**, 201-207.

Mahdi, L. K., Ogunniyi, A. D., LeMessurier, K. S. & Paton, J. C. (2008). Pneumococcal virulence gene expression and host cytokine profiles during pathogenesis of invasive disease. *Infect Immun* **76**, 646-657.

Mahdi, L. K., Wang, H., Van der Hoek, M. B., Paton, J. C. & Ogunniyi, A. D. (2012). Identification of a novel pneumococcal vaccine antigen preferentially expressed during meningitis in mice. *J Clin Invest* **122**, 2208-2220.

Mahdi, L. K., Ebrahimie, E., Adelson, D. L., Paton, J. C. & Ogunniyi, A. D. (2013). A transcription factor contributes to pathogenesis and virulence in *Streptococcus pneumoniae*. *PLoS One* **8**, e70862.

Maier, T., Guell, M. & Serrano, L. (2009). Correlation of mRNA and protein in complex biological samples. *FEBS Lett* **583**, 3966-3973.

Makino, S., van Putten, J. P. & Meyer, T. F. (1991). Phase variation of the opacity outer membrane protein controls invasion by *Neisseria gonorrhoeae* into human epithelial cells. *Embo J* **10**, 1307-1315.

Malley, R., Henneke, P., Morse, S. C., Cieslewicz, M. J., Lipsitch, M., Thompson, C. M., Kurt-Jones, E., Paton, J. C., Wessels, M. R. & Golenbock, D. T. (2003). Recognition of pneumolysin by Toll-like receptor 4 confers resistance to pneumococcal infection. *Proc Natl Acad Sci U S A* **100**, 1966-1971.

Malley, R., Trzcinski, K., Srivastava, A., Thompson, C. M., Anderson, P. W. & Lipsitch, M. (2005). CD4+ T cells mediate antibody-independent acquired immunity to pneumococcal colonization. *Proc Natl Acad Sci U S A* **102**, 4848-4853.

Malley, R. & Anderson, P. W. (2012). Serotype-independent pneumococcal experimental vaccines that induce cellular as well as humoral immunity. *Proc Natl Acad Sci U S A* **109**, 3623-3627.

Manco, S., Hernon, F., Yesilkaya, H., Paton, J. C., Andrew, P. W. & Kadioglu, A. (2006). Pneumococcal neuraminidases A and B both have essential roles during infection of the respiratory tract and sepsis. *Infect Immun* **74**, 4014-4020.

Mangold, M., Siller, M., Roppenser, B., Vlaminckx, B. J., Penfound, T. A., Klein, R., Novak, R., Novick, R. P. & Charpentier, E. (2004). Synthesis of group A streptococcal virulence factors is controlled by a regulatory RNA molecule. *Mol Microbiol* **53**, 1515-1527.

Mann, B., van Opijnen, T., Wang, J., Obert, C., Wang, Y. D., Carter, R., McGoldrick, D. J., Ridout, G., Camilli, A., Tuomanen, E. I. & Rosch, J. W. (2012). Control of virulence by small RNAs in *Streptococcus pneumoniae*. *PLoS Pathog* **8**, e1002788.

Manso, A. S., Chai, M. H., Atack, J. M., Furi, L., De Ste Croix, M., Haigh, R., Trappetti, C., Ogunniyi, A. D., Shewell, L. K., Boitano, M., Clark, T. A., Korlach, J., Blades, M., Mirkes, E., Gorban, A. N., Paton, J. C., Jennings, M. P. & Oggioni, M. R. (2014). A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. *Nat Commun* **5**, 5055.

Margolis, E. & Levin, B. R. (2007). Within-host evolution for the invasiveness of commensal bacteria: an experimental study of bacteremias resulting from *Haemophilus influenzae* nasal carriage. *J Infect Dis* **196**, 1068-1075.

Marinus, M. G. & Morris, N. R. (1973). Isolation of deoxyribonucleic acid methylase mutants of *Escherichia coli* K-12. *J Bacteriol* **114**, 1143-1150.

Marion, C., Aten, A. E., Woodiga, S. A. & King, S. J. (2011a). Identification of an ATPase, MsmK, which energizes multiple carbohydrate ABC transporters in *Streptococcus pneumoniae*. *Infect Immun* **79**, 4193-4200.

Marion, C., Burnaugh, A. M., Woodiga, S. A. & King, S. J. (2011b). Sialic acid transport contributes to pneumococcal colonization. *Infect Immun* **79**, 1262-1269.

Marks, L. R., Reddinger, R. M. & Hakansson, A. P. (2012). High levels of genetic recombination during nasopharyngeal carriage and biofilm formation in *Streptococcus pneumoniae*. *MBio* **3**, 00200-00212.

Marriott, H. M., Gascoyne, K. A., Gowda, R., Geary, I., Nicklin, M. J., Iannelli, F., Pozzi, G., Mitchell, T. J., Whyte, M. K., Sabroe, I. & Dockrell, D. H. (2012). Interleukin-1beta regulates CXCL8 release and influences disease outcome in response to *Streptococcus pneumoniae*, defining intercellular cooperation between pulmonary epithelial cells and macrophages. *Infect Immun* **80**, 1140-1149.

Martin, B., Humbert, O., Camara, M., Guenzi, E., Walker, J., Mitchell, T., Andrew, P., Prudhomme, M., Alloing, G., Hakenbeck, R. & et al. (1992). A highly conserved

repeated DNA element located in the chromosome of *Streptococcus pneumoniae*. *Nucleic Acids Res* **20**, 3479-3483.

Martinelli, L. K., Ducati, R. G., Rosado, L. A., Breda, A., Selbach, B. P., Santos, D. S. & Basso, L. A. (2011). Recombinant *Escherichia coli* GMP reductase: kinetic, catalytic and chemical mechanisms, and thermodynamics of enzyme-ligand binary complex formation. *Mol Biosyst* **7**, 1289-1305.

Martner, A., Skovbjerg, S., Paton, J. C. & Wold, A. E. (2009). *Streptococcus pneumoniae* autolysis prevents phagocytosis and production of phagocyte-activating cytokines. *Infect Immun* **77**, 3826-3837.

Mascher, T., Zahner, D., Merai, M., Balmelle, N., de Saizieu, A. B. & Hakenbeck, R. (2003). The *Streptococcus pneumoniae cia* regulon: CiaR target sites and transcription profile analysis. *J Bacteriol* **185**, 60-70.

Masip, L., Veeravalli, K. & Georgiou, G. (2006). The many faces of glutathione in bacteria. *Antioxid Redox Signal* **8**, 753-762.

Matthias, K. A., Roche, A. M., Standish, A. J., Shchepetov, M. & Weiser, J. N. (2008). Neutrophil-toxin interactions promote antigen delivery and mucosal clearance of *Streptococcus pneumoniae*. *J Immunol* **180**, 6246-6254.

Maus, U. A., Srivastava, M., Paton, J. C., Mack, M., Everhart, M. B., Blackwell, T. S., Christman, J. W., Schlondorff, D., Seeger, W. & Lohmeyer, J. (2004). Pneumolysin-induced lung injury is independent of leukocyte trafficking into the alveolar space. *J Immunol* **173**, 1307-1312.

Mayer, M. P. & Bukau, B. (2005). Hsp70 chaperones: cellular functions and molecular mechanism. *Cell Mol Life Sci* **62**, 670-684.

McAllister, L. J., Tseng, H. J., Ogunniyi, A. D., Jennings, M. P., McEwan, A. G. & Paton, J. C. (2004). Molecular analysis of the *psa* permease complex of *Streptococcus pneumoniae*. *Mol Microbiol* **53**, 889-901.

McClain, M. S., Blomfield, I. C. & Eisenstein, B. I. (1991). Roles of fimB and fimE in site-specific DNA inversion associated with phase variation of type 1 fimbriae in *Escherichia coli*. *J Bacteriol* **173**, 5308-5314.

McCluskey, J., Hinds, J., Husain, S., Witney, A. & Mitchell, T. J. (2004). A two-component system that controls the expression of pneumococcal surface antigen A (PsaA) and regulates virulence and resistance to oxidative stress in *Streptococcus pneumoniae*. *Mol Microbiol* **51**, 1661-1675.

McCullers, J. A. & Tuomanen, E. I. (2001). Molecular pathogenesis of pneumococcal pneumonia. *Front Biosci* **6**, D877-889.

McKessar, S. (2003). The characterisation of phase variation and a novel fimbrial protein in *Streptococcus pneumoniae*. In *School of Pharmaceutical, Molecular and Biomedical Sciences*. Adelaide: University of South Australia.

Meisel, A., Bickle, T. A., Kruger, D. H. & Schroeder, C. (1992). Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature* **355**, 467-469.

Melin, M., Di Paolo, E., Tikkanen, L., Jarva, H., Neyt, C., Kayhty, H., Meri, S., Poolman, J. & Vakevainen, M. (2010a). Interaction of pneumococcal histidine triad proteins with human complement. *Infect Immun* **78**, 2089-2098.

Melin, M., Trzcinski, K., Meri, S., Kayhty, H. & Vakevainen, M. (2010b). The capsular serotype of *Streptococcus pneumoniae* is more important than the genetic background for resistance to complement. *Infect Immun* **78**, 5262-5270.

Meyer, R. R. & Laine, P. S. (1990). The single-stranded DNA-binding protein of *Escherichia coli*. *Microbiol Rev* **54**, 342-380.

Meyers, L. A., Levin, B. R., Richardson, A. R. & Stojiljkovic, I. (2003). Epidemiology, hypermutation, within-host evolution and the virulence of *Neisseria meningitidis*. *Proc Biol Sci* **270**, 1667-1677.

Mirsaeidi, M. & Schraufnagel, D. E. (2014). Pneumococcal vaccines: understanding centers for disease control and prevention recommendations. *Ann Am Thorac Soc* **11**, 980-985.

Mitchell, A. M. & Mitchell, T. J. (2010). *Streptococcus pneumoniae*: virulence factors and variation. *Clin Microbiol Infect* **16**, 411-418.

Mollerach, M., Lopez, R. & Garcia, E. (1998). Characterization of the *galU* gene of *Streptococcus pneumoniae* encoding a uridine diphosphoglucose pyrophosphorylase: a gene essential for capsular polysaccharide biosynthesis. *J Exp Med* **188**, 2047-2056.

Molzen, T. E., Burghout, P., Bootsma, H. J., Brandt, C. T., van der Gaast-de Jongh, C. E., Eleveld, M. J., Verbeek, M. M., Frimodt-Moller, N., Ostergaard, C. & Hermans, P. W. (2010). Genome-wide identification of Streptococcus pneumoniae genes essential for bacterial replication during experimental meningitis. *Infect Immun* **79**, 288-297.

Morens, D. M., Taubenberger, J. K. & Fauci, A. S. (2008). Predominant role of bacterial pneumonia as a cause of death in pandemic influenza: implications for pandemic influenza preparedness. *J Infect Dis* **198**, 962-970.

Morgan, P. J., Hyman, S. C., Rowe, A. J., Mitchell, T. J., Andrew, P. W. & Saibil, H. R. (1995). Subunit organisation and symmetry of pore-forming, oligomeric pneumolysin. *FEBS Lett* **371**, 77-80.

Morona, J. K., Morona, R. & Paton, J. C. (1997). Molecular and genetic characterization of the capsule biosynthesis locus of *Streptococcus pneumoniae* type 19B. *J Bacteriol* **179**, 4953-4958.

Morona, J. K., Morona, R. & 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 U S A* **103**, 8505-8510.

Moxon, E. R. & Murphy, P. A. (1978). *Haemophilus influenzae* bacteremia and meningitis resulting from survival of a single organism. *Proc Natl Acad Sci U S A* **75**, 1534-1536.

Muckerman, C. C., Springhorn, S. S., Greenberg, B. & Lacks, S. A. (1982). Transformation of restriction endonuclease phenotype in *Streptococcus pneumoniae*. *J Bacteriol* **152**, 183-190.

Nelson, A. L., Roche, A. M., Gould, J. M., Chim, K., Ratner, A. J. & Weiser, J. N. (2007). Capsule enhances pneumococcal colonization by limiting mucus-mediated clearance. *Infect Immun* **75**, 83-90.

Ng, E. W., Costa, J. R., Samiy, N., Ruoff, K. L., Connolly, E., Cousins, F. V. & D'Amico, D. J. (2002). Contribution of pneumolysin and autolysin to the pathogenesis of experimental pneumococcal endophthalmitis. *Retina* **22**, 622-632.

Novak, R., Braun, J. S., Charpentier, E. & Tuomanen, E. (1998). Penicillin tolerance genes of *Streptococcus pneumoniae*: the ABC-type manganese permease complex Psa. *Mol Microbiol* **29**, 1285-1296.

Ochoa, T. J., Egoavil, M., Castillo, M. E., Reyes, I., Chaparro, E., Silva, W., Campos, F. & Saenz, A. (2010). Invasive pneumococcal diseases among hospitalized children in Lima, Peru. *Rev Panam Salud Publica* **28**, 121-127.

Oggioni, M. R., Trappetti, C., Kadioglu, A., Cassone, M., Iannelli, F., Ricci, S., Andrew, P. W. & Pozzi, G. (2006). Switch from planktonic to sessile life: a major event in pneumococcal pathogenesis. *Mol Microbiol* **61**, 1196-1210.

Ogunniyi, A. D., Folland, R. L., Briles, D. E., Hollingshead, S. K. & Paton, J. C. (2000). Immunization of mice with combinations of pneumococcal virulence proteins elicits enhanced protection against challenge with *Streptococcus pneumoniae*. *Infect Immun* **68**, 3028-3033.

Ogunniyi, A. D., Giammarinaro, P. & 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., LeMessurier, K. S., Graham, R. M., Watt, J. M., Briles, D. E., Stroeher, U. H. & Paton, J. C. (2007). Contributions of pneumolysin, pneumococcal surface protein A (PspA), and PspC to pathogenicity of *Streptococcus pneumoniae* D39 in a mouse model. *Infect Immun* **75**, 1843-1851.

Ogunniyi, A. D., Grabowicz, M., Mahdi, L. K., Cook, J., Gordon, D. L., Sadlon, T. A. & Paton, J. C. (2009). Pneumococcal histidine triad proteins are regulated by the Zn2+-dependent repressor AdcR and inhibit complement deposition through the recruitment of complement factor H. *Faseb J* **23**, 731-738.

Ogunniyi, A. D., Mahdi, L. K., Jennings, M. P., McEwan, A. G., McDevitt, C. A., Van der Hoek, M. B., Bagley, C. J., Hoffmann, P., Gould, K. A. & Paton, J. C. (2010). Central role of manganese in regulation of stress responses, physiology, and metabolism in *Streptococcus pneumoniae*. *J Bacteriol* **192**, 4489-4497.

Ogunniyi, A. D., Mahdi, L. K., Trappetti, C., Verhoeven, N., Mermans, D., Van der Hoek, M. B., Plumptre, C. D. & Paton, J. C. (2012). Identification of genes that contribute to the pathogenesis of invasive pneumococcal disease by in vivo transcriptomic analysis. *Infect Immun* **80**, 3268-3278.

Oliveira, L., Madureira, P., Andrade, E. B., Bouaboud, A., Morello, E., Ferreira, P., Poyart, C., Trieu-Cuot, P. & Dramsi, S. (2011). Group B streptococcus GAPDH is released upon cell lysis, associates with bacterial surface, and induces apoptosis in murine macrophages. *PLoS One* **7**, e29963.

Orihuela, C. J., Gao, G., McGee, M., Yu, J., Francis, K. P. & Tuomanen, E. (2003). Organ-specific models of *Streptococcus pneumoniae* disease. *Scand J Infect Dis* **35**, 647-652.

Orihuela, C. J., Radin, J. N., Sublett, J. E., Gao, G., Kaushal, D. & Tuomanen, E. I. (2004). Microarray analysis of pneumococcal gene expression during invasive disease. *Infect Immun* **72**, 5582-5596.

Orihuela, C. J., Mahdavi, J., Thornton, J., Mann, B., Wooldridge, K. G., Abouseada, N., Oldfield, N. J., Self, T., Ala'Aldeen, D. A. & Tuomanen, E. I. (2009). Laminin receptor initiates bacterial contact with the blood brain barrier in experimental meningitis models. *J Clin Invest* **119**, 1638-1646.

Overweg, K., Pericone, C. D., Verhoef, G. G., Weiser, J. N., Meiring, H. D., De Jong, A. P., De Groot, R. & Hermans, P. W. (2000). Differential protein expression in phenotypic variants of *Streptococcus pneumoniae*. *Infect Immun* **68**, 4604-4610.

Pallares, R., Gudiol, F., Linares, J., Ariza, J., Rufi, G., Murgui, L., Dorca, J. & Viladrich, P. F. (1987). Risk factors and response to antibiotic therapy in adults with bacteremic pneumonia caused by penicillin-resistant pneumococci. *N Engl J Med* **317**, 18-22.

Pancholi, V. & Fischetti, V. A. (1992). A major surface protein on group A streptococci is a glyceraldehyde-3-phosphate-dehydrogenase with multiple binding activity. *J Exp Med* **176**, 415-426.

Pancholi, V. & Fischetti, V. A. (1997). Regulation of the phosphorylation of human pharyngeal cell proteins by group A streptococcal surface dehydrogenase: signal transduction between streptococci and pharyngeal cells. *J Exp Med* **186**, 1633-1643.

Pancholi, V. & Chhatwal, G. S. (2003). Housekeeping enzymes as virulence factors for pathogens. *Int J Med Microbiol* **293**, 391-401.

Papenfort, K. & Vogel, J. (2010). Regulatory RNA in bacterial pathogens. *Cell Host Microbe* **8**, 116-127.

Parker, D., Soong, G., Planet, P., Brower, J., Ratner, A. J. & Prince, A. (2009). The NanA neuraminidase of *Streptococcus pneumoniae* is involved in biofilm formation. *Infect Immun* **77**, 3722-3730.

Paton, J. C., Rowan-Kelly, B. & Ferrante, A. (1984). Activation of human complement by the pneumococcal toxin pneumolysin. *Infect Immun* **43**, 1085-1087.

Paton, J. C., Morona, J. K. & Morona, R. (1997). Characterization of the capsular polysaccharide biosynthesis locus of *Streptococcus pneumoniae* type 19F. *Microb Drug Resist* **3**, 89-99.

Perez-Dorado, I., Galan-Bartual, S. & Hermoso, J. A. (2012). Pneumococcal surface proteins: when the whole is greater than the sum of its parts. *Mol Oral Microbiol* **27**, 221-245.

Pericone, C. D., Overweg, K., Hermans, P. W. & Weiser, J. N. (2000). Inhibitory and bactericidal effects of hydrogen peroxide production by *Streptococcus pneumoniae* on other inhabitants of the upper respiratory tract. *Infect Immun* **68**, 3990-3997.

Pericone, C. D., Bae, D., Shchepetov, M., McCool, T. & Weiser, J. N. (2002). Short-sequence tandem and nontandem DNA repeats and endogenous hydrogen peroxide production contribute to genetic instability of *Streptococcus pneumoniae*. *J Bacteriol* **184**, 4392-4399.

Pericone, C. D., Park, S., Imlay, J. A. & Weiser, J. N. (2003). Factors contributing to hydrogen peroxide resistance in *Streptococcus pneumoniae* include pyruvate oxidase

(SpxB) and avoidance of the toxic effects of the fenton reaction. *J Bacteriol* **185**, 6815-6825.

Pestova, E. V., Havarstein, L. S. & Morrison, D. A. (1996). Regulation of competence for genetic transformation in *Streptococcus pneumoniae* by an auto-induced peptide pheromone and a two-component regulatory system. *Mol Microbiol* **21**, 853-862.

Peterson, S., Cline, R. T., Tettelin, H., Sharov, V. & Morrison, D. A. (2000). Gene expression analysis of the *Streptococcus pneumoniae* competence regulons by use of DNA microarrays. *J Bacteriol* **182**, 6192-6202.

Peterson, S. N., Sung, C. K., Cline, R., Desai, B. V., Snesrud, E. C., Luo, P., Walling, J., Li, H., Mintz, M., Tsegaye, G., Burr, P. C., Do, Y., Ahn, S., Gilbert, J., Fleischmann, R. D. & Morrison, D. A. (2004). Identification of competence pheromone responsive genes in *Streptococcus pneumoniae* by use of DNA microarrays. *Mol Microbiol* **51**, 1051-1070.

Pettigrew, M. M., Fennie, K. P., York, M. P., Daniels, J. & Ghaffar, F. (2006). Variation in the presence of neuraminidase genes among *Streptococcus pneumoniae* isolates with identical sequence types. *Infect Immun* **74**, 3360-3365.

Pillutla, R. C., Ahnn, J. & Inouye, M. (1996). Deletion of the putative effector region of Era, an essential GTP-binding protein in *Escherichia coli*, causes a dominant-negative phenotype. *FEMS Microbiol Lett* **143**, 47-55.

Plumptre, C. D., Ogunniyi, A. D. & Paton, J. C. (2012). Polyhistidine triad proteins of pathogenic streptococci. *Trends Microbiol* **20**, 485-493.

Plumptre, C. D., Ogunniyi, A. D. & Paton, J. C. (2013). Surface association of Pht proteins of *Streptococcus pneumoniae*. *Infect Immun* **81**, 3644-3651.

Plumptre, C. D., Eijkelkamp, B. A., Morey, J. R., Behr, F., Counago, R. M., Ogunniyi, A. D., Kobe, B., O'Mara, M. L., Paton, J. C. & McDevitt, C. A. (2014a). AdcA and AdcAII employ distinct zinc acquisition mechanisms and contribute additively to zinc homeostasis in *Streptococcus pneumoniae*. *Mol Microbiol* **91**, 834-851.

Plumptre, C. D., Hughes, C. E., Harvey, R. M., Eijkelkamp, B. A., McDevitt, C. A. & Paton, J. C. (2014b). Overlapping functionality of the Pht proteins in zinc homeostasis of *Streptococcus pneumoniae*. *Infect Immun* **82**, 4315-4324.

Potter, A. J., Trappetti, C. & Paton, J. C. (2013). *Streptococcus pneumoniae* uses glutathione to defend against oxidative stress and metal ion toxicity. *J Bacteriol* **194**, 6248-6254.

Pouttu, R., Puustinen, T., Virkola, R., Hacker, J., Klemm, P. & Korhonen, T. K. (1999). Amino acid residue Ala-62 in the FimH fimbrial adhesin is critical for the

adhesiveness of meningitis-associated *Escherichia coli* to collagens. *Mol Microbiol* **31**, 1747-1757.

Price, K. E. & Camilli, A. (2009). Pneumolysin localizes to the cell wall of *Streptococcus pneumoniae*. *J Bacteriol* **191**, 2163-2168.

Puyet, A., Ibanez, A. M. & Espinosa, M. (1993). Characterization of the *Streptococcus pneumoniae* maltosaccharide regulator MalR, a member of the LacI-GalR family of repressors displaying distinctive genetic features. *J Biol Chem* **268**, 25402-25408.

Qin, Y., Polacek, N., Vesper, O., Staub, E., Einfeldt, E., Wilson, D. N. & Nierhaus, K. H. (2006). The highly conserved LepA is a ribosomal elongation factor that back-translocates the ribosome. *Cell* **127**, 721-733.

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P. & Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**, 341.

Quin, L. R., Carmicle, S., Dave, S., Pangburn, M. K., Evenhuis, J. P. & McDaniel, L. S. (2005). In vivo binding of complement regulator factor H by *Streptococcus pneumoniae*. *J Infect Dis* **192**, 1996-2003.

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

Ramnath, M., Beukes, M., Tamura, K. & Hastings, J. W. (2000). Absence of a putative mannose-specific phosphotransferase system enzyme IIAB component in a leucocin A-resistant strain of *Listeria monocytogenes*, as shown by two-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *Appl Environ Microbiol* **66**, 3098-3101.

Ramos-Montanez, S., Tsui, H. C., Wayne, K. J., Morris, J. L., Peters, L. E., Zhang, F., Kazmierczak, K. M., Sham, L. T. & Winkler, M. E. (2008). Polymorphism and regulation of the spxB (pyruvate oxidase) virulence factor gene by a CBS-HotDog domain protein (SpxR) in serotype 2 *Streptococcus pneumoniae*. *Mol Microbiol* **67**, 729-746.

Ramos-Sevillano, E., Urzainqui, A., Campuzano, S., Moscoso, M., Gonzalez-Camacho, F., Domenech, M., Rodriguez de Cordoba, S., Sanchez-Madrid, F., Brown, J. S., Garcia, E. & Yuste, J. (2015). Pleiotropic effects of cell wall amidase LytA on *Streptococcus pneumoniae* sensitivity to the host immune response. *Infect Immun* **83**, 591-603.

Regev-Yochay, G., Trzcinski, K., Thompson, C. M., Lipsitch, M. & Malley, R. (2007). SpxB is a suicide gene of *Streptococcus pneumoniae* and confers a selective advantage in an in vivo competitive colonization model. *J Bacteriol* **189**, 6532-6539.

Ren, B., McCrory, M. A., Pass, C., Bullard, D. C., Ballantyne, C. M., Xu, Y., Briles, D. E. & Szalai, A. J. (2004a). The virulence function of *Streptococcus pneumoniae* surface protein A involves inhibition of complement activation and impairment of complement receptor-mediated protection. *J Immunol* **173**, 7506-7512.

Ren, B., Szalai, A. J., Hollingshead, S. K. & Briles, D. E. (2004b). Effects of PspA and antibodies to PspA on activation and deposition of complement on the pneumococcal surface. *Infect Immun* **72**, 114-122.

Ricci, S., Gerlini, A., Pammolli, A., Chiavolini, D., Braione, V., Tripodi, S. A., Colombari, B., Blasi, E., Oggioni, M. R., Peppoloni, S. & Pozzi, G. (2013). Contribution of different pneumococcal virulence factors to experimental meningitis in mice. *BMC Infect Dis* **13**, 444.

Rioux, S., Neyt, C., Di Paolo, E., Turpin, L., Charland, N., Labbe, S., Mortier, M. C., Mitchell, T. J., Feron, C., Martin, D. & Poolman, J. T. (2010). Transcriptional regulation, occurrence and putative role of the Pht family of *Streptococcus pneumoniae*. *Microbiology* **157**, 336-348.

Rioux, S., Neyt, C., Di Paolo, E., Turpin, L., Charland, N., Labbe, S., Mortier, M. C., Mitchell, T. J., Feron, C., Martin, D. & Poolman, J. T. (2011). Transcriptional regulation, occurrence and putative role of the Pht family of *Streptococcus pneumoniae*. *Microbiology* **157**, 336-348.

Ritz, D. & Beckwith, J. (2001). Roles of thiol-redox pathways in bacteria. *Annu Rev Microbiol* **55**, 21-48.

Robbins, J. B., Austrian, R., Lee, C. J., Rastogi, S. C., Schiffman, G., Henrichsen, J., Makela, P. H., Broome, C. V., Facklam, R. R., Tiesjema, R. H. & et al. (1983). Considerations for formulating the second-generation pneumococcal capsular polysaccharide vaccine with emphasis on the cross-reactive types within groups. *J Infect Dis* **148**, 1136-1159.

Roberts, R. J. (1990). Restriction enzymes and their isoschizomers. *Nucleic Acids Res* **18 Suppl**, 2331-2365.

Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinaite, J., Blumenthal, R. M., Degtyarev, S., Dryden, D. T., Dybvig, K., Firman, K., Gromova, E. S., Gumport, R. I., Halford, S. E., Hattman, S., Heitman, J., Hornby, D. P., Janulaitis, A., Jeltsch, A., Josephsen, J., Kiss, A., Klaenhammer, T. R., Kobayashi, I., Kong, H., Kruger, D. H., Lacks, S., Marinus, M. G., Miyahara, M., Morgan, R. D., Murray, N. E., Nagaraja, V., Piekarowicz, A., Pingoud, A., Raleigh, E., Rao, D. N., Reich, N., Repin, V. E., Selker, E. U., Shaw, P. C., Stein, D. C., Stoddard, B. L., Szybalski, W., Trautner, T. A., Van Etten, J. L., Vitor, J. M., Wilson, G. G. & Xu, S. Y. (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res* **31**, 1805-1812.

Roberts, R. J., Vincze, T., Posfai, J. & Macelis, D. (2010). REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res* **38**, D234-236.

Rodnina, M. V., Savelsbergh, A., Katunin, V. I. & Wintermeyer, W. (1997). Hydrolysis of GTP by elongation factor G drives tRNA movement on the ribosome. *Nature* **385**, 37-41.

Rojas-Espinosa, O., Mendez-Navarrete, I. & Estrada-Parra, S. (1972). Presence of C1q-reactive immune complexes in patients with leprosy. *Clin Exp Immunol* **12**, 215-223.

Rosenow, C., Ryan, P., Weiser, J. N., Johnson, S., Fontan, P., Ortqvist, A. & Masure, H. R. (1997). Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of *Streptococcus pneumoniae*. *Mol Microbiol* **25**, 819-829.

Ross, M. G., Russ, C., Costello, M., Hollinger, A., Lennon, N. J., Hegarty, R., Nusbaum, C. & Jaffe, D. B. (2013). Characterizing and measuring bias in sequence data. *Genome Biol* **14**, R51.

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H., Johnson, K., Milgrew, M. J., Edwards, M., Hoon, J., Simons, J. F., Marran, D., Myers, J. W., Davidson, J. F., Branting, A., Nobile, J. R., Puc, B. P., Light, D., Clark, T. A., Huber, M., Branciforte, J. T., Stoner, I. B., Cawley, S. E., Lyons, M., Fu, Y., Homer, N., Sedova, M., Miao, X., Reed, B., Sabina, J., Feierstein, E., Schorn, M., Alanjary, M., Dimalanta, E., Dressman, D., Kasinskas, R., Sokolsky, T., Fidanza, J. A., Namsaraev, E., McKernan, K. J., Williams, A., Roth, G. T. & Bustillo, J. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475**, 348-352.

Rueda, A. M., Serpa, J. A., Matloobi, M., Mushtaq, M. & Musher, D. M. (2010). The spectrum of invasive pneumococcal disease at an adult tertiary care hospital in the early 21st century. *Medicine (Baltimore)* **89**, 331-336.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. & Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944-945.

Saluja, S. K. & Weiser, J. N. (1995). The genetic basis of colony opacity in *Streptococcus pneumoniae*: evidence for the effect of box elements on the frequency of phenotypic variation. *Mol Microbiol* **16**, 215-227.

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

Sanger, F. & Coulson, A. R. (1978). The use of thin acrylamide gels for DNA sequencing. *FEBS Lett* **87**, 107-110.

Schachern, P. A., Tsuprun, V., Ferrieri, P., Briles, D. E., Goetz, S., Cureoglu, S., Paparella, M. M. & Juhn, S. (2014). Pneumococcal PspA and PspC proteins: potential vaccine candidates for experimental otitis media. *Int J Pediatr Otorhinolaryngol* **78**, 1517-1521.

Schilling, J. D., Mulvey, M. A. & Hultgren, S. J. (2001). Structure and function of *Escherichia coli* type 1 pili: new insight into the pathogenesis of urinary tract infections. *J Infect Dis* **183 Suppl 1**, S36-40.

Schmeck, B., Gross, R., N'Guessan, P. D., Hocke, A. C., Hammerschmidt, S., Mitchell, T. J., Rosseau, S., Suttorp, N. & Hippenstiel, S. (2004). *Streptococcus pneumoniae*-induced caspase 6-dependent apoptosis in lung epithelium. *Infect Immun* **72**, 4940-4947.

Scott, J. R. (1970). Clear plaque mutants of phage P1. *Virology* **41**, 66-71.

Seki, M., Iida, K., Saito, M., Nakayama, H. & Yoshida, S. (2004). Hydrogen peroxide production in *Streptococcus pyogenes*: involvement of lactate oxidase and coupling with aerobic utilization of lactate. *J Bacteriol* **186**, 2046-2051.

Sell, S. M., Eisen, C., Ang, D., Zylicz, M. & Georgopoulos, C. (1990). Isolation and characterization of *dnaJ* null mutants of *Escherichia coli*. *J Bacteriol* **172**, 4827-4835.

Seshadri, R., Hendrix, L. R. & Samuel, J. E. (1999). Differential expression of translational elements by life cycle variants of *Coxiella burnetii*. *Infect Immun* **67**, 6026-6033.

Shak, J. R., Vidal, J. E. & Klugman, K. P. (2013). Influence of bacterial interactions on pneumococcal colonization of the nasopharynx. *Trends Microbiol* **21**, 129-135.

Shakhnovich, E. A., King, S. J. & 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. & 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.

Shewell, L. K., Harvey, R. M., Higgins, M. A., Day, C. J., Hartley-Tassell, L. E., Chen, A. Y., Gillen, C. M., James, D. B., Alonzo, F., 3rd, Torres, V. J., Walker, M. J., Paton, A. W., Paton, J. C. & Jennings, M. P. (2014). The cholesterol-dependent cytolysins

pneumolysin and streptolysin O require binding to red blood cell glycans for hemolytic activity. *Proc Natl Acad Sci U S A* **111**, E5312-5320.

Shinde, D., Lai, Y., Sun, F. & Arnheim, N. (2003). Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)n and (A/T)n microsatellites. *Nucleic Acids Res* **31**, 974-980.

Skov Sorensen, U. B., Blom, J., Birch-Andersen, A. & Henrichsen, J. (1988). Ultrastructural localization of capsules, cell wall polysaccharide, cell wall proteins, and F antigen in pneumococci. *Infect Immun* **56**, 1890-1896.

Smyth, G. K. & Speed, T. (2003). Normalization of cDNA microarray data. *Methods* **31**, 265-273.

Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**, Article3.

Smyth, G. K. (2005). Limma: linear models for microarray data. In *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, pp. 397-420. Edited by G. R, C. V, D. S, I. R & H. W. New York: Springer.

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. & Prince, A. (2006). Bacterial neuraminidase facilitates mucosal infection by participating in biofilm production. *J Clin Invest* **116**, 2297-2305.

Spellerberg, B., Cundell, D. R., Sandros, J., Pearce, B. J., Idanpaan-Heikkila, I., Rosenow, C. & Masure, H. R. (1996). Pyruvate oxidase, as a determinant of virulence in *Streptococcus pneumoniae*. *Mol Microbiol* **19**, 803-813.

Spratt, B. G. & Greenwood, B. M. (2000). Prevention of pneumococcal disease by vaccination: does serotype replacement matter? *Lancet* **356**, 1210-1211.

Srikhanta, Y. N., Maguire, T. L., Stacey, K. J., Grimmond, S. M. & Jennings, M. P. (2005). The phasevarion: a genetic system controlling coordinated, random switching of expression of multiple genes. *Proc Natl Acad Sci U S A* **102**, 5547-5551.

Srikhanta, Y. N., Dowideit, S. J., Edwards, J. L., Falsetta, M. L., Wu, H. J., Harrison, O. B., Fox, K. L., Seib, K. L., Maguire, T. L., Wang, A. H., Maiden, M. C., Grimmond, S. M., Apicella, M. A. & Jennings, M. P. (2009). Phasevarions mediate random switching of gene expression in pathogenic *Neisseria*. *PLoS Pathog* **5**, e1000400.

Srikhanta, Y. N., Gorrell, R. J., Steen, J. A., Gawthorne, J. A., Kwok, T., Grimmond, S. M., Robins-Browne, R. M. & Jennings, M. P. (2011). Phasevarion mediated epigenetic gene regulation in *Helicobacter pylori*. *PLoS One* **6**, e27569.

Srivastava, A., Henneke, P., Visintin, A., Morse, S. C., Martin, V., Watkins, C., Paton, J. C., Wessels, M. R., Golenbock, D. T. & Malley, R. (2005). The apoptotic response to pneumolysin is Toll-like receptor 4 dependent and protects against pneumococcal disease. *Infect Immun* **73**, 6479-6487.

Steinfort, C., Wilson, R., Mitchell, T., Feldman, C., Rutman, A., Todd, H., Sykes, D., Walker, J., Saunders, K., Andrew, P. W., Boulnois, G. J. & Cole, P. J. (1989). Effect of *Streptococcus pneumoniae* on human respiratory epithelium in vitro. *Infect Immun* **57**, 2006-2013.

Steinmoen, H., Teigen, A. & Havarstein, L. S. (2003). Competence-induced cells of *Streptococcus pneumoniae* lyse competence-deficient cells of the same strain during cocultivation. *J Bacteriol* **185**, 7176-7183.

Stern, A., Brown, M., Nickel, P. & Meyer, T. F. (1986). Opacity genes in *Neisseria gonorrhoeae*: control of phase and antigenic variation. *Cell* **47**, 61-71.

Stern, A. & Meyer, T. F. (1987). Common mechanism controlling phase and antigenic variation in pathogenic neisseriae. *Mol Microbiol* **1**, 5-12.

Streisinger, G. & Owen, J. (1985). Mechanisms of spontaneous and induced frameshift mutation in bacteriophage T4. *Genetics* **109**, 633-659.

Stroeher, U. H., Paton, A. W., Ogunniyi, A. D. & Paton, J. C. (2003). Mutation of *luxS* of *Streptococcus pneumoniae* affects virulence in a mouse model. *Infect Immun* **71**, 3206-3212.

Swiatlo, E., Champlin, F. R., Holman, S. C., Wilson, W. W. & Watt, J. M. (2002). Contribution of choline-binding proteins to cell surface properties of *Streptococcus pneumoniae*. *Infect Immun* **70**, 412-415.

Syk, A., Norman, M., Fernebro, J., Gallotta, M., Farmand, S., Sandgren, A., Normark, S. & Henriques-Normark, B. (2014). Emergence of hypervirulent mutants resistant to early clearance during systemic serotype 1 pneumococcal infection in mice and humans. *J Infect Dis* **210**, 4-13.

Szczelkun, M. D. (2011). Translocation, switching and gating: potential roles for ATP in long-range communication on DNA by Type III restriction endonucleases. *Biochem Soc Trans* **39**, 589-594.

Takaya, A., Tomoyasu, T., Matsui, H. & Yamamoto, T. (2004). The DnaK/DnaJ chaperone machinery of *Salmonella enterica* serovar Typhimurium is essential for invasion of epithelial cells and survival within macrophages, leading to systemic infection. *Infect Immun* **72**, 1364-1373.

Taniai, H., Iida, K., Seki, M., Saito, M., Shiota, S., Nakayama, H. & Yoshida, S. (2008). Concerted action of lactate oxidase and pyruvate oxidase in aerobic growth of *Streptococcus pneumoniae*: role of lactate as an energy source. *J Bacteriol* **190**, 3572-3579.

Terrasse, R., Tacnet-Delorme, P., Moriscot, C., Perard, J., Schoehn, G., Vernet, T., Thielens, N. M., Di Guilmi, A. M. & Frachet, P. (2012). Human and pneumococcal cell surface glyceraldehyde-3-phosphate dehydrogenase (GAPDH) proteins are both ligands of human C1q protein. *J Biol Chem* **287**, 42620-42633.

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. & Fraser, C. M. (2001). Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293**, 498-506.

Tilley, S. J., Orlova, E. V., Gilbert, R. J., Andrew, P. W. & Saibil, H. R. (2005). Structural basis of pore formation by the bacterial toxin pneumolysin. *Cell* **121**, 247-256.

Tomasz, A. (1965). Control of the competent state in Pneumococcus by a hormone-like cell product: an example for a new type of regulatory mechanism in bacteria. *Nature* **208**, 155-159.

Tomczyk, S., Bennett, N. M., Stoecker, C., Gierke, R., Moore, M. R., Whitney, C. G., Hadler, S. & Pilishvili, T. (2014). Use of 13-valent pneumococcal conjugate vaccine and 23-valent pneumococcal polysaccharide vaccine among adults aged >/=65 years: recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Morb Mortal Wkly Rep* **63**, 822-825.

Tong, H. H., Blue, L. E., James, M. A. & 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.

Trappetti, C., Kadioglu, A., Carter, M., Hayre, J., Iannelli, F., Pozzi, G., Andrew, P. W. & 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.

Trappetti, C., Gualdi, L., Di Meola, L., Jain, P., Korir, C. C., Edmonds, P., Iannelli, F., Ricci, S., Pozzi, G. & Oggioni, M. R. (2011a). The impact of the competence quorum sensing system on *Streptococcus pneumoniae* biofilms varies depending on the experimental model. *BMC Microbiol* **11**, 75.

Trappetti, C., Ogunniyi, A. D., Oggioni, M. R. & Paton, J. C. (2011b). Extracellular matrix formation enhances the ability of *Streptococcus pneumoniae* to cause invasive disease. *PLoS One* **6**, e19844.

Trappetti, C., Potter, A. J., Paton, A. W., Oggioni, M. R. & Paton, J. C. (2011c). LuxS mediates iron-dependent biofilm formation, competence, and fratricide in Streptococcus pneumoniae. *Infect Immun* **79**, 4550-4558.

Tseng, H. J., McEwan, A. G., Paton, J. C. & Jennings, M. P. (2002). Virulence of *Streptococcus pneumoniae*: PsaA mutants are hypersensitive to oxidative stress. *Infect Immun* **70**, 1635-1639.

Tu, A. H., Fulgham, R. L., McCrory, M. A., Briles, D. E. & Szalai, A. J. (1999). Pneumococcal surface protein A inhibits complement activation by *Streptococcus pneumoniae*. *Infect Immun* **67**, 4720-4724.

Uchiyama, S., Carlin, A. F., Khosravi, A., Weiman, S., Banerjee, A., Quach, D., Hightower, G., Mitchell, T. J., Doran, K. S. & Nizet, V. (2009). The surface-anchored NanA protein promotes pneumococcal brain endothelial cell invasion. *J Exp Med* **206**, 1845-1852.

van der Poll, T. & Opal, S. M. (2009). Pathogenesis, treatment, and prevention of pneumococcal pneumonia. *Lancet* **374**, 1543-1556.

van der Woude, M. W. & Baumler, A. J. (2004). Phase and antigenic variation in bacteria. *Clin Microbiol Rev* **17**, 581-611, table of contents.

van Ginkel, F. W., McGhee, J. R., Watt, J. M., Campos-Torres, A., Parish, L. A. & Briles, D. E. (2003). Pneumococcal carriage results in ganglioside-mediated olfactory tissue infection. *Proc Natl Acad Sci U S A* **100**, 14363-14367.

van Ham, S. M., van Alphen, L., Mooi, F. R. & van Putten, J. P. (1993). Phase variation of *H. influenzae* fimbriae: transcriptional control of two divergent genes through a variable combined promoter region. *Cell* **73**, 1187-1196.

van Rossum, A. M., Lysenko, E. S. & Weiser, J. N. (2005). Host and bacterial factors contributing to the clearance of colonization by *Streptococcus pneumoniae* in a murine model. *Infect Immun* **73**, 7718-7726.

Walther, P., Wootton, S. & Schade, L. L. (2013). The future of forensics has arrived: The application of next generation sequencing. In *Forensic Magazine*.

Wang, Z., Gerstein, M. & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57-63.

Ware, D., Jiang, Y., Lin, W. & Swiatlo, E. (2006). Involvement of *potD* in *Streptococcus pneumoniae* polyamine transport and pathogenesis. *Infect Immun* **74**, 352-361.

Waters, L. S. & Storz, G. (2009). Regulatory RNAs in bacteria. *Cell* **136**, 615-628.

Watson, M. E., Jr., Jarisch, J. & Smith, A. L. (2004). Inactivation of deoxyadenosine methyltransferase (dam) attenuates *Haemophilus influenzae* virulence. *Mol Microbiol* **53**, 651-664.

Ween, O., Gaustad, P. & Havarstein, L. S. (1999). Identification of DNA binding sites for ComE, a key regulator of natural competence in *Streptococcus pneumoniae*. *Mol Microbiol* **33**, 817-827.

Ween, O., Teigen, S., Gaustad, P., Kilian, M. & Havarstein, L. S. (2002). Competence without a competence pheromone in a natural isolate of *Streptococcus infantis*. *J Bacteriol* **184**, 3426-3432.

Weinberger, D. M., Malley, R. & Lipsitch, M. (2011). Serotype replacement in disease after pneumococcal vaccination. *Lancet* **378**, 1962-1973.

Weinrauch, Y. & Lacks, S. A. (1981). Nonsense mutations in the amylomaltase gene and other loci of *Streptococcus pneumoniae*. *Mol Gen Genet* **183**, 7-12.

Weiser, J. N., Maskell, D. J., Butler, P. D., Lindberg, A. A. & Moxon, E. R. (1990). Characterization of repetitive sequences controlling phase variation of *Haemophilus influenzae* lipopolysaccharide. *J Bacteriol* **172**, 3304-3309.

Weiser, J. N. (1993). Relationship between colony morphology and the life cycle of *Haemophilus influenzae*: the contribution of lipopolysaccharide phase variation to pathogenesis. *J Infect Dis* **168**, 672-680.

Weiser, J. N., Austrian, R., Sreenivasan, P. K. & Masure, H. R. (1994). Phase variation in pneumococcal opacity: relationship between colonial morphology and nasopharyngeal colonization. *Infect Immun* **62**, 2582-2589.

Weiser, J. N., Markiewicz, Z., Tuomanen, E. I. & Wani, J. H. (1996). Relationship between phase variation in colony morphology, intrastrain variation in cell wall physiology, and nasopharyngeal colonization by *Streptococcus pneumoniae*. *Infect Immun* **64**, 2240-2245.

Weiser, J. N. (1998). Phase variation in colony opacity by *Streptococcus pneumoniae*. *Microb Drug Resist* **4**, 129-135.

Weiser, J. N. & Kapoor, M. (1999). Effect of intrastrain variation in the amount of capsular polysaccharide on genetic transformation of *Streptococcus pneumoniae*: implications for virulence studies of encapsulated strains. *Infect Immun* **67**, 3690-3692.

Weiser, J. N., Bae, D., Epino, H., Gordon, S. B., Kapoor, M., Zenewicz, L. A. & Shchepetov, M. (2001). Changes in availability of oxygen accentuate differences in capsular polysaccharide expression by phenotypic variants and clinical isolates of *Streptococcus pneumoniae*. *Infect Immun* **69**, 5430-5439.

Weiser, J. N. (2010). The pneumococcus: why a commensal misbehaves. *J Mol Med (Berl)* **88**, 97-102.

Wellmer, A., Zysk, G., Gerber, J., Kunst, T., Von Mering, M., Bunkowski, S., Eiffert, H. & Nau, R. (2002). Decreased virulence of a pneumolysin-deficient strain of *Streptococcus pneumoniae* in murine meningitis. *Infect Immun* **70**, 6504-6508.

Westerink, M. A., Schroeder, H. W., Jr. & Nahm, M. H. (2012). Immune Responses to pneumococcal vaccines in children and adults: Rationale for age-specific vaccination. *Aging Dis* **3**, 51-67.

Weyand, N. J. & Low, D. A. (2000). Regulation of Pap phase variation. Lrp is sufficient for the establishment of the phase off *pap* DNA methylation pattern and repression of *pap* transcription in vitro. *J Biol Chem* **275**, 3192-3200.

Willems, R., Paul, A., van der Heide, H. G., ter Avest, A. R. & Mooi, F. R. (1990). Fimbrial phase variation in *Bordetella pertussis*: a novel mechanism for transcriptional regulation. *Embo J* **9**, 2803-2809.

Wilson, G. G. (1991). Organization of restriction-modification systems. *Nucleic Acids Res* **19**, 2539-2566.

Wilson, R., Cohen, J. M., Jose, R. J., de Vogel, C., Baxendale, H. & Brown, J. S. (2014). Protection against *Streptococcus pneumoniae* lung infection after nasopharyngeal colonization requires both humoral and cellular immune responses. *Mucosal Immunol* **2014**, 95.

Wren, B. W. (1991). A family of clostridial and streptococcal ligand-binding proteins with conserved C-terminal repeat sequences. *Mol Microbiol* **5**, 797-803.

Wyres, K. L., Lambertsen, L. M., Croucher, N. J., McGee, L., von Gottberg, A., Linares, J., Jacobs, M. R., Kristinsson, K. G., Beall, B. W., Klugman, K. P., Parkhill, J., Hakenbeck, R., Bentley, S. D. & Brueggemann, A. B. (2013). Pneumococcal capsular switching: a historical perspective. *J Infect Dis* **207**, 439-449.

Xu, G., Kiefel, M. J., Wilson, J. C., Andrew, P. W., Oggioni, M. R. & Taylor, G. L. (2011). Three *Streptococcus pneumoniae* sialidases: three different products. *J Am Chem Soc* **133**, 1718-1721.

Yeo, Z. X., Chan, M., Yap, Y. S., Ang, P., Rozen, S. & Lee, A. S. (2012). Improving indel detection specificity of the Ion Torrent PGM benchtop sequencer. *PLoS One* **7**, e45798.

Yesilkaya, H., Spissu, F., Carvalho, S. M., Terra, V. S., Homer, K. A., Benisty, R., Porat, N., Neves, A. R. & Andrew, P. W. (2009). Pyruvate formate lyase is required for pneumococcal fermentative metabolism and virulence. *Infect Immun* **77**, 5418-5427.

Zahner, D., Grebe, T., Guenzi, E., Krauss, J., van der Linden, M., Terhune, K., Stock, J. B. & Hakenbeck, R. (1996). Resistance determinants for beta-lactam antibiotics in laboratory mutants of *Streptococcus pneumoniae* that are involved in genetic competence. *Microb Drug Resist* **2**, 187-191.

Zhang, J. R., Mostov, K. E., Lamm, M. E., Nanno, M., Shimida, S., Ohwaki, M. & Tuomanen, E. (2000). The polymeric immunoglobulin receptor translocates pneumococci across human nasopharyngeal epithelial cells. *Cell* **102**, 827-837.

Zhu, J. & Shimizu, K. (2004). The effect of pfl gene knockout on the metabolism for optically pure D-lactate production by *Escherichia coli*. *Appl Microbiol Biotechnol* **64**, 367-375.

Zhu, L. & Lau, G. W. (2011). Inhibition of competence development, horizontal gene transfer and virulence in *Streptococcus pneumoniae* by a modified competence stimulating peptide. *PLoS Pathog* **7**, e1002241.

Zysk, G., Schneider-Wald, B. K., Hwang, J. H., Bejo, L., Kim, K. S., Mitchell, T. J., Hakenbeck, R. & Heinz, H. P. (2001). Pneumolysin is the main inducer of cytotoxicity to brain microvascular endothelial cells caused by *Streptococcus pneumoniae*. *Infect Immun* **69**, 845-852.

Appendices

Appendix A. Differential gene expression of *S. pneumoniae* as determined by microarray analysis

Gene name and/or description[a]	TIGR4 annotation[a]	Fold change[b] and *p*-value[c] in:					
		D39		WCH16		WCH43	
		Fold-change	*p*-value	Fold-change	*p*-value	Fold-change	*p*-value
Transposons							
IS630-Spn1, transposase Orf1	0299	2.53	7.84E-08				
IS630-Spn1, transposase Orf2	0300	2.01	1.01E-06				
IS1167, transposase; K07485 transposase	0865			-1.49	5.50E-04		
IS1167, transposase; K07485 transposase	1692			-1.35	2.70E-03		
Unknown							
hypothetical protein	0029	9.85	4.89E-11				
hypothetical protein	0031	11.27	2.90E-13				
hypothetical protein	0055	2.60	3.08E-08				
hypothetical protein	0077			1.41	5.71E-04		
hypothetical protein	0096	4.14	2.51E-07			2.32	1.28E-02
hypothetical protein	0097			3.11	3.51E-06		
hypothetical protein	0098			2.58	8.09E-06		
hypothetical protein	0099			2.01	3.55E-03	1.68	1.45E-02
hypothetical protein	0115					1.51	1.93E-02
hypothetical protein	0122	2.89	1.22E-07				
ABC transporter, substrate-binding protein	0148	2.76	1.37E-09				
hypothetical protein	0191	2.90	1.39E-07				
hypothetical protein	0270			-1.40	1.11E-03		
hypothetical protein	0276	2.67	1.00E-08				
hypothetical protein (CAAX amino acid protease family protein)	0288			-1.80	1.44E-04		
hypothetical protein	0293	4.74	1.30E-08				
hypothetical protein	0311	3.22	7.31E-07				
hypothetical protein	0385			1.71	6.48E-05		
hypothetical protein	0389			1.56	5.50E-04		
hypothetical protein	0404	2.80	5.79E-08				
hypothetical protein	0429	3.67	5.98E-08				
hypothetical protein	0448					-1.59	1.62E-02
hypothetical protein	0449	-2.26	2.97E-07				
hypothetical protein	0487	2.87	8.59E-07				
BlpT protein, fusion	0524	3.51	1.10E-08				
hypothetical protein	0534			-1.48	1.89E-03		
hypothetical protein	0552	3.29	9.95E-09				
hypothetical protein	0565	-2.79	2.96E-08				
hypothetical protein	0639	-2.32	1.70E-07				
hypothetical protein	0677	2.76	3.38E-07				
hypothetical protein	0682	-3.34	1.82E-06				
hypothetical protein	0742			-2.28	6.39E-04		
hypothetical protein	0748	2.59	5.07E-08				
hypothetical protein	0781	2.68	5.37E-09				
hypothetical protein	0782	3.67	5.90E-08				
hypothetical protein; putative ABC transport system permease protein	0787			-1.50	7.63E-05		

Continues on following page

Appendix A continued

Gene name and/or description[a]	TIGR4 annotation[a]	Fold change[b] and *p*-value[c] in:					
		D39		WCH16		WCH43	
		Fold-change	*p*-value	Fold change	*p*-value	Fold-change	*p*-value
hypothetical protein	0800					-1.72	1.93E-02
4-methyl-5(b-hydroxyethyl)-thiazole monophosphate biosynthesis protein, putative	0804	-2.78	2.36E-10				
hemolysin-related protein	0834	-2.10	7.48E-07				
hypothetical protein	0858	-2.50	1.58E-06				
hypothetical protein	0859	-2.34	2.46E-07				
hypothetical protein; GAF domain-containing protein	0864			1.40	6.29E-04		
NifU family protein	0870	-2.60	6.79E-08	-2.10	1.31E-04		
hypothetical protein	0910	2.46	1.09E-07	1.85	1.51E-05		
hypothetical protein	0911					1.71	1.50E-02
hypothetical protein	0924	3.48	6.19E-11				
hypothetical protein	0947			-1.36	4.66E-04		
hypothetical protein	0956					1.51	3.92E-02
hypothetical protein	0990	-2.36	7.49E-08				
hypothetical protein	1007			-1.65	4.82E-04		
hypothetical protein	1027	4.54	1.40E-08				
helicase, putative	1028	5.30	1.90E-07				
hypothetical protein	1041	-3.28	1.26E-06				
hypothetical protein; putative ABC transport system substrate-binding protein	1069	-3.82	6.63E-10				
hypothetical protein; putative ABC transport system permease protein	1070			-1.55	1.61E-03		
ABC transporter, ATP-binding protein	1071	2.03	1.26E-06	-1.69	3.61E-07		
hypothetical protein	1083	-3.21	4.21E-07				
hypothetical protein	1153	-2.18	1.95E-08				
hypothetical protein	1232	-2.14	5.58E-08	1.80	9.47E-06		
hypothetical protein	1261	2.36	1.70E-06	1.42	3.09E-04		
ABC transporter, ATP-binding protein	1282			-1.48	1.51E-05		
hypothetical protein	1284			1.42	2.24E-05		
hypothetical protein (putative lantibiotic synthetase)	1344	2.23	5.01E-08				
hypothetical protein	1353			-1.73	1.11E-03		
ABC transporter ATP-binding protein/permease	1357			-1.36	4.94E-04		
ABC transporter, ATP-binding protein	1358			-1.42	3.30E-05		
hypothetical protein	1364	-3.41	1.07E-07				
hypothetical protein	1454	2.59	1.65E-06				
hypothetical protein	1465			1.51	3.60E-06		
hypothetical protein	1473	-2.38	4.11E-08				
hypothetical protein	1480	3.80	1.50E-09				
hypothetical protein	1481	3.80	2.76E-07				
ABC transporter, ATP-binding protein	1553	2.02	1.05E-06				
hypothetical protein	1561	2.39	1.15E-08				
ABC transporter, ATP-binding protein	1553	2.02	1.05E-06				
hypothetical protein	1561	2.39	1.15E-08				
hypothetical protein	1562	2.54	1.28E-08				
hypothetical protein	1612			1.42	9.57E-05		
hypothetical protein	1628	3.39	7.18E-07				

Continues on following page

Appendix A - continued

Gene name and/or description[a]	TIGR4 annotation[b]	Fold change[b] and *p*-value[c] in:					
		D39		WCH16		WCH43	
		Fold change	*p*-value	Fold-change	*p*-value	Fold-change	*p*-value
hypothetical protein	1634	3.14	8.59E-07				
hypothetical protein (putative ABC transport system permease protein)	1652	3.10	8.12E-09				
hypothetical protein	1660	-2.51	4.19E-08				
hypothetical protein	1705			1.57	3.73E-04		
hypothetical protein	1706			1.43	7.79E-04		
hypothetical protein	1707					1.58	3.87E-02
hypothetical protein	1720	2.20	1.09E-06				
hypothetical protein	1746	-2.23	6.33E-10				
hypothetical protein	1775	2.75	1.29E-06				
hypothetical protein	1801	2.64	2.75E-07				
hypothetical protein	1810	3.60	3.33E-09				
hypothetical protein	1862			-1.33	2.02E-03		
hypothetical protein	1914			3.09	2.63E-09		
hypothetical protein	1925			1.32	7.39E-05		
hypothetical protein	1926			1.43	2.15E-03		
hypothetical protein	1938	2.42	8.29E-07				
hypothetical protein	1945	4.60	2.23E-07				
hypothetical protein	1947	3.43	9.28E-08	1.33	1.89E-03		
hypothetical protein	1972			1.40	2.35E-04		
hypothetical protein	2017	4.73	1.38E-10				
hypothetical protein	2047			2.50	1.50E-03		
hypothetical protein	2049	8.08	4.39E-08	2.19	3.87E-04		
hypothetical protein	2057	-2.11	1.39E-07				
LysM domain-containing protein	2063			2.93	1.28E-03		
hypothetical protein	2081	-2.01	4.91E-07				
hypothetical protein	2102	2.92	1.01E-07				
hypothetical protein	2115	4.31	1.69E-08				
hypothetical protein (SPFH domain/band 7 family protein)	2156			1.97	1.46E-03		
hypothetical protein	2177	5.69	1.85E-07	1.76	6.05E-05		
hypothetical protein	2182					2.91	2.58E-06
hypothetical protein	2187			-1.47	8.50E-04		
hypothetical protein	2209	2.17	7.05E-07	1.45	3.81E-05		
orf47; hypothetical protein	SpR6 0181	9.36	2.36E-10				
hypothetical protein	SpR6 1213			-1.58	2.46E-04		

[a]Gene name and annotation according to KEGG (http://www.kegg.com/) website

[b]A positive fold change indicates up-regulation in the O form and a negative fold change indicates up-regulation in the T form

[c]The *p-value* used here is the adjusted *p*-value of four biological replicates

Appendix B. Raw data of the comparison of D39O vs. D39T gene regulation using DNA microarray analysis (Section 4.2.2).

Spot ID[a]	Fold change	T vs. O	*p*-value[b]	Bayesian
SpTIGR4-1908 (5I23)	25.79	Down	1.82E-11	23.62
SpTIGR4-1088 (3N16)	25.71	Down	4.49E-10	18.02
SpTIGR4-0125 (1M4)	19.59	Down	2.49E-10	18.89
SpTIGR4-1907 (5P22)	15.60	Down	1.22E-07	10.35
SpTIGR4-0030 (1F4)	12.52	Down	1.82E-11	23.27
SpTIGR4-0954 (3P11)	11.36	Down	8.01E-08	10.92
SpTIGR4-0031 (1G4)	11.27	Down	2.90E-13	28.21
SpTIGR4-0029 (1E4)	9.85	Down	4.89E-11	21.75
SpTIGR4-2050 (6O4)	9.80	Down	3.76E-08	12.01
SpR6-0181 (6I19)	9.36	Down	2.36E-10	19.19
SpTIGR4-2053 (6J5)	9.19	Down	3.37E-09	15.23
SpTIGR4-2046 (6K4)	8.29	Down	3.37E-09	15.25
SpTIGR4-2049 (6N4)	8.08	Down	4.39E-08	11.79
SpTIGR4-0837 (3C9)	7.89	Down	3.61E-11	22.32
SpTIGR4-1659 (5P3)	6.98	Up	2.36E-10	19.36
SpTIGR4-2048 (6M4)	6.89	Down	7.03E-07	8.19
SpTIGR4-1035 (3A22)	6.88	Up	2.73E-09	15.50
SpTIGR4-1941 (6B3)	6.46	Down	2.00E-09	15.92
SpTIGR4-0862 (3D12)	6.35	Up	8.00E-10	17.06
SpTIGR4-1089 (3O16)	6.29	Down	1.36E-09	16.40
SpTIGR4-2019 (6H12)	5.94	Down	5.90E-08	11.29
SpTIGR4-2177 (6E20)	5.69	Down	1.85E-07	9.84
SpTIGR4-0033 (1A5)	5.43	Down	5.07E-08	11.55
SpTIGR4-2051 (6P4)	5.31	Down	2.52E-07	9.43
SpTIGR4-1028 (3B21)	5.30	Down	1.90E-07	9.81
SpTIGR4-2208 (6D24)	4.99	Down	2.86E-08	12.39
SpTIGR4-1508 (4I21)	4.78	Up	5.79E-08	11.36
SpR6-0317 (6P19)	4.74	Up	1.22E-07	10.34
SpTIGR4-0293 (1M13)	4.74	Down	1.30E-08	13.43
SpTIGR4-2017 (6F12)	4.73	Down	1.38E-10	20.21
SpTIGR4-1174 (4D3)	4.72	Up	4.96E-11	21.47
SpTIGR4-1945 (6F3)	4.60	Down	2.23E-07	9.59
SpTIGR4-1027 (3A21)	4.54	Down	1.40E-08	13.33
SpTIGR4-1160 (4F1)	4.47	Up	2.77E-08	12.44
SpTIGR4-1161 (4G1)	4.39	Up	4.95E-07	8.59
SpTIGR4-2115 (6P12)	4.31	Down	1.69E-08	13.09
SpTIGR4-1510 (4K21)	4.16	Up	2.49E-10	18.91
SpTIGR4-1004 (3B18)	4.14	Up	7.09E-07	8.16
SpTIGR4-0096 (1H12)	4.14	Down	2.51E-07	9.43
SpTIGR4-0231 (1G17)	4.12	Up	2.62E-07	9.37
SpTIGR4-0869 (3K1)	4.10	Up	4.49E-10	17.99
SpTIGR4-1201 (4G6)	4.06	Up	8.00E-10	17.13
SpTIGR4-1511 (4L21)	4.04	Up	5.26E-09	14.68
SpTIGR4-1162 (4H1)	4.01	Up	1.08E-07	10.51
SpTIGR4-2207 (6C24)	4.01	Down	1.55E-08	13.20
SpTIGR4-0848 (3F10)	4.00	Up	3.46E-09	15.14
SpTIGR4-1266 (4O2)	3.97	Down	2.77E-08	12.44
SpR6-1060 (7D1)	3.97	Up	7.01E-09	14.29
SpTIGR4-1274 (4O3)	3.91	Up	6.31E-10	17.58
SpTIGR4-1069 (3K14)	3.82	Up	6.63E-10	17.46
SpR6-0316 (6O19)	3.80	Up	5.31E-09	14.65
SpTIGR4-1480 (4M17)	3.80	Down	1.50E-09	16.26
SpTIGR4-1481 (4N17)	3.80	Down	2.76E-07	9.29
SpTIGR4-1509 (4J21)	3.76	Up	7.49E-09	14.21

Continues on following page

Appendix B continued

Spot ID[a]	Fold change	T vs. O	*p*-value[b]	Bayesian
SpTIGR4-1507 (4P20)	3.76	Up	4.49E-10	17.99
SpTIGR4-0807 (3E5)	3.76	Up	4.01E-08	11.92
SpTIGR4-1003 (3A18)	3.73	Up	9.57E-07	7.80
SpTIGR4-1200 (4F6)	3.73	Up	1.67E-10	19.92
SpTIGR4-1778 (5G18)	3.68	Up	1.92E-08	12.93
SpTIGR4-0782 (3D2)	3.67	Down	5.90E-08	11.29
SpTIGR4-0429 (2E6)	3.67	Down	5.98E-08	11.27
SpTIGR4-1068 (3J14)	3.66	Up	4.89E-11	21.63
SpTIGR4-2201 (6E23)	3.65	Down	2.10E-07	9.69
SpR6-0318 (6I20)	3.64	Up	2.36E-10	19.23
SpTIGR4-1193 (4G5)	3.64	Down	1.00E-06	7.74
SpTIGR4-1512 (4M21)	3.64	Up	3.63E-07	8.97
SpTIGR4-0603 (2A16)	3.64	Down	2.36E-10	19.33
SpR6-0322 (6M20)	3.61	Up	1.28E-08	13.49
SpTIGR4-1810 (5G22)	3.60	Down	3.33E-09	15.28
SpTIGR4-1383 (4D17)	3.59	Up	6.21E-08	11.22
SpTIGR4-0102 (1N1)	3.56	Down	3.11E-08	12.21
SpR6-0319 (6J20)	3.53	Up	3.46E-09	15.15
SpTIGR4-0524 (2K6)	3.51	Down	1.10E-08	13.70
SpTIGR4-0924 (3J8)	3.48	Down	6.19E-11	21.12
SpTIGR4-1661 (5J4)	3.46	Up	1.33E-09	16.47
SpTIGR4-0976 (3F14)	3.46	Up	1.40E-08	13.32
SpTIGR4-1202 (4H6)	3.45	Up	8.81E-09	13.97
SpTIGR4-1947 (6H3)	3.43	Down	9.28E-08	10.71
SpTIGR4-1364 (4A15)	3.41	Up	1.07E-07	10.54
SpTIGR4-0697 (2O15)	3.40	Up	3.46E-09	15.17
SpTIGR4-1628 (5A12)	3.39	Down	7.18E-07	8.14
SpTIGR4-1992 (6E9)	3.37	Down	3.84E-08	11.98
SpTIGR4-0275 (1C23)	3.35	Down	9.94E-09	13.84
SpTIGR4-0338 (1J19)	3.35	Down	2.70E-08	12.50
SpTIGR4-0682 (2P13)	3.34	Up	1.82E-06	7.03
SpTIGR4-1586 (5G6)	3.34	Up	2.07E-08	12.82
SpTIGR4-0675 (2I13)	3.33	Up	9.03E-10	16.91
SpTIGR4-0552 (2N9)	3.29	Down	9.95E-09	13.82
SpTIGR4-1041 (3G22)	3.28	Up	1.26E-06	7.48
SpR6-0321 (6L20)	3.27	Up	8.12E-09	14.10
SpTIGR4-2218 (6N13)	3.27	Up	3.08E-10	18.63
SpTIGR4-2006 (6C11)	3.26	Down	1.37E-06	7.36
SpTIGR4-1084 (3J16)	3.24	Up	2.49E-07	9.45
SpTIGR4-0311 (1O15)	3.22	Down	7.31E-07	8.12
SpTIGR4-1083 (3I16)	3.21	Up	4.21E-07	8.80
SpTIGR4-1980 (6A8)	3.16	Down	1.28E-08	13.48
SpTIGR4-1002 (3H17)	3.15	Up	8.70E-09	14.00
SpTIGR4-0897 (3O4)	3.15	Up	7.01E-09	14.29
SpTIGR4-1634 (5G12)	3.14	Down	8.59E-07	7.92
SpTIGR4-1906 (5O22)	3.14	Down	3.46E-09	15.16
SpR6-1042 (7C1)	3.12	Up	1.40E-08	13.35
SpTIGR4-1273 (4N3)	3.12	Up	1.30E-08	13.44
SpTIGR4-1652 (5I3)	3.10	Down	8.12E-09	14.10
SpTIGR4-1888 (5M20)	3.09	Up	1.35E-07	10.22
SpTIGR4-2077 (6J8)	3.04	Down	5.01E-08	11.57
SpTIGR4-1559 (5D3)	3.04	Up	1.40E-08	13.31
SpTIGR4-1542 (5C1)	3.03	Up	7.31E-10	17.29
SpTIGR4-0825 (3G7)	3.00	Up	3.08E-08	12.23
SpTIGR4-1960 (6E5)	2.99	Up	4.71E-08	11.68
SpTIGR4-0717 (2K18)	2.97	Up	5.37E-09	14.61
SpTIGR4-0846 (3D10)	2.94	Up	2.71E-08	12.48
SpTIGR4-0028 (1D4)	2.93	Down	2.96E-08	12.33

Continues on following page

Appendix B continued

Spot ID[a]	Fold change	T vs. O	*p*-value[b]	Bayesian
SpTIGR4-1858 (5O16)	2.93	Down	2.45E-07	9.48
SpTIGR4-2102 (6K11)	2.92	Down	1.01E-07	10.61
SpTIGR4-0419 (2C5)	2.91	Up	2.60E-07	9.39
SpTIGR4-0191 (1O12)	2.90	Down	1.39E-07	10.18
SpTIGR4-1074 (3P14)	2.89	Down	5.64E-07	8.44
SpTIGR4-0718 (2L18)	2.89	Up	6.43E-08	11.17
SpTIGR4-0403 (2C3)	2.89	Down	1.73E-07	9.93
SpTIGR4-0122 (1J4)	2.89	Down	1.22E-07	10.33
SpR6-0315 (6N19)	2.88	Up	4.71E-08	11.68
SpTIGR4-2219 (6O13)	2.88	Up	4.81E-07	8.64
SpTIGR4-0081 (1A11)	2.87	Down	1.25E-09	16.56
SpTIGR4-1296 (4M6)	2.87	Down	3.29E-07	9.08
SpTIGR4-1709 (5J10)	2.87	Up	2.49E-10	19.02
SpTIGR4-0487 (2N1)	2.87	Down	8.59E-07	7.92
SpTIGR4-2169 (6E19)	2.86	Up	8.01E-08	10.92
SpR6-0320 (6K20)	2.86	Up	5.97E-07	8.38
SpTIGR4-1666 (5O4)	2.82	Up	8.00E-10	17.08
SpTIGR4-1244 (4A12)	2.81	Up	2.90E-08	12.37
SpTIGR4-1469 (4J16)	2.81	Up	2.49E-10	18.99
SpTIGR4-0669 (2C24)	2.80	Down	2.61E-07	9.38
SpTIGR4-0404 (2D3)	2.80	Down	5.79E-08	11.36
SpTIGR4-1662 (5K4)	2.79	Up	1.22E-07	10.33
SpR6-0581 (6O21)	2.79	Up	4.11E-09	14.94
SpTIGR4-0565 (2K11)	2.79	Up	2.96E-08	12.33
SpTIGR4-0804 (3B5)	2.78	Up	2.36E-10	19.18
SpTIGR4-2221 (6I14)	2.78	Up	8.24E-08	10.88
SpTIGR4-0148 (1L7)	2.76	Up	1.37E-09	16.37
SpTIGR4-0677 (2K13)	2.76	Down	3.38E-07	9.05
SpTIGR4-1225 (4G9)	2.75	Up	5.90E-08	11.30
SpTIGR4-1775 (5D18)	2.75	Down	1.29E-06	7.44
SpTIGR4-1890 (5O20)	2.74	Up	7.41E-07	8.10
SpTIGR4-1637 (5J1)	2.71	Down	2.21E-09	15.78
SpTIGR4-0847 (3E10)	2.71	Up	8.00E-10	17.14
SpTIGR4-1176 (4F3)	2.71	Up	2.59E-09	15.60
SpTIGR4-0781 (3C2)	2.68	Down	5.37E-09	14.62
SpTIGR4-0446 (2F8)	2.68	Up	1.78E-06	7.06
SpTIGR4-0276 (1D23)	2.67	Down	1.00E-08	13.80
SpTIGR4-0865 (3G12)	2.66	Up	1.13E-06	7.59
SpTIGR4-1033 (3G21)	2.65	Up	3.08E-08	12.24
SpTIGR4-1801 (5F21)	2.64	Down	2.75E-07	9.30
SpTIGR4-0589 (2C14)	2.63	Up	1.08E-07	10.50
SpTIGR4-0100 (1L1)	2.62	Down	4.80E-07	8.65
SpTIGR4-1365 (4B15)	2.61	Up	1.26E-06	7.47
SpTIGR4-0870 (3L1)	2.60	Up	6.79E-08	11.11
SpTIGR4-1246 (4C12)	2.60	Up	2.10E-08	12.79
SpTIGR4-0055 (1G7)	2.60	Down	3.08E-08	12.23
SpTIGR4-0748 (2J22)	2.59	Down	5.07E-08	11.55
SpTIGR4-1454 (4K14)	2.59	Down	1.65E-06	7.16
SpTIGR4-0581 (2C13)	2.58	Up	2.14E-07	9.66
SpTIGR4-0381 (1M24)	2.58	Down	4.49E-10	18.07
SpTIGR4-1850 (5O15)	2.57	Down	1.83E-06	7.02
SpTIGR4-0803 (3A5)	2.57	Up	1.83E-07	9.86
SpTIGR4-1996 (6A10)	2.56	Down	9.66E-07	7.78
SpTIGR4-1966 (6C6)	2.56	Up	5.64E-09	14.54
SpTIGR4-1474 (4O16)	2.56	Up	2.16E-07	9.64
SpTIGR4-1359 (4D14)	2.55	Up	5.01E-07	8.58
SpR6-1995 (7A7)	2.55	Up	2.23E-07	9.59
SpTIGR4-1082 (3P15)	2.54	Up	1.94E-07	9.77

Continues on following page

Appendix B continued

Spot ID[a]	Fold change	T vs. O	*p*-value[b]	Bayesian
SpTIGR4-1562 (5G3)	2.54	Down	1.28E-08	13.49
SpTIGR4-0057 (1A8)	2.54	Down	1.09E-07	10.49
SpTIGR4-1923 (5P24)	2.53	Up	4.40E-07	8.75
SpTIGR4-0299 (1K14)	2.53	Down	7.84E-08	10.95
SpTIGR4-0556 (2J10)	2.53	Up	5.68E-08	11.40
SpTIGR4-0515 (2J5)	2.53	Down	2.77E-08	12.43
SpTIGR4-1638 (5K1)	2.52	Down	1.91E-09	15.99
SpTIGR4-1354 (4G13)	2.52	Up	1.22E-07	10.34
SpTIGR4-1660 (5I4)	2.51	Up	4.19E-08	11.84
SpTIGR4-0858 (3H11)	2.50	Up	1.58E-06	7.20
SpTIGR4-1610 (5G9)	2.49	Down	1.92E-08	12.92
SpTIGR4-1517 (4J22)	2.48	Up	2.19E-09	15.81
SpTIGR4-0508 (2K4)	2.48	Up	1.28E-06	7.45
SpTIGR4-0845 (3C10)	2.48	Up	1.36E-09	16.41
SpTIGR4-1946 (6G3)	2.47	Down	2.11E-08	12.78
SpTIGR4-0910 (3L6)	2.46	Down	1.09E-07	10.48
SpTIGR4-0855 (3E11)	2.46	Up	5.83E-10	17.69
SpTIGR4-0013 (1E2)	2.45	Up	3.08E-08	12.26
SpTIGR4-1449 (4N13)	2.45	Down	7.65E-07	8.06
SpTIGR4-0857 (3G11)	2.45	Up	1.63E-08	13.14
SpTIGR4-1701 (5J9)	2.44	Up	3.95E-08	11.94
SpTIGR4-1013 (3C19)	2.44	Up	2.93E-07	9.22
SpTIGR4-1700 (5I9)	2.43	Up	9.79E-08	10.65
SpTIGR4-1938 (6G2)	2.42	Down	8.29E-07	7.97
SpTIGR4-1241 (4F11)	2.42	Up	4.04E-08	11.89
SpTIGR4-1889 (5N20)	2.42	Up	1.85E-07	9.84
SpTIGR4-1306 (4O7)	2.42	Up	1.10E-08	13.68
SpTIGR4-1624 (5E11)	2.41	Down	5.79E-08	11.35
SpTIGR4-1602 (5G8)	2.41	Down	7.19E-10	17.34
SpTIGR4-2189 (6A22)	2.40	Up	1.76E-08	13.04
SpR6-1194 (7B3)	2.40	Up	2.75E-07	9.30
SpTIGR4-1561 (5F3)	2.39	Down	1.15E-08	13.62
SpTIGR4-1961 (6F5)	2.39	Up	1.51E-07	10.09
SpTIGR4-1473 (4N16)	2.38	Up	4.11E-08	11.87
SpTIGR4-0021 (1E3)	2.37	Down	5.79E-08	11.36
SpTIGR4-1745 (5F14)	2.37	Up	3.93E-10	18.29
SpTIGR4-0990 (3D16)	2.36	Up	7.49E-08	11.01
SpTIGR4-0313 (1I16)	2.36	Down	1.59E-06	7.19
SpTIGR4-0789 (3C3)	2.36	Down	4.12E-07	8.83
SpTIGR4-1261 (4J2)	2.36	Down	1.70E-06	7.12
SpTIGR4-1646 (5K2)	2.35	Down	5.90E-08	11.31
SpTIGR4-0757 (2K23)	2.35	Up	3.71E-07	8.94
SpTIGR4-0859 (3A12)	2.34	Up	2.46E-07	9.47
SpTIGR4-0639 (2E20)	2.32	Up	1.70E-07	9.96
SpTIGR4-1513 (4N21)	2.32	Up	6.50E-07	8.27
SpTIGR4-0827 (3A8)	2.32	Down	1.08E-07	10.50
SpTIGR4-1664 (5M4)	2.32	Up	3.17E-07	9.13
SpTIGR4-0264 (1H21)	2.31	Up	1.76E-08	13.03
SpTIGR4-0370 (1J23)	2.31	Down	8.34E-08	10.86
SpTIGR4-1887 (5L20)	2.31	Up	8.55E-08	10.81
SpTIGR4-0644 (2B21)	2.30	Up	8.34E-08	10.85
SpTIGR4-1475 (4P16)	2.29	Up	3.08E-08	12.24
SpTIGR4-2195 (6G22)	2.28	Down	4.71E-08	11.69
SpTIGR4-0503 (2N3)	2.27	Up	3.23E-07	9.11
SpTIGR4-1072 (3N14)	2.27	Down	5.90E-08	11.30
SpTIGR4-1397 (4B19)	2.26	Up	1.93E-07	9.79
SpTIGR4-2016 (6E12)	2.26	Down	2.76E-07	9.29
SpTIGR4-0449 (2A9)	2.26	Up	2.97E-07	9.20

Continues on following page

Appendix B continued

Spot ID[a]	Fold change	T vs. O	*p*-value[b]	Bayesian
SpTIGR4-0376 (1P23)	2.25	Up	2.17E-07	9.62
SpTIGR4-1746 (5G14)	2.23	Up	6.33E-10	17.54
SpTIGR4-0526 (2M6)	2.23	Down	2.61E-07	9.37
SpTIGR4-1344 (4M12)	2.23	Down	5.01E-08	11.58
SpTIGR4-0079 (1G10)	2.22	Up	4.01E-08	11.91
SpTIGR4-1212 (4B8)	2.22	Up	2.54E-08	12.58
SpTIGR4-1001 (3G17)	2.21	Up	2.17E-07	9.62
SpTIGR4-1001 (3G17)	2.21	Up	2.17E-07	9.62
SpTIGR4-1499 (4P19)	2.21	Down	5.90E-08	11.30
SpTIGR4-1720 (5M11)	2.20	Down	1.09E-06	7.64
SpTIGR4-1245 (4B12)	2.20	Up	3.37E-10	18.49
SpTIGR4-1100 (3J18)	2.20	Up	7.74E-07	8.04
SpTIGR4-1575 (5D5)	2.20	Up	5.90E-07	8.39
SpTIGR4-1313 (4N8)	2.19	Down	4.61E-07	8.70
SpTIGR4-1169 (4G2)	2.19	Down	1.77E-06	7.07
SpTIGR4-1153 (3O24)	2.18	Up	1.95E-08	12.89
SpTIGR4-2209 (6E24)	2.17	Down	7.05E-07	8.17
SpTIGR4-1151 (3M24)	2.16	Up	2.14E-07	9.66
SpTIGR4-1665 (5N4)	2.16	Up	7.05E-07	8.18
SpTIGR4-0996 (3B17)	2.15	Down	4.40E-07	8.75
SpTIGR4-1405 (4B20)	2.14	Down	8.00E-10	17.11
SpTIGR4-1232 (4F10)	2.14	Up	5.58E-08	11.43
SpTIGR4-0362 (1J22)	2.13	Up	1.73E-06	7.10
SpTIGR4-2203 (6G23)	2.13	Up	4.50E-08	11.75
SpTIGR4-1115 (3I20)	2.12	Up	1.26E-06	7.47
SpTIGR4-1368 (4E15)	2.11	Up	3.87E-07	8.89
SpTIGR4-2057 (6N5)	2.11	Up	1.39E-07	10.18
SpTIGR4-1128 (3N21)	2.10	Up	1.71E-07	9.94
SpTIGR4-0834 (3H8)	2.10	Up	7.48E-07	8.09
SpTIGR4-1159 (4E1)	2.10	Down	5.01E-08	11.58
SpTIGR4-1355 (4H13)	2.09	Up	8.99E-08	10.75
SpTIGR4-1550 (5C2)	2.09	Up	4.91E-07	8.61
SpTIGR4-1242 (4G11)	2.09	Up	1.29E-06	7.43
SpTIGR4-0841 (3G9)	2.09	Down	1.44E-06	7.31
SpTIGR4-0083 (1C11)	2.08	Down	4.80E-07	8.65
SpTIGR4-1389 (4B18)	2.08	Up	2.73E-09	15.51
SpTIGR4-0236 (1D18)	2.06	Up	2.66E-07	9.34
SpTIGR4-2204 (6H23)	2.06	Up	1.22E-07	10.36
SpTIGR4-1744 (5E14)	2.06	Up	8.50E-09	14.04
SpTIGR4-1998 (6C10)	2.05	Down	2.51E-07	9.43
SpTIGR4-1226 (4H9)	2.04	Up	6.01E-08	11.26
SpTIGR4-0568 (2N11)	2.04	Up	1.78E-06	7.06
SpTIGR4-1071 (3M14)	2.03	Up	1.26E-06	7.47
SpTIGR4-1851 (5P15)	2.03	Down	4.71E-08	11.66
SpTIGR4-0739 (2I21)	2.02	Down	6.50E-07	8.28
SpTIGR4-1460 (4I15)	2.02	Up	1.20E-07	10.38
SpTIGR4-1553 (5F2)	2.02	Up	1.05E-06	7.68
SpTIGR4-1482 (4O17)	2.02	Down	8.34E-08	10.85
SpTIGR4-0300 (1L14)	2.01	Down	1.01E-06	7.73
SpTIGR4-0509 (2L4)	2.01	Up	5.51E-07	8.47
SpTIGR4-2081 (6N8)	2.01	Up	4.91E-07	8.61
SpTIGR4-1362 (4G14)	2.01	Down	6.00E-07	8.37

[a]Spot ID labelled on microarray slide
[b]Adjusted *p*-value across four biological replicates

Appendix C. Raw data of the comparison of WCH16O vs. WCH16T gene regulation using DNA microarray analysis (Section 4.2.3)[a].

Spot ID[b]	Fold change	T vs. O	*p*-value[c]	Bayesian
SpTIGR4-1249 (4F12)	3.78	Down	2.63E-09	18.42
SpTIGR4-0097 (1I1)	3.11	Down	3.51E-06	9.59
SpTIGR4-1914 (5O23)	3.09	Down	2.63E-09	18.23
SpTIGR4-2063 (6L6)	2.93	Down	1.28E-03	1.42
SpTIGR4-1915 (5P23)	2.75	Down	6.29E-09	17.14
SpTIGR4-0921 (3O7)	2.72	Up	7.48E-05	5.30
SpTIGR4-2050 (6O4)	2.70	Down	8.02E-05	5.14
SpTIGR4-0107 (1K2)	2.69	Down	8.93E-04	1.89
SpTIGR4-0918 (3L7)	2.68	Up	1.17E-05	7.93
SpTIGR4-0098 (1J1)	2.58	Down	8.09E-06	8.46
SpTIGR4-1913 (5N23)	2.56	Down	6.26E-08	14.41
SpTIGR4-2051 (6P4)	2.54	Down	2.66E-04	3.53
SpTIGR4-0919 (3M7)	2.53	Up	1.72E-05	7.26
SpTIGR4-2047 (6L4)	2.50	Down	1.50E-03	1.21
SpTIGR4-2053 (6J5)	2.36	Down	3.87E-04	3.07
SpTIGR4-0742 (2L21)	2.28	Up	6.39E-04	2.32
SpTIGR4-0287 (1G24)	2.27	Up	2.37E-06	10.24
SpTIGR4-2052 (6I5)	2.25	Down	3.26E-04	3.26
SpTIGR4-0869 (3K1)	2.23	Up	1.46E-09	19.78
SpTIGR4-2046 (6K4)	2.20	Down	1.02E-03	1.75
SpTIGR4-2049 (6N4)	2.19	Down	3.87E-04	3.06
SpTIGR4-0922 (3P7)	2.15	Up	5.63E-04	2.51
SpTIGR4-0870 (3L1)	2.10	Up	1.31E-04	4.53
SpTIGR4-2143 (6D16)	2.09	Up	2.13E-03	0.73
SpTIGR4-0916 (3J7)	2.08	Up	2.49E-06	10.05
SpTIGR4-0099 (1K1)	2.01	Down	3.55E-03	0.03
SpTIGR4-2156 (6H17)	1.97	Down	1.46E-03	1.25
SpTIGR4-2048 (6M4)	1.94	Down	1.47E-03	1.23
SpTIGR4-1814 (5C23)	1.92	Up	1.11E-03	1.60
SpTIGR4-1231 (4E10)	1.92	Down	5.97E-08	14.61
SpTIGR4-1812 (5A23)	1.91	Up	2.46E-04	3.68
SpTIGR4-0386 (2B1)	1.90	Down	2.24E-05	6.90
SpTIGR4-1815 (5D23)	1.89	Up	8.02E-04	2.04
SpTIGR4-2173 (6A20)	1.88	Down	1.01E-08	16.48
SpTIGR4-2176 (6D20)	1.88	Down	5.28E-06	8.97
SpTIGR4-0871 (3M1)	1.88	Up	4.06E-05	6.11
SpTIGR4-2175 (6C20)	1.87	Down	8.02E-05	5.15
SpTIGR4-0910 (3L6)	1.85	Down	1.51E-05	7.52
SpTIGR4-1811 (5H22)	1.84	Up	1.21E-03	1.49
SpTIGR4-0868 (3J1)	1.84	Up	1.27E-07	13.60
SpTIGR4-2207 (6C24)	1.83	Down	1.43E-03	1.28
SpTIGR4-1230 (4D10)	1.82	Down	3.13E-04	3.34
SpTIGR4-0387 (2C1)	1.80	Down	1.66E-04	4.22
SpTIGR4-1004 (3B18)	1.80	Up	1.55E-05	7.39
SpTIGR4-0288 (1H24)	1.80	Up	1.44E-04	4.38
SpTIGR4-1232 (4F10)	1.80	Down	9.47E-06	8.26
SpTIGR4-2091 (6P9)	1.79	Down	1.11E-03	1.61
SpTIGR4-0017 (1A3)	1.78	Up	2.30E-03	0.62
SpTIGR4-0731 (2I20)	1.77	Up	1.11E-03	1.62
SpTIGR4-2177 (6E20)	1.76	Down	6.05E-05	5.63
SpTIGR4-2216 (6L13)	1.75	Down	1.96E-03	0.87
SpTIGR4-0515 (2J5)	1.74	Down	2.60E-04	3.56
SpTIGR4-1174 (4D3)	1.73	Up	1.89E-03	0.93
SpTIGR4-1353 (4F13)	1.73	Up	1.11E-03	1.61

Continues on following page

Appendix C continued

Spot ID[b]	Fold change	T vs. O	*p*-value[c]	Bayesian
SpTIGR4-0716 (2J18)	1.72	Up	3.26E-04	3.27
SpTIGR4-1353 (4F13)	1.73	Up	1.11E-03	1.61
SpTIGR4-0716 (2J18)	1.72	Up	3.26E-04	3.27
SpTIGR4-0385 (2A1)	1.71	Down	6.48E-05	5.53
SpR6-1060 (7D1)	1.71	Up	2.86E-06	9.85
SpTIGR4-2045 (6J4)	1.71	Down	7.79E-04	2.08
SpTIGR4-1655 (5L3)	1.70	Down	5.94E-05	5.67
SpTIGR4-2240 (6L16)	1.70	Down	2.85E-05	6.57
SpTIGR4-1003 (3A18)	1.69	Up	1.12E-05	8.01
SpTIGR4-0771 (3A1)	1.69	Down	4.00E-06	9.31
SpTIGR4-1071 (3M14)	1.69	Up	3.61E-07	12.18
SpTIGR4-0390 (2F1)	1.69	Down	3.41E-03	0.08
SpTIGR4-2174 (6B20)	1.69	Down	1.72E-05	7.23
SpTIGR4-0286 (1F24)	1.69	Down	7.82E-05	5.19
SpTIGR4-1586 (5G6)	1.68	Up	4.78E-04	2.79
SpTIGR4-0717 (2K18)	1.67	Up	2.24E-05	6.92
SpTIGR4-2135 (6D15)	1.67	Up	5.94E-04	2.44
SpTIGR4-0281 (1A24)	1.66	Down	9.98E-06	8.16
SpTIGR4-0766 (2L24)	1.66	Up	3.05E-04	3.38
SpTIGR4-1002 (3H17)	1.66	Up	2.51E-04	3.62
SpTIGR4-1007 (3E18)	1.65	Up	4.82E-04	2.77
SpTIGR4-2201 (6E23)	1.64	Down	2.50E-04	3.64
SpTIGR4-1960 (6E5)	1.64	Up	1.98E-07	12.94
SpTIGR4-1816 (5E23)	1.63	Up	1.06E-03	1.69
SpTIGR4-1907 (5P22)	1.62	Down	1.52E-03	1.18
SpTIGR4-0102 (1N1)	1.62	Down	7.76E-06	8.54
SpTIGR4-0642 (2H20)	1.61	Up	5.07E-04	2.68
SpTIGR4-2142 (6C16)	1.61	Up	2.79E-03	0.36
SpTIGR4-1975 (6D7)	1.60	Down	2.55E-03	0.49
SpTIGR4-0760 (2N23)	1.60	Up	1.33E-04	4.49
SpTIGR4-2111 (6L12)	1.59	Down	1.09E-04	4.78
SpTIGR4-0867 (3I1)	1.59	Up	5.52E-04	2.54
SpTIGR4-0075 (1C10)	1.59	Down	5.97E-04	2.42
SpR6-1213 (7F3)	1.58	Up	2.46E-04	3.67
SpTIGR4-1969 (6F6)	1.57	Down	1.97E-07	13.05
SpTIGR4-0423 (2G5)	1.57	Up	1.43E-03	1.28
SpTIGR4-0808 (3F5)	1.57	Up	1.71E-04	4.18
SpTIGR4-0981 (3C15)	1.57	Down	5.43E-04	2.59
SpTIGR4-0426 (2B6)	1.57	Up	3.08E-03	0.21
SpTIGR4-2134 (6C15)	1.57	Up	1.79E-04	4.11
SpTIGR4-1705 (5N9)	1.57	Down	3.73E-04	3.12
SpTIGR4-2044 (6I4)	1.57	Down	2.66E-07	12.56
SpTIGR4-1587 (5H6)	1.57	Up	4.71E-04	2.83
SpTIGR4-0120 (1P3)	1.57	Up	1.17E-05	7.89
SpTIGR4-0784 (3F2)	1.56	Down	1.16E-03	1.55
SpTIGR4-0389 (2E1)	1.56	Down	5.50E-04	2.56
SpTIGR4-1271 (4L3)	1.56	Down	5.07E-04	2.67
SpTIGR4-1944 (6E3)	1.56	Down	1.09E-04	4.75
SpTIGR4-1886 (5K20)	1.55	Up	3.04E-03	0.26
SpTIGR4-1906 (5O22)	1.55	Down	2.50E-04	3.63
SpTIGR4-1228 (4B10)	1.55	Up	5.39E-05	5.80
SpTIGR4-1070 (3L14)	1.55	Up	1.61E-03	1.11
SpTIGR4-1943 (6D3)	1.54	Down	1.09E-04	4.77
SpTIGR4-0438 (2F7)	1.53	Down	1.35E-05	7.68
SpTIGR4-2126 (6C14)	1.53	Down	2.46E-04	3.69
SpTIGR4-1961 (6F5)	1.52	Up	2.49E-06	10.07
SpTIGR4-0718 (2L18)	1.52	Up	2.00E-03	0.83
SpTIGR4-1607 (5D9)	1.51	Down	2.44E-03	0.55

Continues on following page

Appendix C continued

Spot ID[b]	Fold change	T vs. O	*p*-value[c]	Bayesian
SpTIGR4-0237 (1E18)	1.51	Up	1.35E-05	7.68
SpTIGR4-0743 (2M21)	1.51	Down	1.40E-03	1.31
SpTIGR4-1268 (4I3)	1.51	Down	2.07E-03	0.77
SpTIGR4-0966 (3D13)	1.51	Down	4.73E-04	2.81
SpTIGR4-1229 (4C10)	1.51	Up	5.94E-04	2.44
SpTIGR4-0441 (2A8)	1.51	Up	6.72E-05	5.45
SpTIGR4-1647 (5L2)	1.51	Down	1.19E-03	1.52
SpTIGR4-1465 (4N15)	1.51	Down	3.60E-06	9.51
SpTIGR4-2237 (6I16)	1.50	Up	3.35E-03	0.10
SpTIGR4-0787 (3A3)	1.50	Up	7.63E-05	5.25
SpTIGR4-0119 (1O3)	1.49	Up	2.00E-03	0.84
SpTIGR4-0366 (1N22)	1.49	Up	3.07E-03	0.23
SpTIGR4-0836 (3B9)	1.49	Up	5.50E-04	2.56
SpTIGR4-2239 (6K16)	1.49	Down	2.26E-03	0.64
SpR6-0321 (6L20)	1.49	Up	1.12E-04	4.72
SpTIGR4-1625 (5F11)	1.49	Up	3.25E-03	0.15
SpTIGR4-0534 (2M7)	1.48	Up	1.89E-03	0.92
SpTIGR4-1072 (3N14)	1.48	Up	4.00E-06	9.30
SpTIGR4-1282 (4O4)	1.48	Up	1.51E-05	7.50
SpR6-0320 (6K20)	1.48	Up	1.76E-03	1.01
SpTIGR4-2187 (6G21)	1.47	Up	8.50E-04	1.95
SpTIGR4-0541 (2L8)	1.47	Up	1.06E-03	1.70
SpR6-0345 (6O20)	1.47	Down	2.06E-03	0.78
SpTIGR4-0572 (2J12)	1.47	Up	1.11E-03	1.61
SpTIGR4-0106 (1J2)	1.46	Down	7.42E-04	2.15
SpTIGR4-1778 (5G18)	1.46	Up	1.40E-03	1.32
SpTIGR4-2209 (6E24)	1.45	Down	3.81E-05	6.20
SpTIGR4-1074 (3P14)	1.45	Up	8.47E-04	1.96
SpTIGR4-0173 (1M10)	1.45	Down	9.59E-04	1.82
SpTIGR4-0101 (1M1)	1.45	Down	5.69E-05	5.73
SpTIGR4-1926 (6C1)	1.43	Down	2.15E-03	0.71
SpTIGR4-1706 (5O9)	1.43	Down	7.79E-04	2.08
SpTIGR4-0019 (1C3)	1.43	Down	1.51E-03	1.19
SpTIGR4-0532 (2K7)	1.43	Up	2.35E-04	3.78
SpTIGR4-1861 (5J17)	1.43	Up	1.29E-04	4.56
SpTIGR4-0041 (1A6)	1.43	Up	6.13E-04	2.39
SpTIGR4-1612 (5A10)	1.42	Down	9.57E-05	4.94
SpTIGR4-2208 (6D24)	1.42	Down	2.18E-03	0.69
SpTIGR4-1261 (4J2)	1.42	Down	3.09E-04	3.36
SpTIGR4-1267 (4P2)	1.42	Down	2.30E-03	0.62
SpTIGR4-1358 (4C14)	1.42	Up	3.30E-05	6.40
SpTIGR4-1284 (4I5)	1.42	Down	2.24E-05	6.88
SpTIGR4-1846 (5K15)	1.41	Up	2.31E-05	6.83
SpTIGR4-1266 (4O2)	1.41	Down	3.05E-03	0.25
SpTIGR4-1697 (5N8)	1.41	Up	6.48E-05	5.52
SpTIGR4-0077 (1E10)	1.41	Down	5.71E-04	2.49
SpTIGR4-1972 (6A7)	1.40	Down	2.35E-04	3.81
SpTIGR4-1698 (5O8)	1.40	Up	1.36E-04	4.45
SpTIGR4-1968 (6E6)	1.40	Down	1.55E-05	7.40
SpTIGR4-0605 (2C16)	1.40	Down	1.95E-03	0.88
SpTIGR4-0270 (1F22)	1.40	Up	1.11E-03	1.62
SpR6-0322 (6M20)	1.40	Up	2.46E-04	3.67
SpTIGR4-0864 (3F12)	1.40	Down	6.29E-04	2.35
SpTIGR4-2103 (6L11)	1.40	Down	1.21E-03	1.48
SpTIGR4-1446 (4K13)	1.40	Down	2.36E-04	3.76
SpTIGR4-0617 (2G17)	1.39	Down	1.46E-03	1.25
SpTIGR4-1292 (4I6)	1.39	Down	1.06E-03	1.69
SpTIGR4-0893 (3K4)	1.39	Down	1.55E-05	7.41

Continued on following page

Appendix C continued

Spot ID[b]	Fold change	T vs. O	*p*-value[c]	Bayesian
SpTIGR4-2112 (6M12)	1.39	Down	2.36E-05	6.79
SpTIGR4-0508 (2K4)	1.39	Up	2.92E-03	0.30
SpTIGR4-1624 (5E11)	1.39	Down	7.81E-05	5.21
SpTIGR4-1777 (5F18)	1.38	Up	2.35E-04	3.77
SpR6-0111 (6O17)	1.38	Down	1.52E-03	1.18
SpTIGR4-0361 (1I22)	1.38	Up	1.89E-03	0.92
SpTIGR4-1035 (3A22)	1.38	Up	3.81E-05	6.19
SpTIGR4-0210 (1B15)	1.38	Up	1.30E-03	1.40
SpTIGR4-0211 (1C15)	1.38	Up	2.03E-03	0.81
SpTIGR4-1699 (5P8)	1.38	Up	5.50E-04	2.55
SpTIGR4-1551 (5D2)	1.37	Up	3.23E-04	3.29
SpTIGR4-2092 (6I10)	1.37	Down	3.42E-05	6.34
SpTIGR4-1941 (6B3)	1.36	Down	6.24E-04	2.36
SpTIGR4-1110 (3L19)	1.36	Up	3.16E-03	0.18
SpTIGR4-0947 (3I11)	1.36	Up	4.66E-04	2.86
SpTIGR4-1466 (4O15)	1.36	Down	1.02E-04	4.87
SpTIGR4-1357 (4B14)	1.36	Up	4.94E-04	2.73
SpTIGR4-0362 (1J22)	1.36	Up	1.14E-03	1.57
SpTIGR4-1457 (4N14)	1.35	Down	1.74E-04	4.15
SpTIGR4-0841 (3G9)	1.35	Down	1.19E-03	1.51
SpTIGR4-1360 (4E14)	1.35	Up	5.30E-04	2.62
SpTIGR4-1845 (5J15)	1.35	Up	5.03E-04	2.70
SpTIGR4-2016 (6E12)	1.35	Up	2.35E-04	3.77
SpTIGR4-1692 (5I8)	1.35	Up	2.70E-03	0.42
SpTIGR4-1942 (6C3)	1.34	Down	4.71E-04	2.83
SpTIGR4-0953 (3O11)	1.34	Up	6.72E-05	5.47
SpTIGR4-1460 (4I15)	1.34	Up	2.61E-03	0.46
SpTIGR4-1068 (3J14)	1.33	Down	7.48E-05	5.29
SpTIGR4-1200 (4F6)	1.33	Up	2.51E-04	3.61
SpTIGR4-1947 (6H3)	1.33	Down	1.89E-03	0.92
SpTIGR4-1591 (5D7)	1.33	Down	2.79E-03	0.37
SpTIGR4-0644 (2B21)	1.33	Up	4.94E-04	2.73
SpTIGR4-1862 (5K17)	1.33	Up	2.02E-03	0.82
SpTIGR4-0709 (2K17)	1.32	Up	1.96E-03	0.87
SpTIGR4-2170 (6F19)	1.32	Up	1.31E-04	4.52
SpTIGR4-1293 (4J6)	1.32	Up	3.07E-03	0.23
SpTIGR4-0618 (2H17)	1.32	Down	2.05E-03	0.79
SpTIGR4-1925 (6B1)	1.32	Down	7.39E-05	5.33
SpTIGR4-1171 (4A3)	1.32	Down	1.84E-03	0.97
SpTIGR4-1008 (3F18)	1.32	Down	2.47E-03	0.53
SpTIGR4-1269 (4J3)	1.31	Down	2.03E-03	0.80
SpTIGR4-0776 (3F1)	1.31	Up	6.89E-04	2.23
SpTIGR4-0084 (1D11)	1.31	Up	2.60E-03	0.47
SpTIGR4-1865 (5N17)	1.31	Down	7.60E-04	2.12
SpTIGR4-0212 (1D15)	1.30	Up	2.53E-03	0.50
SpTIGR4-1998 (6C10)	1.30	Down	8.47E-04	1.97
SpTIGR4-1359 (4D14)	1.30	Up	3.07E-03	0.22
SpTIGR4-1734 (5C13)	1.30	Up	2.35E-04	3.81
SpTIGR4-0852 (3B11)	1.30	Down	4.66E-04	2.86
SpTIGR4-0435 (2C7)	1.30	Down	8.12E-04	2.03

[a]Highlighted rows indicate IDs with statistical cut-offs described in Section 4.2.3
[b]Spot ID labelled on microarray slide
[c]Adjusted *p*-value across four biological replicates

Appendix D. Raw data of the comparison of WCH43O vs. WCH43T gene regulation using DNA microarray analysis (Section 4.2.4)[a].

Spot ID[b]	Fold change	T vs. O	*p*-value[c]	Bayesian
SpTIGR4-0715 (2I18)	15.98	Up	3.86E-08	15.48
SpTIGR4-0712 (2N17)	5.88	Up	3.27E-07	13.19
SpTIGR4-1121 (3O20)	3.30	Down	1.83E-04	6.81
SpTIGR4-2182 (6B21)	2.91	Down	2.58E-06	10.95
SpTIGR4-1685 (5J7)	2.65	Down	2.58E-06	10.84
SpTIGR4-2167 (6C19)	2.36	Down	3.74E-02	0.32
SpTIGR4-1197 (4C6)	2.32	Down	1.28E-02	2.26
SpTIGR4-0096 (1H12)	2.32	Down	1.28E-02	2.19
SpTIGR4-2206 (6B24)	1.97	Down	8.76E-04	5.12
SpTIGR4-2148 (6A17)	1.92	Down	1.45E-02	1.88
SpTIGR4-0060 (1D8)	1.89	Down	2.60E-02	0.74
SpTIGR4-2108 (6I12)	1.75	Down	8.76E-04	5.05
SpTIGR4-2055 (6L5)	1.75	Down	3.90E-02	0.17
SpTIGR4-1996 (6A10)	1.74	Down	1.45E-02	1.74
SpTIGR4-0800 (3F4)	1.72	Up	1.93E-02	1.12
SpTIGR4-0911 (3M6)	1.71	Down	1.50E-02	1.61
SpTIGR4-2054 (6K5)	1.69	Up	1.45E-02	1.75
SpTIGR4-0099 (1K1)	1.68	Down	1.45E-02	1.70
SpTIGR4-1714 (5O10)	1.63	Up	1.33E-02	2.07
SpTIGR4-1661 (5J4)	1.62	Up	2.48E-02	0.83
SpTIGR4-0448 (2H8)	1.59	Up	1.62E-02	1.47
SpTIGR4-1707 (5P9)	1.58	Down	3.87E-02	0.21
SpTIGR4-0805 (3C5)	1.57	Up	1.93E-02	1.19
SpTIGR4-1580 (5A6)	1.56	Down	1.93E-02	1.12
SpTIGR4-0061 (1E8)	1.54	Down	3.25E-02	0.49
SpTIGR4-0023 (1G3)	1.51	Down	7.42E-03	2.92
SpTIGR4-0115 (1K3)	1.51	Down	1.93E-02	1.11
SpTIGR4-0956 (3J12)	1.51	Down	3.92E-02	0.13
SpTIGR4-0736 (2N20)	1.41	Down	4.16E-02	0.05
SpTIGR4-1008 (3F18)	1.34	Down	3.75E-02	0.28

[a]Highlighted rows indicate IDs with statistical cut-offs described in Section 4.2.4
[b]Spot ID labelled on microarray slide
[c]Adjusted *p*-value across four biological replicates

Appendix E. Combined gene expression changes of D39, WCH16 and WCH43.

Table showing the combined fold-changes, adjusted *p*-values and Bayesian values of all three strains used to generate the top 50 differentially regulated genes between O and T, as determined by microarray data. This was used to generate the heat-map (Figure 4.3).

TIGR4 annotation[a]	Fold change[b]	Adjusted *p*-value	Bayesian
2050	4.23	0.00379	3.44
2053	3.78	0.00424	2.67
2052	3.47	0.00648	1.61
2051	3.34	0.00648	1.71
2046	3.23	0.00396	3.29
2049	2.84	0.00648	1.68
1907	2.77	0.00424	3.00
0954	2.43	0.00694	1.28
1915	2.09	0.00648	1.60
2207	2.07	0.00332	4.20
1941	2.07	0.00707	1.20
2201	2.06	0.00332	3.90
2177	2.04	0.00648	1.84
1266	2.04	0.00424	2.97
2208	1.98	0.00664	1.50
0099	1.90	0.00262	4.77
1914	1.85	0.00694	1.36
0515	1.77	0.00335	3.68
1906	1.74	0.00262	4.99
2176	1.71	0.00262	4.69
1947	1.64	0.00648	1.62
2105	1.58	0.00648	1.67
1624	1.57	0.00332	3.83
1017	1.52	0.00707	1.21
0911	1.51	0.00648	1.82
1295	1.50	0.00424	2.64
0669	1.48	0.00648	1.60
0381	1.45	0.00707	1.19
0841	1.40	0.00648	1.91
1457	1.40	0.00424	2.66
0839	1.39	0.00612	2.06
1174	-1.87	0.00648	1.71
0870	-1.81	0.00648	1.55
0871	-1.71	0.00707	1.16
0717	-1.68	0.00527	2.24
1002	-1.66	0.00648	1.56
1661	-1.65	0.00648	1.60
1960	-1.58	0.00694	1.39
1071	-1.55	0.00694	1.30
0716	-1.52	0.00694	1.31
1961	-1.52	0.00527	2.32
1890	-1.51	0.00707	1.18

Continues on following page

Appendix E continued

TIGR4 annotation[a]	Fold change[b]	Adjusted *p*-value[c]	Bayesian
1575	-1.47	0.00694	1.41
1664	-1.46	0.00424	2.83
1891	-1.41	0.00463	2.51
1745	-1.41	0.00694	1.43
1860	-1.40	0.00424	2.77
1576	-1.40	0.00424	2.79
0969	-1.40	0.00694	1.30
1699	-1.30	0.00527	2.26

[a]Gene name and annotation according to KEGG (http://www.kegg.com/) website

[b]A positive fold change indicates up-regulation in the O form and a negative fold-change indicates up-regulation in the T form

[c]Adjusted *p*-value of 4 biological replicates

Publication and Conference Presentation

Manso, A. S., **Chai, M. H**., Atack, J.M., Furi, L., De Ste Croix, M., Haigh, R., Trappetti, C., Ogunniyi, A. D., Shewell, L. K., Boitano, M., Clark, T.A., Korlach, J., Blades, M., Mirkes, E., Gorban, A. N., Paton, J. C., Jennings, M. P. & Oggioni, M. R. (2014). A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. *Nat Commun* **5**, 5055.

Chai., M., Ogunniyi, A.D., Van der Hoek, M. B., and Paton, J. C. (23rd – 26th June 2011). Regulation of gene expressions in *Streptococcus pneumoniae* phenotypic variants. Poster presentation, 10th European Meeting on the Molecular Biology of the Pneumococcus, Amsterdam, The Netherlands.