University of Southampton

Faculty of Natural and Environmental Sciences

Centre for Biological Sciences

Genetic and phenotypic diversification within biofilms formed by clinically relevant strains of *Streptococcus pneumoniae*

By

Nicholas William Vere Churton

Thesis submitted for the degree of Doctor of Philosophy

September 2014

Supervisors: Dr. J.S. Webb, Dr. S.C. Clarke

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

Centre for Biological Sciences

Doctor of Philosophy

GENETIC AND PHENOTYPIC DIVERSIFICATION WITHIN BIOFILMS FORMED BY CLINICALLY RELEVANT STRAINS OF STREPTOCOCCUS PNEUMONIAE

By Nicholas William Vere Churton

Streptococcus pneumoniae is a commensal human pathogen and the causative agent of invasive pneumococcal disease. Carriage of the pneumococcus in the nasopharynx of humans is thought to be mediated by biofilm formation. Isogenic populations of *S. pneumoniae* grown under biofilm conditions frequently give rise to morphological colony variants, including small colony variant (SCV) phenotypes. This work employs phenotypic characterisation and whole genome sequencing coupled with ultra-pure liquid chromatography mass spectrophotometry (UPLC/MS_E) of biofilm-derived *S. pneumoniae* serotype 22F pneumococcal colony morphology variants to investigate the diversification during biofilm formation. Phenotypic profiling revealed that SCVs exhibit reduced growth rates, reduced capsule expression, altered metabolic profiles and increased biofilm formation compared to the parent strain. Whole genome sequencing of 12 SCVs from independent biofilm experiments revealed that all SCVs studied had mutations within the DNA-directed RNA polymerase delta subunit (RpoE). Mutations included four large-scale deletions ranging from 51-264 basepairs (bp), one insertion resulting in a coding frameshift and seven nonsense single nucleotide substitutions that result in a truncated gene product. UPLC/MS_E of the SCVs revealed up-regulation of a common sub-set of stress-inducible proteins which are part of an interaction network consisting of the 60 kDa chaperonin, chaperone protein DnaK, cell division protein FtsZ and manganese superoxide dismutase. This work links mutations in the *rpo*E gene to SCV formation and enhanced biofilm development in *S. pneumoniae*, with important implications for colonisation, carriage and persistence of the organism. Furthermore, consistent mutation within the pneumococcal *rpo*E gene presents an unprecedented level of parallel evolution in pneumococcal biofilm development. This work has given insight into the genetic diversity which may arise during pneumococcal colonisation which in turn may help inform future drug and vaccine design.

Table of Contents

List of figures

List of tables

DECLARATION OF AUTHORSHIP

I, Nicholas William Vere Churton, declare that this thesis entitled "Genetic and phenotypic diversification within biofilms formed by clinically relevant strains of *Streptococcus pneumoniae*" and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University;
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- Where I have consulted the published work of others, this is always clearly attributed;
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- None of this work has been published before submission

Signed: ……

Date: ……

Acknowledgements

Firstly I would like to thank my supervisors Dr. Jeremy Webb and Dr. Stuart Clarke for allowing me to undertake this project and for their guidance and support.

Special thanks go to Dr. Robert Howlin and Dr. Ray Allan for their help in protocol design and implementation. In addition thanks are due to Caroline Duignan, Connor Frapwell, Dr. Rebecca Gladstone, Dr. Johanna Jefferies, Dr. Richard Edwards, Dr. Luanne Hall-Stoodley and Professor Saul Faust and Professor Bill Keevil.

I would like to thank Professor Saheer Gharbia, Dr. Raju Misra, Dr. Nicola Thorne, Dr. Lisa Fox and Dr. Chloe Bishop for their whole genome sequencing services at the Department of Bioanalysis and Horizon Technologies, Public Health England, Colindale, UK. Thanks are also due to Alice Still and Jenna Alnajar for additional whole genome sequencing services at the Academic Unit of Clinical and Experimental Sciences, University of Southampton, UK.

Thanks are due to Dr. David Johnston and the Southampton Biomedical Imaging Unit for support imaging the biofilms. I would also like to thank Dr. Paul Skipp and the Centre for Proteomic Research for performing the label-free mass spectrophotometry. Thanks are also due the staff and laboratories of the Southampton National Institute of Health Research (NIHR) Wellcome Trust Clinical Research Facility and NIHR Respiratory Biomedical Research Unit.

I would like to thank my funders. This work is funded by a BBSRC/Pfizer CASE studentship with additional funding from the EPSRC. In particular I would like to thank Dr. Mei Lee and the entire Pfizer Vaccines department at Pfizer Ltd. Walton Oaks, Surrey. The Vaccines team made me feel extremely welcome during my 3-month industrial placement (October 2013 to December 2013). This placement gave me great insight into the promotional and non-promotional marketing strategy of pharmaceutical products.

Last but not least, I would like to thank my family, in particular my wife Katrina, for supporting me over the last four years.

Publications and conference contributions

Publications:

Elita Jauneikaite, Johanna Mary Carnon Jefferies, **Nicholas William Vere Churton**, Raymond Tzer Pin Lin, Martin Lloyd Hibberd and Stuart Charles Clarke (2014) *Genetic diversity of Streptococcus pneumoniae causing meningitis and sepsis in Singapore during the first year of PCV7 implementation* Emerging Microbes & Infections 3, e39; doi:10.1038/emi.2014.37

Conferences:

Society for General Microbiology Spring conference, Harrogate International Centre 11th-14th April 2011, "Informing vaccine policy through the molecular characterisation of pneumococci in Singapore". Poster presentation.

Medicine & Biological Sciences Postgraduate Conference, Southampton General Hospital 8th June 2011, "Genetic diversification within biofilms formed by clinically relevant strains of *Streptococcus pneumoniae*". Poster presentation.

Institute for Life Sciences: Biofilm and Microbial communities conference, Chilworth Manor, Southampton 12th September 2011, "Genetic diversification within biofilms formed by clinically relevant strains of *Streptococcus pneumoniae*". Poster presentation.

Centre for Biological Sciences Postgraduate Conference Symposium, Life Sciences Building 85 14th June 2012, "Genetic diversification within biofilms formed by clinically relevant strains of *Streptococcus pneumoniae*". Poster presentation.

The 15th Conference in Genomics and Proteomics of Human Pathogens, HPA Colindale 21st-22nd June 2012, "Genetic and phenotypic diversification within biofilms formed by clinically relevant strains of *Streptococcus pneumoniae*". Poster presentation.

Centre for Biological Sciences Postgraduate Conference Symposium, Life Sciences Building 85 19th June 2013, "Genetic diversification within biofilms formed by clinically relevant strains of *Streptococcus pneumoniae*". Oral presentation.

Best Practice conference, Birmingham NEC 16th-17th October 2013. Pfizer attendee.

Federation of infection societies: Action on Infection, Birmingham ICC 11th-13th November 2013. Pfizer attendee.

Awards:

Received the Dr. Christine McCartney *Best Poster Prize* at the 15th Conference in Genomics and Proteomics of Human Pathogens on 22nd June 2012.

Definitions and abbreviations

ABC	ATP-binding cassette
AHL	Acyl-homoserine lactone
ATP	Adenosine triphosphate
BHI	Brain heart infusion
BLAST	Basic Local Alignment Search Tool
CAT	Casein tryptone media
CBA	Columbia blood agar
CDC	Centers for Disease Control
CDM	Chemically defined medium
CF	Cystic fibrosis
CFU	Colony forming unit
CI	Confidence interval
CLSM	Confocal laser scanning microscopy
CSP	Competence stimulating peptide
CV	Crystal violet
DdNTP	Dideoxy nucleoside triphosphates
DNA	Deoxyribonucleic acid
DTT	DL-Dithiothreitol
EDIC/EF	Episcopic differential interference contrast microscopy/with epi-fluorescence
EPS	Extracellular polymeric substance
GO	Oxidative repair system
GP	General Practitioner
GS	Genome Sequencer
HBSS	Hank's balanced salt solution
HCl	Hydrochloric acid
HQNO	4-hydroxy-2-heptylquinoline-N-oxide
IGV	Integrative Genomics Viewer
IPD	Invasive pneumococcal disease
KO	Knockout
LCV	Large colony variant
LDS	lithium dodecyl sulphate
LMV	Large mucoid variant
MCV	Medium colony variant
MLST	Multi locus sequence typing
M.I.C	Minimum inhibitory concentration
MMV	Medium mucoid variants

MOPS	3-(N-morpholino)propanesulfonic acid
MRSA	Methicillin-resistant *Staphylococcus aureus*
NGS	Next generation sequencing
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PCV	Pneumococcal conjugate vaccine
PCV7	Pneumococcal conjugate vaccine 7 (Prevenar™)
PCV13	Pneumococcal conjugate vaccine 13 (Prevenar13™)
PGM	Personal Genome Machine
PHiD-CV	Pneumococcal *Haemophilus influenzae* protein D conjugate vaccine (Synflorix™)
PIA	Polysaccharide intracellular adhesin
PLM	PCR ligation mutagenesis
PMN	Polymorphonuclear leukocyte
PPV	Pneumococcal polysaccharide vaccine (Pneumovax™)
PTS	Phosphotransferase system
RNA	Ribonucleic acid
ROS	Reactive oxygen species
SCV	Small colony variant
SEM	Scanning electron microscopy
SMV	Small mucoid variant
SNP	Single nucleotide polymorphism
ST	Sequence type
TCV	Typical colony variant
TD-SCV	Thymidine-dependent small colony variant
TEAB	Triethylammonium bicarbonate
THY	Todd-Hewitt broth with yeast extract
TSB	Tryptone soya broth
UHS	University Hospital Southampton
UPLC/MS_E	Ultra-pure liquid chromatography mass spectrophotometry
URT	Upper respiratory tract
v/v	Volume/volume
WGS	Whole genome sequencing
WT	Wild type
w/v	Weight /volume

Chapter 1

General introduction

1 Chapter 1: Introduction

Streptococcus pneumoniae is a commensal human pathogen which colonises the nasopharynx. Attachment to the nasopharynx is thought to be mediated by the formation of biofilms (Marks *et al*., 2012b). Biofilms are aggregation of bacteria, adhered to a surface encased in a matrix of extracellular polymeric substances (EPS) (Allegrucci *et al*., 2006). Pneumococcal biofilm-mediated growth is thought to act as a survival mechanism, facilitating protection from antibiotics and the immune system whilst simultaneously providing an environment for genetic diversity to arise within the bacterial population (Allegrucci and Sauer, 2007). The aim of this thesis is to investigate the level of genetic diversity across the whole genome of the organism, to assess whether genotype-phenotype relationships exist, and to evaluate whether adaptive biofilm mutations may facilitate long-term persistence and vaccine escape. In turn, this work will help to better understand the significance of biofilms in the colonisation and spread of *S. pneumoniae*.

Genetic variation, in bacteria, can arise through gene deletions, gene acquisition and gene duplication via mechanisms such as horizontal gene transfer (Hausner and Wuertz, 1999), transformation (Li *et al*., 2001), phase variation (Drenkard and Ausubel, 2002) and adaptive mutation (Foster, 2007). This variation leads to an increased phenotypic and genetic diversity within the population. Dispersal of cells from a biofilm can lead to the dissemination of variants into the surrounding environment (Boles *et al*., 2004) and therefore the biofilm can be thought of a reservoir for facilitating bacterial diversity. The variation which bacteria undergo during the biofilm mode of growth is thought to act as a survival strategy, with the premise being that those species with the highest genetic diversity have the greatest ability to withstand perturbations within the environment (Boles *et al*., 2004). The mechanisms by which bacteria attain such diversity still remain unclear, but are of fundamental interest in understanding how bacteria in biofilms persist despite continual antimicrobial therapy and vaccine use.

Evidence of biofilm-mediated diversity can be seen in the emergence of colony morphology variants from isogenic populations of bacteria grown under biofilm conditions. This phenomenon of colony morphology variation has been reported in a range of human pathogens including *Pseudomonas aeruginosa* (Boles *et al*., 2004), *S. pneumoniae* (Allegrucci and Sauer, 2007) and *Staphylococcus aureus* (Mitchell *et al*., 2010b). The fact that colony morphology variants have been observed in multiple bacterial species suggests that this

morphological change is not unique to one pathogen and that common underlying mechanisms may be responsible for this variation. Variation in colony morphology has been shown to correlate to genetic mutations, including genes involved in capsule production (Allegrucci and Sauer, 2007) and the SOS response (Boles *et al*., 2004). Variation in morphology has also been associated with release of exogenous signals from surrounding bacteria (Mitchell *et al*., 2010b). This variation is of particular relevance to the design of vaccines as the capsule of many pathogens acts as a major target for vaccine design. It is therefore of interest to discover the genetic mutations involved in these phenotypes and the mechanisms by which they may occur.

There have been many notable reviews regarding colony variation (Haussler, 2004, Proctor *et al*., 2006, Melter and Radojevic, 2010). This chapter will introduce the pathogen *S. pneumoniae* and look at the fundamental characteristics of biofilm formation, highlighting the role of biofilms in generating colony variation in *S. pneumoniae*, *S. aureus* and *P. aeruginosa*, and discuss the role of biofilms in colonisation and long-term persistence. This chapter will also discuss some of the mechanisms which can contribute to biofilm-derived diversity and discuss how the use of next generation sequencing (NGS) can provide a rapid and reliable method of identifying genetic changes within the organism.

1.1 *Streptococcus pneumoniae*

S. pneumoniae is a Gram-positive, encapsulated bacterium found growing asymptomatically on the upper respiratory tract (URT) of healthy individuals. Often referred to as 'the pneumococcus', *S. pneumoniae* has been implicated in numerous human acute infections including pneumonia, sinusitis and otitis media, in addition to causing invasive diseases such as meningitis, septicaemia and endocarditis (Hall-Stoodley *et al*., 2006, Mitchell and Mitchell, 2010, Nistico *et al*., 2011) and remains a major cause of morbidity and mortality worldwide (O'Brien *et al*., 2009).

S. pneumoniae is the model organism used in this thesis. The reasons for choosing this pathogen as a model organism are three-fold. Firstly, colonisation of asymptomatic pneumococci in the URT is indicative of biofilm formation (Marks *et al*., 2012b). Secondly, pneumococcal biofilm research is a promising field of research with previous examples of biofilm-derived diversity (Allegrucci and Sauer, 2007, Allegrucci and Sauer, 2008, Waite *et al*., 2001, Domenech *et al*., 2009, McEllistrem *et al*., 2007). Thirdly, pneumococcal vaccine design has been based on providing immunity to serotypes implicated in invasive disease; such as meningitis and septicaemia. This work aims to understand the role that genetic diversification can play in pneumococcal persistence and in doing so will help contribute to pneumococcal vaccine development; preventing pneumococcal carriage and acute infections such as otitis media.

1.1.1 Pneumococcal burden of disease

In a 2000 study, pneumococcal burden of disease was estimated with approximately 14.5 million episodes of serious pneumococcal disease occurring worldwide causing approximately 826,000 deaths in children aged 1–59 months (O'Brien *et al*., 2009). Highest levels of pneumococcal disease (60 %) were recorded in developing countries (predominantly Africa and Asia) where a routine vaccination program was not in place (O'Brien *et al*., 2009). A total of 46 serogroups and 95 serotypes of *S. pneumoniae* have been described to date (Oliver *et al*., 2013). Globally around 20 serotypes are implicated in >80 % of pneumococcal disease worldwide and of those, 13 cause 70-75 % of invasive disease in children under 5 years of age (WHO, 2007). One of the most prominent risk factors associated with pneumococcal disease is

age, with children <2 years and adults >65 years of age being more susceptible (Sleeman *et al*., 2001). Additional risk factors include individuals with HIV, chronic respiratory tract infections, heart disease, and asplenia (Schutze *et al*., 2002, Boikos and Quach, 2013, von Mollendorf *et al*., 2014), as well as lifestyle choices such as smoking and alcohol abuse and day-care attendance (Ahl *et al*., 2014, Cruickshank *et al*., 2014). Surveillance from Public Health England estimate that in England and Wales over 5000 cases of pneumococcal disease are reported annually with highest reports associated with winter months. Furthermore pneumococcal disease is estimated to account for ~40,000 hospitalisations due to pneumococcal pneumonia and ~630,000 GP consultations due to pneumococcal otitis media (Melegaro *et al*., 2006). The pathogenic route of infection of *S. pneumoniae* begins with the inhalation of airborne droplets (Bogaert *et al*., 2004). From here the pathogen can spread locally to the ears and sinuses causing otitis media and sinusitis, respectively. Alternatively spread to the lungs can result in pneumonia, empyema and if translocated to the blood, septicaemia (Bogaert *et al*., 2004). Transmission and survival of the pathogen is mediated by its repertoire of virulence factors (Kadioglu *et al*., 2008).

1.1.2 Pneumococcal virulence factors

Pneumococcal virulence factors can be broadly distinguished into protein virulence factors and polysaccharide virulence factors. One of the major virulence factors of *S. pneumoniae* is the capsular polysaccharide (Kadioglu *et al*., 2008). The capsule forms the outermost layer of the bacterium, spanning ~200-400 nm in thickness and is covalently attached to the cell wall (Sorensen *et al*., 1988). The polysaccharide capsule protects the bacterium from phagocytosis by preventing complement C3b opsonisation (Hyams *et al*., 2010). Antibodies form against the capsule in a type-specific manner; as a result the polysaccharide sugars are the primary target used in pneumococcal vaccines to develop immunity against invasive pneumococcal disease (IPD) (Harro *et al*., 2010). The antibody response to the capsule also forms the basis for the bacterium's nomenclature, in the form of serotyping. Capsular serotypes are controlled at the capsular polysaccharide synthesis (*cps*) locus, an extensive cluster of genes unique to each serotype. Only *cpsA* is conserved between serotypes (Bentley *et al*., 2006). In addition to the polysaccharide capsule the pneumococcal surface consists of common cell wall polysaccharides teichoic acid (TA) and lipoteichoic acid (LTA) or F-antigen (Sorensen *et al*., 1988). Phase variation of surface structures can result in

visible changes in colony morphology in the form of opaque and transparent capsule variants and furthermore have been correlated with altered rates of polysaccharide production and teichoic acid (Kim and Weiser, 1998), autolysis and adherence (Allegrucci *et al*., 2006). A graphical representation of the pneumococcal outer surface can be seen in Figure 1-1.

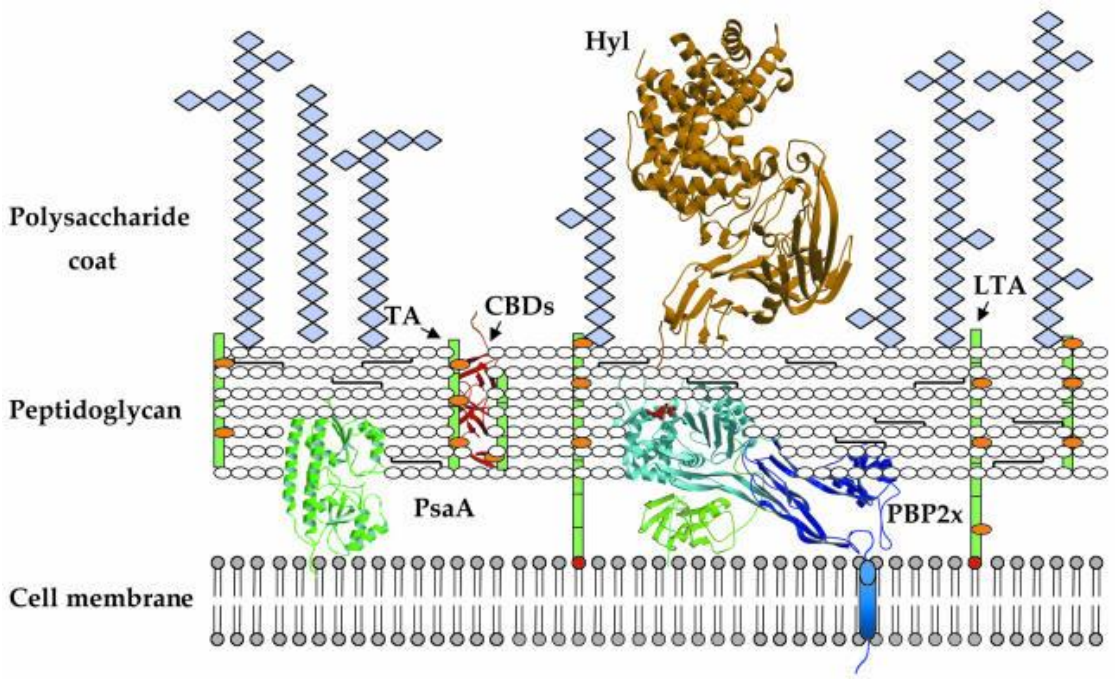


Figure 1-1: Illustration of the pneumococcal outer surface

The pneumococcal cell surface can be broadly divided into three distinct layers, the cell membrane, the peptidoglycan layer and the polysaccharide outer capsule. The peptidoglycan layer is connected to the cell membrane by means of lipoteichoic acids (LTAs) (green rectangles) via a C-terminal fatty acyl group (red circle). Teichoic acids (TAs) (green rectangles) are linked to the peptidoglycan layer via a phosphodiester linkage. Both TAs and LTA are rich in choline (orange circles). Choline-binding proteins are linked to TAs and LTAs via choline-binding domains (CBDs) (red ribbon diagram). Penicillin-binding proteins (PBPs) are situated in the periplasmic space. Attachment of the pneumococcal surface antigen (PsaA) below the peptidoglycan layer is mediated via a LXXC motif, whereas attachment of the Hyaluronate lyase (Hyl) above the peptidoglycan layer is mediated via a LPXTG motif. Figure reproduced from Di Guilmi and Dessen (2002). Permission was granted for the reproduction of this image from John Wiley & Sons via the Copyright Clearance Center's RightsLink® service (License number 3434160347264).

In addition to the polysaccharide capsule, *S. pneumoniae* has a multitude of virulence factors which aids its infectivity including pneumolysin, a 53 kDa toxin which binds to cholesterol in host cell membranes and induces lysis (Nollmann *et al*., 2004). Release of pneumolysin is mediated, in part, by the LytA lysis of pneumococcal cells; LytA is a 36 kDa protein which degrades peptidoglycan resulting in the liberation of the pneumolysin toxin (Kadioglu *et al*., 2008). Factors which counteract the immune system include pneumococcal surface protein A (PspA) which inhibits the host complement system (Hammerschmidt *et al*., 1999) and choline binding protein A (PspC) which binds to human secretory component (Rosenow *et al*., 1997). PspC also recognises epithelial sialic acid residues which mediates pneumococcal adherence and promotes infectivity into nasopharyngeal cells (Dave *et al*., 2004).

Additional virulence factors include neuraminidases (NanA and NanB) which cleave host cell glycans (Mitchell, 2000) and aid cell attachment to the substratum (Soong *et al*., 2006). Furthermore *S. pneumoniae* neuraminidase (NanA) is thought to be important in colonisation of the URT (Brittan *et al*., 2012) and biofilm formation (Parker *et al*., 2009). *S. pneumoniae* is considered a human-specific pathogen where transfer between individuals is mediated via the inhalation of respiratory droplets. *S. pneumoniae* is a fastidious pathogen and is known to readily desiccate and die outside of the human host. A recent study of bacteriological cultures from items in a day-care centre was able to show high levels of viable *S. pneumoniae* and *Streptococcus pyogenes* (Marks *et al*., 2013b). The ability of these pathogens to survive outside of the human host was due to biofilm growth on the investigated fomites (inanimate objects that can facilitate the transmission of infectious agents from person to person). This study highlights the importance of biofilm formation in transmission, nasopharyngeal colonization and persistence of pneumococci within the population.

1.1.3 Pneumococcal vaccines

Vaccination against *S. pneumoniae* is the principle method of protecting infants and adults against pneumococcal infection. With over 90 different serotypes to protect against, vaccine design has focused predominantly on those serotypes which are most prevalent in invasive pneumococcal disease. The pneumococcal polysaccharide vaccine (PPV Pneumovax™; Merck), which covers 23 serotypes, has been licenced and has been in use in the UK and USA for over 25 years for those at risk of pneumococcal infection. The polysaccharides are T-cell-independent antigens meaning that exposure to B-cells result in the production of antibodies without the presence of T-cells. As such PPV is limited in its use as it does not provide long-term protection in infants under 2 years old (Stein, 1992). At present three, recently licenced, conjugate vaccines exist for *S. pneumoniae* (pneumococcal conjugate vaccine 7 (PCV7, Prevenar™), pneumococcal conjugate vaccine 13 (Prevenar13™) and the pneumococcal *Haemophilus influenzae* protein D conjugate vaccine (Synflorix™). Prevenar™ has been a part of the routine childhood immunization programme since September 2006 in the UK and since 2000 in the US. Serotypes covered in Prevenar™ include 4, 6B, 9V, 14, 18C, 19F and 23F due to their prevalence in invasive disease in the US (Hausdorff *et al*., 2009). Serotypes within PCV7 are covalently linked to a non-toxic diphtheria variant carrier protein (CRM_{197}) which increases the vaccine efficacy by stimulating a T-cell helper response. Despite the reduction in the number of serotypes covered compared to Pneumovax, the ability to produce a T-cell response allows Prevenar™ to protect infants under the age of 2. The 13-valent PCV13 (Prevenar13™), developed by Pfizer, includes the seven serotypes of PCV7, in addition to serotypes 1, 3, 5, 6A, 7F, and 19A, has recently been licensed and in use in the UK childhood vaccine schedule since February 2010. Both Prevenar™ and Prevenar13™ are administered in three injections at 2, 4 and 22 months. The development of a PCV15 conjugate vaccine is in progress; PCV15 includes the previous 13 serotypes form Prevenar13™ in addition to serotype 22F and 33F and is currently undergoing pre-clinical evaluation (Skinner *et al*., 2011).

The pneumococcal *H. influenzae* protein D conjugate vaccine otherwise known as Synflorix™ is a 10-valent pneumococcal conjugate vaccine. The ten pneumococcal serotypes include 1, 4, 5, 6B, 7F, 9V, 14, 18C, 19F, and 23F. Eight of the 10 serotypes (1, 4, 5, 6B, 7F, 9V, 14 and 23F) are conjugated to the non-typeable *H. influenzae* (NTHi) protein D whereas serotypes 18C and 19F are conjugated to tetanus toxoid and to diphtheria toxoid, respectively. Conjugation to the NTHi protein D is thought to provide added protection against otitis media (OM) caused by NTHi but does not substitute routine immunization with the *H. influenzae*

type b vaccines (GlaxoSmithKline – 2009 Product Monograph). Synflorix™ has been developed by GlaxoSmithKline, and has been licenced in over 60 countries since March 2009.

Vaccine development is a dynamic process influenced greatly by on-going epidemiological studies which assess for changes in serotype distribution as a result of vaccine implementation (Clarke *et al*., 2006, Tocheva *et al*., 2010). In the UK, the use of PCV7 has shown a marked decrease in pneumococcal infection in both vaccinated and non-vaccinated populations (Gladstone *et al*., 2011), however with over 90 serotypes existing for *S. pneumoniae*, the majority of pneumococcal strains are not included in current pneumococcal vaccines. One of the major concerns in pneumococcal vaccine development is will vaccinating against the most common serotypes select for non-vaccine serotypes in the nasopharynx, leading to the increased carriage and transmission of non-vaccine serotypes in the wider population (Spratt and Greenwood, 2000)? Changes in serotype among natural isolates of *S. pneumoniae* has been shown to occur by recombination at the capsular biosynthetic locus (Coffey *et al*., 1998), a question of particular interest to this project and to biofilm research in general, is can genetic variation during biofilm growth alter the serotype of a bacterium?

1.2 Role of biofilms in clinical disease

Over the past decade a wealth of research has been performed using planktonic cultures as models for bacterial growth and development. It is now acknowledged however that in nature the predominant mode of bacterial growth is not as single-celled entities but as attached communities of bacteria called biofilms (Stoodley *et al.*, 2002). Biofilms are complex aggregations of bacteria, often consisting of multiple species, which adhere to inert or living surfaces and are encased in an EPS matrix which can protect the bacteria from the surrounding environment (Costerton *et al.*, 1995). It is estimated that 65 % of all human infections involve biofilm formation (Potera, 1999), making biofilms a major contributor to chronic disease. As a result, biofilm research is of paramount importance to better understand the role of bacteria in clinical diseases.

1.2.1 Biofilm development

Biofilms are defined as aggregations of bacteria, adhered to an inert or living surface encased in an EPS matrix (Costerton *et al*., 1995), however the word 'biofilm' is a broad term to describe the attachment of microorganism to a surface and may include such examples as dental plaque, marine biofouling and chronic wound infections. In the context of this thesis, this work will focus on the clinical significance of biofilms in the human host. A generalised model of biofilm development, shared by numerous species, has been established (Costerton *et al*., 1995), and consists of five key stages; initial reversible attachment, irreversible attachment, microcolony formation, biofilm maturation and finally dispersal of the biofilm (Figure 1-2) (Stoodley *et al*., 2002). At each stage the cells exhibit distinct phenotypes. This distinction was shown by examining the protein expression of *P. aeruginosa* biofilms using two-dimensional gel electrophoresis (Sauer *et al*., 2002). An average of approximately 525 proteins differed in regulation between each of the five stages of development. In particular proteins related to motility, alginate production, and quorum sensing were shown to change expression during biofilm development. Strikingly, when planktonic cells were compared with mature biofilm cells, a six-fold or greater change in expression was observed in over 800 proteins, constituting more than 50 % of the entire proteome (Sauer *et al*., 2002). This indicated the marked difference between the biofilm and planktonic lifestyles.

Evidence for this model of development has been primarily obtained *in vitro* using single-species biofilms on an abiotic surface, namely *P. aeruginosa*, and as such should be used with caution when applying to biofilms of different species within the clinical setting. Nonetheless biofilm formation has been observed in clinically relevant respiratory pathogens known to colonise the human naso/oropharynx (Webb *et al*., 2003a, Hall-Stoodley and Stoodley, 2009, Murphy *et al*., 2009). The stages of biofilm formation are described in the following sections.

1.2.1.1 Reversible/irreversible attachment

Biofilm formation begins with the first planktonic bacteria adhering to a solid inert or living surface. Initial attachment between the bacteria and the substratum is influenced by van der Waals forces, electrostatic attractions of the substratum and cell surface hydrophobicity (Marshall *et al*., 1971), however a subsequent change in gene expression in these colonising bacteria leads to the down-regulation of motility genes such as flagella and the up-regulation of attachment genes such as pili leading to irreversible attachment to the substratum (Sauer *et al*., 2002) (Figure 1-2). Cell to cell signalling can play an important role in biofilm formation; this signalling process known as quorum sensing and is mediated in Gram-negative bacteria via the secretion of acyl-homoserine lactones (AHL) (Parsek *et al*., 1999) and by the secretion of pheromone peptides in gram-positive bacteria (Suntharalingam and Cvitkovitch, 2005). Quorum sensing allows bacteria to determine the local density of the surrounding bacterial population and co-ordinate gene expression, specifically in the later stages of biofilm formation (Sauer *et al*., 2002).

1.2.1.2 Microcolony formation and biofilm maturation

Once attached to the surface, a foci of cells form densely-packed aggregates called microcolonies either via clonal expansion or cell motility (Klausen *et al*., 2003a, Klausen *et al*., 2003b). Cells within a microcolony grow under suboptimal conditions and can exhibit functional heterogeneity (Hall-Stoodley and Stoodley, 2009) leading to a variety of metabolic states (Sauer *et al*., 2002). It is this heterogeneity that provides an important survival strategy allowing bacteria to establish multiple niches within the biofilm (Costerton *et al*., 1999).

Adaptation of this kind includes the development of phenotypic variants, allowing bacteria to adapt easily to changes in the environment (Costerton *et al*., 1999). Structurally, biofilms have been shown to contain voids in which liquid and nutrients can circulate (deBeer *et al*., 1996), in addition to different niches, in which cells exhibit different patterns of gene expression (Margolis *et al*., 2010).

Mature biofilms can exhibit 'mushroom' and 'tower' morphologies which are encased in an EPS matrix consisting of proteins, extracellular DNA and carbohydrates (Costerton *et al*., 1999). The EPS is thought to have an important role in stabilising the architecture of the biofilm and protecting the biofilm from antimicrobial treatment, macrophage engulfment and the shear forces which may arise in fluidic environments (Stoodley *et al*., 2002). Although a vital part of the biofilm structure, very little is still know about the EPS. In *P. aeruginosa*, three exopolysaccharides, (Psl, Pel, and alginate) are known to contribute to biofilm formation (Branda *et al*., 2005) with disruption of either one affecting biofilm formation (Ryder *et al*., 2007). Extracellular DNA (eDNA) has also been shown to be an important factor in biofilm formation and maturation, with eDNA thought to play an important role in cell to cell communication and potential diversification (Allesen-Holm *et al*., 2006). Direct visualisation of eDNA in *P. aeruginosa* flow-cell biofilms suggests the eDNA is primarily located on the stalks of the mushroom shaped microcolonies and its release may be driven be quorum sensing-controlled factors (Allesen-Holm *et al*., 2006).

Cells within a biofilm often exhibit a reduced metabolism and more diverse gene expression compared to their planktonic counterparts resulting in a significant increase in recalcitrance to antimicrobial compounds (Stewart and Costerton, 2001). Causes of bacterial persistence within a biofilm have been attributed to phenotypic heterogeneity and slower growth of bacteria within the biofilm (Hoiby *et al*., 2010), delayed antibiotic penetration through the EPS matrix, (Costerton *et al*., 1999) and the presence of persister cells (Lewis, 2010). The biofilm mode of growth can be considered paradoxical in the sense that the beneficial antimicrobial protection is counterbalanced by a myriad of factors which hinder bacterial growth, leading to the hypothesis that biofilm growth acts not as a virulence factor, but as a survival strategy for bacteria (Boles *et al*., 2004). Detrimental factors associated with biofilm growth include steep oxygen and pH gradients and differing nutrient availability which occur due to the production of toxic metabolites and the biofilm's densely-packed nature (Debeer et al., 1994, Sternberg et al., 1999). As a result bacteria growing within a biofilm exhibit reduced growth rates in order to optimise their survival in a nutrient depleted environment. Oxygen and chemical gradients create micro-environments within the biofilms

resulting in differing levels of metabolic activity for the various microbial species (Debeer et al., 1994). In order to adapt to the multiple environmental niches within the biofilm, bacteria exhibit many phenotypes and functional heterogeneity regarding metabolism and replication (Ehrlich *et al*., 2005). For example RpoS is a sigma factor found in many Gram-negative bacilli, including *P. aeruginosa* and *Escherichia coli* (Cochran *et al*., 2000, Sheldon *et al*., 2012). RpoS is up-regulated in response to environmental stresses such as nutrient depletion in planktonic cultures (Henggearonis, 1993) and is thought to be a relevant factor in biofilm development (Stoodley *et al*., 2002).

1.2.1.3 Dispersal

Although considered sessile, biofilms give rise to non-sessile planktonic bacteria that are dispersed once a mature biofilm has been established in order to re-colonise additional surfaces (Costerton *et al*., 1999). Dispersal can be triggered by physical factors such as shear forces (Stoodley *et al*., 2001) or via environmental cues such as nutrient depletion (Sauer *et al*., 2004) and genetically programmed cell death (Webb *et al*., 2003b). The latter was observed in *P. aeruginosa* using confocal microscopy and the *Bac*Light live/dead stain. In this, the authors observed high levels of cell death within the biofilm triggering the release of surviving cells from the centre of the biofilm (Webb *et al*., 2003b). This process of dispersal is thought to contribute to the spread of infection and persistence of the organism.

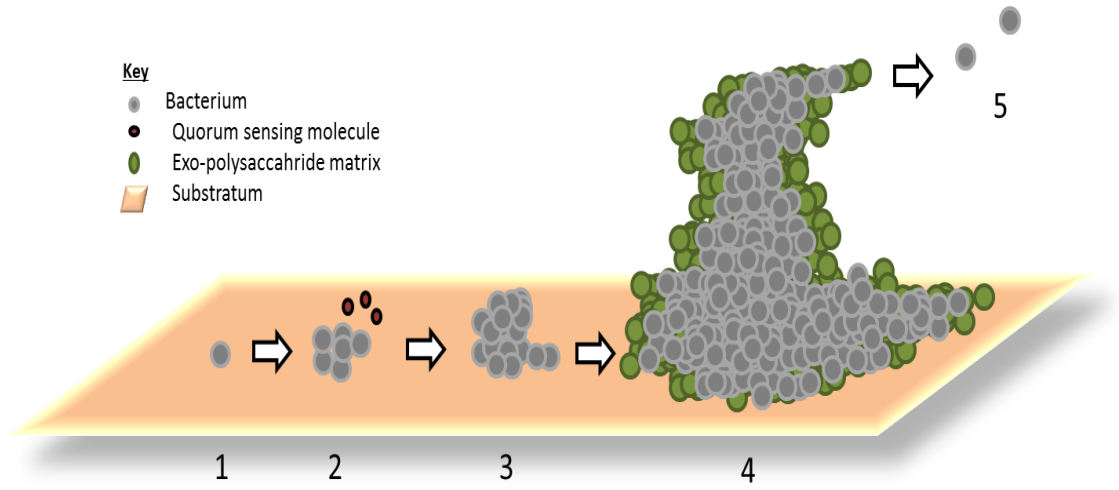


Figure 1-2: Stages of biofilm development

Planktonic bacteria adhere to a substratum (1) and undergo reversible attachment. Change in gene expression leads to a down regulation of motility genes and up regulation of attachment genes leading to irreversible attachment of the substratum (2). Quorum sensing signalling mediates the attachment of further bacteria (3) and matrix production leading to a mature biofilm and microcolony formation (4). Dispersal of cells from the biofilm enables colonization of new surfaces (5). Figure adapted from Stoodley *et al.* (2002). Additional permission was not required from the Annual Review of Microbiology journal for the republishing of this image in accordance with Copyright Clearance Center's RightsLink® service.

1.2.2 Genetic variation within biofilms formed by clinically relevant pathogens

An important hypothesis in biofilm research is the concept that genetic variation within a bacterial biofilm community facilitates bacteria to withstand perturbations in the environment. This concept of mutation for the benefit of survival is termed the *insurance hypothesis* (Boles *et al*., 2004). Originally the *insurance hypothesis* describes how ecological communities that are genetically diverse have a greater ability to withstand environmental changes (Yachi and Loreau, 1999), however the same principal may apply to bacterial communities within a biofilm environment. The Gram-negative opportunistic pathogen, *P. aeruginosa*, has been extensively studied in biofilm research due to its prevalence in the lungs of cystic fibrosis (CF) patients. After initial colonization of the lungs by *S. aureus* and *H. influenzae*, *P. aeruginosa* has the highest prevalence (~80 %) in the lungs of CF patients (Donlan and Costerton, 2002). Once biofilms are established, this pathogen persists despite antimicrobial treatment and is the cause of death for most sufferers of cystic fibrosis. *P. aeruginosa* has also been recorded in many other chronic infections including otitis media, and urinary tract infections in addition to colonisation on catheters and medical implants (Donlan and Costerton, 2002). In 2004, Boles *et al.* showed that after only 2-7 days of biofilm growth in a drip-flow reactor, *P. aeruginosa* underwent extensive variation in colony morphology. The authors described two variants termed 'mini' and 'wrinkly' due to their respective small and irregular morphologies. Phenotypic analysis revealed that the 'wrinkly' phenotype exhibited increased attachment and cell cluster rates and a reduced detachment rate leading to accelerated biofilm production. In contrast, the 'mini' phenotype exhibited hyper-detachment leading to an increased ability to disseminate. The 'wrinkly' phenotype also exhibited an increased tolerance to hydrogen peroxide compared to wild type cells indicating an increased resistance to oxidative stress (Boles *et al*., 2004). The hereditability of the variants suggested that they were a result of genetic changes. The authors showed that strains deficient in the *recA* gene, a gene which is a central part of the SOS response which produces genetic change via recombination and induces error prone DNA-polymerases (Boles *et al*., 2004), showed a marked decrease in the small colony phenotype leading to the conclusion that the SOS response plays a part in generating morphology via a *recA*-dependant mechanism. Critically no apparent genomic analysis was performed on the 'mini' or 'wrinkly' variants generated in the biofilms to show any mutations in the *recA* gene. Although not performed in the latter study, comparative genetic analysis of the variants would have been useful to determine if different mutations in the *recA* gene were responsible for eliciting the two

respective variants or alternatively to determine any recombination events which may have occurred as a result of *recA* expression. Nevertheless the phenotypic analysis did show that the colony variants generated under biofilm growth had different phenotypic variation which concurred with their original *insurance hypothesis* theory that diversity within a population may aid survival. To date a range of colony variants of *P. aeruginosa* have been observed and include mucoid (Deretic *et al*., 1994), small-colony variants (Boles *et al*., 2004, Webb *et al*., 2004), lipopolysaccharide-deficient (Dasgupta *et al*., 1994), hyperpiliated (Deziel *et al*., 2001), and antibiotic resistant colonies (Drenkard and Ausubel, 2002).

A characteristic trait of biofilm formation is the development of 3-dimensional foci of densely-packed aggregates of bacteria called microcolonies. These microcolonies are subject to nutritional deficiencies, steep oxygen and pH gradients (DeBeer *et al*., 1994, Sternberg *et al*., 1999). Due to the differing physiochemical environments of microcolonies within a biofilm, there is intense competition and elevated mutation rates. In order to investigate the role of microcolonies in the generation of genetic diversity Conibear *et al.* (2009) grew biofilms of transformed *P. aeruginosa* cells with a green fluorescent protein gene containing a +1 frame shift mutation; the premise being that the transformed cells would only fluoresce if a mutation occurred which reverted the frame shift mutation to the wild type sequence. By visualising 7-10 day old biofilms *in situ* using confocal microscopy the authors were able to determine what regions of the biofilm were more prone to mutations. They observed that fluorescence-inducing mutations were only present in microcolonies but not in other regions of the biofilm or planktonic cultures, thus providing evidence to suggest that microcolonies are key to generating genetic diversity and evolutionary change. The authors also showed that mutator phenotypes, such as cells deficient in DNA repair (Δ*mutS* and Δ*mutL* repair genes) have an enhanced ability to form microcolonies. This led the authors to propose a model of microevolution, similar to that seen in tumour development, where selected mutated bacterial cells expand clonally in number to develop the microcolony structure (Conibear *et al*., 2009).

One aspect of biofilm development and diversification which is important to co-evolutionary dynamics is the bacterial host's interaction with bacteriophage (Betts *et al*., 2014). In 2004, Webb *et al*. showed that exposure of *P. aeruginosa* to filamentous bacteriophage Pf4 produced a small colony variant as a result of phase variation (Webb *et al*., 2004). This paper highlights the fact that the diversity seen from biofilm cultures may be a result of mutations in the phage genome rather than mutations in the host genome and would be an interesting avenue to investigate further using whole genome sequencing (WGS). Bacteriophage may not be solely responsible for the biofilm-derived variation. In 2005, Kirisits

et al. reported the isolation of a 'sticky' colony variant which displayed hyperadherence, increased hydrophobicity and reduced swimming and twitching phenotypic characteristics (Kirisits *et al*., 2005). The transcriptional profiling in this study revealed that the most down-regulated locus was associated with bacteriophage production, suggesting that the variation seen in the study was not a result of phage. An interesting point made in this study is that despite the similarity in appearance of colony variants on solid medium the phenotypic properties of the variants may differ, indeed the phenotypic profile of the 'sticky' phenotype differed in terms of twitching motility when compared to Haussler *et al.* (Haussler *et al*., 2003). In addition Kirisits *et al.* noted the variation in swimming motilities of isolates with the wild type colony morphology from a 6-day biofilm, compared to the wild type parent strain. This observation emphasises the point that the morphological phenotype on solid medium can be a useful marker for genetic change but cannot conclusively determine the metabolic or genetic phenotype of the colony and that numerous mutations may go unnoticed due to the similar morphology to the wild type parent strain.

The Gram-positive bacterium *S. aureus* has been well-documented for its methicillin-resistant (MRSA) strains in the hospital environment. In addition to MRSA, numerous *S. aureus* strains are commonly found in infections associated with cystic fibrosis, rhinosinusitis, and chronic otitis media, in which the bacterium grows as biofilms (Gotz, 2002). *S. aureus* has also been reported colonising prosthetic equipment such as cochlear implants (Kos *et al*., 2009). The production of a 'slime-like' substance called polysaccharide intercellular adhesin (PIA) is regarded as a trademark of biofilm growth in most *Staphylococci* (Gotz, 2002) and has been shown to protect *Staphylococci* against the human innate immune system (Vuong *et al*., 2004). The biosynthesis of PIA is controlled by the expression of the *ica* operon (Cramton *et al*., 1999) and interestingly has been shown to be induced by anoxic conditions in *S. aureus* (Cramton *et al*., 2001). Other genes of particular relevance to biofilm formation include biofilm-associated protein (Cucarella *et al*., 2001) and the accumulation associated protein. More recently an alternative mechanism has been described for biofilm formation which is independent of the *ica* operon. This mechanism involves genes SarA and *agr* which are involved in initial attachment and dispersal, respectively (Gotz, 2002).

SCVs of *S. aureus* have also been reported when grown in biofilms (Kahl *et al*., 1998, Hausner and Wuertz, 1999). *S. aureus* SCVs can be broadly grouped into two classes based on their auxotrophic phenotypes; SCVs have either been shown to exhibit defective electron transport or be unable to synthesise thymidine. In 2002 Sadowska *et al.* tested *S. aureus* SCVs for their susceptibility to defensins, adhesion to the A549 cell line and killing activity of

professional phagocytes. The authors' results showed that the adhesion to the A549 cells was higher for SCVs than the 8325-4 parental strains. The uptake of SCVs by granulocytes was lower than ingestion of normal strains, but SCVs were killed with equal potency (Sadowska *et al*., 2002). These data suggest that SCVs of *S. aureus* have enhanced biofilm production and increased recalcitrance to the host immune system. Admittedly this paper does not specify that the SCVs were obtained from biofilms, simply that they were extracted from routine sputum extracts, however SCVs have been reported to develop during the biofilm mode of growth in *S. aureus* (Mitchell *et al*., 2010a). Taken together these data suggests that SCVs may contribute to the long-term persistence of *S. aureus* infections.

S. aureus and *H. influenzae* are common pathogens infecting the bronchi in CF patients. Infections with these bacteria are thought to initiate chronic damage of the lung epithelium allowing adhesion and the subsequent colonization of *P. aeruginosa*. Once established, *P. aeruginosa* dominates the bacterial flora in CF infection, however *S. aureus* does persist and there is growing evidence for co-habitation between these two pathogens as SCVs of both *S. aureus* and *P. aeruginosa* have been isolated from patients with cystic fibrosis (Schneider *et al*., 2008). In a recent study, Mitchell *et al.* (2010b) showed that exposure to the exoproduct 4-hydroxy-2-heptylquinoline-N-oxide (HQNO) produced by *P. aeruginosa* induced the advent of SCVs and increased biofilm formation in *S. aureus*. Interestingly an increase in biofilm formation was only observed in normal *S. aureus* strains ATCC 29213, Newman, Newbould, CF03-L, CF07-L and CF1A-L, whereas SCV strains NewbouldhemB, CF03-S, CF07-S, and CF1D-S showed no increase in biofilm formation. The authors showed that upon application of HQNO, sigma factor B (SigB) increased expression of fibronectin-binding protein A and the *sar*A genes leading to increase biofilm and SCV formation. SigB has previously been shown to be an essential regulator in *ica*-independent biofilm formation (Lauderdale *et al*., 2009). Furthermore in 2008 Kim *et al.* identified biofilm-producing variants in the *S. aureus* 483 strain and showed, via sequence analysis, that these variants were a result of mutations in the SigB operon; three point mutations in the *rsb*U gene and two point mutations in the *rsb*W gene resulting in a change in amino acid sequence of the respective gene products (Kim *et al*., 2008). The SigB operon consists of four genetic components, *rsb*UVW*sig*B, and regulates a number of virulence factors, and has now been shown to regulate the *ica* operon. The results support the idea that *P. aeruginosa* influences the pathogenicity of *S. aureus* by inducing the selection of SCVs which have been shown to form biofilms more readily than non SCVs (Sadowska *et al*., 2002). These results also show how species within a biofilm can induce pathogenicity in adjacent bacterial species (Mitchell *et al*., 2010).

In a previous study Hoffman *et al.*, (2006) showed that *S. aureus* exposed to exoproduct HQNO were protected from aminoglycosides antibiotics, including tobramycin and in 2010 Mitchell *et al.* showed that the HQNO-mediated protection is also under the control of sigma factor B (Mitchell *et al*., 2010a). Taken together these studies indicate that SigB mediated HQNO signalling is an important signal mechanism for interspecies interactions within a biofilm community.

The following section will focus specifically on *S. pneumoniae*, the model organism used in this thesis. It will cover the burden of pneumococcal disease, pneumococcal biofilm formation and research methodology, biofilm-mediated diversity in pneumococci and causes of diversification and persistence of the organism.

1.3 **Pneumococcal biofilms and research methodologies**

Pathogenic bacteria undergo a lifecycle alternating between planktonic and biofilm modes of growth (Figure 1-2) (Stoodley *et al*., 2002). *S. pneumoniae* exists in both an asymptomatic "carriage state" in the nasopharynx and in a "virulent state" in the blood during pneumococcal invasion and carriage of *S. pneumoniae* is known to facilitate transmission of the pathogen and precedes pneumococcal invasion (Simell *et al*., 2012). Colonisation of the nasopharynx is thought to be mediated, in part, by biofilm formation (Marks *et al*., 2012a). Furthermore, biofilm formation is commonly observed in infections with *S. pneumoniae* (Hall-Stoodley *et al*., 2006, Nistico *et al*., 2011) and is often associated with *H. influenzae* and *Moraxella catarrhalis* (Margolis *et al*., 2010). This section aims to describe the various methods of biofilm culture and discuss their merits and limitations. The methods used to culture and study pneumococcal biofilms *in vitro* range from basic microtitre plate assays to more complex continuous flow cell systems. Recent work has also cultured pneumococci on a substratum of epithelial cells to better mimic colonisation within the host (Marks *et al*., 2012a).

A recent review by Domenech *et al.* (2012) highlighted the fact that different groups researching pneumococcal biofilms have used a range of biofilm culture methods and as a result, care is required when direct comparisons are made between groups. The literature reveals that pneumococcal biofilms have been cultured in brain heart infusion broth (BHI), Todd-Hewitt broth (THB), Tryptone soya broth (TSB), Todd-Hewitt broth containing 0.5 % weight /volume (w/v) yeast extract (THY), chemically defined medium (CDM), and C medium with 0.08 % w/v yeast extract (CY) (not to mention the various sugar supplementations), in addition to being cultured at 34 ^{o}C, 35 ^{o}C, 37 ^{o}C, 34 ^{o}C in 5 % CO_2, 37 ^{o}C in 5 % CO_2 and 37 ^{o}C in 10 % CO_2 (Moscoso *et al*., 2006, Garcia-Castillo *et al*., 2007, Hall-Stoodley *et al*., 2008, Lizcano *et al*., 2010, Tapiainen *et al*., 2010, Drago *et al*., 2012, Marks *et al*., 2012a). This disparity in culture conditions in addition to the difference in biofilm culture methods will inevitably affect the growth and metabolism of the pneumococci and highlights the problem with direct comparisons between different groups.

In vitro pneumococcal biofilm research falls broadly into two categories; those who perform large scale high throughput studies on numerous pneumococcal isolates of various serotypes, sequence types and antibiotic resistance (Garcia-Castillo *et al*., 2007, Tapiainen *et al*., 2010, Camilli *et al*., 2011) and those who perform more in-depth molecular studies on specific serotypes (Waite *et al*., 2001, Allegrucci and Sauer, 2007, Domenech *et al*., 2009). The

high throughput studies tend to use high throughput screening methods such as the 96-well plate assay and crystal violet staining to determine biofilm formation whereas the more in-depth studies tend to use more complex biofilm continuous flow culture methods with microscopy-based methods of quantification. Examples of pneumococcal biofilm formation can be seen in Figure 1-3.

This section extends the review by Domenech *et al.* (2012) by looking in-depth at the methods used in pneumococcal biofilm studies, the implications they may have had on the results and looking at the future of pneumococcal biofilm research.

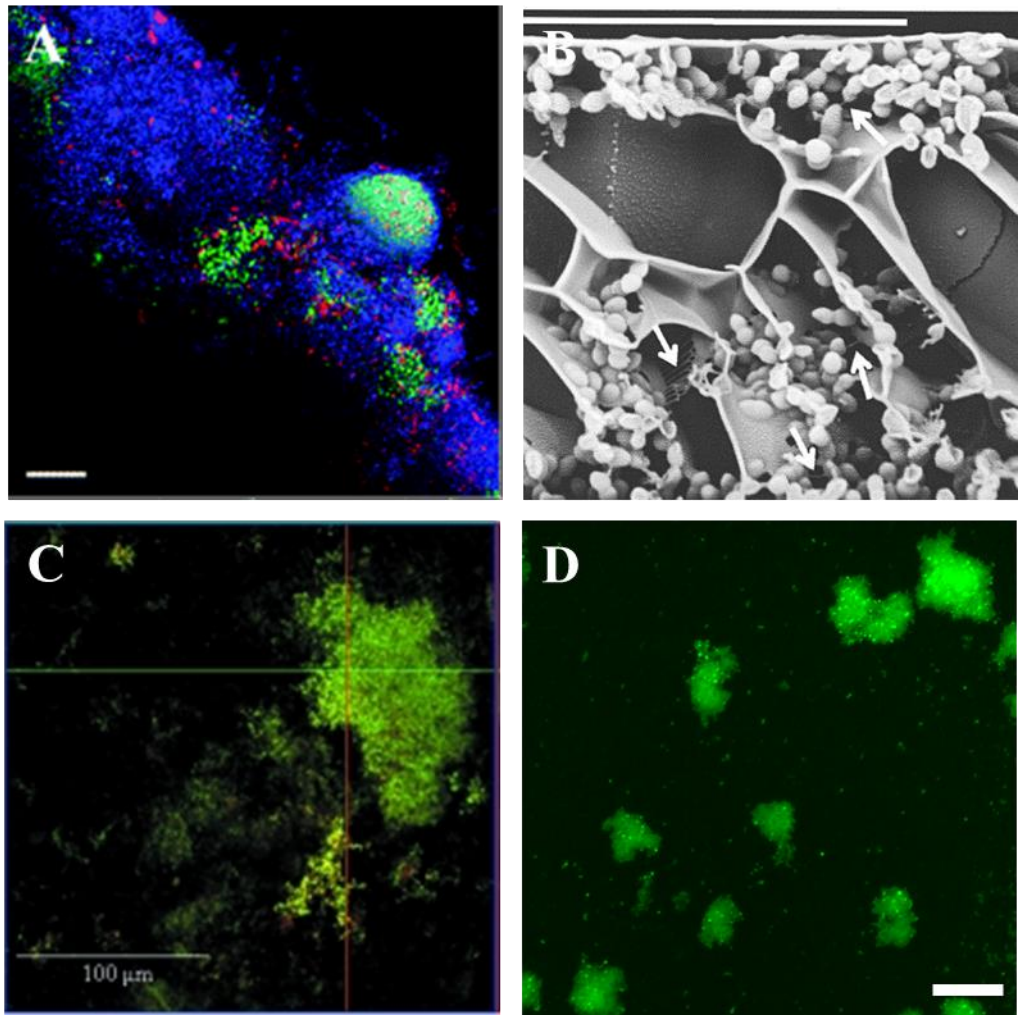


Figure 1-3: Pneumococcal biofilm formation

Pneumococcal biofilm formation has been visualised on a range of surfaces including on adenoid tissue (A), green stain indicates live bacteria, blue stain indicates the mucosa layer, scale bar 10 µm. Low-temperature scanning electron microscopy of pneumococcal strain R6 biofilm on glass (B) white arrows indicate filamentous attachment, scale bar 20 µm. Confocal laser scanning microscopy of a small colony variant of serotype 19F after 6 days of growth under flow conditions (C). Epi-fluorescent microscopy of 24-hour biofilms of serotype 14 grown under static conditions in a MatTek plate, scale bars 25 µm. Permission was granted from the American Society for Microbiology for the reproduction of these images via the Copyright Clearance Center's RightsLink® service; Image A was adapted from Nistico *et al.* (2011), image B was adapted from Moscoso *et al.* (2006) and image C was adapted from Allegucci and Sauer (2008). Image D, is unpublished data from this thesis.

1.3.1.1 Microtitre plate assay

The 96-well microtitre plate assay is a commonly used method to assess biofilm formation as it allows low-cost, high throughput results in a short time scale. Isolates are grown to an equivalent cell density, pipetted into the wells and allowed to adhere for the designated time period. Quantification of the biofilm is determined optically by staining the biofilm with a low percentage of crystal violet (0.1-1 %). The crystal violet binds to the biomass of the biofilm and after re-solubilisation the absorbance is read on an optical density plate reader. Biofilms with a greater biomass have a higher optical density reading and *vice versa*. Despite its experimental simplicity there are some notable limitations; firstly the space in each well limits the surface area for the inoculated cells to form a biofilm, therefore the cells would have less space to adhere and competition between cells for attachment will likely be increased. Secondly the limited volume of medium is likely to create a competition for nutrients which may result in increased cell death. Thirdly due to the close proximity of the cells within each well, the effect of quorum sensing and the build-up of waste products will likely have an effect on biofilm formation. Indeed this phenomenon was seen but not extensively discussed by Moscoso *et al.* (2006). Fourthly, without the presence of flow, the force of gravity will result in the deposition of dead, dormant or non-motile cells to the surface, which may either increase the biomass (if viable) or increase the build-up of waste products (if non-viable), and ultimately, may affect the overall quantification of the biofilm.

To better understand the biofilm forming properties of *S. pneumoniae* Moscoso *et al.* (2006) tested the biofilm forming properties of 40 pneumococcal strains of different serotypes. Using gene mutants the authors highlighted the importance of LytA, LytB, and LytC in biofilm formation in addition to PspA, CbpA and PcpA. To optimise biofilm formation a range of media were used including CY medium (C medium supplemented with 0.08 % yeast extract), THB supplemented with 0.5 % w/v yeast extract, and casein tryptone (CAT) medium. In addition to the chemically defined media Cden and CDM supplemented with 5 µg mL^{-1} of choline chloride, 50 µg mL^{-1} of asparagine and 250 µg mL^{-1} of sodium pyruvate. Biofilms were cultured on either 96-well flat bottom plates or glass bottom dishes and incubated for 6 hours at 34 ^{o}C under static conditions. The authors reported that Cden and CDM provided the best media for biofilm formation whereas THB + 0.5 % w/v yeast extract resulted in weak biofilm formation. This report was interesting as numerous studies have used THB as their sole medium for biofilm formation, albeit at different concentrations and without the supplement of yeast (Allegrucci *et al*., 2006, Hall-Stoodley *et al*., 2008, Sanchez *et al*., 2010). However this result

does coincide with the notion that rich media results in poor biofilm formation (Marks *et al*., 2012b). Using confocal laser scanning microscopy (CLSM) the authors showed that pneumococcal biofilms form 3D architecture approximately 20-30 μm in depth and that encapsulated clinical isolates formed weak biofilms. This is consistent with the observation that pneumococcal biofilms can be divided into architectural groups depending on the strain used (Allegrucci *et al*., 2006) and that the capsule production highly influences biofilm formation (Hall-Stoodley *et al*., 2006, Qin *et al*., 2013). In addition the authors used low temperature scanning electron microscopy to show the 'honeycomb' structure of an R6 biofilm (Figure 1-3B), indicating the architectural complexity of pneumococcal biofilms, a result which has been verified in additional strains cultured on epithelial cell lines (Marks *et al*., 2012a).

In order to establish a connection between colonization and biofilm formation Munoz-Elias *et al.* (2010) conducted an *in vitro* screen of 6500 *mariner* transposon mutants in an acapsular *S. pneumoniae* serotype 4 TIGR strain to reveal 69 biofilm-defective mutants. These mutations included genes encoding pili, choline binding proteins, ATP-binding cassette (ABC) and phosphotransferase system (PTS) transporters and efflux pumps, in addition to transcriptional regulator and conserved proteins of unknown function. The results showed that mutants defective in biofilm production were also defective in colonization of the nasopharynx. These data are relevant as it reveals that biofilm formation likely plays a vital role in nasopharyngeal carriage. Furthermore these data show that although the gene expression is known to be significantly different between planktonic and biofilm culture, there exists subsets of genes which are crucial for both modes of growth.

Those studies which aim to harvest cells directly from the biofilm are better suited to using a larger plate assay such as a 6-well plate assay or a continuous flow system, as accurately harvesting cells from a 96-well plate is challenging without the use of a sonication device. Hall-Stoodley *et al.* (2008) used three plate-based assays to quantify biofilm formation of a number of clinical pneumococci. To assess initial attachment 1×10^7 CFU mL^{-1} of *S. pneumoniae* were inoculated into 48-well plates and incubated at 37 ^{o}C in 5 % CO_2 for 24 hours before being quantified using 0.5 % w/v crystal violet. To quantify CFU counts from the biofilm a larger 6-well plate assay was used where 1×10^8 CFU mL^{-1} was used as an inoculum standard, and grown for 1, 3 and 6 days, replacing the medium daily with fresh 1:5 diluted THB. For confocal microscopy 1×10^8 CFU mL^{-1} was inoculated into MatTek culture plates with 1mL of 1:5 diluted THB and grown for 1, 3 and 6 days before being analysed using COMSTAT computer software (Heydorn *et al*., 2000). The benefit of using MatTek culture plates is that the central coverslip provides a clear thin surface to image through, providing quality confocal images.

Data from all three culture methods were combined to generate a biofilm forming Index scoring method. This scoring method has not been widely used but it use should be merited as it prevents bias that can occur as a result of using a single culture method.

In 2009, a comprehensive study was undertaken by Trappetti *et al.* to determine the effect of different sugars on biofilm growth in both a static biofilm model and in a murine model. Biofilms of pneumococcal strains D39 and TIGR4 were cultured in 96-well microtitre plates. The addition of competence stimulating peptides, CSP1 and CSP2, were used for D39 and TIGR4 strains, respectively. A total of 27 sugars were separately added to the TSB biofilm culture medium at a concentration of 0.2 % w/v. Of these sugars, sialic acid was shown to significantly increase the attachment of bacteria to the wells after 24 hours, the concentration of which is similar to that seen in human saliva. The same result was shown using soy and yeast based media suggesting that it was sialic acid that was responsible for the increase in cell attachment. Furthermore intranasal administration of sialic acid was shown to significantly increase pneumococcal counts in the MF1 mouse nasopharynx and was accompanied with infection of the lungs. This observation was confirmed in independent experiments using both the D39 and TIGR4 pneumococcal strains. This work highlights the importance of the microtitre plate in determining novel features involved in biofilm formation and also highlights the importance of supporting initial *in vitro* findings in an *in vivo* environment (Trappetti *et al*., 2009).

As mentioned the benefit of the microtitre plate assay is that is allows high throughput analysis of numerous samples. Such a study was performed by Garcia-Castillo *et al.* (2007), who investigated the difference in biofilm forming ability of pneumococcal isolates from CF samples and those from non-CF blood cultures. Biofilms were cultured for 20 hours at 37 ^{o}C on polystyrene pegs on the lid of the 96-well microtitre plate. The addition of polystyrene pegs offers the benefit of providing a surface area for biofilm formation which is suspended in the medium, thus preventing the build-up of dead or dormant planktonic pneumococci on the bottom of the well which might affect the overall quantification of the biofilm. The authors reported that 80 % of CF isolates were capable of forming biofilms compared to 50 % of non-CF blood isolates. The authors also reported no relationship between the biofilm forming ability of the isolate and the serotype or clone of the isolate. This result is maybe due to the fact that blood isolates, in order to avoid the immune system, would have to up regulate the capsule genes which have previously been shown to inhibit biofilm formation (Moscoso *et al*., 2006). In contrast the CF lung is notorious for harbouring biofilm of numerous pathogens including *P.* aeruginosa, *S. aureus* and *Burkholderia cepacia* (Costerton, 2001).

To determine whether a difference in biofilm forming properties was the cause of diverse clinical outcomes, Tapiainen *et al.* (2010) tested the biofilm properties of 204 clinical samples isolated from children; 106 samples from the nasopharynx, 43 samples from middle ear with effusion and 55 samples from blood. Biofilm quantification was determined optically by using the microtitre plate assay and staining with crystal violet. Bacterial suspensions were diluted 1:100 in BHI supplemented with 10 % volume/volume (v/v) horse serum and 0.5 % w/v glucose and biofilms were cultured in 96-well plates for 18 hours at 35 ^{o}C and 5 % CO_2. The author's revealed that unlike *E. coli* and *P. aeruginosa*, in which isolates from certain clinical infections have shown to produce more efficient biofilms *in vitro* (Martinez-Solano *et al*., 2008, Salo *et al*., 2009), isolates of *S. pneumoniae* showed no difference in the ability to form biofilms regardless of the type of clinical infection or site of isolation. The authors did note however that serotypes 14 and 33 were the most efficient biofilm formers. These results differ somewhat from Garcia-Castillo *et al.* (2007) who reported that pneumococci isolated from the CF lung produced better biofilms than non-CF blood samples, and that no difference in serotype and biofilm forming ability was observed. Despite the similarity in method used the difference could simply be due to the different sites of isolation, upper respiratory tract *vs.* CF lung, and the fact the pneumococcus is a highly diverse organism (Bentley *et al*., 2006, Hiller *et al*., 2010, Croucher *et al*., 2011). Alternatively it could highlight the limitation of the crystal violet assay as a means for biofilm quantification, as neither paper used a microscopy-based method of quantification and the architecture and viability of the biofilms cultured were not assessed.

The clinical significance of biofilm formation in invasive disease had also been investigated by Lizcano *et al.* who, in 2010, tested the biofilm-forming ability of 30 invasive and 22 non-invasive clinical isolates of serotype 6A and 6B and showed that the ability to form early biofilms *in vitro* does not reflect the true virulence potential seen following intranasal and intratracheal challenge of BALB/cJ mice. These results are not surprising considering that the biofilm mode of growth is considered to be a survival mechanism rather than a virulence factor and similar results have been verified in a flow cell system (Sanchez *et al*., 2011b). Trappetti *et al.* (2013) performed a similar experiment where the authors tested the biofilm forming ability of 12 blood and 13 ear serotype 3 pneumococcal isolates and related this to virulence in 6-week old CD-1 mice. They reported that blood isolates formed robust biofilms at pH 7.4 with the addition of ferric nitrate-supplemented medium, however were unable to colonise the mouse nasopharynx. In contrast, ferric nitrate-supplemented medium was shown to have an inhibitory effect on biofilm formation of ear isolates at pH 6.8. Ear isolates were

shown to be readily capable of colonising the mouse nasopharynx and capable of infecting the ears. Although Trappetti *et al.* (2013) used fewer isolates, both papers (Lizcano *et al*., 2010, Trappetti *et al*., 2013) surpass previous work by supporting the *in vitro* biofilm data with *in vivo* mouse studies. The role of ferric-supplemented medium was investigated previously by Trappetti *et al.* (2011b) who showed that biofilms form better in the presence of Iron (Fe) (III) and that the quorum sensing enzyme S-ribosylhomocysteine lyase (encoded for by the LuxS gene) regulates iron uptake in a static biofilm model. In this model pneumococci at a concentration of 1 x 10^5 CFU mL^{-1} were inoculated into 24-or 96-well microtitre plates in CY medium supplemented with increasing concentrations of Fe(III) (10, 50, 100, and 200 µM), Fe(II), or hemin. Biofilms were also grown with the with the Fe(III) chelator deferoxamine to see the effect of biofilm formation in the absence of iron. All biofilms were cultured at 37 ^{o}C in 5 % CO_2. The importance of LuxS was confirmed in a further study using a bioreactor and immobilised human respiratory epithelial cells (HREC) as a substratum (Vidal *et al*., 2013). Here the authors noted that biofilm biomass was significantly increased when grown on HREC compared to the abiotic surface highlighting the effect the substratum can have on biofilm formation. What is clear from this work is that biofilm formation plays a key role in colonisation and persistence; however its role in virulence is far less defined.

In 2011, Camilli *et al.* used the 96-well plate assay to determine whether the serotype or genetic background of 98 clinical pneumococcal isolates affected biofilm growth. The authors noted that antibiotic-susceptible isolates were able to form thicker biofilms when compared to their resistant counterparts; however there was no relationship between this phenomenon and the serotype of the isolates. Using the 96-well plate assay allows high throughput screening; however it does not take into account the architecture of the biofilm such as number or thickness of the microcolonies. Indeed these results agree with previous studies using plate based assays (Garcia-Castillo *et al*., 2007) but contradict other papers (Allegrucci *et al*., 2006, Hall-Stoodley *et al*., 2008) which have used a microscopy-based method of quantification. The only exception was serotype 3 which was classed as a non-biofilm former. This result is consistent with Hall-Stoodley *et al* (2008) but is an interesting report considering many papers have used serotype 3 as a model serotype (Waite *et al*., 2001, Allegrucci and Sauer, 2007, McEllistrem *et al*., 2007, Domenech *et al*., 2009). One suggestion for this disparity is the methods used; Allegrucci *et al.* (2007) used a continuous flow system whereas Camilli *et al.* (2011) used a static system. This may suggest one of two hypotheses, either that a continuous flow method may be a better method for biofilm formation or that the continuous flow method selects specifically for variants within a serotype 3 population that

are better adapted to form biofilms. This is a clear indication of how different methods can produce varying results and highlights the need for a consistent method of culture to be established for achieving reproducible results and accurate comparisons.

Interestingly, Camilli *et al.* (2011) reported that THB supplemented with 1 % w/v glucose produced the strongest biofilm formation. This coincides to an extent with Moscoso *et al.* (2006), who suggested that the additional glucose in CDM was the most likely cause for the extra biofilm growth in this medium, but also contradicts Moscoso *et al.* (2006) who reported that both THB and the supplementation of minimal media with 0.8 % w/v glucose failed to grow biofilm. The age difference in biofilm culture; 6 hours (Moscoso *et al*., 2006) and 18 hours (Camilli *et al*., 2011) may also contribute to differences seen in these two papers. Regardless both papers suggest however that glucose may be an important factor in biofilm development.

The microtitre plate assay has also been used to determine the role of the host complement defence system in response to pneumococcal biofilm formation; an avenue of research which has yet to be extensively studied (Domenech *et al*., 2013b). Using the 96-well microtitre plate assay, described previously (Moscoso *et al*., 2006), the authors showed a reduced deposition of the C3b complement component to pneumococcal biofilms. C3b is a key complement of the component system responsible for the optimisation and ultimate clearance of pathogens. A reduction in binding of C-reactive protein and C1q was also observed indicating the pneumococcal biofilms are capable of avoiding the classical complement pathways. This is the first evidence that biofilm formation is capable of directly avoiding the immune response and as such is clear evidence for why pneumococcal infection persists over a long time frame.

The importance of using different i*n vitro* biofilm models has been highlighted previously (Trappetti *et al*., 2011a). Three biofilm culture methods, 2 96-well microtitre plate assays and 1 continuous culture system were used to better understand the role of quorum sensing in biofilm formation. The ability to form biofilms in competence gene mutants differed between the static and continuous model. The continuous culture system had previously been developed by the Centers for Disease Control (CDC) (Donlan *et al*., 2004). It consisted of a biofilm reactor with eight removable rods, which in turn contained three polycarbonate coupons (1.3 cm in diameter) to allow cell attachment. The reactor contained 400 mL of BHI supplemented with 0.5 % w/v casein and 0.2 % w/v yeast extract. The inoculum was allowed to attach during a 12 hour batch culture period before 10 % BHI broth was pumped at a rate of

0.5 mL per minute using a Materflex peristaltic pump. The reactor was incubated at 35 ^{o}C with a gas mixture of 85 % nitrogen, 5 % oxygen and 10 % CO_2. The difference in biofilm-forming ability of the two methods used is likely to be due to numerous factors including the surface area of attachment, the difference in growth media constituents and the effect of flow. The authors do report the consistent finding that encapsulated pneumococci formed weaker biofilm. Interestingly the authors state that neither model used indicated superiority over the other and that the use of varying models is important for understanding the true importance of pneumococcal biofilms.

The difference in biofilm-forming ability differs between published studies. A recent paper by (Domenech *et al*., 2013a) showed using a 96-plate model that clinical isolates of pneumococcal serotypes 19F and 19A, as well as isogenic transformants of 19F and 19A, are good biofilm formers compared to serotypes 19B and 19C and that the difference in the biochemical structure of the capsule may account for the difference in biofilm formation *in vitro*. These results, in turn, may help explain the prevalence of these serotypes in pneumococcal carriage. This is one of the first studies to highlight the importance of biofilm study in the context of pneumococcal epidemiology and emergence of serotypes post-vaccination.

1.3.1.2 Continuous flow systems

To date, three primary methods of continuous flow culture have been used to study pneumococcal biofilms; Sorbarod filters (Budhani and Struthers, 1997), once through continuous flow cell systems (Allegrucci *et al*., 2006) and a bioreactor developed by the CDC (Donlan *et al*., 2004). Continuous flow systems have the benefit of being able to form more established biofilms, to study the effect of shear forces and nutrient content and facilitate continuous real-time biofilm observation without the need for sacrificial wells. However they are costly and require expertise in assembly. They are also not suited for high throughput screening of numerous samples. Nevertheless they have been instrumental in understanding the genetic and proteomic diversity which can arise as a result of growth under biofilm conditions.

In order to investigate the capsule variation seen in serotype 3 isolates, Waite *et al.* (2001) followed a similar protocol described previously by Budhani and Struthers (1997), by culturing biofilms in Sorbarod filters at 34 ^{o}C. A Sorbarod is a paper sleeve made up of

compacted cellulose fibres; Sorbarod filters measure 10 mm in diameter by 20 mm in length and are placed within lengths of silicone to provide a substratum for cell attachment. In the case of both groups, the system is fed with BHI broth by a Watson Marlow 205U peristaltic pump at a flow rate of 0.1 mL per minute. Sample collection was achieved by collecting the effluent using special adaptors connected to glass receptacle bottles. Interestingly each group used slightly different culture conditions; Budhani and Struthers (1997) incubated their sorbarod biofilms at 37 ^{o}C whereas Waite *et al.* (2001) incubated the biofilm at 34 ^{o}C in atmospheric conditions, with the principle being that 34 ^{o}C is more likely to simulate nasopharyngeal carriage (Keck *et al*., 2000). In addition the inoculum differed, with Budhani and Struthers inoculating a 3 mL inoculum of a 10–12 hour broth culture compared to Waite *et al.* who used 2 mL of a 10^{6} CFU mL^{-1} culture. Waite *et al.* (2001) reported the generation of random tandem duplications within the *cap*3A gene, similar results were also reported by McEllistrem *et al.* (2007), who used a plate based biofilm method using 0.2 m/25 mm filters as a substratum. Each biofilm filter was seeded with 10^{6} colonies and incubated overnight at 34 °C on blood agar plates with 5 % v/v Sheep blood. Sorbarod filters have been shown to produce reproducible results but the nature of the cellulose fibres does not facilitate direct visualisation of the biofilm compared to a flow cell method.

The appearance of a small non-mucoid variant (SCV), a medium mucoid (MMV) and a large mucoid variant (LMV) from *S. pneumoniae* serotype 3 biofilms was reported in 2007 by Allegrucci and Sauer. To assess colony variation Allegrucci *et al*. (2007) cultured biofilms in a diluted THB broth under flow conditions in a once-through continuous tube reactor system with the addition of a flow cell to visualise the biofilm. The flow cell contained an aluminium chamber measuring 4 mm x 1.3 cm x 5 cm covered by glass surfaces, one which acted as the substratum and the other as a coverslip for visualisation. Biofilms were incubated at 37 °C in 5 % CO_2 with a flow rate of 0.014 mL per minute for 1, 3, 6, and 9 days (Allegrucci *et al*., 2006). Using this flow cell model the authors were also able to distinguish 3 main classes of pneumococcal biofilm architecture in various serotypes (Allegrucci *et al*., 2006). Group I biofilms consisted of highly structured biofilms with distinct microcolonies and water channels microcolonies ranged from 40-150 µm in diameter and 90-15 µm in height. This group was characteristic of serotype 3 and 23. Group II biofilms were shown to be less structured with smaller microcolonies and were characteristic of serotypes 9, 14, 11, 6 and 18. Group III produced flat biofilms with small microcolonies <20 µm in diameter and were characteristic of serotype 19 in addition to other uncharacterised serotypes. This work contradicts that of Garcia-Castillo *et al.* (2007) which showed no difference in the ability of different serotypes

using the microtitre assay, but supports the work of others (Hall-Stoodley *et al*., 2008, Tapiainen *et al*., 2010) who also noted differences between serotype and biofilm formation. This disparity is most likely due to the differences in biofilm culture used.

To assess mature biofilm growth the once-through continuous tube reactor system, used by Allegrucci *et al.* (2006) was also employed by Sanchez *et al* (Sanchez *et al*., 2010, Sanchez *et al*., 2011a, Sanchez *et al*., 2011b). Their system was inoculated with mid-log phase growth and allowed to adhere to the glass substratum for 2 hours before culturing at 37 °C in 5 % CO_2 for 3 days. Biofilms were also cultured in a meter-long section of Masterflex silicone tubing with an internal diameter of 0.89 mm. These biofilms were cultured at 37 °C in 5 % CO_2 for 3 days with a flow rate of 0.035 mL per minute. Using this method the authors have been able to highlight the importance of the pneumococcal serine rich repeat protein (PsrP) in biofilm formation (Sanchez *et al*., 2010), showing that the virulent TIGR4 strain formed biofilms of a densely packed nature with large bacterial aggregates unlike the PsrP mutant (T4Δ*psr*P) which formed biofilms with a 'patchier phenotype' and smaller aggregates. The authors also observed biofilm-like aggregates in the lungs of TIGR4 infected 6 week old BALB/cj mice which were not seen with the T4Δ*psr*P mutant. To examine the effect of PsrP in early biofilm formation the authors inoculated a 96-well plate with 10^6 CFUs in THB and cultured biofilms for 2, 4, 6, 8, 18 and 24 hours at 37 ^oC in 5 % CO_2 before quantifying the biofilm by staining the biomass with 0.1 % w/v crystal violet. The authors showed no difference in the biofilm forming ability using the microtitre plate assay suggesting that PsrP is not involved in the initial attachment and early biofilm formation. Ideally performing experiments on a 1-day biofilm in the flow cell system would have facilitated a direct comparison to account for the differences in the biofilm culture methods. Sanchez *et al.* (2011a) have also shown that the difference in protein profiles between planktonic and biofilm samples may provide valuable insight into future vaccine design and that biofilm cell lysate react less to human convalescent sera compared to planktonic sample (Sanchez *et al*., 2011a) strengthening the concept that biofilm growth itself is not a virulence mechanism but a mechanism for survival and persistence (Oggioni *et al*., 2006). A follow-up paper in the same year supported this by showing that biofilm-derived bacteria adhere better to *in vitro* cell lines compared to planktonic cells but are attenuated in a mouse infection model (Sanchez *et al*., 2011b). The authors noted that the proteomic and microarray data obtained in their studies disagreed with previous literature (Allegrucci *et al*., 2006) and suggested that this was due to the difference in strains used and the large proportion of dead bacteria observed within the mature biofilms. The literature published by Sanchez *et al.* is particularly noteworthy as the authors have not limited

themselves to a single culture model and have tried to bridge the gap between *in vitro* and *in vivo* models.

1.3.1.3 *in vivo* biofilm models

In recent years an increasing number of groups have used *in vivo* models to study to role of *S. pneumoniae* infection, including the mouse, rat and chinchilla models (Oggioni *et al*., 2006, Trappetti *et al*., 2009, Weimer *et al*., 2010, Sanchez *et al*., 2011b). These models predominantly study the role of biofilms in otitis media (Weimer *et al*., 2010, Yadav *et al*., 2012) but have also been used to study colonisation of the lung and nasopharynx (Trappetti *et al*., 2009, Marks *et al*., 2012a). The benefit of *in vivo* models is that the pneumococcal cells grow on cellular substratum akin to what they would in the human host and as such provide a more representative model as they take into account host-bacterial interactions. Moreover those groups that have utilised *in vivo* models have also used *in vitro* models, allowing a comparative assessment between *in vitro* and *in vivo* biofilms. The disadvantage of *in vivo* models is that they have a high associated cost, require expertise and the ethical considerations make them less suited for large scale, high throughput studies.

One of the first uses of a mouse model of infection, in the context of biofilm formation, was described by Oggioni *et al.* in 2006. Here the authors revealed that the gene expression pattern seen in biofilm cells grown *in vitro* for 18 hours resembled the gene expression seen in the mouse meningitis and pneumonia infection models; with increased expression of neuraminidase (*nan*A and *nan*B), competence genes (*com*A and *com*X), metalloproteinases and oxidative stress genes (Oggioni *et al*., 2006). An additional gene expression pattern was seen in pneumococci harvested from blood which differed from that seen in biofilm cells suggesting that biofilm does not play a role in blood-based pneumococcal infections. Biofilm formation was achieved using the 96-well microtitre assay and was reported to be dependent on the addition of synthetic competence stimulating peptide (CSP). The addition of CSP has not been necessary for other groups using the plate-based biofilm assays, however the inability of CSP mutants to form biofilms in this study suggests CSP is essential for biofilm formation. Addition of CSP into the mouse model also showed increased virulence in causing pneumonia. Moreover, when mice underwent intracranial, intranasal and intravenous challenge with biofilm-derived bacteria, the bacteria showed increased efficiency in inducing pneumonia and meningitis compared to planktonic bacteria, whereas planktonic cultures were more virulent in

the sepsis model. Although direct visualisation of biofilm was not observed within the mouse model this work reveals two important points, namely that pneumococci express a distinct set of genes when under biofilm and non-biofilm conditions and that similarities in gene expression between *in vitro* biofilm assays and *in vivo* infection models suggest that pneumococcal pneumonia and meningitis infection are likely to involve biofilm formation.

Biofilms are considered to be inherently resistant to antimicrobial agents which has been attributed to the phenotypic heterogeneity and slower growth of bacteria within the biofilm (Hoiby *et al*., 2010), delayed antibiotic penetration through the EPS (Costerton *et al*., 1999), and the role of persister cells in facilitating long-term survival (Lewis, 2010). To understand the basis of increased antibiotic resistance of pneumococci in carriage and pneumococcal colonisation and to understand why antimicrobial treatment fails to irradiate the bacterium from the nasopharynx, Marks *et al.* (2012a) developed a model whereby pneumococcal biofilms of non-invasive strain EF3030 were cultured on a substratum of H292 epithelial cells throughout the entire biofilm growth period, in addition to using a murine infection model to observe biofilm-mediated colonisation directly *in vivo*. The authors used a murine model to colonise BALB/cByJ with 5 x 10^6 CFUs to correlate the similarities seen *in vitro* and *in vivo*; by analysing infected tissue using scanning electron microscopy (SEM) the authors were able to visualise bacterial colonisation in the nasopharynx. Interestingly the authors noted that the posterior region harboured large aggregates of pneumococci in complex organised structures embedded in a matrix whereas the anterior region was colonised more sparsely over the ciliated epithelium. Furthermore pneumococci colonising the nasopharynx were shown to be more resistant to gentamicin and penicillin due to biofilm formation rather than acquired resistance. In conjunction with the murine infection model pneumococci were seeded into polystyrene 24-well plates containing a confluent layer of H292 mucoepidermoid bronchial carcinoma cells on glass coverslips and primary ciliated tracheobronchial epithelial cells from healthy volunteers on Transwell® inserts. Biofilms were subsequently cultured in CDM at 34 ^{o}C with 5 % CO_2 to mimic nasopharyngeal carriage. Biofilm structure and cell viability were determined using SEM and CLSM. For long-term biofilm work H292 cells and differentiated epithelial cells were fixed with 4 % w/v paraformaldehyde to prevent pneumococcal killing of the substratum. Biofilms were shown to form on both primary ciliated tracheobronchial epithelial cells and the H292 carcinoma line with a comparable morphology seen *in vivo*. By comparison to biofilms formed on glass, biofilms formed on cell lines showed rapid biofilm formation after 6 hours with more advanced architecture including increased biomass, cellular towers with a 'honeycomb' structure and well developed water channels. In

addition when treated with gentamicin, biofilms grown on plastic and glass were shown to be more sensitive over a 48 hour period. It should be noted that the advanced biofilm formation on epithelial cells seen for strain EF3030 (a serotype 19 strain isolated from otitis media patient) was not mirrored by all the other clinical stains tested, however each strain's ability to form biofilms on *in vitro* cell lines and the ability to colonise the nasopharynx was well correlated suggesting that the epithelial substratum model is an excellent model to study pneumococcal colonisation. Previously attachment assays using cells lines had been performed on biofilm bacteria to assess differences in attachment (Sanchez *et al*., 2011b), but to date, this is the first paper to grow pneumococcal biofilm on a cellular substratum *in vitro*. Using a similar model Marks *et al.* (2012b) have confirmed that high levels of genetic recombination occur during nasopharyngeal carriage which may play a vital role in pneumococcal infection and vaccine design (Marks *et al*., 2012b), as well as showing that biofilm formation enhances the survival pneumococci on contaminated fomites (Jefferies *et al*., 2011, Marks *et al*., 2013b) which may facilitate pneumococcal colonisation and persistence. Previously, this group and others have only been able to culture biofilms on immobilised cell lines due to pneumococcal killing of the substratum (Marks *et al*., 2012a, Vidal *et al*., 2013), however recently Marks *et al.* (2013a) were able to develop a model whereby biofilms were cultured on live epithelial cells. This was achieved by first culturing biofilms on immobilised NCI-H292 bronchial carcinoma cells for 48 hours at 34 ^{o}C before washing and transferring the biofilm cells in RPMI 1640 media with 2 % w/v foetal bovine serum to live HRECs. Using this model the authors were able to show that infection of the epithelial substratum with Influenza A virus resulted in pneumococcal dispersal *in vitro* and increased dissemination from the nasopharynx *in vivo*. Furthermore addition of norepinephrine, ATP and glucose to 48-hour biofilms grown on immobilised HRECs showed an increase in cell dispersal. Increased dispersal was also observed when biofilms were cultured under temperatures simulating fever (38.5 ^{o}C). Taken together these data may help explain why pneumococcal invasive disease is increased in conjunction with increased influenza virus infection (Jansen *et al*., 2008) and moreover highlights the adaptability of biofilm cells to fluctuations within the environment.

The chinchilla model of infection was used by Reid *et al.*, in 2009, to determine the role of biofilms in otitis media; 0.1 mL suspensions containing 40 CFUs, 4 CFUs, or 1 CFUs were inoculated via transbullar injection into chinchillas and cultured for 1, 3, 7 or 12 days before dispatching the animals to assess the middle ear cavity. The authors observed that after 12 days of infection dense aggregates akin to biofilm was present in the middle ear, further analysis using SEM revealed similar biofilm structures to that seen *in vitro*. To compare *in vivo*

observations with *in vitro* culture, TIGR4 pneumococci were cultured as biofilms for 24 hours in a continuous flow system using a Stovall glass flow cell in THY medium at 37 ^{o}C with flow rate of 5 mL per minute. Notably biofilms formed *in vivo* were shown to contain more matrix than those seen *in vitro* (Reid *et al*., 2009). Recently Yadav *et al.* (2013) have used both an *in vitro* system and the *in vivo* rat model to study biofilms in the context of otitis media. *In vitro* biofilm formation was determined using the 96-well plate assay as previously described with the exception that TSB medium was supplemented with 0 %, 1 %, and 2 % glucose. Biofilms were cultured for 6, 12, 18, and 24 hours at 37 °C in 5 % CO_2. Using this model the authors reported that addition of glucose into the medium enhanced biofilm formation at 18 and 24 hours, a consistent finding reported previously by other *in vitro* studies (Moscoso *et al*., 2006, Camilli *et al*., 2011). For *in vivo* studies, Sprague-Dawley rats were inoculated with 3×10^7 CFUs of *S. pneumoniae* into the middle ear cavity. Animals were infected for 1 or 2 weeks before being dispatched and the middle ear mucosa underwent SEM to determine the presence of biofilm; 50 % of rats dispatched at 1 week showed high biofilm formation in middle ear mucosa, predominantly located on the non-ciliated epithelium. Interestingly this number dropped in the group dispatched at 2 weeks, however the consistent finding that biofilm formed predominantly on the non-ciliated epithelium was consistent in both groups. Both the rat and chinchilla models have been instrumental in directly observing pneumococcal biofilms in the middle ear thus highlighting the role of biofilms in pneumococcal otitis media infections.

1.3.1.4 Multispecies models

S. pneumoniae has been implicated as a major component of polymicrobial infections and is often associated with *Moraxella catarrhalis* and *Haemophilus influenzae* in otitis media (Murphy *et al*., 2009). The regularity in which these three pathogens are reported indicates that interspecies-interactions are essential to the success of their infectivity. Indeed penicillin-sensitive pneumococcal strains have shown to be protected from amoxicillin and benzylpenicillin antibiotics when co-cultured with β-lactamase-producing *M. catarrhalis* in a continuous biofilm system (Budhani and Struthers, 1998). This is clear evidence that multispecies co-culture can have synergistic effects which are beneficial for surrounding bacteria and furthermore highlights the fact that previous literature which uses only a single culture system doesn't provide a realistic interpretation of biofilms *in vivo*. The chinchilla model was subsequently used by Weimer *et al.* (2010) who showed that coinfection with *H.*

influenzae promotes pneumococcal biofilm formation and furthermore increases the number of pneumococcal transparent colonies which have a reduced propensity to cause systemic infection. The increase in transparent colonies is consistent with biofilm production as transparent colonies exhibit increased adherence to human epithelial cells and enhanced colonisation of the nasopharynx (Allegrucci and Sauer, 2008). It is also consistent with the notion suggested by Sanchez *et al.* (2011b) that death of opaque cell variants within the biofilm provides building components for the EPS thus facilitating the survival of the transparent phenotype.

Previously the inhibition of *H. influenzae* by hydrogen peroxide produced by the *S. pneumoniae spx*B gene (Pericone *et al*., 2000) is evidence for the concept that *S. pneumoniae* acts as a protector against *H. influenzae* and *S. aureus.* However recent work from Cope *et al* (2011), looking at the chronic rhinosinusitis (CRS) model of infection, suggests that co-culture of *S. pneumoniae* with NTHi may actually be synergistic. The authors' co-culture *in vitro* biofilm experiments revealed that the NTHi type IV pili, a gene essential for NTHi biofilm formation, was only expressed in co-culture and that the *S. pneumoniae* pyruvate oxidase gene was also up-regulated in co-culture, which may aid survival when other substrates are unavailable. Interestingly the pneumococcal pneumolysin and pneumococcal adherence factor A genes were down-regulated in co-culture. The authors' confirmed these results with transcriptional analysis in excised human CRS tissue. Furthermore the authors' showed that the interactions between *H. influenzae* and *S. pneumoniae* were influenced by physical contact of the bacteria and undefined chemical mechanisms in the supernatant of the co-cultured biofilms. Taken together these data suggest that synergistic and antagonistic interactions mechanisms involved in co-culture biofilms are complex and that further work is needed to better understand the interactions involved in polymicrobial biofilms.

1.4 Colony morphology variation in *S. pneumoniae* biofilms

Since the publication of Avery *et al.'s* seminal work (Avery *et al*., 1944) revealing that the factor responsible for the transformation of *S. pneumoniae* from a smooth to a rough variant was DNA, numerous papers have described phenotypic variation of *S. pneumoniae*. However relatively few papers have considered the variation in the context of biofilm and those that have, have limited their studies to serotypes 3 and 19 (Waite *et al*., 2001, Allegrucci and Sauer, 2007, McEllistrem *et al*., 2007, Allegrucci and Sauer, 2008, Domenech *et al*., 2009). Previously two phase variable colony phenotypes, termed transparent and opaque, were described for *S. pneumoniae* (Overweg *et al*., 2000). The transparent phenotype has increased cell wall carbohydrates and reduced capsular polysaccharide, resulting in an increased adherence to human epithelial cells and enhanced colonisation of the nasopharynx (Waite *et al*., 2001). Whereas the opaque phenotype has no decrease in the capsular polysaccharide and as such exhibits an increased resistance to opsonophagocytosis (Cundell *et al*., 1995). Both phase variants have been directly observed in pneumococcal infections (Arai *et al*., 2011) with the opaque variant being the dominant variant (Serrano *et al*., 2006).

Previously non-phase variable variants observed in pneumococcal biofilms exhibited varied colony size and mucoid appearance on blood agar (Allegrucci and Sauer, 2007). In a 2007 paper, Allegrucci and Sauer reported the appearance of a SCV a medium mucoid (MMV) and a large mucoid variant (LMV) from *S. pneumoniae* serotype 3 biofilms. The SCV variant appeared early in biofilm development and exhibited auto-aggregation, hyper-adhesion, and formed biofilms with distinct microcolony structure (Figure 1-3C). In contrast, LMVs appeared later in biofilm development and exhibited poor aggregation, poor adherence and formed flat unstructured biofilms. Based on previous work by Waite *et al*., (2001) Allegrucci and Sauer identified a 7kb deletion in the *cps3DSU* operon which was implicated in the formation of the SCV variants.

As a follow-up to their 2007 study, Allegrucci and Sauer (2008) reported the presence of several colony variants of *S. pneumoniae* serotype 19 when grown under biofilm conditions. The variants included SCV, small mucoid variants (SMV), MMV and LMV (Allegrucci and Sauer, 2008). Sequence analysis revealed that all biofilm variants had single nucleotide polymorphisms (SNPs) in *cps19F* gene. The variants emerged only upon surface-associated growth and were shown to be non-revertible, with the exception of SCVs. SCVs were shown to revert to SMVs suggesting that the variants were phase-variable; however the lack of reversion to a wild type phenotype meant that the variants were distinguishable from the phase-variable

transparent and opaque colony phenotypes (Overweg *et al*., 2000). The SCV-SMV reversion was attributed to partial restoration of the wild type *cps*19F sequence, however due to the presence of a transition in the ribosomal binding site of the *cps19F* in all the revertants, the authors noted that the partial restoration may have been coincidental and that full genome sequencing may elucidate further mutations to explain the variations. The authors' data suggests the emergence of non-phase-variable colony variants is dependent on environmental conditions intrinsic to biofilms. The authors' suggested that the cause of the non-phase-variable colony variants was oxidative stress. To test this hypothesis the authors grew biofilms in medium supplemented with 10 mM sodium thiosulphate or catalase to eliminate hydrogen peroxide; growth under these conditions yielded a significant reduction in variants formed by wild type biofilms. *S. pneumoniae* is capable of producing hydrogen peroxide when grown aerobically in liquid culture via activation of the pyruvate oxidase-encoding gene, *spx*B. The TIGR4-delta *spx*B mutant is defective in hydrogen peroxide production. When TIGR4-delta was grown under biofilm conditions, the authors noted a significant reduction in acapsular variant formation. The authors also noted that the concentration of hydrogen peroxide is a crucial factor in the diversification of *S. pneumoniae* into SCV, SMV, MMV and LMV, with lower concentrations resulting in the formation of SCV, whereas higher concentrations resulted in the formation of SMV, MMV and LMV in a sequential manner. These data suggest that hydrogen peroxide plays an important role in generating the observed phenotypic variation. Production of hydrogen peroxide by *S. pneumoniae* has previously been shown to inhibit the growth of neighbouring bacteria (Pericone *et al*., 2000). Taken together these data suggest that in addition to inhibiting the growth of competitors, production of hydrogen peroxide by *S. pneumoniae* contributes to generating genetic diversity within a biofilm community.
Hydrogen peroxide has previously been shown to induce mucoid colony variants in *P. aeruginosa* (Mathee *et al*., 1999). Sequence analysis revealed that the mucoid variants had mutations in the *mucA22* incurred as a result of hydrogen peroxide (Mathee *et al*., 1999). This work suggests that this phenomenon of colony morphology variance to form biofilms is not unique to *S. pneumoniae* and is a useful marker for indicating phenotypic and genetic variation.

1.5 Causes of heterogeneity

There is a long standing debate over whether mutations under selection arise due to adaptive mutation or selected for under the laws of natural selection (Roth *et al*., 2006). Discussed below are some of the key mechanisms involved in generating genetic variation within the cell.

1.5.1 Adaptive mutation and phase variation

Phase variation and adaptive mutation are thought to be key genetic mechanisms involved in generating the phenotypic variation seen in biofilms. Adaptive mutation is a recombination-dependent form of genome mutation, initiated in response to an environmental stress (McKenzie *et al*., 2000). The environmental stresses may not be directly mutagenic but instead activate mechanisms which cause mutations (Roth *et al*., 2006) . Phase variation is a mechanism by which bacteria are able to generate a reversible phenotypically heterogeneous population in order to respond to changing environmental conditions. Phase variation is often referred to as an 'on-off' switch in phenotype (Henderson *et al*., 1999) and occurs at relatively higher rates compared to other mechanisms responsible for changes in phenotype, such as point mutations (Chia *et al*., 2008). Inverted DNA segments, mobile transposons and expression of genes via homologous recombination are all factors responsible for the switching mechanism of phase variation. For example *Staphylococcus epidermidis* has been shown to produce variants expressing the PIA, a gene responsible for biofilm development. The insertion of insertion sequence 256 (IS256) into the *icaA* and *icaC* genes results in the inactivation of the *ica* genes resulting in variants that are PIA-negative (Ziebuhr *et al*., 1999), however the insertion of IS256 was shown to be reversible whereby PIA expression was restored. In addition Webb *et al*. (2004) have shown that activation of the lysogenic filamentous bacteriophage Pf4 in *P. aeruginosa* results in a small colony variant as a result of phase variation (Webb *et al*., 2004).

1.5.2 Mutation frequency and hypermutation

In every generation there are numerous replication errors including mismatched base pairing, transversions and transitions. Due to the short replication time of bacteria, these replication errors are highly prone to occur. Mutations conferring beneficial traits such as antibiotic resistance will be selected for under the appropriate environmental conditions. Hypermutable strains are strains which have an increased spontaneous mutation rate due to defective DNA repair systems (Oliver and Mena, 2010). Strains deficient in the DNA mismatch repair system, containing mutations in the *mut*S, *mut*L and *mut*U genes, have been the best characterised to date (Conibear *et al*., 2009), however, a recent report has shown that mutations in the oxidative repair (GO) system genes (*mut*M, *mut*Y and *mut*T) also result in an increased mutation frequency (Mandsberg *et al*., 2009).

Hypermutation has been observed in *P. aeruginosa* (Driffield *et al*., 2008), *S. aureus* (Prunier *et al*., 2003) and *H. influenzae* (Watson *et al*., 2004) showing that this phenomenon is not specific to one type of pathogen. It has been shown that spontaneous mutation is increased in bacteria during biofilm growth (Driffield *et al*., 2008). In 2008, Driffield *et al.* (2008) showed that *P. aeruginosa* cultures grown under biofilm conditions exhibited a 105-fold increase in mutation frequency compared with planktonic cultures. Furthermore transcriptomics revealed a down regulation of genes conferring protection against oxidative stress in biofilm cultures including *ahp*C, *kat*A, and *sod*B. The authors concluded that this down-regulation may enhance the rate of mutation within the biofilm leading to an increase proportion of hypermutators which in turn may result in increased numbers of antibiotic-resistant mutants (Driffield *et al*., 2008). Although most research of this kind has focused on monospecies of *P. aeruginosa*, hypermutation may also play a major role in generating diversity within a multi-species biofilm community. As one example, Watson *et al.* (2004) identified *mutS* mutants in hypermutable *H. influenzae* isolated from a microbial consortium of CF sputum.

In 2008 Besier *et al.* revealed that thymidine-dependent small colony variants (TD-SCVs) of *S. aureus* exhibit significantly higher rates of mutation than both wild type isolates derived from CF patients and non-CF-derived wild type isolates (Besier *et al*., 2008). Using sequence analysis the authors revealed that frameshift mutations in the *mutL* mismatch repair gene were responsible for the observed hypermutation. Significantly higher resistance to non-beta-lactam antibiotics was observed in the TD-SCVs isolates compared to the CF and non-CF

derived wild type isolates, implicating hypermutation as the cause for this increase in antibiotic resistance.

1.5.3 Oxidative stress

Oxidative stress is the stress applied to bacteria via the manifestation of reactive oxygen species (ROS). For example in the lung, ROS are produced by polymorphonuclear leukocytes (PMNs) as a result of the host immune response which can result in an oxidatively damaged form of guanosine called 7,8-dihydro-8-oxodeoxyguanosine (8-oxodG) (Michaels and Miller, 1992). This form of guanosine can be repaired by the cells oxidative repair system (GO) (Michaels and Miller, 1992). ROS have been shown to have a mutagenic effect on DNA which can result in a higher mutation frequency and antibiotic resistance (Ciofu *et al*., 2005), in addition to increased alginate production in biofilms (Mathee *et al*., 1999). Moreover an increased mutation frequency has been linked to defective oxidative repair system in *P. aeruginosa* (Mandsberg *et al*., 2009).

In the past few years a number of articles have highlighted the role of oxidative stress in generating diversity. In 2008 Boles and Singh studied the role of oxidative stress on generating diversity in a *P. aeruginosa* biofilm community. The authors noted that cells within a biofilm can incur double-stranded breaks due to oxidative stress which are subsequently repaired via recombinatorial DNA repair mechanism. During this repair process, mutations arise due to the efficacy of the DNA repair system. Their hypothesis was that mutations which arise from the repair mechanism may promote adaptation to environmental changes within the biofilm community (Boles and Singh, 2008). Variants isolated from biofilms were observed over-expressing pyomelanin which has previously been shown to protect cells against oxidative stress (Nosanchuk and Casadevall, 2003). Interestingly the authors eliminated the possibility of the SOS response as being responsible for generating diversity by using a SOS-deficient strain with a *lexA* point mutation and showing no additional variation compared to the wild type. The authors showed that biofilms containing cells defective in the recombinatorial repair gene *recC* produced fewer antibiotic resistant variants compared to the wild type when low concentrations of gentamicin antibiotic was applied to the biofilm. This suggested that oxidative stress may be a factor which promotes antibiotic resistance in biofilms.

In 2009, Mandsberg *et al.* set out to test the hypothesis that mutations in the oxidative repair system correlated with an increased mutation frequency in *P. aeruginosa* biofilms. To test their hypothesis, the authors constructed *mut*T, *mut*Y, and *mut*M mutants in the *P. aeruginosa* strain PAO1 and tested for mutation frequency, levels of 8-oxodG after exposure to PMNs and antibiotic resistance in the mutants compared to wild type *P. aeruginosa* strain PAO1. Their results showed that mutations in the *mut*T and *mut*Y genes exhibited the largest increase in mutation frequency compared to the PAO1 wild type control, with a 28-fold and 7.5-fold increase respectively. The increased mutation frequency observed in the *mut*M was negligible and not considered significant. After exposure to PMNs the GO mutants exhibited higher levels of 7,8-dihydro-8-oxodeoxyguanosine compared to the wild type strain, and moreover resistance to antibiotics was more frequent. Increased resistance to beta-lactamase was attributed to an over-expression of the MexCD-OprJ efflux-pump. The authors concluded that oxidative stress may be a key factor in the development of antibiotic resistance in CF patients and that possible antioxidant-therapy would be beneficial (Mandsberg *et al*., 2009). Although data in this paper and its conclusions are sound, to date no GO system mutants have been reported from *P. aeruginosa* isolates extracted from CF patient sputum, so the relevance of this study to biofilms *in vitro* and *in vivo* remains unclear.

1.5.4 SOS system

The bacterial SOS response is a cell cycle check-point control in which the cell cycle is arrested and DNA repair and mutagenesis are induced (Gotoh *et al*., 2010). One of the central features of the SOS response is the de-repression of over 20 repressed genes under the control of the LexA repressor, which is responsible for the repression of the LexA regulon (McKenzie *et al*., 2000). The LexA regulon includes *rec*A, *rec*N, and *ruv*AB, (used for recombination and repair), uvrAB and uvrD (used in nucleotide excision repair) in addition to the error-prone DNA polymerase genes *din*B which encodes pol IV and *umu*DC which encodes pol V and DNA polymerase II (McKenzie *et al*., 2000). A functional SOS response is required to prevent damage due to DNA damaging agents (Gotoh *et al*., 2010). To date no LexA analogue has been identified in *S. pneumoniae* and it would appear that *S. pneumoniae* does not have an SOS response analogous to that characterised in *E. coli* (Cirz *et al*., 2006) suggesting that the variation seen in *S. pneumoniae* biofilms is a result of other mechanisms.

1.5.5 Transformation

Transformation is the uptake, incorporation and expression of exogenous DNA by competent bacterial strains. Transformation in *S. pneumoniae* is mediated by the extracellular concentration of an exported peptide pheromone called competence stimulating peptide (CSP) (Yang *et al*., 2010). A two-component regulatory system is responsible for the reception of CSP pheromones. CSP binds to the histidine kinase receptor, *com*D which in turn activates the response regulator, *com*E via phosphorylation. This induces the transcription of early *com* genes including the alternative sigma factor *com*X. *com*X induces the transcription of late *com* genes. This cascade of gene expression then facilitates the uptake and incorporation of exogenous DNA (Suntharalingam and Cvitkovitch, 2005). Interesting, however, DNA microarray analysis, performed by Petersen *et al.*, in 2004, revealed that in *S. pneumoniae,* 23 of 124 CSP-inducible genes are required for transformation whereas 67 were individually dispensable for transformation (Peterson *et al*., 2004) suggesting many CSP-inducible genes have alternative functions. Indeed the expression of CPS-induced genes includes proteins involved in the synthesis of bacteriocins and has been linked to a phenomenon known as *pneumococcal fratricide* (Havarstein *et al*., 2006). *Pneumococcal fratricide* refers to the phenomenon in which competent pneumococcal cells express bacteriocins and other killing factors which result in the lysis of non-competent pneumococci in the surrounding environment (Claverys *et al*., 2007). The fact that CPS mediates both *pneumococcal fratricide* and DNA uptake has led to the hypothesis that *pneumococcal fratricide* is both a method of out competing neighbouring cells as well as a method of attaining genetic material such as virulence factors from the resulting cell lysate (Claverys *et al*., 2007).

Interestingly research has shown that *S. pneumoniae* cells unable to produce CSP are also defective in biofilm formation (Oggioni *et al*., 2006). Moreover, Petersen *et al.* (Petersen et al., 2004) reported a 10-fold increase in competence in *S. intermedius* biofilm-associated cells compared to planktonic cells, suggesting that CSP is important for streptococci biofilm formation. Transformation is thought to provide a selective advantage to competent cells, allowing bacteria to acquire novel genes from donor cells (Suntharalingam and Cvitkovitch, 2005). Due to the close proximity of cells within a biofilm, transformation may be a key determinant in acquiring heterogeneity during biofilm development.

1.6 **DNA sequencing platforms**

Over the past 5 years NGS has revolutionised the way prokaryote genomes are sequenced; allowing detailed high-throughput data generation in a matter of days. The majority of work is currently performed on a large scale, however three bench-top sequencers have recently been released allowing rapid sequencing on a much smaller scale; these sequencers include the Roche 454 Genome Sequencer (GS) Junior™, the Illumina MiSeq™ and the Life Technologies Ion Torrent Personal Genome Machine™ (PGM) (Loman *et al*., 2012a). A recent article compared the ability of the three bench-top sequencers to sequence the Shiga-toxin-producing *E. coli* strain responsible for the 2011 outbreak in Germany (Loman *et al*., 2012b). This article reported that all three sequencers were capable of producing useful draft genomes but that neither sequencer was capable of accurately generating a complete genome. Furthermore the Miseq™ was found to have the highest throughput and lowest error rate, but short reads and long paired end sequencing time. The Roche 454 GS Junior™ had the lowest throughput but had the longest reads. The Ion PGM™ had the shortest run time and the fastest throughput per hour. Both the GS Junior™ and Ion PGM™ were prone to errors in homopolymeric regions. The authors concluded that taken together these data favoured the MiSeq™ as the more reliable sequencer (Loman *et al*., 2012b).

This section discusses the different platforms for the three bench-top next generation sequencers; Roche 454 GS Junior™, the Illumina MiSeq™ and the Life Technologies Ion Torrent PGM™ in addition to traditional Sanger sequencing.

1.6.1 Sanger sequencing

Prior to the advent of NGS, sequencing was performed using a method developed by pioneer Frederick Sanger in 1977. This method employs the chain-terminating dideoxynucleotide method of DNA sequencing (Sanger *et al*., 1977). Chain-terminating dideoxynucleotide (ddNTPs) are nucleotides without a 3'-hydroxyl group. Absence of this group prevents additional nucleotides from being added to the extended nucleotide chain. DNA templates are amplified via a polymerase chain reaction (PCR) or vector based method (Shendure *et al*., 2008). During template sequencing this results in DNA fragments of varying sizes, with each fragment ending in a ddNTP. Modern versions of Sanger sequencing uses a dye-terminator sequencing and capillary electrophoresis method of sequencing. This process involves the elongation of the template DNA with ddNTP where each ddNTPs (ddA, ddG, ddC or ddT) are labelled with a fluorescent dye which have the same excitation wavelength but emit different wavelengths of light, fluorescent energy resonance transfer is subsequently used to identify each base (Shendure *et al*., 2008). Capillary electrophoresis uses an electrophoresis field to separate the amplified DNA chain-terminated sequences based on their size to charge ratio, through a capillary, and the sequence is read by a fluorescence detector which records the correct base and displays the sequence on a chromatogram. In recent years, Sanger sequencing has been replaced by NGS which provides shorter DNA sequence reads in a relatively shorter time scale. Sanger sequencing still remains the 'Gold Standard' for DNA sequencing especially for sequencing reads greater than 500 nucleotides and is often used to verify novel mutations (Chan, 2009).

1.6.2 Illumina sequencing – MiSeq™

The MiSeq™ was released in late 2011 as a cheaper lower-throughput version of the Illumina HiSeq. The Illumina MiSeq™ sequencing workflow uses the Solexa chemistry and consists of library preparation, cluster generation and finally sequencing. Library preparation involves the fragmentation of prepared genomic DNA, the sheared ends are repaired and adenylated. Unique adaptor oligonucleotides are added to both ends of the fragment, fragments are then purified. Amplification of these fragments occurs via isothermic amplification on a flow cell in a process called 'cluster generation'. The flow cell contains a dense lawn of oligonucleotides bound to the surface which bind to adaptor nucleotides bound

to the fragments, which were added in the library preparation stage. The fragments hybridise to the oligonucleotides on the flow cell. The fragments are copied and covalently bound to the flow cell surface via 'bridge' amplification. Each fragment is clonally amplified. Reverse stands are cleaved and washed away. The fragment ends are blocked that the DNA sequencing primer is hybridised to the fragments which serve as DNA templates. The clusters are sequenced simultaneously; all four fluorescently labelled, reversely-terminated nucleotides are added simultaneously and compete to bind to the corresponding nucleotide on the DNA template. After each round of synthesis the nucleotides are excited by a laser releasing a unique colour depending on the nucleotide which is recorded by the MiSeq™. The fluorescent label and blocking group are removed allowing binding of another nucleotide in the next round of synthesis (information regarding this workflow can be found on the Illumina website http://www.illumina.com – Last accessed 09 August 2014).

1.6.3 Roche 454 sequencing – GS Junior™

The Roche GS Junior™, released in 2010 uses the 454 sequencing platform and was developed as a cheaper lower-throughput version of the 454 GS FLX+ sequencer. The 454 sequencing workflow consists of library preparation, emulsion-PCR and finally sequencing. Library preparation is similar to the MiSeq™ in the sense that the DNA is fragmented and adaptor fragments are added to both ends of the fragment to acts as templates for the sequencing primers and polymerase enzyme. Each single-stranded DNA library fragment is immobilised onto a DNA Capture Bead. The library is then clonally amplified in an emulsion-PCR. The fragments are loaded onto a picotiter plate for sequencing so that one well contains one bead and packing beads are added to each well for stability. Sequencing consists of the genome sequencer serially flowing single nucleotides across each of the wells. Once attached to the complementary region of the DNA template the nucleotides release a chemiluminescent signal which is recorded by a charge coupled device camera on the GS sequencer. As multiple nucleotides can bind in regions of homopolymers, the intensity of the signal represents the proportion of bases incorporated (information regarding this workflow can be found on the GS Junior website http://www.gsJunior.com/instrument-workflow.php – Last accessed 09 August 2014).

1.6.4 SOLiD – Ion Torrent PGM™

The Ion Torrent SOLiD PGM™ was released in early 2011. The sequencing workflow consists of enzyme and sample preparation, PCR and substrate preparation, ligation imaging and data analysis. Library preparation can either exist as a fragment library or a mate-paired library. Target DNA is sheared to a specific size and adaptors are ligated to the end of the fragments; a fragment library incorporated a single piece of DNA whereas a mate-paired library incorporates two pieces of DNA that are a known distance apart. Fragments are then clonally amplified onto beads in an emulsion-PCR reaction. The beads are enriched and covalently attached to a glass slide with 1, 4 or 8 samples on a single slide. Sequencing of the fragments involves the use of specific ligases and di-based probes fluorescently labelled with 4 dyes. Each represents four of 16 possible nucleotides. The complementary probe hybridises to the template sequence and is ligated; after fluorescence is recorded the dye is cleaved allowing a new probe to bind, the synthesis continues for 35 bases. The synthesised stand is then removed and the process is repeated with one base off set. This primer reset process is repeated five times. The system uses exact call chemistry to sequence the relevant nucleotides by using a four-different colour system to accurately map out all possible combination of sequence (information regarding this workflow can be found on the Ion Torrent website http://www.lifetechnologies.com/uk/en/home/brands/ion-torrent.html – Last accessed 09 August 2014).

1.6.5 Next generation sequencing applications

Genome sequencing has numerous applications with regards vaccine research and a recent paper highlights the ability to track vaccine (PCV7) escape variants from a subset of isolates, in the US (Golubchik *et al*., 2012). The authors identified five independent examples of capsular switching where serotype 4 was replaced by serotype 19A (referred to as P1-P5). Throughout the 5 samples they estimated up to 27 fragments (ranging from 0.04-44kb) were transferred from donor and recipient strains, supporting their theory that multiple fragments maybe transferred in a single recombination episode. The authors further reported that the genotype of escape variant P1 has been recovered from 175 individuals, by 2007, and has become extremely prevalent in post-vaccinated (PCV7) population across the US (Golubchik *et al*., 2012). This work shows the limitation of vaccine efficacy by highlighting the occurrence of capsular switching and the need for constant monitoring of serotype distribution. The colonisation of the nasopharynx provides a major reservoir for *S. pneumoniae*, facilitating recombination and subsequent transmission (Marks *et al*., 2012b). Considering colonisation is thought to involve the formation of biofilms the high mutation rates seen in biofilms may escalate the emergence of vaccine escape variants in the population.

Each platform has its merits and limitations (Loman *et al*., 2012b). The sequencing performed in this work will employ the Roche GS Junior™ and Illumina MiSeq™ sequencers. Roche GS Junior™ sequencer will provide longer contigs to generate overlapping reads and create the wild type reference sequence. Both paired end and shot gun sequencing will be employed to ensure a high coverage. The wild type reference will also be sequenced on the Illumina MiSeq™ to get accurate base calling and the two data sets will be combined. To gain accurate base calling with deeper coverage for SNP analysis all variant samples will be sequenced on the MiSeq™ using a pair-end protocol.

1.7 Project aims

The overall aim of this project is to assess the biofilm formation in clinically relevant strains of *S. pneumoniae* and furthermore assess, using conventional microbiology phenotyping and next generation WGS coupled with mass spectrometry to investigate relationships between the genomic, proteomic and morphological plasticity arising within subpopulations of bacteria during biofilm development. The underlying hypothesis of this work is that pneumococci grown under biofilm conditions exhibit phenotypic and genetic diversity compared to the parent strain which aid colonisation and persistence of the organism. The aims and key findings of each chapter are outlined below.

1.7.1 Chapter 3

Colonisation of the nasopharynx is thought to be mediated, in part, by biofilm formation (Marks *et al*., 2012a). This chapter aimed to characterise the biofilm forming properties of four clinical pneumococcal serotypes (serotypes 1 (ST306), 3, (ST3205), 14 (ST124) and 22F (ST433)), isolated from the third year of an on-going paediatric pneumococcal carriage study (October 2008 to March 2009) at University Hospital Southampton (UHS), to assess whether each serotype was capable of forming biofilms under static conditions and whether there was a significant difference in the biofilm forming ability of the various serotypes. Furthermore this chapter aimed to assess the phenomenon of colony morphology variance in clinical pneumococci. Previous work in pneumococcal biofilms has shown that colony morphology variation is a marker for genetic variation (Allegrucci and Sauer, 2007, Allegrucci and Sauer, 2008). This work will use stable colony morphology variation as a marker for genetic variation. As such this thesis will not use the opaque and transparent classification of pneumococcal variation (Overweg *et al*., 2000) as the objective of the project is to observe stable genetic changes within biofilms of *S. pneumoniae*, as opposed to non-phase variable changes. This chapter presents evidence of biofilm-derived variation as a common phenomenon among the four serotypes studied. Evidence of which has not previously been reported in these serotypes.

1.7.2 Chapter 4

This chapter aimed to characterise the biofilm-derived variants of serotype 22F using conventional microbiology phenotyping to assess the differences in phenotypes between the morphological variants and the parent strain. This chapter presents evidence that biofilm-derived SCVs exhibit reduced growth rates, reduced capsule expression, altered metabolic profiles and increased biofilm formation compared to the parent strain. Furthermore the use of API Rapid ID 32 Strep assay presents an original and effective use of determining phenotypic diversity among biofilm-derived variants.

1.7.3 Chapter 5

This chapter aimed to investigate relationships between the morphological plasticity observed in biofilm-derived colonies and the underlying genetic diversity. Through the use of novel next generation WGS, this chapter documents the repertoire of mutations observed in biofilm-derived colony variants and relate these mutations to the phenotypic diversity seen. The WGS of 12 SCVs from independent biofilm experiments revealed that all SCVs studied had mutations within the DNA-directed RNA polymerase delta subunit (*rpo*E). Mutations included four large-scale deletions ranging from 51-264 bp, one insertion resulting in a coding frameshift and seven nonsense single nucleotide substitutions that result in a truncated gene product. Furthermore this chapter presents evidence of an unprecedented level of parallel evolution in pneumococcal biofilm development which suggests this is an important driving force of mutation in pneumococcal biofilm development.

1.7.4 Chapter 6

Using conventional microbiology phenotyping and next generation WGS, this work observed that biofilm-derived small colony variants of *S. pneumoniae* exhibit increased biofilm formation, reduced capsule expression and slower growth phenotype which has been attributed to mutations within the DNA-directed RNA polymerase delta subunit (RpoE). The aim of this chapter was to understand the translational effect of the *rpo*E mutations by

employing UPLC/MS_E analysis. This chapter presents evidence of the consistent up-regulation of a common sub-set of stress-inducible proteins which are part of an interaction network consisting of the 60 kDa chaperonin, chaperone protein DnaK, cell division protein FtsZ and manganese superoxide dismutase and the down regulation of metabolic proteins. This chapter thus presents novel evidence of the role of *rpo*E in stress response pathways.

Chapter 2

Materials and Methods

2 Chapter 2: Materials and Methods

2.1 Bacterial strains and growth conditions

Clinical isolates of *S. pneumoniae* serotype 1 (ST306), 3, (ST3205), 14 (ST124) and 22F (ST433) were obtained from the third year (October 2009 - March 2010) of an on-going nasopharyngeal carriage study in children aged 4 years and under at University Hospital Southampton (UHS), UK. Isolates were stored at -80 ^{o}C in glycerol stock consisting of 50 % brain heart infusion (BHI) (Oxoid) and 50 % glycerol (Sigma). Capsular serotyping was performed using multiplex PCR (Pai *et al*., 2006). Multiplex PCR is the process whereby genomic regions of interest are amplified from pools of primer sequences in a single PCR experiment. Multiplex PCR was used to amplify capsular regions to define the serotype of the clinical isolates. Forward and reverse primers were used as previously described (Pai *et al*., 2006). Multi locus sequence typing (MLST) is the process of sequencing seven pneumococcal housekeeping genes to define clonal lineages; genes tested for the MLST of *S. pneumoniae* include: *aro*E (shikimate dehydrogenase), *gdh* (glucose-6-phosphate dehydrogenase) *gki* (glucose kinase) *rec*P (transketolase), *spi* (signal peptidase I), *xpt* (xanthine phosphoribosyltransferase) and *ddl* (D-alanine-D-alanine ligase) (Enright and Spratt, 1998). Each combination of alleles defines a particular pneumococcal clone or 'sequence type' (ST). MLST was performed using Qiagen Genomic Services and the MLST website www.mlst.net to assign sequence type prior to biofilm experimentation.

2.2 Growth curves

Frozen stocks of all isolates were streaked onto Columbia blood agar + 5 % v/v horse blood (Oxoid) and incubated overnight at 37 ^{o}C in 5 % CO_2. Colony growth was transferred using a sterile swab to 2 mL of culture medium (BHI, casein tryptone [CAT] or Todd-Hewitt with 5 % yeast extract [THY] broth). The suspension was centrifuged for 3 minutes at 918 x *g* and 1 mL was transfer to 10 mL of culture medium (1:10 dilution) before being incubated at 37 ^{o}C in 5 % CO_2. Growth curves were performed over an 8-10 hour time period. The optical density was measured at 600 nm (OD_{600} for BHI and THY), or 550 nm (OD_{550} for CAT media) hourly to establish what OD corresponded to the exponential phase of growth. A 100 µL aliquot of culture was taken every two hours to perform viable counts. Growth curves for each strain were performed in triplicate on separate days.

2.3 Colony forming unit analysis

Frozen stocks of all serotypes were streaked onto Columbia blood agar + 5 % v/v horse blood (Oxoid) and incubated at 37 ^{o}C in 5 % CO_2. Colony growth was transferred using a sterile swab to 2 mL of culture medium (BHI, CAT or THY broth). The suspension was centrifuged for 3 minutes at 918 x *g* and 1 mL was transfer to 9 mL of culture medium (1:10 dilution) before being incubated at 37 ^{o}C in 5 % CO_2. Using previous growth curves for reference, at 3 points during exponential growth 100 µL samples were taken and serially diluted to 10^{-6} in BHI. A 20 µL aliquot of the 10^{-2} to 10^{-6} dilutions were spotted 5 times on Columbia blood agar and incubated at 37 ^{o}C in 5 % CO_2 overnight in accordance with the Miles and Misra methodology (Miles *et al*., 1938). CFU counts in BHI medium were analysed using regression analysis to generate a calibration curve of OD_{600} vs. CFU. The ODs corresponding to 10^{8} cellsmL^{-1} are as follows, serotype 1 (OD_{600} 0.17), serotype 3 (OD_{600} 0.4), serotype 14 (OD_{600} 0.22) and serotype 22F (OD_{600} 0.4). All cultures were incubated at 37 ^{o}C in 5 % CO_2. CFU counts for each strain were performed in triplicate.

2.4 Biofilm culture and colony variation

Approximately 1 x 10^{8} CFU mL^{-1} of each serotype [serotype 1 (ST306), 3, (ST3205), 14 (ST124) and 22F (ST433)] was inoculated into 6-well tissue culture plates (Nunc) as previously described (Hall-Stoodley *et al*., 2008), for a total of 9 days. Per experiment, triplicate biofilms were grown under static conditions at 37 ^{o}C in 5 % CO_2 for 1, 3, 6 and 9 days. Half the culture medium was removed daily and replaced with equal quantities of fresh 1:5 BHI warmed to 37 ^{o}C. At each time point the biofilms were washed twice with fresh warmed 1:5 BHI and the biomass was harvested using a sterile cell scraper, vortexed, diluted to 10^{-6} and 100 µL was plated onto Columbia blood agar (Oxoid) to assess for changes in colony morphology compared to the inoculum. Figure 2-1 depicts the experimental set up. Independent biofilm experiments were performed in duplicate for each serotype tested.

2.5 Assessment of colony morphology

Colony morphology was assessed based on size and regularity of the colony surface: undulated vs. smooth colony surface. For serotypes 1 and 22F, variants were defined as follows: small colony variants (SCVs) (<0.5 mm), medium colony variants (MCVs) (0.5-<1 mm) and typical colony variants (TCVs) (≥1 mm). For serotypes 14 variants were defined as follows: SCVs (0.1-<0.3 mm), MCVs (0.3-<0.6 mm) and typical colony variants (TCVs) (≥0.6 mm). Due to the mucoid nature of serotype 3, colony variants were defined as; SCVs (<1 mm), MCVs (1-<2 mm) and TCVs (≥2 mm) in accordance with previous literature (Allegrucci and Sauer, 2007). Colonies

were visualised under a Leica MZ 16F stereomicroscope and images were taken using a Leica digital camera with microscope attachment with a 5 mm graticule scale. Using the graticule scale as a reference each colony was quantified using the ImageJ analysis software (http://rsbweb.nih.gov/ij/). Each colony was measured along two axes; centrally and along the cross-section of the colony, this method of measurement took into account slight irregularities in the colony morphology, an average was taken of the two diameters. A total of nine colony variants from separate triplicate biofilms were measured. Per experiment, biofilm colony variant phenotypes were harvested from triplicate biofilms. Where possible, up to 12 phenotypic variants of each morphology type were selected per time point (Figure 2-1). Variants were sub-cultured and stored at -80 ^{o}C for future phenotypic and genetic analysis. This was done to ensure that variants were selected from different biofilms and avoid selected clones from the same biofilm.

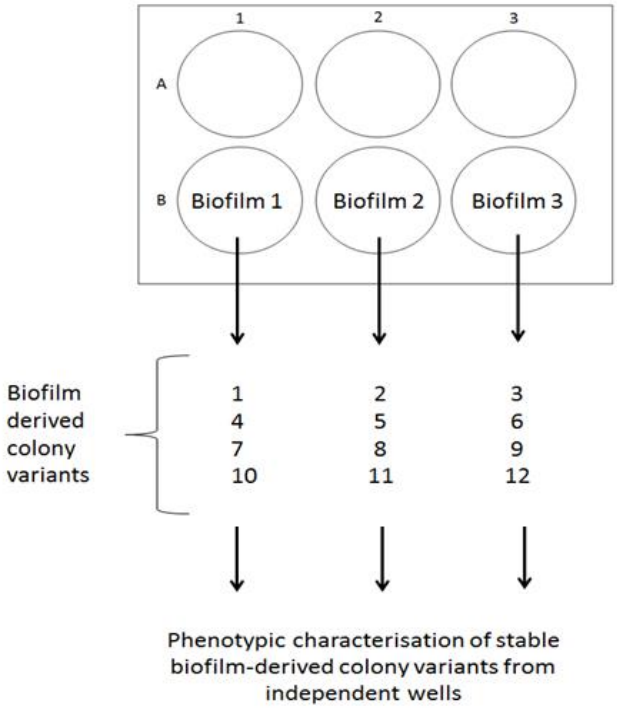


Figure 2-1: Representation of colony harvesting method

Biofilms were grown under static conditions at 37 ^{o}C in 5 % CO_2 for 1, 3, 6 and 9 days. At each time point the biomass was harvested using a sterile cell scraper, vortexed, diluted to 10^{-6} and 100 µL was plated onto Columbia blood agar to assess for changes in colony morphology compared to the inoculum. Per experiment, biofilm colony variant phenotypes were harvested from triplicate biofilms. Up to 12 phenotypic variants of each morphology type were selected per time point. Variants were sub-cultured and stored at -80 ^{o}C for future phenotypic and genetic analysis.

2.6 Visualisation of the biofilm

Biofilm formation was observed using a Nikon Eclipse LV100 episcopic differential interference contrast microscopy/epi-fluorescent (EDIC/EF) microscope and the *Bac*Light live/dead stain (Invitrogen). Biofilms were cultured in MatTek culture plates (MatTek Corporation Ashland, MA) under static biofilm conditions 37 ^{o}C in 5 % CO_2 for 1, 3, 6 and 9 days and stained according to the manufacturer's instructions. Biofilms were washed twice with fresh warmed 1:5 strength BHI. Biofilms were stained with SYTO®9 and propidium iodide, 2 µL per 1 mL of Hank's balanced salt solution (HBSS). SYTO®9 is a green-fluorescent nucleic acid stain and stains viable and non-viable cells by penetrating both intact and damaged cell membranes. Propidium iodide is red fluorescent nucleic acid and stains cells with damaged cell membrane; incorporation of propidium iodide displaces the bound SYTO®9. Therefore, those cells that fluoresce green are referred to as live viable cells whereas cells that fluoresce red are referred to as non-viable dead cells. A minimum of ten fields of view were taken for triplicate biofilms to get an accurate indication of biofilm growth. Propidium iodide and SYTO®9 images were merged to determine the overall cell surface coverage of the biofilm. All images were taken under epi-fluorescence filters at 400x magnification with the Image-Pro 6.2 software (Media Cybernetics) and analysed using ImageJ analysis software (http://rsbweb.nih.gov/ij/). All images underwent deconvolution using the No Neighbor algorithm incorporated into the Image-Pro 6.2 software prior to percentage surface quantification. Ten fields of view were taken for triplicate biofilms to get a reliable indication of biofilm growth. Biofilm formation was also observed using CLSM and the *Bac*Light live/dead stain (Invitrogen) for the wild type (WT) 22F and a biofilm-derived SCV isolate. All images were taken using a 63x objective on a Leica TCS SP5 confocal laser scanning microscope on a Leica DMI6000 inverted microscope frame. Five fields of view were taken per triplicate biofilm to get an accurate indication of biofilm growth. Z-Scans were performed every 0.3 µm on each field of view. Analysis of the confocal images was performed using COMSTAT 1 (Heydorn *et al*., 2000). Threshold levels for each image stack were determined using the COMSTAT program looktif.m; these values were inputted into the comstat.m program to generate data on biomass, maximum thickness average thickness, surface area and surface area to volume ratio.

2.7 Crystal violet initial attachment assay

Crystal violet assay was performed as previously described (Hall-Stoodley *et al*., 2008). Isolates were cultured in BHI to an OD of ~1 x 10^8 CFU mL^{-1}. A 1:10 dilution into 1:5 BHI was added to each well of a 48 well micro-titre plates (Nunc) (100 µL of culture into 900 µL of 1:5 BHI – thus creating an inoculum of ~1 x 10^7 CFU mL^{-1}) and cells were allowed to adhere for 24 hours. The

biofilms were subsequently washed twice with 1:5 BHI and stained with a 1 % crystal violet solution (Pro-Lab Diagnostics) for 15 minutes. The stain was removed and the biofilms were washed three times with distilled water and 1 mL of 95 % ethanol was added to each well to solubilize the stain. The absorbance was read on the FLUOstar Optima plate reader at 560 nm. Six replicates for each serotype and negative control were performed per plate and averages were taken. Readings were subtracted from the negative control sample average to give a final absorbance value. A total of 3 independent experiments were performed using three independent inocula on separate days. Viable counts of all inocula were performed.

2.8 API Rapid ID 32 Strep assay

The API Rapid ID 32 Strep assay (bioMérieux) was performed according to the manufacturer's instructions. Isolates were streaked onto Columbia blood agar + 5 % v/v horse blood (Oxoid) and incubated for 18 hours at 37 ^{o}C in 5 % CO_2. Using a sterile swab the growth was removed from the plate into a 2 mL ampule of API suspension medium and the turbidity was adjusted to McFarland No. 4. A 55 µL aliquot of the suspension was added to each well of the API Rapid ID 32 Strep strip and incubated for 4 hours at 37 ^{o}C. Results were recorded visually using the table of identification in the manufacturer's protocol booklet. The chemical reactions in the API strip are divided into blocks of three with each well designated a score of 1, 2, or 4 respectively. A positive reaction for all three reactions in a block would score a maximum of 7 (1 +2 +4 = 7). Each block is scored to generate an 11-digit code (see Figure 8-1 and Table 8-1 in Appendix 1). This code was entered into the online database, API*web* (https://apiweb.biomerieux.com) to verify the identity of the organism.

2.9 DNA extraction for polymerase chain reaction

S. pneumoniae DNA was extracted using the QIAamp DNA mini kit (QIAGEN) according to the manufacturer's instructions with minor modifications. Frozen stocks of each isolate were streaked onto Columbia blood agar + 5 % v/v horse blood (Oxoid) and incubated at 37 ^{o}C in 5 % CO_2. Colony growth was transferred using a sterile swab to 200 µl of lysis buffer (10 mM Tris pH8, 100 mM ethylenediaminetetraacetic acid pH8, 0.5% w/v sodium dodecyl sulphate) and incubated at 37 ^{o}C for 1 hour. 20 µl proteinase K was added and samples were incubated at 56 ^{o}C for 1 hour. The following steps were performed as described in the Qiagen QIAamp mini kit protocol; 200 µl Buffer AL was added and samples were incubated at 70 ^{o}C for 10 minutes. 200 µl of ethanol was added to the samples and centrifuged briefly. The total volume was transferred to a spin column and centrifuged at 2500 x *g* for 1 minute, the filtrate was discarded. 500 µl of buffer AW1 was added to the column and centrifuged at 2500 x *g* for

1 minute, the filtrate was discarded. 500 µl of buffer AW2 was added to the column and centrifuged at 10000 x *g* for 3 minutes, the filtrate was discarded. The column was centrifuged at 10000 x *g* for 3 minutes again to dry the column. The column was placed over a fresh microfuge tube and 200 µl of ultrapure water was added. The column was centrifuged at 2500 x *g* for 1 minute to elute the DNA.

2.10 Polymerase chain reaction of serotype 22F

To confirm all biofilm-derived variants were *S. pneumoniae* and assess changes in serotype from the serotype 22F parent, genomic DNA from each variant underwent PCR for the 22F gene at 643 basepairs (bp) and the pneumococcal *cps*A gene at 160 bp. Forward and reverse primers were used as previously described (Pai *et al*., 2006): 22F-forward 5'-GAGTATAGCCAGATTATGGCAGTTTTATTGTC-3', 22F-reverse 5'-CTCCAGCACTTGCGCTGGAAACAACAGACAAC-3, *cps*A-forward 5'-GCAGTACAGCAGTTTGTTGGACTGACC-3' *cps*A-reverse 5'-GAATATTTTCATTATCAGTCCCAGTC-3'. Forward and reverse primers were diluted to 100 µM. PCR reaction mix consisted of MyTaq™ Red Mix x2 concentration (Bioline), magnesium chloride (2 µM), forward and reverse primers (0.5 µM) and the total volume was made up 12.5 µL with UltraPure™ DNase/RNase-free distilled water (Life Technologies). The PCR program consisted of three stages: stage 1: one cycle of 94 ^{o}C for 4 minutes; stage 2: 35 cycles of 94 ^{o}C for 45 seconds (melting), 54 ^{o}C for 45 seconds (primer binding) and 72 ^{o}C for 2 minutes and 30 seconds (elongation); Stage 3: 10 ^{o}C. PCR products were run out on 2 % w/v agar gel for 1.5 hours at 125 V, stained with GelRed™ (Biotium) and visualised under Ultraviolet light. Bioline hyperladder™ IV 100-1000 bp was used to assess size of the bands.

2.11 Slide agglutination

Frozen stocks of all isolates were streaked onto Columbia blood agar + 5 % horse blood (Oxoid) and incubated at 37 ^{o}C in 5 % CO_2 for 18 hours. Half the colony growth was transferred using a sterile swab to 5 mL of Todd-Hewitt broth (THB) and cultured overnight at 37 ^{o}C in 5 % CO_2. The overnight liquid culture was centrifuged for 30 minutes at 5000 x *g* on an Eppendorf 5804R centrifuge to pellet the cells. The supernatant was discarded and the pellet was re-suspended in 200 µL of THB by vortexing each pellet. This suspension was used to perform the slide agglutination assay; 22F antiserum (factor serum 22b) was obtained from Statens Serum institute (Denmark). A 10 µL aliquot of each variant suspension were pipetted onto a glass slide and 10 µL of 22F antisera subsequently added. As a negative control, 22A antiserum (factor serum 22c) was tested in parallel with 22F. The glass slide was gently rocked for 30

seconds to mix reagent until agglutination (clumping with clearing background) of the suspension was recorded.

2.12 Gram staining

A single well-isolated colony was mixed with distilled water on a microscope slide, allowed to air dry and the cells were subsequently heat fixed by passing the underside of the slide though a gentle flame. Cells were stained with Gram's crystal violet stain (Sigma-Aldrich) for one minute, washed with distilled water and stained with Gram's iodine solution (Sigma-Aldrich). The stain was decolourised with 95 % ethanol and subsequently washed with distilled water. Slides were counterstained with Gram's safranin solution (Sigma-Aldrich) washed with distilled water and allowed to air dry. Slides were examined under brightfield microscopy using a 100x objective on an Olympus BX51 microscope.

2.13 Antibiotic susceptibility

The minimum inhibitory concentration (M.I.C) of antibiotics effective against the phenotypic variants was tested using E-test strips (Oxoid). Four antibiotics, commonly used in pneumococcal treatment were selected [erythromycin, tetracycline, ceftriaxone, and penicillin G]. The M.I.C assay was performed by spread-plate culture inhibition: a lawn of each biofilm variant was plated onto Columbia blood agar + 5 % v/v horse blood (Oxoid). Triplicate plates were left to dry for 5 minutes. A single antibiotic E-test strip applied to the centre of the agar plate using sterile forceps. Plates were incubated for 18 hours at 37 ^{o}C in 5 % CO_2. The M.I.C was measured by reading the concentration on the E-test strip at the maximum point of the zone of inhibition.

2.14 Capsule staining

Quantification of the capsule was adapted and performed according to Hammerschidt *et al.* (2005) and Rukke *et al.* (2012). Variants were grown on Columbia blood agar + 5 % v/v horse blood (Oxoid) overnight at 37 ^{o}C and 5 % CO_2. The growth was transferred to 10 mL of BHI and cultured to an OD equating to approximately 10^8 CFU mL^{-1} (OD_{600} ~ 0.3-4). Five mL cultures were centrifuged at 2500 x *g* for 10 minutes on a Heraeus megafuge 1.0 to pellet the cells. The supernatant was discarded and the pellet was washed twice with 0.5 mL PBS, before being re-suspended into 5 mL of sterile distilled water. To stain for the capsule, 250 µL of re-suspended cells was transferred to a clean microfuge tube and 1 mL of Stain-all solution [20 mg of Stains-all (1-Ethyl-2-[3-(1-ethylnaphtho[1,2-d]thiazolin-2-ylidene)-2-methylpropenyl]naphtho[1,2-d]thiazolium bromide, 3,3′ -Diethyl-9-methyl-4,5,4′ ,5′ -dibenzothiacarbocyanine) (Sigma-Aldrich), 60 µL of glacial acetic acid in 100 mL of 50 % formamide] was added. The absorbance

was measured at 640 nm and subtracted from the negative control (250 µL of sterile distilled water stained with 1 mL of Stains-all).

2.15 Rifampicin mutation frequency

The rifampicin assay was performed as previous described (Lee *et al*., 2010). Isolates were streaked onto Columbia blood agar (Oxoid) from frozen -80 ^{o}C glycerol stocks and incubated overnight at 37 ^{o}C in 5 % CO_2 for 18 hours. Three to five colonies from this overnight culture were re-suspended in 10 mL of BHI broth and incubated for 6–8 hours at 37 ^{o}C in 5 % CO_2, until the cell density reached an OD_{600} ~0.1 (~1 x 10^{10} CFU mL^{-1}). A 100 µl aliquot was diluted for total viable counts. The remaining culture was centrifuged for 10 minutes at 2500 x *g* on a Heraeus megafuge 1.0 as previously described (Lee *et al*., 2010). The pellet was re-suspended in 500 µL of BHI broth and plated onto triplicate blood agar plates containing 5 % v/v defibrinated horse blood and 2 µg mL^{-1} rifampicin (Sigma-Aldrich). The assay was performed in triplicate for each sample, using independent bacterial cultures. The mutation frequency values were reported as the proportion of rifampicin-resistant colonies (after 48 hours of incubation in a 5 % CO_2 atmosphere) compared to the total viable cell counts. A mutator strain refers to an isolate with a mutation frequency ≥7.5 × 10^{-8} (Morosini *et al*., 2003).

2.16 Genomic DNA extraction

Genomic DNA was extracted using the Genomic tip 100/G protocol (Qiagen). Frozen stocks of all isolates were streaked onto Columbia blood agar + 5 % horse blood (Oxoid) and incubated overnight at 37 ^{o}C in 5 % CO_2. A single colony was picked and re-streaked onto CBA and incubated overnight at 37 ^{o}C in 5 % CO_2. The resulting colonies were cultured in BHI and incubated at 37 ^{o}C in 5 % CO_2 to approximately 1 x 10^{9} CFU mL^{-1}, and centrifuged to create a pellet. The pellet was re-suspended and DNA was extracted according to the Genomic tip 100/G protocol (Qiagen). The pellet was re-suspended in 3.5 mL of Buffer B1 (with 7 µl of 100 mg/mL RNase A) in a 15 mL centrifugation tube and mixed thoroughly by vortexing to homogenise the suspension. 80 µl of lysozyme (100 mg/mL) and 100µl Qiagen proteinase K stock solution was added to the sample and incubated at 37 ^{o}C for 1 hour and 30 minutes. 1.2 mL of Buffer B2 was added incubate at 50 ^{o}C for 30 minutes. For each sample, a Qiagen Genomic-tip 100/G column was placed over an open 50 mL centrifugation tube and the column was equilibrated with 4 mL of Buffer QBT and allowed to empty by gravity flow. Each sample was added to the column and allowed to empty by gravity flow. The column was washed twice with 7.5 mL of Buffer QC. The column was placed over a clean 10mL centrifugation tube, and the DNA was eluted with 5 mL of pre-warmed (50 ^{o}C) Buffer QF. DNA

was precipitated by adding 3.5 mL (0.7 volumes) room-temperature isopropanol to the eluted DNA. Precipitated DNA was recovered by inverting the tube 10 to 20 times and centrifuged immediately at 5000 x *g* on an Eppendorf 5804R centrifuge. Supernatant was discarded. The DNA pellet was washed with 4 mL of 70% ethanol centrifuged immediately at 5000 x *g*. Supernatant was discarded. Pellet was air dried and re-suspended in ultra-pure water. DNA was transferred to a microfuge tube and stored at -20 ^{o}C.

2.17 Whole genome sequencing

Biofilm-derived variants and the parent wild type 22F strain underwent WGS using the Roche GS Junior™ 454 sequencer or the Illumina MiSeq™. The parent 22F strain was sequenced both as shot gun and paired-end on the Roche GS Junior™ and on the Illumina MiSeq™ using the 150 base, paired-end protocol. Variant samples were sequenced on the Illumina MiSeq™ only using the 150 or 250 base, paired-end protocol. For Roche GS Junior™ 454 paired-end sequencing, genomic DNA was sheared and GS Titanium Library Paired End Adaptors were added to the fragments, emPCR was performed using the Roche GS Junior™ Titanium emPCR Kit (Lib L) and samples were loaded onto the Roche GS Junior™ PicoTiterPlate according to the manufacturer's instructions and the Roche GS Junior™ Titanium Sequencing Kit was used to sequence the samples on the Roche GS Junior™. For Illumina MiSeq™ sequencing, all samples were sheared and libraries were prepared using the Nextera library prep kit. Illumina MiSeq™ reagent cartridge and sequencing reagents were used according to the manufacturer's instructions. Paired-end sequencing of the wild type strain acted as the template for comparing the biofilm-derived variants. Newbler Metrics Results recorded this template to have 86x coverage across the genome. Illumina sequence data were assembled *de novo* using Velvet and the Velvet optimiser script to determine the K-mer value required to achieve 20x coverage of the whole genome (average coverage 49x). Assemblies were assessed for quality using the assemblathon script (Earl *et al*., 2011); average contig length was 341 bp and satisfactory n50 scores were generated (average n50 = 36272). Selected assembly metrics data can be seen in Table 8-2, in Appendix 2. For the parent 22F strain, Roche 454 sequence data was assembled *de novo* using Newbler GSassembler and Illumina sequence data was combined to generate a more comprehensive reference sequence. Newbler Metrics results recorded the reference to have 86x coverage across the genome. The parent strain *de novo* assembled scaffolds were aligned against strain D39 reference in Mauve (Darling *et al*., 2010), in order to arrange the scaffold. D39 (serotype 2) was selected due to the relative relatedness to serotype 22F (Mavroidi *et al*., 2007). All gaps were removed to create a pseudoreference, all annotations were based using this pseudoreference. Annotation of the parent strain *de novo*

assembly was achieved using RAST (Rapid Annotation using Subsystem Technology) the online annotation service (Overbeek *et al*., 2014). Sequencing, assembly and annotation of the parent strain were performed courtesy of Professor Saheer Gharbia's group at Public Health England, Colindale (Department of Bioanalysis and Horizon Technologies). *De novo* assembly of the biofilm-derived variant sequences and assembly metrics were performed by the author.

2.18 *in silico* serotyping and multi-locus sequence typing (MLST) of biofilm-derived variants

All sequences were assembled *de novo* as described above. Stereotyping of the pneumococcal isolates was achieved using assembled contigs and an in-house *in silico* PCR script designed in conjunction with the Sanger Institute Cambridge. The *in silico* primers used can be found in Table 8-3 in Appendix 3. To determine the sequence type, the assembled contigs were uploaded to the Centre for Genomic Epidemiology online service (Larsen *et al*., 2012). Alternatively an additional method of typing was employed using the short read sequence typing script (SRST2) (http://katholt.github.io/srst2/) an updated version of the previous short read sequence typing script (Inouye *et al*., 2012).

2.19 Mutation analysis

In order to identify mutations in the biofilm-derived variants, sequence data for each isolate were mapped against the annotated parent strain reference. Illumina fastq reads were mapped against the parent 22F strain sequence using STAMPY (Lunter and Goodson, 2011). The output sequence alignment mapping (SAM) files were converted to binary alignment mapping (BAM) files using SAMtools and variant discovery was performed using SAMtools mpileup, bcftools and gatk annotator (Li *et al*., 2009, McKenna *et al*., 2010) to generate variant call files (.vcf) containing the list of SNPs present and the position within the genome. Variants were filtered based on quality score (q-value), SNPS with q-values <20 were rejected. Alternatively variant call files (.vcf) were generated using SNPtree1.1 (Leekitcharoenphon *et al*., 2012) and manually curated. Initial SNP calling and filtering was performed courtesy of Dr. Raju Misra at Public Health England, Colindale (Department of Bioanalysis and Horizon Technologies). Each SNP was manually curated by mapping the Illumina fastq files against the WT reference using Bowtie2 and SAMtools to generate bam files (.bam) and indexed bam files (.bai). Bam files were subsequently visualised using Integrative Genomics Viewer (Robinson *et al*., 2011, Thorvaldsdottir *et al*., 2013) to determine coverage for each variant. Furthermore the percentage confidence of each mutation was calculated as the coverage of the identified variant base call divided by the total base call coverage at the variant position. Mutations underwent a rigorous selection criteria to avoid mis-calling; only variant positions that were

present in both in the mapping data and *de novo* assemblies and had >20x coverage and a percentage confidence >90% were classed as confidence mutations. Functional effects of the mutations were determined using the ExPASy research portal (Artimo *et al*., 2012) and the Ensembl SMART database (Schultz *et al*., 1998, Letunic *et al*., 2012).

2.20 Polymerase chain reaction of *rpo*E

To confirm the validity of the *rpo*E deletions in the SCVs, genomic DNA from each SCV underwent PCR for the *rpo*E gene. Forward primer 5'-GAGGAGAAACGCTTTGGAATTAGAAG-3' and reverse primer 5'-GCTAACTCTTATTCCTCGCTGGTTTC-3' were designed. All primers were run on PrimerBLAST to confirm specificity to the *S. pneumoniae rpo*E gene prior to PCR. PCR program consisted of three stages; stage 1: one cycle of 94 ^{o}C for 4 minutes, stage 2: 35 cycles of 94 ^{o}C for 45 seconds, 52 ^{o}C for 45 seconds and 72 ^{o}C for 2 minutes and 30 seconds, Stage 3: 10 ^{o}C pause. PCR products were run out on 2 % agar gel for 1.5 hours at 125 V, stained with GelRed^{TM} and visualised under ultraviolet light. Bioline hyperladder^{TM} IV was used to assess size of the bands. The full PCR product revealed a band at 609 bp.

2.21 Phylogenetic analysis of *rpo*E alleles in carriage isolates

To assess the functional relation of the 22 *rpo*E alleles identified in carriage isolates, the amino acid (aa) sequence of the *rpo*E alleles were assessed phylogenetically using the online resource phylogeny.fr (http://phylogeny.lirmm.fr/phylo_cgi/index.cgi) (Dereeper *et al*., 2008). Sequences were aligned using MUSCLE and the alignment was curated using Gblocks (Castresana, 2000). The phylogenetic tree was reconstructed using PhyML and maximum likelihood method using the Blosum62 substitution model. The tree was constructed using TreeDyn (Chevenet *et al*., 2006) and rooted against the *Streptococcus mitis rpo*E gene.

2.22 Pneumococcal proteome extraction

All isolates were cultured in 40 mL of BHI at 37 ^{o}C in 5 % CO_2 to an OD of ~ 10^8 CFU mL^{-1} (~OD 0.3-0.4). Viable counts were taken to verify equal CFU mL^{-1} and purity of culture. Cultures were centrifuged at 3200 x *g* for 5 minutes to pellet the cells and the pellet was washed twice with 1 mL of 100 mM triethylammonium bicarbonate (TEAB) buffer (Sigma- Aldrich). All samples were suspended in 1 mL of 100 mM TEAB + 4 M guanidine hydrochloride extraction buffer and lysed mechanically twice at 50x for 5 minutes on the Qiagen TissueLyser LT, using MP lysis matrix B tubes (containing 0.1 mm silica spheres); samples were incubated on ice for 30 seconds between lysing sessions. Samples were visualised a under light microscope at 1000x magnification, to verify efficiency of mechanical lysis. Samples were transferred to sterile microfuge tubes and centrifuged at 918 x *g* for 3 minutes to pellet cells (where the

pellet contains the insoluble fraction and the supernatant contains the soluble fraction), samples were filter sterilized through a 0.22 μm membrane to remove any live cells or residual silica spheres. The supernatant was removed and precipitated overnight at -20 ^{o}C in 6x volumes of 100 % ethanol. Samples were subsequently centrifuged at 5000 x *g* for 15 minutes at 4 ^{o}C. The pellet was washed once with 100 % ethanol at 5000 x *g* for 5 minutes and re-suspended in an equal volume of 250 μL of 100 mM TEAB and stored at -20 ^{o}C. Protein samples from each isolate were extracted in triplicate on separate days.

2.23 Bradford's assay

Protein concentration was determined as described by Marion Bradford (Bradford, 1976). Twenty (20) μL of each protein sample was added to a 96-well plate and 180 μL of Coomassie Brilliant Blue was added to each sample. To quantify the protein, 20 μL of triplicate BSA standards (0.03125 - 2 mg mL^{-1}) were stained in unison to generate a Bradford's assay regression curve. Triplicate blanks of 100 mM TEAB were used. Samples were stained for 5 minutes before reading the optical density (OD) at 595 nm on Biochrom EZ Read 400 microplate reader. Average OD readings were subtracted from the blank well average and protein concentration was determined using a regression curve equation.

2.24 1-Dimensional gel

Thirty (30) μL of each protein sample was added to 15 μL of Nu-PAGE LDS sample buffer (4x) and 6 μL of Nu-PAGE reducing agent and incubated at 70 ^{o}C for 10 minutes before 15 μg of protein was loaded onto a standard 12 % Bis-tris Nu-PAGE pre-cast polyacrylamide gel in MOPS buffer with 10 μL Novex® Sharp Pre-stained Protein Standards to determine band weight. The gel was run for 45 minutes at 200 V in an Invitrogen Novex™ mini-cell tank and subsequently stained with SafeBLUE Protein Stain (Invitrogen) for two hours and de-stained in distilled water overnight.

2.25 UPLC/MS$_E$ mass spectrometry

All protein samples were normalised to the lowest sample concentration value in 300 μL of 100 mM TEAB + 0.5 % RapiGest and sent to the University of Southampton Centre for Proteomic Research for preparation and label-free UPLC/MS$_E$ Mass spectrometry using the Waters SYNAPT®G2-S HDMS™ system. Enolase 1 from *Saccharomyces cerevisiae* was used as an internal standard. Three biological replicates of each sample tested were analysed and each biological replicate was sequenced in three technical replicates. Prior to analysis all protein concentrations were normalised by dividing the protein concentration by the total protein concentration detected in each technical replicate. The exclusion criteria were as follows:

proteins present within a single MS technical replicate were excluded from analysis. Proteins were further excluded if present in only one biological replicate. Mass spectrometry data were subsequently analysed in Microsoft Excel for increased and decreased fold changes relative to the WT 22F parent strain. Protein function was determined using UniProt (http://www.uniprot.org/), Gene ontology (http://www.geneontology.org/) and the STRING 9.1 online database (http://string-db.org/) (Jensen *et al*., 2009).

2.26 Statistical analysis

All Statistical analysis was performed on Minitab™ 16.2 Statistical Software (www. minitab.com). Normality of the data was determined using the Anderson-darling test and equality of variance was determined using the F-test. Where needed normality and equal variance was achieved after square root transformation of the datasets. Unless stated otherwise one-way ANOVA was used to determine a significant difference between serotypes. Alternatively a 2-sample t-test was used to determine a significant difference between colony variants datasets and the corresponding wild type parent strain datasets. Correlation analysis was performed using the Pearson product-moment correlation coefficient analysis.

Chapter 3

Characterisation of biofilms formed by carriage isolates of *Streptococcus pneumoniae*

3 Chapter 3: Characterisation of biofilms formed by carriage isolates of *Streptococcus pneumoniae*

3.1 Introduction

S. pneumoniae is a causative agent of meningitis, sepsis, pneumonia and otitis media, and remains a major cause of morbidity and mortality worldwide (O'Brien *et al*., 2009). Pneumococcal carriage is known to precede pneumococcal disease (Simell *et al*., 2012) and it is generally considered that growth of *S. pneumoniae* in the human nasopharynx is a biofilm-mediated phenomenon (Marks *et al*., 2012b). Moreover pneumococcal biofilm formation has been directly observed in pneumococcal infections (Hall-Stoodley *et al*., 2006). Respiratory bacteria grown under biofilm conditions can result in rapid genetic diversification and the emergence of colony morphology variants that can exhibit increased adhesion, increased dispersal, and increased recalcitrance to oxidative stress and antibiotics (Boles *et al*., 2004). *S pneumoniae* is a highly diverse organism, comprising of over 90 different serotypes. Previously pneumococcal biofilm studies have focused predominantly on serotype 3 (Waite *et al*., 2001, Allegrucci and Sauer, 2007, McEllistrem *et al*., 2007, Allegrucci and Sauer, 2008, Domenech *et al*., 2009). This work will investigate the level of diversity that can arise during biofilm growth in additional *S. pneumoniae* strains to better understand the level of diversity that may arise during pneumococcal colonisation.

This chapter aims to characterise the biofilm forming ability of four different pneumococcal serotypes of clinical relevance, to assess whether each serotype was capable of forming biofilms under static conditions and whether there was a significant difference in the biofilm forming ability of the various serotypes. The serotypes of interest include serotypes 1 (ST306), 3, (ST3205), 14 (ST124) and 22F (ST433). All isolates were taken from the third year of an on-going paediatric pneumococcal carriage study (October 2008 to March 2009) at UHS. The rationale for choosing this year is three-fold; firstly, all isolates of interest were present in this carriage year, thus standardising the time in -80 ^{o}C storage. Secondly, isolates extracted via the nasal swabs are likely to have formed biofilms in the nasopharynx and so should provide adequate biofilm models. Finally all serotypes have relevance to both carriage and/or disease; serotype 1, specifically ST306 (Normark *et al*., 2001, Kirkham *et al*., 2006) is a virulent clone which increased in prevalence leading up to the introduction of PCV7 (Jefferies *et al*., 2010). Serotype 3 is a well-studied model serotype for biofilm formation and will provide a useful

biofilm-comparison (Allegrucci and Sauer, 2007). Prior to PCV7 serotype 14, specifically ST124, remained a virulent clone in IPD and carriage (Clarke *et al*., 2006) however has declined in recent years since implementation of the conjugate vaccine (Jefferies *et al*., 2010). Serotype 14 strains have also been shown to grow consistently as biofilms *in vitro* (Tapiainen *et al*., 2010) and as such are a useful model to use. Serotype 22F has been seen to increase in carriage and pneumococcal disease in the USA and the UK since the introduction of PCV7 in 2000 (Jacobs *et al*., 2008, Tocheva *et al*., 2011).

3.2 Materials and Methods

See sections 2.1 to 2.7 and 2.26 for more detail.

3.3 Results

3.3.1 Standardisation of inoculum

To standardise the number of cells inoculated into the biofilm, growth curves and colony forming unit regression curves were performed. Cultures of all isolates were incubated at 37 ^{o}C in 5 % CO_2 in BHI over a 6-8 hour time period as described in materials and methods sections 2.2 and 2.3. Growth curves for each strain were performed in triplicate on separate days (Figure 3-1). All serotypes exhibited similar growth curves with comparable lag phases and exponential phases. Using these growth curves as a reference, at three points during exponential growth 100 µL samples were taken and serially diluted to 10^{-8} in BHI. Twenty (20) µL of the 10^{-4} to 10^{-8} dilutions were spotted on Columbia blood agar and incubated at 37 ^{o}C in 5 % CO_2 overnight. CFU counts were analysed using regression analysis to generate a calibration curve of OD_{600} vs. CFU as depicted in Figure 3-2. The ODs corresponding to 10^{8} CFU mL^{-1} are as follows, serotype 1 (OD_{600} 0.17), serotype 3 (OD_{600} 0.4), serotype 14 (OD_{600} 0.22) and serotype 22F (OD_{600} 0.45). CFU counts were performed in 3 independent experiments. These OD_{600} values were used to seed all subsequent biofilm experiments.

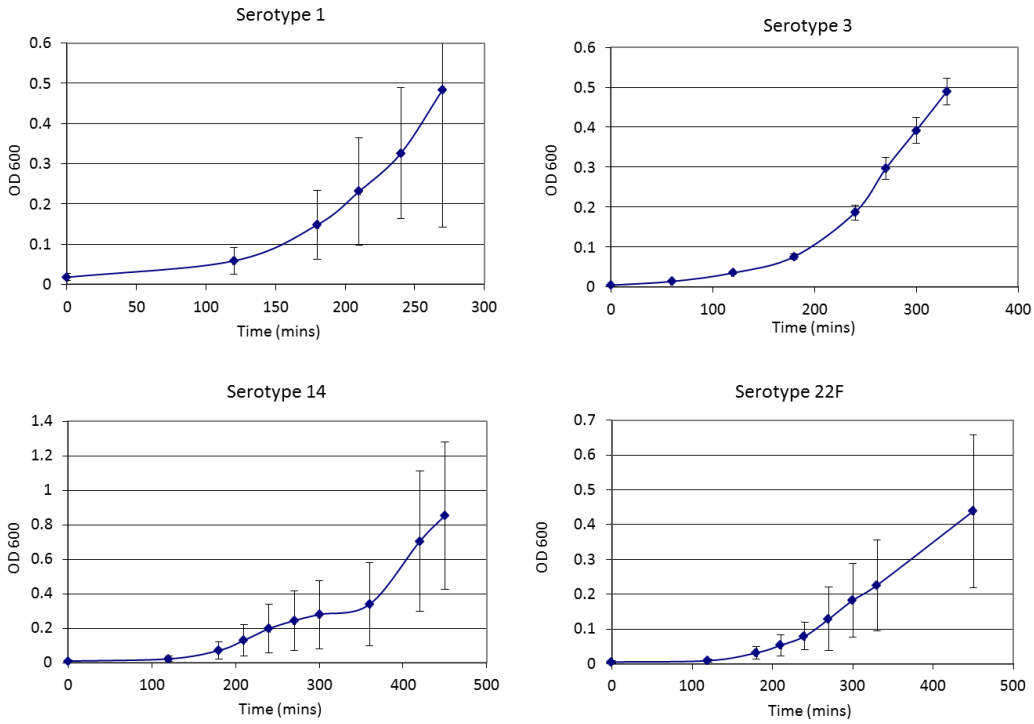


Figure 3-1: Growth curve of the pneumococcal serotypes

Growth curves of pneumococcal serotypes grown in BHI broth for a total of 6-8 hours. OD_{600} readings and were taken every hour in order to establish what OD corresponded to the exponential phase of growth. Growth curves for each serotype were performed in triplicate on separate days. Error bars represent standard error.

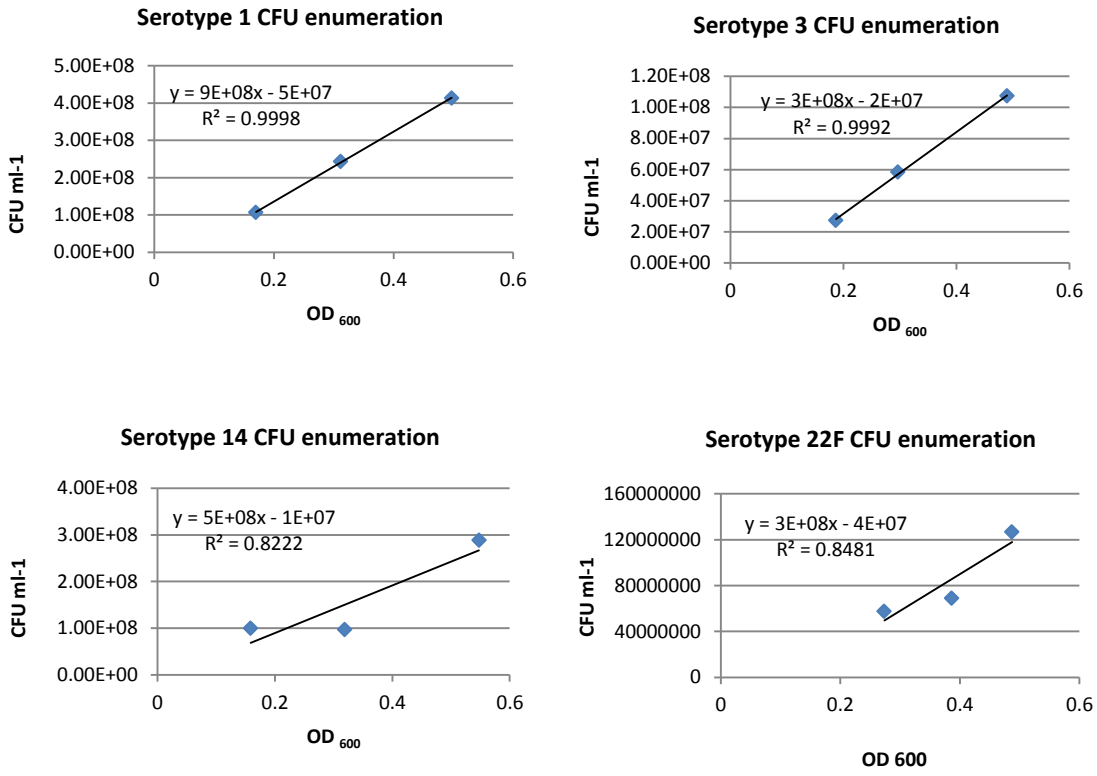


Figure 3-2: Colony Forming Unit enumeration and regression analysis of the pneumococcal serotypes

CFU enumeration of pneumococcal serotypes grown in BHI broth for a total of 6-8 hours. Using previous growth curves for reference, at three points during the exponential growth phase, 100 µL samples were taken and viable counts were performed. CFU counts were analyzed using regression analysis to generate a calibration curve of OD_{600} vs. CFU. The approximate optical densities corresponding to 10^8 CFU mL^{-1} are as follows, serotype 1 (OD_{600} 0.17), serotype 3 (OD_{600} 0.4), serotype 14 (OD_{600} 0.22) and serotype 22F (OD_{600} 0.45).

3.3.2 Initial attachment assay of pneumococcal serotypes

To assess the difference in initial attachment of the four different serotypes, each serotype was grown under biofilm conditions in 48 well plates for 24 hours and subsequently stained with crystal violet to assess initial attachment (Figure 3-3) as described in section 2.7. All datasets had a normal distribution and equal variance. One-way ANOVA revealed a significant difference in biofilm formation between the various serotypes ($P = 0.004$). Tukey's pairwise comparison revealed serotype 1 differed significantly from serotype 14 but not serotype 3 or 22F, serotype 3 differed significantly from serotype 14 and 22F but not serotype 1, serotype 14 differed significantly from serotype 1 and 3 but not 22F and serotype 22F differed significantly from serotype 3 only. Based on these data serotypes 14 and 22F had the greatest biofilm formation after 24 hours whereas serotype 3 had the least attachment after 24 hours.

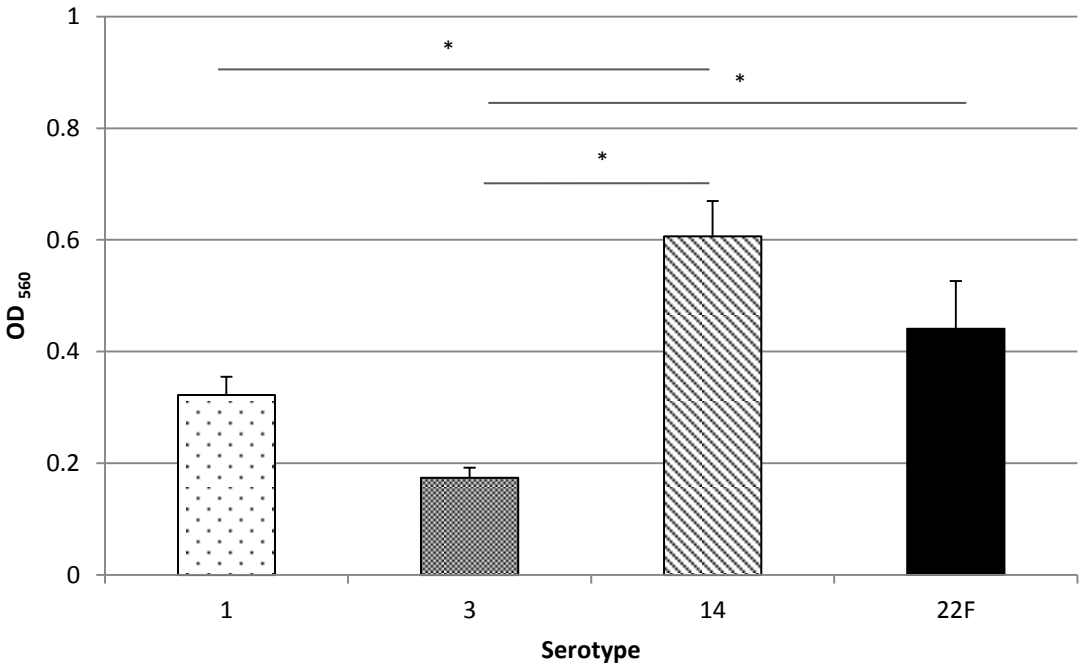


Figure 3-3: Crystal violet staining of 24 hour pneumococcal biofilms

Pneumococcal biofilms were cultured in 48-well plates in 1:5 BHI at 37 ^{o}C in 5 % CO_2 and allowed to attach for 24 hours before being stained with crystal violet and absorbance read on the FLUOstar Optima plate reader at 560 nm. Six replicates for each serotype and negative control were performed per plate and averages were taken. Readings were subtracted from the negative control sample average to give a final absorbance value. A total of three independent experiments were performed using three independent inocula. Error bars represent standard error. One-way ANOVA with Tukey's comparison was performed to assess the difference in attachment between serotypes. Asterisks denote a significant difference at the 95 % confidence level.

3.3.3 Biofilm formation of pneumococcal serotypes

To establish a more comprehensive assessment of biofilm formation over a longer time course biofilms were seeded with 1 x 10^8 CFU mL^{-1} of the relevant serotype and inoculated into MatTek culture plates, cultured for 1, 3, 6 and 9 days and imaged using an EF microscopy and the *Bac*Light live/dead stain as described in section 2.6.

Figure 3-4 depicts the percentage surface coverage of the four pneumococcal serotypes over the 9-day time course. Statistically all datasets had a normal distribution (Anderson-Darling) and equal variance. One-way ANOVA revealed a significant difference in biofilm formation between the various serotypes at day 1 ($p = <0.001$), day 3 ($p = <0.001$) day 6 ($p = <0.001$) and day 9 ($p = <0.001$). Tukey's pairwise comparison at the 95 % confidence level revealed that by day 1, serotype 1 differed from serotype 3 and 22F; serotype 3 differed significantly from serotypes 1, 14, and 22F; serotype 14 differed from serotype 3 only and serotype 22F differed significantly from serotypes 1, 3 but not serotype 14. By day 3, serotype 3 differed significantly from serotypes 1, 14, and 22F; however no significant difference was seen between 1, 14 and 22F. By day 6, serotype 3 differed significantly from serotypes 1, 14 and 22F, serotypes 1 and 14 differed significantly from serotypes 3, and 22F but not from each other. Serotype 22F differed significantly from serotypes 1, 3 and 14. By day 9, serotypes 1 and 3 decreased in confluency. Tukey's pairwise comparison revealed that serotype 3 differed significantly from serotypes 1, 14 and 22F, serotype 14 differed significantly from serotypes 3 and 22F but not from serotype 1. Serotype 22F differed from serotypes 1, 3 and 14.

Visually, serotype 1 biofilms produced a monolayer of cells with large microcolonies (Figure 3-5B) throughout the 9-day time course reaching a stable viable cell confluency of 35 % by day 6 (Figure 3-4). Serotype 3 remained a poor biofilm former throughout the 9-day time course reaching a maximum confluency of only 12 % at day 6 (Figure 3-4). Figure 3-6D reveals minimal cell attachment; cell clustering is evident however no structured microcolonies were present by day 9. Serotype 14 produce dense biofilms with large microcolonies (Figure 3-7), similarly to serotype 1, serotype 14 reached a maximum confluency of 35 % by day 6 (Figure 3-4). Serotype 22F appeared to be the best biofilm former, reaching a maximum confluency of 65 % by day 9 (Figure 3-4); visually the biofilm consisted of a dense monolayer of cells interspersed with smaller microcolonies (Figure 3-8B). These data suggest that 22F is consistently the most prolific biofilm former over the 9-day time course, followed by serotype 14 and 1 respectively. Serotype 3 however formed minimal biofilm throughout the 9 days.

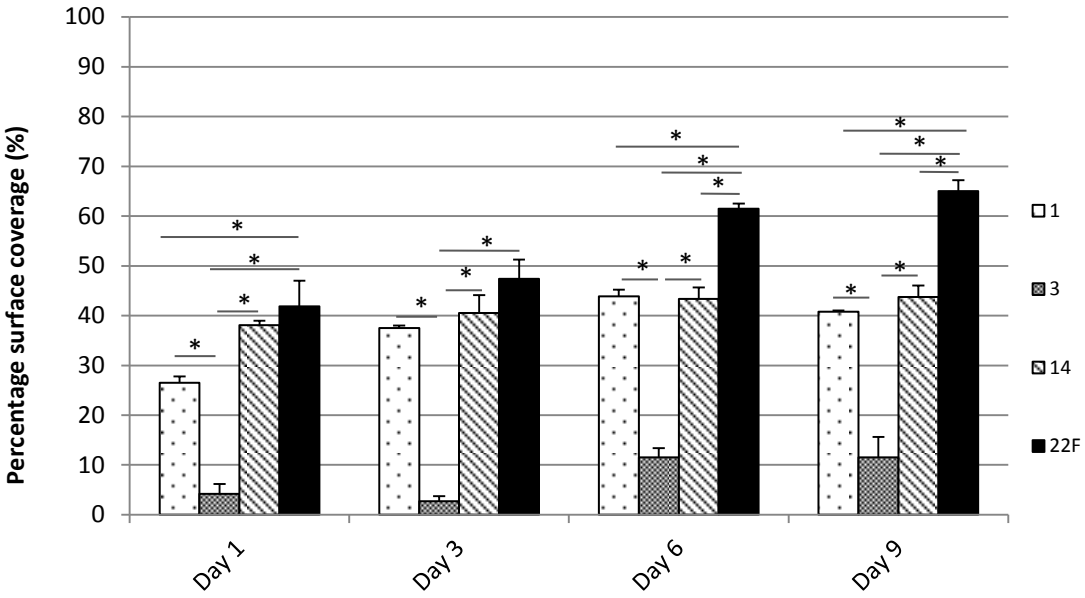


Figure 3-4: Percentage surface coverage of pneumococcal biofilms

Biofilm formation was observed using a Nikon Eclipse LV100 epi-fluorescent microscope and the *Bac*Light live/dead stain. Biofilms were cultured in MatTek culture plates under static biofilm conditions for 1, 3, 6 and 9 days. Images were taken under epi-fluorescence filters at 400x magnification with the Image-Pro 6.2 software and analysed using ImageJ analysis software. A minimum of ten fields of view were taken for triplicate biofilms. SYTO®9 and propidium iodide images were merged to give the total percentage coverage. Error bars represent standard error. One-way ANOVA with Tukey's comparison was performed to assess the difference in percentage surface coverage of each serotype at each time point. Asterisks denote a significant difference at the 95 % confidence level.

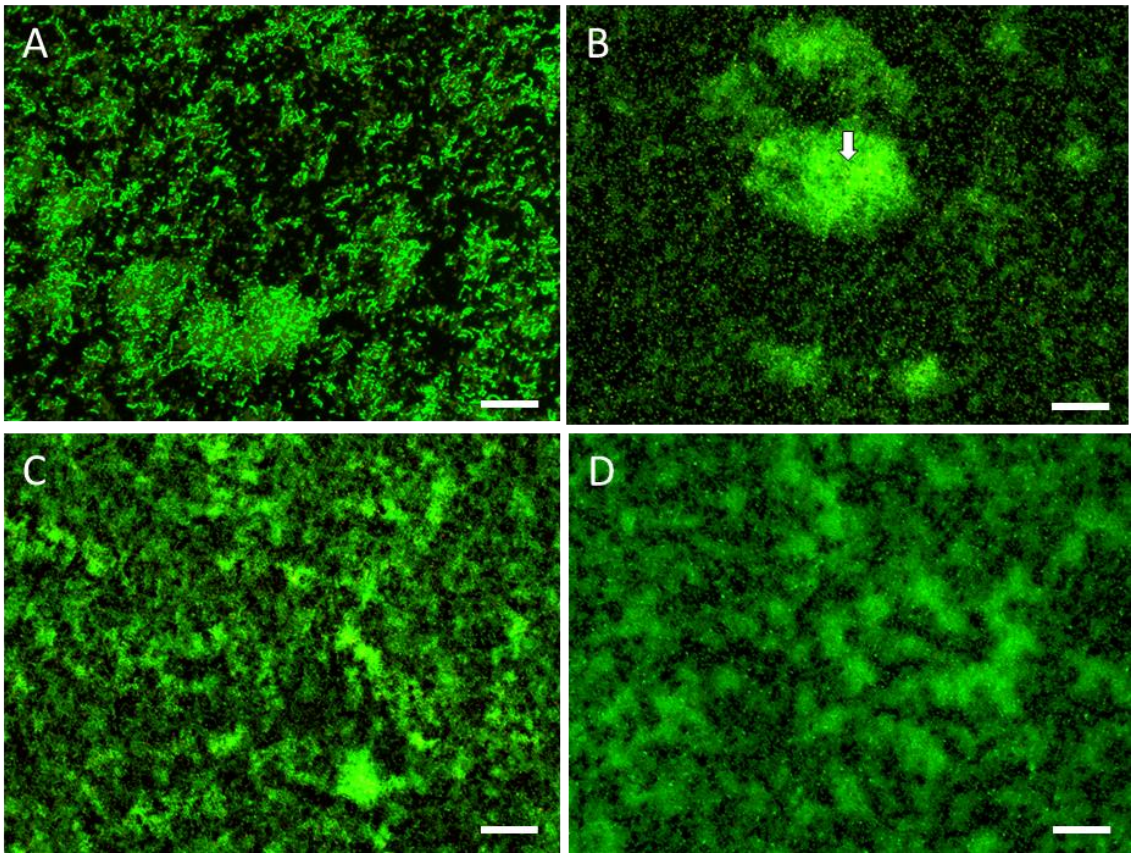


Figure 3-5: Pneumococcal biofilms of Serotype 1

Biofilm formation of serotype 1 pneumococci was observed using a Nikon Eclipse LV100 epi-fluorescence microscope and the *Bac*Light live/dead stain (Invitrogen). Biofilms were cultured in MatTek culture plates under static biofilm conditions for 1 (A), 3 (B), 6 (C) and 9 days (D). Images were taken under epi-fluorescence filters at 400x magnification. Green cells indicate viable cells (live), red cells indicate cells with a compromised membrane (dead). White arrows indicate microcolonies. Scale bar = 20 µm.

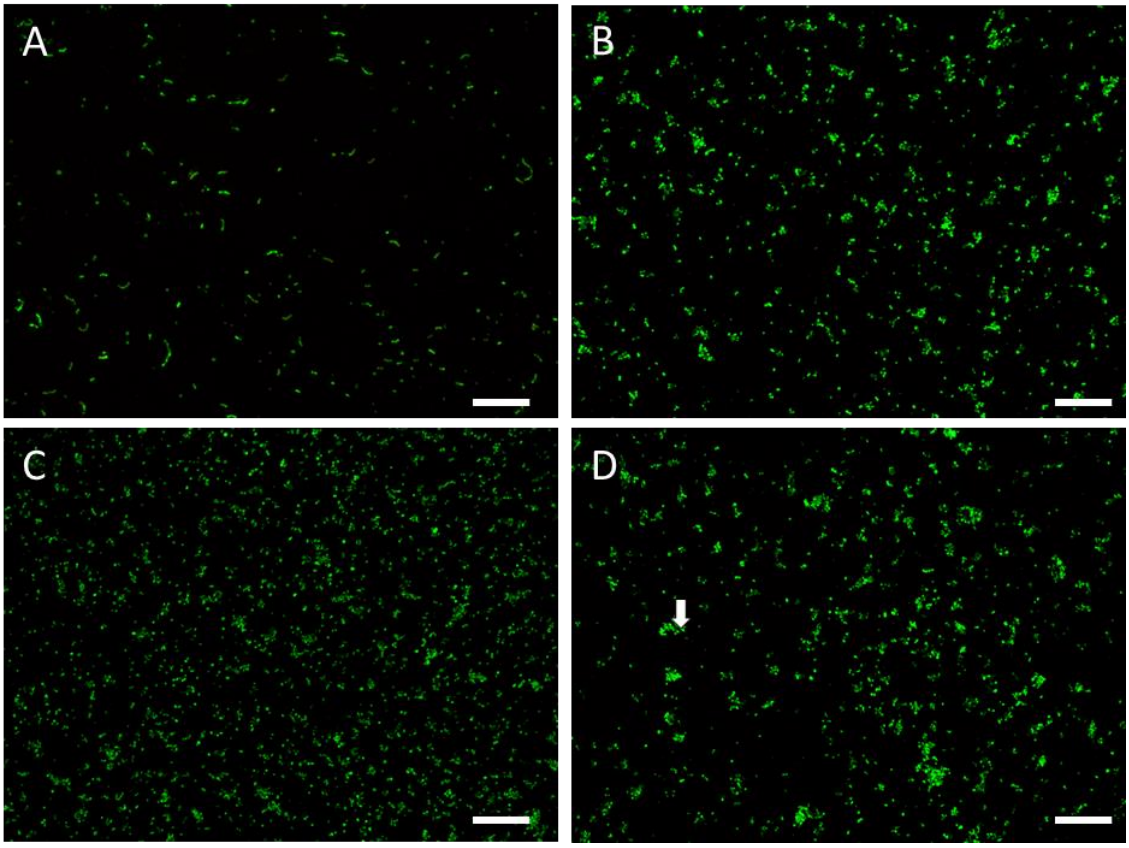


Figure 3-6: Pneumococcal biofilms of Serotype 3

Biofilm formation of serotype 3 pneumococci was observed using a Nikon Eclipse LV100 epi-fluorescence microscope and the *Bac*Light live/dead stain (Invitrogen). Biofilms were cultured in MatTek culture plates under static biofilm conditions for 1 (A), 3 (B), 6 (C) and 9 days (D). Images were taken under epi-fluorescence filters at 400x magnification. Green cells indicate viable cells (live), red cells indicate cells with a compromised membrane (dead). White arrows indicate microcolonies. Scale bar = 20 µm.

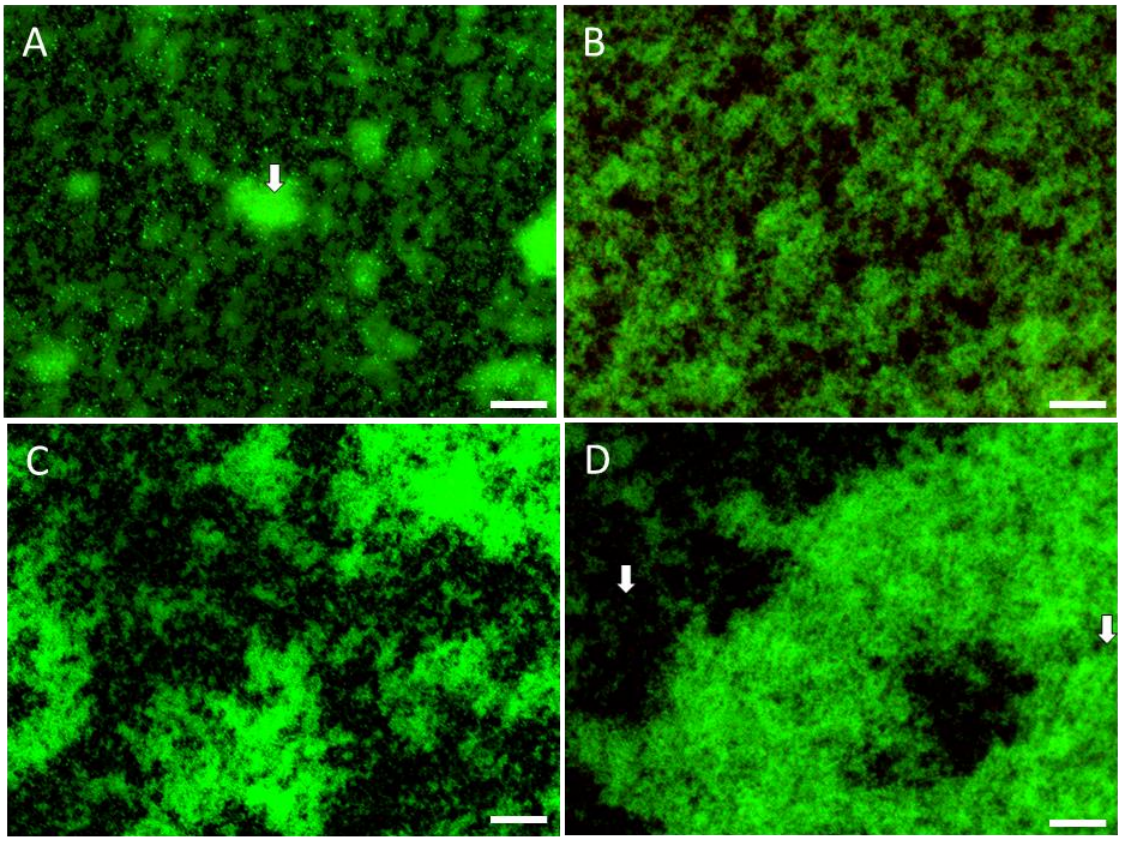


Figure 3-7: Pneumococcal biofilms of Serotype 14

Biofilm formation of serotype 14 pneumococci was observed using a Nikon Eclipse LV100 epi-fluorescence microscope and the *Bac*Light live/dead stain (Invitrogen). Biofilms were cultured in MatTek culture plates under static biofilm conditions for 1 (A), 3 (B), 6 (C) and 9 days (D). Images were taken under epi-fluorescence filters at 400x magnification. Green cells indicate viable cells (live), red cells indicate cells with a compromised membrane (dead). White arrows indicate microcolonies. Scale bar = 20 µm.

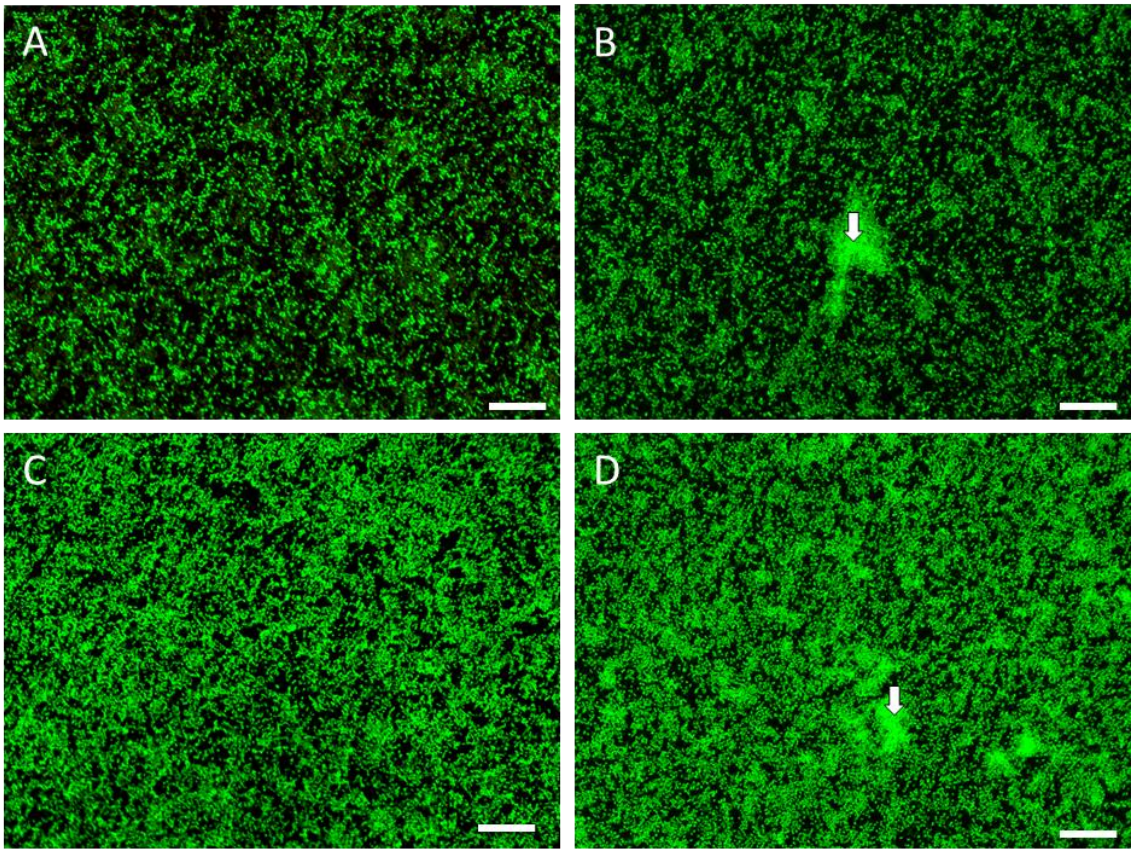


Figure 3-8: Pneumococcal biofilms of Serotype 22F

Biofilm formation of serotype 22F pneumococci was observed using a Nikon Eclipse LV100 epi-fluorescence microscope and the *Bac*Light live/dead stain (Invitrogen). Biofilms were cultured in MatTek culture plates under static biofilm conditions for 1 (A), 3 (B), 6 (C) and 9 days (D). Images were taken under epi-fluorescence filters at 400x magnification. Green cells indicate viable cells (live), red cells indicate cells with a compromised membrane (dead). White arrows indicate microcolonies. Scale bar = 20 µm.

To assess the total viable cell counts from the pneumococcal biofilms, each biofilm was seeded with 1 x 10^8 cells mL^{-1} of the pneumococcal serotype of interest and inoculated into 6-well plates and cultured under static biofilm conditions for 1, 3, 6 and 9 days. At each time point the biomass was harvested and plated onto Columbia blood agar to assess the viable cell counts from the pneumococcal biofilms. What is clear from Figure 3-9 is that the viable cells culturable from the biofilms remain consistent throughout the 9-day time course. Serotype 3 had the lowest culturable cell counts which is consistent with both the crystal violet assay and percentage surface coverage assay. Viable counts increased in all serotypes over the 9 day time course with the exception of serotype 14 which see a reduction in viable cells (7 x 10^5) at day 6, by day 9 however culturable cells recovers back to 2 x 10^6. A correlation analysis over the 9-day time course of biofilm CFU vs. biofilm percentage surface coverage, using the Pearson product-moment correlation coefficient, showed a strong positive correlation in serotypes 1 and 22F (r = 0.83 and r = 0.8, respectively) indicating that over time, as the biofilm increases in percentage surface coverage an increase is also observed in the viable cell counts. In contrast a weak negative correlation was seen in serotypes 3 and 14 (r = -0.1 and r = -0.23, respectively) indicating that over time, the increase in biofilm percentage surface does not necessarily correlate with an increase in viable cell counts from the biofilm.

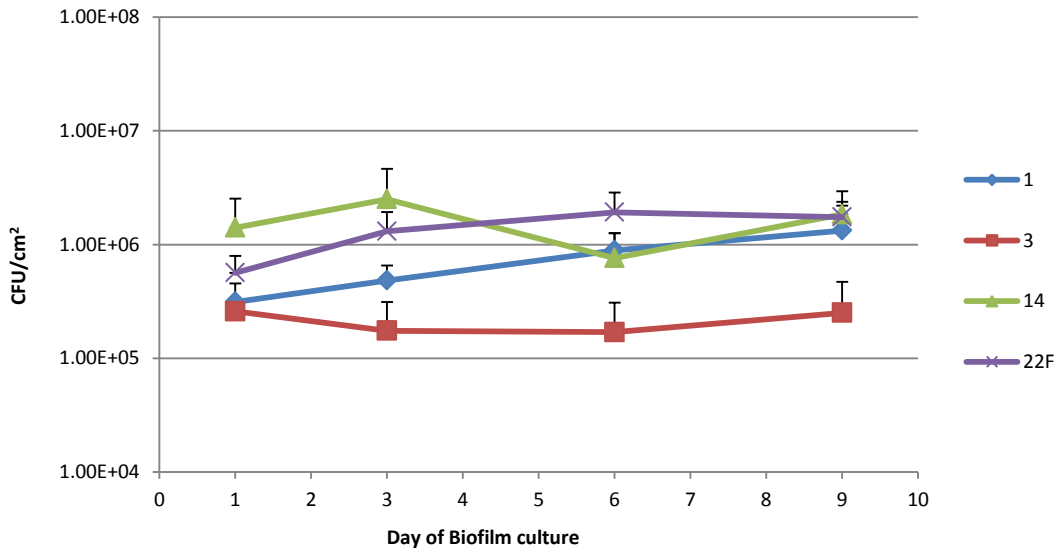


Figure 3-9: Colony Forming Unit enumeration of pneumococcal biofilms

Biofilms were seeded with 1 x 10^8 CFU mL^{-1} of the relevant serotype and inoculated into 6-well plates and cultured under static biofilm conditions for 1, 3, 6 and 9 days. Cells were extracted from triplicate pneumococcal biofilms, from duplicate experiments at time points 1, 3, 6 and 9 days, using a cell scraper, vortexed and plated onto Columbia blood agar to assess total viable cells in the biofilms. Error bars represent standard error.

3.3.4 Biofilm-derived colony morphology variance

After 3 days of biofilm growth three distinct colony variants were observed which were classified as small colony variants (SCVs), medium colony variants (MCVs) and typical colony variants (TCVs) as described in section 2.5.

Figure 3-10 shows that colony variation was observed in all four serotypes tested and shows that SCVs and MCVs represented approximately 1 % and 3 % of the population respectively, suggesting that the mutations responsible for the SCV and MCV phenotype occur at relatively low frequency. MCVs and SCVs were only observed after 3 days of biofilm culture in serotypes 1 and 14. MCVs were observed in day 1 22F biofilms, however SCVs were only observed after 3 days of biofilm culture. Colony variation was only observed at day nine in serotype 3 biofilms. In each serotype tested, multiple colony variants were not seen in the starting inoculum. Variants from each serotype remained sensitive to optochin, indicating the variants were *S. pneumoniae* and not contamination and secondly that resistance against optochin had not occurred as a result of biofilm growth. It should be noted that the TCV phenotype resembled that of the wild type parent morphology, but due to potential unseen genetic changes TCVs isolated from biofilms could not be classed as having wild type morphology. TCVs and MCVs of serotype 1, 14 and 22F exhibited autolysis, shown by the 'crater' or 'draftsman' like appearance of the colonies (Figure 3-11). All variant morphologies had smooth margins, no variants were observed with distinct undulated margins. In order to determine whether this variation in morphology was a direct result of biofilm growth and not simply prolonged growth, planktonic cultures were grown for nine days, sub-culturing daily into fresh BHI broth the maintain cell viability. At 1, 3, 6 and 9 days cells were serially diluted and plated onto blood agar to assess morphology. Over the nine-day time course the wild type morphology remained consistent and the presence of small and medium sized colonies was not observed. These data suggest that the changes in morphology and any genetic mechanisms associated with each morphology type, is a direct result of biofilm and not prolonged growth.

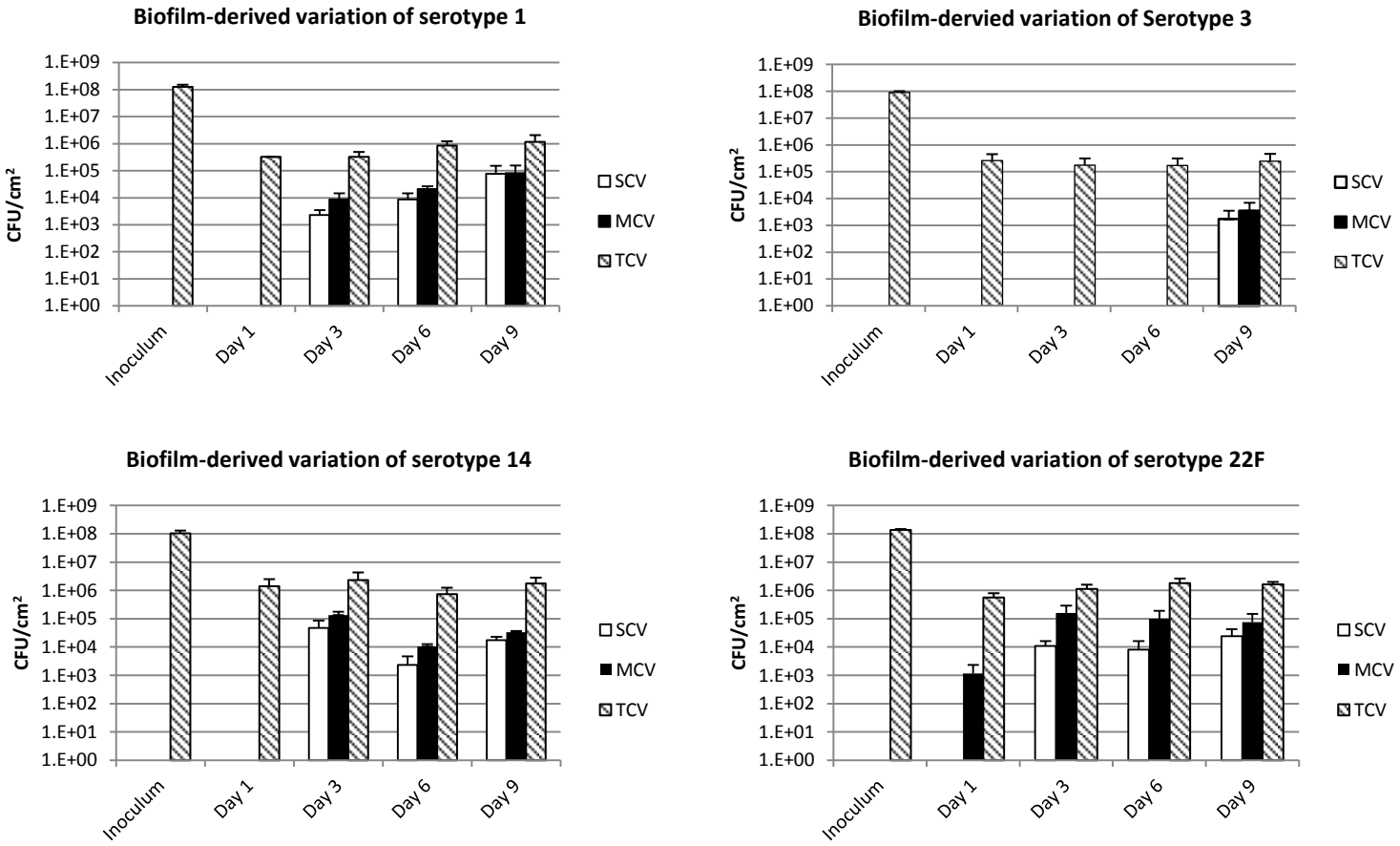


Figure 3-10: Colony Forming Unit enumeration of biofilm-derived pneumococcal variants

Cells were harvested from pneumococcal biofilms at time points 1, 3, 6 and 9 days and plated onto Columbia blood agar to assess for changes in colony morphology. Colony variants were defined as described in the materials and methods section 2.5. All colonies were visualised under a Leica MZ 16F dissection stereomicroscope. Error bars represent standard error.

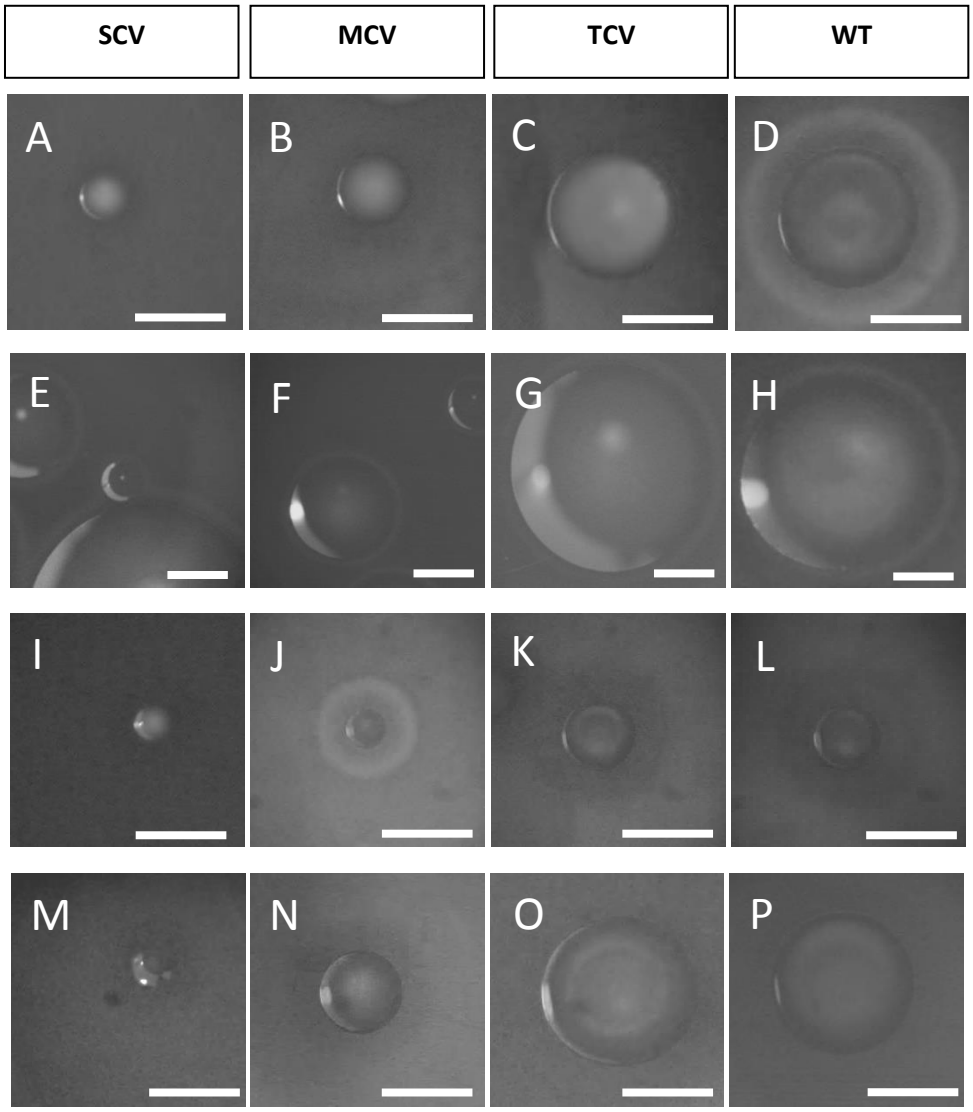


Figure 3-11: Colony morphology variation in pneumococcal biofilms

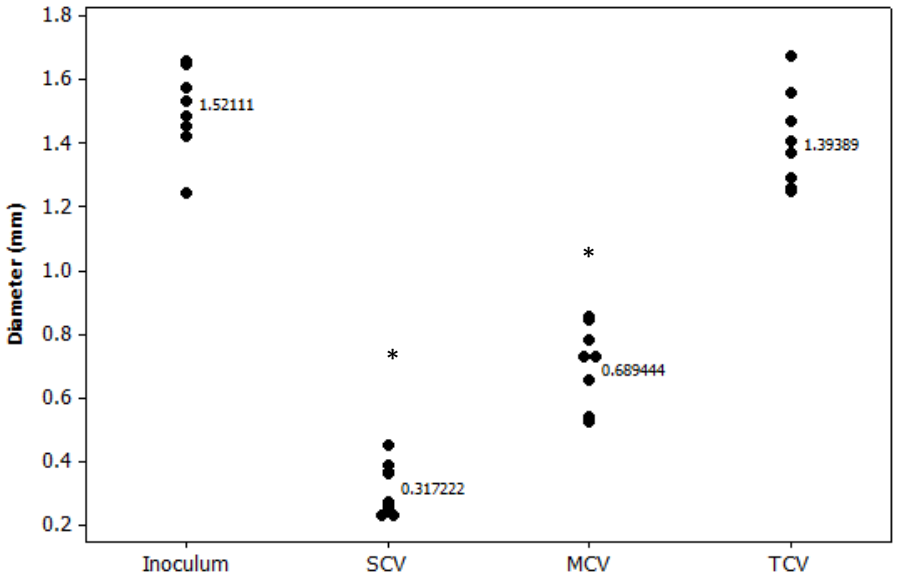
Cells were harvested from pneumococcal biofilms at time points 1, 3, 6 and 9 days and plated onto Columbia blood agar to assess for changes in colony morphology. Colony morphology was assessed based on size and regularity of the colony surface undulated vs. smooth colony surface. For serotypes 1 (A-D) and 22F (M-P), variants were defined as follows; small colony variants (SCVs) (<0.5 mm), medium colony variants (MCVs) (0.5-<1 mm) and wild type/typical colony variants (TCVs) (≥1 mm). For serotypes 14 (I-L), variants were defined as follows; small colony variants (SCVs) (0.1-<0.3 mm), medium colony variants (MCVs) (0.3-<0.6 mm) and wild type/typical colony variants (TCVs) (≥0.6 mm). Due to the mucoid nature of serotype 3 (E-H), colony variants were defined as; small colony variants (SCVs) (<1 mm), medium colony variants (MCVs) (1-<2 mm) and wild type/typical colony variants (TCVs) (≥2 mm) in accordance with previous literature (Allegrucci and Sauer, 2007). Colonies were visualised under a Leica MZ 16F stereomicroscope and images were taken using a Leica digital camera with microscope attachment. White scale bar = 1 mm.

3.3.5 Colony morphology quantification

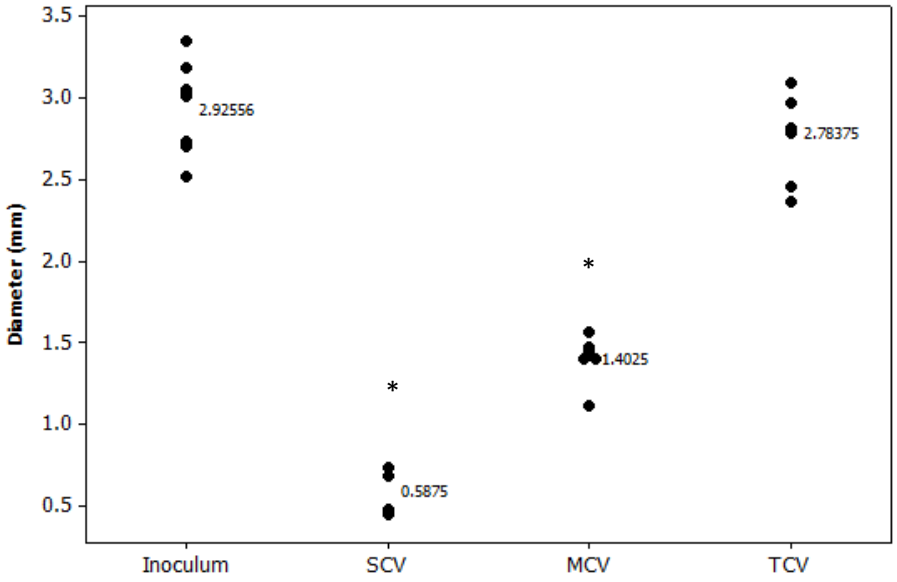
To reduce the subjectivity in the classification method of colony variants, the cross-sectional diameter of the colony variants was quantified. Colony morphology was assessed based on size and regularity of the colony surface undulated vs. smooth colony surface. Colonies were visualised under a Leica MZ 16F stereomicroscope and images were taken using a Leica digital camera with microscope attachment with a 5 mm graticule scale. Statistically all datasets were tested for normal distribution and equal variance. A 2-sample t-test was used to assess whether there was a significant difference between the variants and the wild type parent strain.

Quantification of the biofilm-derived colony variants confirmed the observation that SCVs and MCVs are statistically distinct from the wild type parent strains ($p = <0.001$) based on diameter size alone, and as such may be a marker for genetic change. Using a 2-sample t-test for all serotypes SCVs and MCVs were shown to be significantly different from the parent strains used to inoculate the biofilms ($p = <0.001$). In contrast TCVs were not shown to be significantly different from the parent strains for serotype 1 ($p = 0.078$), serotype 3 ($p = 0.281$), serotype 14 ($p = 0.279$) and serotype 22F ($p = 0.781$) (see Figure 3-12). Furthermore using one-way ANOVA with Tukey's comparison, these data showed that for all serotypes, each variant (SCV, MCV and TCV) had a significantly different mean from each other at the 95 % confidence level. The purpose of quantifying the size of the variants was to remove the subjectivity in their classification. It should be noted that all variants were cultured directly from the biofilm and measured; plates containing cells harvested from the biofilms were incubated for exactly 18 hours to standardise the time in the incubator.

A


Diameter (mm)
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
1.52111
0.689444
0.317222
1.39389
Inoculum
SCV
MCV
TCV

B


Diameter (mm)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
2.92556
1.4025
0.5875
2.78375
Inoculum
SCV
MCV
TCV

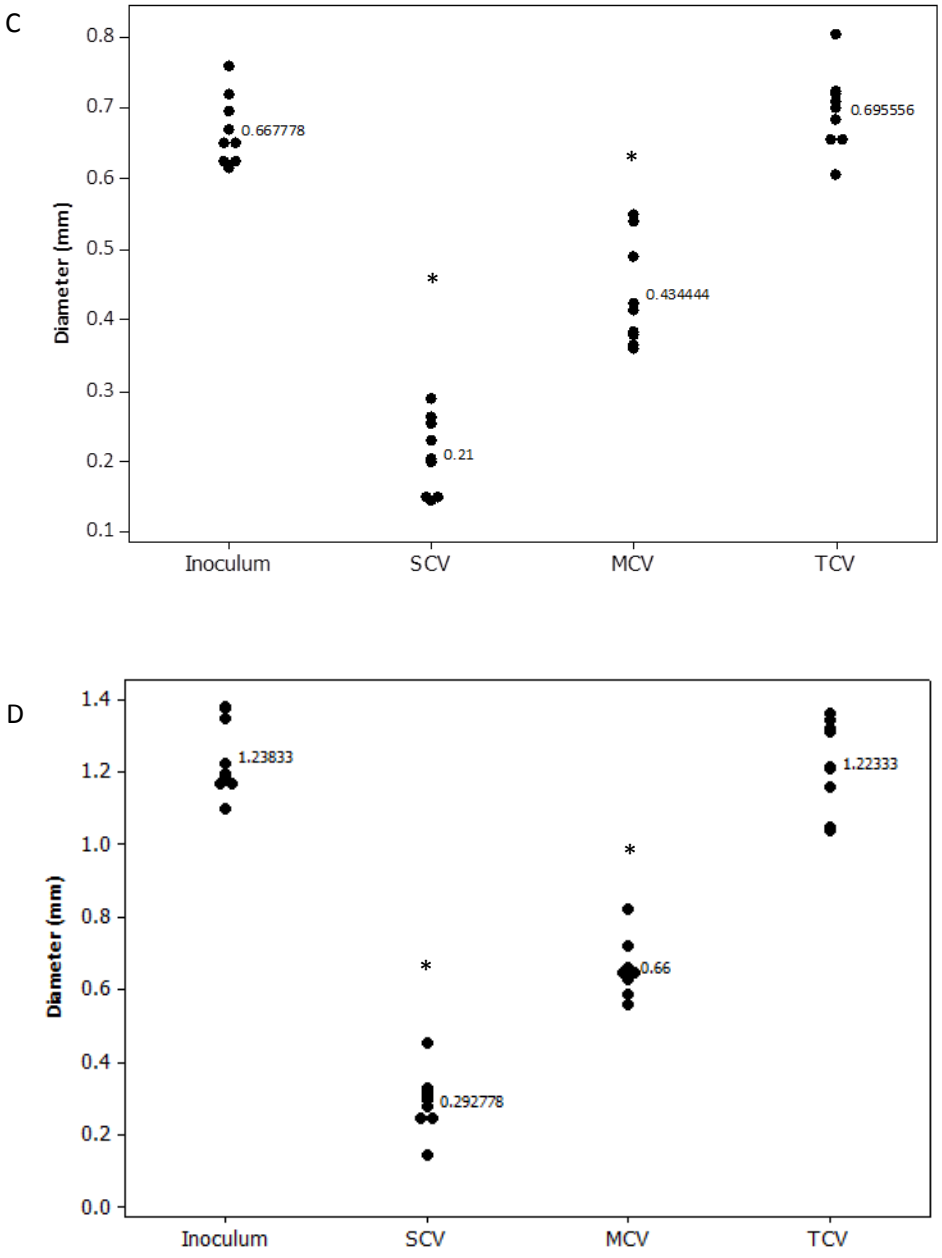


Figure 3-12: Colony quantification of serotype 22F biofilm-derived colony variants

Colony variants harvested from 9 day triplicate biofilms were imaged on a Leica MZ 16F stereomicroscope and quantified against a known graticule scale using ImageJ analysis software. For serotype 1 (A) and 22F (D), variants were defined as follows; small colony variants (SCVs) (<0.5 mm), medium colony variants (MCVs) (0.5-<1 mm) and wild type/typical colony variants (TCVs) (≥1 mm). Due to the mucoid nature of serotype 3 (B), colony variants were defined as; small colony variants (SCVs) (<1 mm), medium colony variants (MCVs) (1-<2 mm) and wild type/typical colony variants (TCVs) (≥2 mm). For serotypes 14 (C), variants were defined as follows; small colony variants (SCVs) (0.1-<0.3 mm), medium colony variants (MCVs) (0.3-<0.6 mm) and wild type/typical colony variants (TCVs) (≥0.6 mm). A 2-sample t-test was used to compare each variant diameter to the WT. Asterisk denote a significant difference from the WT parent (where p= <0.05).

3.4 **Discussion of Results**

The aim of this chapter was to establish the biofilm forming properties of the four pneumococcal serotypes of interest taken from an on-going paediatric pneumococcal carriage study at the UHS [serotype 1 (ST306), 3, (ST3205), 14 (ST124) and 22F (ST433)]. Two assays were employed to measure biofilm formation, crystal violet (CV) staining and direct visualisation using the *Bac*Light live/dead stain and epi-fluorescence microscopy. These data have shown us that all four pneumococcal serotypes are capable of forming biofilm under static conditions when grown for nine days, but that there is a significant difference in the biofilm forming ability between the four serotypes.

Initially growth curves were performed for each serotype to determine the optical density which corresponded to 10^8 CFU. This allowed each of the various pneumococcal biofilms to be seeded with approximately the same number of cell, thus making the results more reliable. *S. pneumoniae* is a nutritionally fastidious facultative anaerobe. Carbohydrates are converted to pyruvate via glycolysis (see Figure 7-1) and are the only nutrients which support growth and cell division in *S. pneumoniae* (Hoskins *et al*., 2001). The use of BHI as a growth medium was a result of growth study data in a collaborative laboratory working on *S. pneumoniae* biofilms; BHI permitted consistent growth of the bacterium and is a well-established growth medium (Kurola *et al.*, 2011, Allan *et al*., 2014). To initiate biofilm growth, biofilms were cultured in a diluted medium of 1:5 strength BHI as previously described (Hall-Stoodley *et al*., 2008). BHI is a rich medium with brain and heart infusion solids, peptone, disodium hydrogen phosphate, D-glucose and sodium chloride as its main constituents, however the availability of glucose within the nasopharynx is thought to be at a low concentration (Philips *et al*., 2003) therefore the rationale for the use of a diluted growth medium was to initiate a stress response in the pneumococci to start developing a biofilm. A rich medium would not have be representative of the host environment.

Each of the pneumococcal serotypes was assessed for their ability to adhere to polystyrene well in a 48-well plate after 24 hours. This assay allowed the rapid comparison of cell attachment and biofilm formation over a 24 hour period. CV staining is a well-established assay used for measuring biofilm formation. To quantify the biofilm biomass the crystal violet dye bound to adherent cells is re-solubilized in 95 % ethanol and measured at an optical density of OD_{560}. Data obtained from the CV initial attachment assay revealed that serotypes 14 and 22F had the greatest biofilm formation after 24 hours whereas serotype 1 and 3 had

the less attachment after 24 hours. It is possible therefore, within 24 hours, to see a hierarchy of biofilm forming ability where serotypes 14 and 22F as 'good' biofilm formers, serotype 1 as a 'moderate' biofilm former and serotype 3 as a 'poor' biofilm former.

In order to get a more comprehensive idea of biofilm formation over a 9-day time course, biofilms of each serotype were stained with SYTO®9 and propidium iodide live/dead stain. Theoretically, SYTO®9 stains both living and dead bacteria whereas propidium iodide penetrates only bacteria with damaged membranes. As a result live bacteria fluoresce green, while bacteria with damaged membranes fluoresce red. However, due to the presence of eDNA in the biofilm matrix there is the possibility that both SYTO®9 and propidium iodide may stain eDNA. It is clear from Figure 3-4 that 22F produced biofilms which were consistently confluent throughout the nine day time course. Serotype 22F appeared to be the best biofilm former, reaching a maximum confluency of 65 % by day 9, whereas serotype 3 appeared to be the poorest biofilm former reaching a maximum confluency of only 12 % at day 6. Interestingly serotypes 1 and 3 decreased in confluency by day 9, this may suggest that by day 9 either the cells within the biofilm had started to die or that a dispersal event had occurred. Statistically these data suggest that by day 9, 22F was the most prolific biofilm former, followed by serotypes 14 and 1 respectively. Serotype 3 however remained a poor biofilm former throughout the 9-day time course.

When the epi-fluorescence percentage coverage data from day 1 is compared with the CV initial attachment we see consistent results and a similar distribution of data. In the CV assay a significant difference was seen between when comparing serotype 3 to 14 and 22F, whereas in the EF microscopy revealed serotype 3 differing significantly from all other serotypes (including serotype 1). The reason for this this slight discrepancy can be accounted for by the different nature of the two different assays. CV staining does not differentiate between living and dead cells, or cells and the biofilm matrix but instead stains the entire biomass including polysaccharide matrix components and eDNA. *Bac*Light live/dead assay, used to determine percentage surface coverage stains both living and dead cells, including eDNA but does not stain other matrix components, and as such, is perhaps a more selective technique. In addition the two assays were performed in different polystyrene plates; the CV assay utilised 48-well plates whereas the EF percentage surface coverage assay utilised MatTek plates. To validate the CV data, the initial attachment assay could be repeated in MatTek plates but for the purpose of this work it was unnecessary. A correlation analysis over the 9-day time course of biofilm CFU vs. biofilm percentage surface coverage showed a strong positive correlation in serotypes 1 and 22F and a weak contrast in serotypes 3 and 14. The

expectation would be that as a biofilm increase in surface coverage (by staining with viability stains) it is increasing in cell mass and viable cells recovery would increase in proportion, however, this was not observed in serotypes 3 and 14. This observation was not unexpected for serotype 3, which produced poor biofilms throughout the 9-day time-course. This weak correlation in serotype 14 may be due to many factors; the decrease in viable CFU counts on day 6 in serotype 14 could be explained by the phenomenon of viable but non-culturable (VBNC) cells (Trevors, 2011), error in the recovery technique, or alternatively, the differences between the two assays could be due to differences in the 6-well plate and the MatTek culture plate substratum (glass vs. polystyrene) and the fact that the percentage surface coverage does not take into account the depth of the biofilm. In addition, the possible binding of SYTO®9 and propidium iodide to eDNA in the biofilm EPS may have resulted in an over-estimation of the biofilm coverage and biofilm viability. Nonetheless these assays have shown that all four serotypes are capable of forming biofilm and that viable cells are recoverable for phenotypic analysis.

Previously, Tapiainen *et al.* (2010) tested the biofilm properties of 204 clinical samples isolated from the nasopharynx, the middle ear and from blood of paediatric patients. This study showed no difference in the biofilm forming ability of the isolated based on site of isolation, but was limited in its biofilm quantification methods and did not explore the relationship of ST to biofilm formation. To further the work in this chapter, additional ST clones and serotypes of *S. pneumoniae* isolated from invasive and non-invasive sites could be assessed to determine the biofilm forming properties of the various isolates. With a greater N number of samples a correlation analysis could be performed to compare ST with serotypes and ST/serotypes with site of infection. Alternatively data from all three culture methods were combined to generate a biofilm forming Index score as seen in Hall-Stoodley *et al.,* (2008). For this to be consistent with previous literature confocal microscopy of each serotype would have to be performed.

All biofilm experiments were performed in static microtitre plates; 6-well plate, 24-well plates or MatTek plates. The rationale for choosing this method of biofilm culture was that it is a well-documented method of biofilm culture and provided a rapid and effective model for biofilm growth (Hall-Stoodley *et al*., 2008). As this work is relevant to the carriage model of pneumococcal infection perhaps a continuous culture method may have also been appropriate as this may mimic, to some extent, the flow of nutrient *in vivo*. Pilot studies were conducted using a once-through continuous flow-cell model. The model was set up as previously described (Allegrucci and Sauer, 2008); biofilms were cultured for up to 6 days and

cells were harvested by pinching along the silicone tubing. Recoverable cell titres from the flow-cell model were low (~10^2) and as such the static microtitre plate assay proved a more consistent model for the isolation of biofilm-derived variants. As the aim of the work was to identify phenotypic variants, this was achieved using the static microtitre plate assay and as such was an acceptable method of biofilm culture. Recent advances in pneumococcal biofilm culture have seen pneumococci grown on a substratum of fixed epithelial cells and *in vivo* (Marks *et al.,* 2013), a future experiment would be to see whether biofilm growth on a more clinically relevant substratum result in the same level of diversity seen in this study. Indeed a point of interest highlighted by Domenech *et al.* (2009) is that the method of biofilm culture may itself select for particular capsular mutations. Type 3 pneumococci have been grown on Sorbarod filters (Waite *et al*., 2001), under continuous flow-cell (Allegrucci and Sauer, 2008) and under a microtitre plate-based system (McEllistrem *et al*., 2007), in each case mutations in the *cap*3A genes differed suggesting that the method used to culture biofilms may modulate the type of mutations seen in the capsular genes. In addition biofilm growth of *P. aeruginosa* in the 2004 study by Boles *et al.* revealed that the generation of variants was dependant on the growth conditions applied to the biofilm (Boles *et al*., 2004). Taken together this suggests that the variation seen in the literature is inconsistent making it difficult to determine how relevant the diversity seen *in vitro* is to what may occur *in vivo*. Alternatively the difference in mutation between the various methods of biofilm culture may simply reflect the dynamic nature of the biofilm itself.

Recently an increasing number of groups have used *in vivo* models and *in vitro* models with an epithelial cell-substratum to study colonisation of the lung and nasopharynx (Trappetti *et al*., 2009, Marks *et al*., 2012a). In retrospect, the polystyrene surface of the 6- and 48-well plates and the glass surface of the MatTek plates used in this chapter are not truly a representative surface to mimic nasopharageal colonisation. To advance the findings in this work further, biofilms could be grown using a model similar to Marks *et al.* (2012a) whereby pneumococcal biofilms were cultured on a substratum of H292 epithelial cells throughout the entire biofilm growth period *in vitro* and in BALB/cByJ mice to observe colonisation directly *in vivo*. By providing a more representative growth substratum the level of phenotypic and genetic diversity could be observed to give better insight into what diversity might arise during colonisation of the human nasopharynx.

The difference in biofilm-forming ability of the four serotypes studies is an interesting phenomenon and may be relevant to the epidemiology of the pneumococcus; prior to the introduction of PCV7, serotype 14 remained one of the most common serotypes isolated for

IPD (Johnson *et al*., 2010) and has been shown to grow consistently as biofilms *in vitro* (Tapiainen *et al*., 2010) so the ability to form good biofilms in this work was not unexpected. There is a scarcity of literature on serotype 22F, the ability to form good biofilms and its increased prevalence in pneumococcal carriage (Jacobs *et al*., 2008), which highlights the need for more work on this serotype. Prior to the introduction of PCV13, serotype 1 was one of the most common serotypes isolated for IPD (Johnson *et al*., 2010), the ability to form moderate biofilms may help persistence of this serotype. The most interesting results were that of serotype 3. As previously mentioned numerous biofilm studies have focused on serotype 3 due it its role in pneumococcal disease and its ease in distinguishing between colony morphology variants (Waite *et al*., 2001, Allegrucci and Sauer, 2007, McEllistrem *et al*., 2007, Allegrucci and Sauer, 2008, Domenech *et al*., 2009). The lack of serotype 3 biofilm formation in this work could suggest that the biofilm forming ability of serotype 3 may vary significantly depending on the clone used. Alternatively the method of biofilm culture may have resulted in the discrepancy. Indeed, previous work in microtitre plates has recorded similar serotype 3 biofilm results as this work (Hall-Stoodley *et al*., 2008, Camilli *et al*., 2011). As mentioned in section 1.3.1.3 this may indicate that a continuous flow method may be a better method for biofilm formation or that the continuous flow method selects specifically for variants within a serotype 3 population that are better adapted to form biofilms.

Epi-fluorescence microscopy is a well-documented method of assessing biofilm growth (Giao *et al*., 2010, Giao and Keevil, 2014) however it is limited by its ability to assess the depth and overall biomass of the biofilm. This work would benefit from the use of CLSM to achieve a more comprehensive assessment of biofilm formation. Crystal violet staining and EF microscopy were sufficient to determine that all four serotype are capable of forming biofilm and determine a hierarchy of biofilm forming ability. Therefore, for the purpose of this work it was not deemed necessary to perform CLSM at this point.

Overall, these data have shown that all four pneumococcal serotypes [serotype 1 (ST306), 3, (ST3205), 14 (ST124), and 22F (ST433)] are capable of forming biofilm under static conditions when grown for nine days. These data indicate that serotype 22F formed good biofilms: with >60 % percentage coverage by day 9, serotypes 1 and 14 forms moderate biofilms: with >40 % coverage by day 9 and serotype 3 formed poor biofilms: with <20 % coverage by day 9. The next stage of the project was to assess the presence of morphological variance when clinical strains of *S. pneumoniae* are grown under biofilm conditions which will be explored in chapter 5. Thus far this work has shown that colonies harvested from multiple replicates of duplicate independent biofilm experiments exhibit morphology variation and

phenotypic variation from the parent strain. Previous work has reported biofilm-derived morphological variation in serotypes 3 and 19F (Allegrucci and Sauer, 2007, Allegrucci and Sauer, 2008). This work extends previous studies by reporting the phenomenon of colony morphology variants harvested from biofilm in previously unstudied serotypes. Nomenclature of the colony morphology was based on the following metrics; TCVs were defined as colonies with diameters equal to or greater than the size of the wild type. SCVs were defined as colonies with diameters less than half the size of the wild type/TCV diameter, and MCVs were defined as colonies with diameters greater than half the size, but less than the full size of the wild type/TCV diameter.

Initial studies to test for reversion of the different morphologies revealed that a high proportion of MCVs and SCVs of serotypes 1 and 3 reverted back to the wild type colonies after a single subculture whereas SCVs of serotype 22F remained stable after multiple passages. The high level of reversion in serotypes was not unexpected and a similar example of variant reversion has been reported (Allegrucci and Sauer, 2008). This may suggest that the variation seen in serotype 1 and 3 are a result of phase variation as opposed to a stable genetic change. The possibility that the colony variation is a result of nutritional deficiencies is unlikely as all sub-cultured variants were well isolated from other colonies and incubated for an equal amount of time. A more plausible reason for the rapid reversion of SCVs to TCVs is that biofilms are densely packed in nature so phase variable mutations which permit small colony growth may be selected for however upon the removal of the selection pressure biofilms inherently produce, growth is immediately reverted to the wild type morphology. Conversely the relative stability of the SCVs seen in 14 and 22F suggests a more permanent change has arisen. As no SCVs were present in planktonic cultures this would suggest the hypothesis that the phenomenon of SCVs is an adaptive response to growth under biofilm conditions (adaptive mutation) as opposed to concept that the mutations responsible for SCVs are present in the initial population and they are selected for under biofilm growth (natural selection). Whole genome sequencing of both SCVs and the planktonic inoculum would help elucidate this by identifying SNPs, deletions of recombination events which may be attributed to the SCV phenotype.

To account for any phenotypic change seen, and to better understand both the phenotypic and genomic diversification during biofilm development, colony variants will be characterised phenotypically in chapter 4 and the whole genomes will be sequenced using next generation sequencing platforms in chapter 5. In doing so this project aims to investigate the

level of diversity which arises in clinical strains of pneumococci and give insight into the phenotypic and genetic diversity which could arise during pneumococcal colonisation to help inform future vaccine design.

Chapter 4

Phenotypic characterisation of biofilm-derived pneumococcal variants

4 Chapter 4: Phenotypic characterisation of biofilm-derived pneumococcal variants

4.1 Introduction

To understand the level of phenotypic and genetic diversity that arises in *S. pneumoniae* as a result of biofilm growth, a pilot study entitled *An Integrated Systems Biology Approach to Bacterial Biofilm Development*, was performed. This study aims to characterise and understand in more detail the phenotypic diversity associated with biofilm-derived colony variants of *S. pneumoniae*. The work assesses the ability of colony variants to establish new biofilms relative to the parent strain, and compares their interaction with surfaces. It also aims to use quantitative proteomics to determine changes to the proteome of variants compared to the parent strain

In achieving these objectives this pilot project will help understand the role of biofilms in driving phenotypic and genetic diversification in new variant lineages. In doing so a causal link between the phenotypic diversity and the genetic diversity of the biofilm-derived variants compared to the parent strain should be determined.

Serotype 22F was selected as the model serotype due to its ability to form good biofilms and due to the relative stability of the morphological variants. As this project aims to identify stable genetic mutations only the SCVs and TCVs morphological variants were selected for phenotypic and genetic analysis. Biofilms were cultured as described in the materials and methods section 2.4 and biofilm variants were harvested as depicted in Figure 2-1. To ensure that the variants were *S. pneumoniae* and not contaminants each variant underwent Gram staining, PCR for the 22F capsule gene and the API Rapid ID 32 strep assay. These assays were initially performed on 10 of each biofilm-derived variant (SCV or TCV) of serotype 22F. Five SCVs and 5 TCVs, from three separate biofilm replicates, were subsequently selected for WGS and proteomic analysis.

4.2 Materials and Methods

See sections 2.1 to 2.15 and 2.26 for more detail.

4.3 **Results**

4.3.1 API Rapid ID 32 Strep assay

The API Rapid ID 32 Strep assay was used to simultaneously assess changes in the phenotypic profile of the biofilm-derived small and typical colony variants of serotype 22F and to confirm that the variants were *S. pneumoniae*. The results of the API Rapid ID 32 Strep assay can be seen in Table 4-1 and Table 4-2. Analysis revealed that 5/10 SCVs were unable to metabolise one or more of the carbon substrates; maltose (MAL), pullulan (PUL), D-trehalose (TRE), and D-saccharose [sucrose] (SAC) (Table 4-1). In contrast, the TCVs tested revealed no deficiency in metabolising the carbon substrates. The results obtained for the wild type 22F parent strain concurred with the *S. pneumoniae* profile from the table of identification in the API ID 32 strep handbook, suggesting the test is accurate and that the differences seen in the biofilm-derived variants are genuine. These data show that despite differences in the API profiles all TCVs and SCVs revealed a high percentage of similarity to *S. pneumoniae* (Table 4-3).

Table 4-1: API Rapid ID 32 Strep assay profile for the biofilm-derived small colony variants [Ψ]

	ADH	BGLU	BGAR	BGUR	aGAL	PAL	RIB	MAN	SOR	LAC	TRE	RAF	SAC	LARA	DARL	CDEX	VP	APPA	BGAL	PyrA	BNAG	GTA	HIP	GLYG	PUL	MAL	MEL	MLZ	MBDG	TAG	BMAN	URE
SCV1 D3	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	-
SCV2 D9	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	-
SCV3 D3	-	-	+	-	+	-	-	-	-	+	-	+	+	-	-	-	-	+	-	-	+	+	-	-	-	+	-	-	-	-	-	-
SCV5 D3	-	-	+	-	+	-	-	-	-	+	-	+	-	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	-
SCV8 D3	-	-	+	-	+	-	-	-	-	+	-	+	+	-	-	-	-	+	-	-	+	+	-	-	-	+	-	-	-	-	-	-
WT	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	-

	ADH	BGLU	BGAR	BGUR	aGAL	PAL	RIB	MAN	SOR	LAC	TRE	RAF	SAC	LARA	DARL	CDEX	VP	APPA	BGAL	PyrA	BNAG	GTA	HIP	GLYG	PUL	MAL	MEL	MLZ	MBDG	TAG	BMAN	URE
SCV3 D9	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	-
SCV5 D9	-	-	+	-	+	-	-	-	-	+	-	+	+	-	-	-	-	+	-	-	+	+	-	-	-	-	-	-	-	-	-	-
SCV6 D9	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	-
SCV7 D9	-	-	+	-	+	-	-	-	-	+	-	+	+	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	-
SCV9 D9	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	-
WT	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	-

[Ψ] Positive symbols indicate active metabolism, negative symbols indicate no metabolism. Results were recorded visually using the table of identification in the manufacturer's protocol booklet. Results which differ from the WT profile are coloured red. The full list of substrates can be seen in Appendix 1

Table 4-2: API Rapid ID 32 Strep assay profile for the biofilm-derived typical colony variants [Ψ]

	ADH	BGLU	BGAR	BGUR	aGAL	PAL	RIB	MAN	SOR	LAC	TRE	RAF	SAC	LARA	DARL	CDEX	VP	APPA	BGAL	PyrA	BNAG	GTA	HIP	GLYG	PUL	MAL	MEL	MLZ	MBDG	TAG	BMAN	URE
TCV1 D9	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	-
TCV2 D9	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	-
TCV3 D9	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	-
TCV4 D9	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	-
TCV5 D9	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	-
WT	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	-

	ADH	BGLU	BGAR	BGUR	aGAL	PAL	RIB	MAN	SOR	LAC	TRE	RAF	SAC	LARA	DARL	CDEX	VP	APPA	BGAL	PyrA	BNAG	GTA	HIP	GLYG	PUL	MAL	MEL	MLZ	MBDG	TAG	BMAN	URE
TCV6 D9	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	-
TCV7 D9	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	-
TCV8 D9	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	-
TCV9 D9	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	-
TCV10 D9	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	-
WT	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	-

[Ψ] Positive symbols indicate active metabolism, negative symbols indicate no metabolism. Results were recorded visually using the table of identification in the manufacturer's protocol booklet. Results which differ from the WT profile are coloured red. The full list of substrates can be seen in Appendix 1.

Table 4-3: API*web* identification of the biofilm-derived variants using the Rapid ID 32 Strep assay tests

Isolate	*APIweb* percentage similarity to *S. pneumonie* (%)[Ψ]	*APIweb* percentage similarity to *S. oralis* (%)	*APIweb* percentage similarity to *S. parasanguinis* (%)
WT 22F	***99.9***	***0.1***	***-***
TCV1D9	99.9	0.1	-
TCV2D9	99.9	0.1	-
TCV3D9	99.9	0.1	-
TCV4D9	99.9	0.1	-
TCV5D9	***99.1***	***0.7***	***-***
TCV6D9	***99.9***	***0.1***	***-***
TCV7D9	***99.9***	***0.1***	***-***
TCV8D9	99.9	0.1	-
TCV9D9	***99.9***	***0.1***	***-***
TCV10D9	***99.9***	***0.1***	***-***
SCV1D3	***99.1***	***0.7***	***-***
SCV2D9	99.1	0.7	-
SCV3D3	98.7	-	0.7
SCV5D3	***96.2***	***-***	***1.4***
SCV8D3	96.2	-	1.4
SCV3D9	***99.1***	***0.7***	***-***
SCV5D9	98.5	-	0.4
SCV6D9	99.1	0.7	-
SCV7D9	***94.4***	***3.2***	***-***
SCV9D9	***99.8***	***-***	***0.1***

[Ψ] Where 99.9 % similarity was not achieved alternative species similarity was recorded. Those isolates highlighted in blue were selected for whole genome sequencing.

4.3.2 Serotyping of the 22F biofilm-derived variants

To assess for changes in serotype each variant tested underwent PCR for the 22F gene at 643 basepairs (bp) and the pneumococcal *cps*A gene at 160 bp. All variants tested were PCR positive for both bands indicating all isolates are *S. pneumoniae* and no bands were present in the negative control (Figure 4-1). As all isolates were positive for the 22F gene, this would indicate that no change in serotype has occurred.

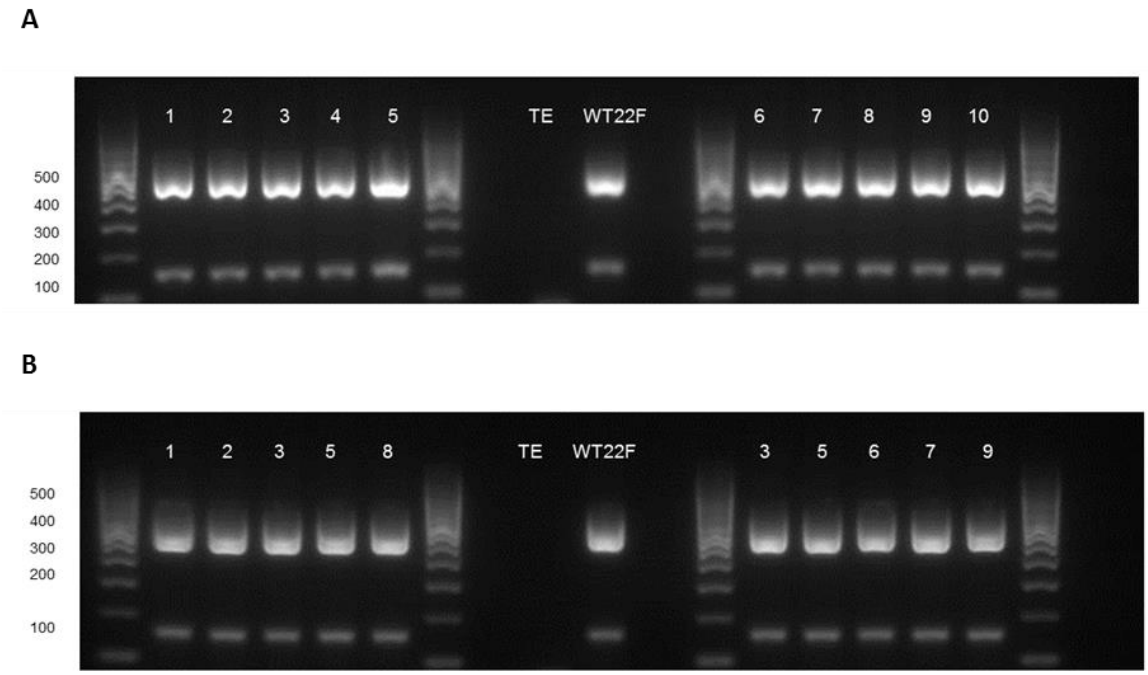


Figure 4-1: *cps*A and 22F PCR of the WT and biofilm-derived SCVs and TCVs

PCR for the presence of both the 22F capsule gene at 643 bp and the *cps*A gene at 160 bp. Gel electrophoresis of the typical colony variants (A) and the small colony variants (B) harvested from serotype 22F biofilms. The PCR products were run out on 2 % agar gel for 1.5 hours at 125 V and stained with Gel Red. Bioline hyperladder^{TM} VI was used to assess size of the bands.

4.3.3 Gram staining of the 22F biofilm-derived variants

All further phenotypic analysis is focused on the variants chosen to undergo genome sequencing (Table 4-3). To confirm that the biofilm-derived variants were *S. pneumoniae* and to corroborate the PCR data, all variants chosen to undergo genome sequencing were stained with Gram's crystal violet to confirm the cells were Gram-positive cocci. All variants were confirmed to be Gram positive and have cocci morphology. Presence of any pink cells was accounted for by over decolourisation of the crystal violet stain with the 95 % ethanol or unwashed Gram's safranin solution.

4.3.4 Initial attachment assay of the 22F biofilm-derived variants

To assess the difference in initial attachment of the biofilm-derived small and typical colony variants of serotype 22F, variants were grown under biofilm condition in 48 well plates for 24 hours and subsequently stained with crystal violet. Samples TCV5D9, TCV6D9 TCV7D9 and SCV9D9 had higher absorbance values than the wild type; all other samples had lower absorbance readings (Figure 4-2). All datasets had a normal distribution and equal variance. A 2-sample t-test revealed no significant difference between the different variants compared to the wild type parent strain with the exception of SCV1D3 (P = 0.026).

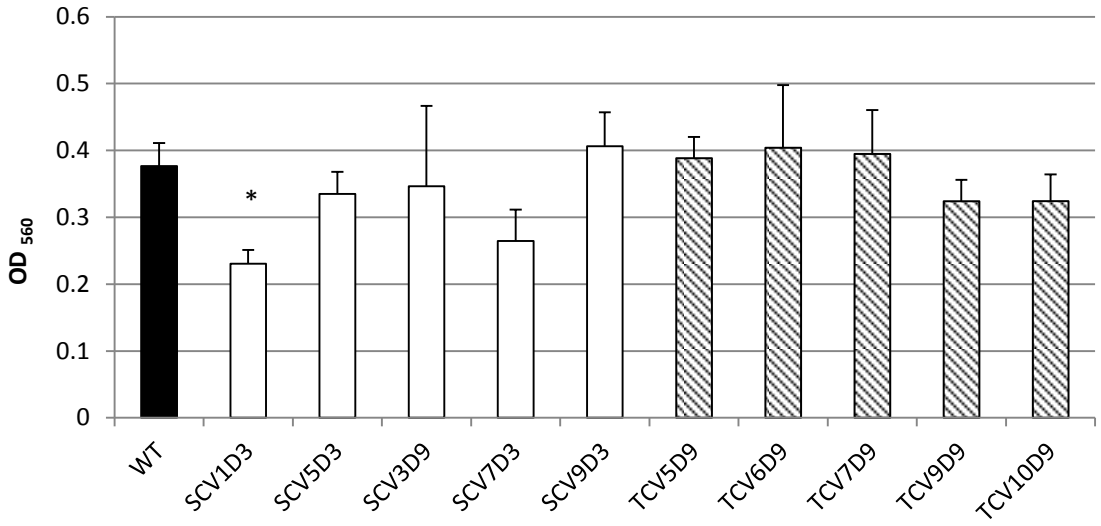


Figure 4-2: Crystal violet staining of 24 hour variant biofilms

Pneumococcal variants were cultured in 48 well plates and allowed to attach for 24 hour before being stained with crystal violet, re-solubilized in 95 % ethanol and read on a FLOUstar Optima plate reader at 560 nm. Six replicates for each serotype and negative control were performed per plate and averages were taken. Readings were subtracted from the control sample average to give a final absorbance value. A total of 3 independent experiments were performed using three independent inocula on separate days. A 2-sample t-test was used to compare each variant to the WT. Asterisks represent significant differences initial attachment compared to the WT parent strain (where p = <0.05). Error bars represent standard error.

4.3.5 Biofilm formation of the 22F biofilm-derived variants

To assess for differences in biofilm forming ability of the pneumococcal variants, biofilms were cultured for a total of 9 days and visualised using epi-fluorescence microscopy. Biofilms were seeded with 1 x 10^8 CFU mL^{-1} of the relevant serotype and inoculated into MatTek culture plates, cultured for 1, 3, 6 and 9 days and imaged using an EDIC/EF microscope and the *Bac*Light live/dead stain. Visually SCVs appeared to have more structured biofilms with larger microcolonies, whereas TCVs appeared to have flatter less structure biofilms, resembling a monolayer of cells. Statistically all datasets had a normal distribution and equal variance. A 2-sample t-test revealed a significant difference compared to the wild type parent strain ($P = <0.05$) at both day 3 and day 6; All SCVs and TCVs had significantly different means compared to the wild type strain by day 3 with the exception of TCV5D9 ($P = 0.714$) (Figure 4-3A) and by day 6 with the exception of TCV5D9 ($P = 0.169$) (Figure 4-3B).

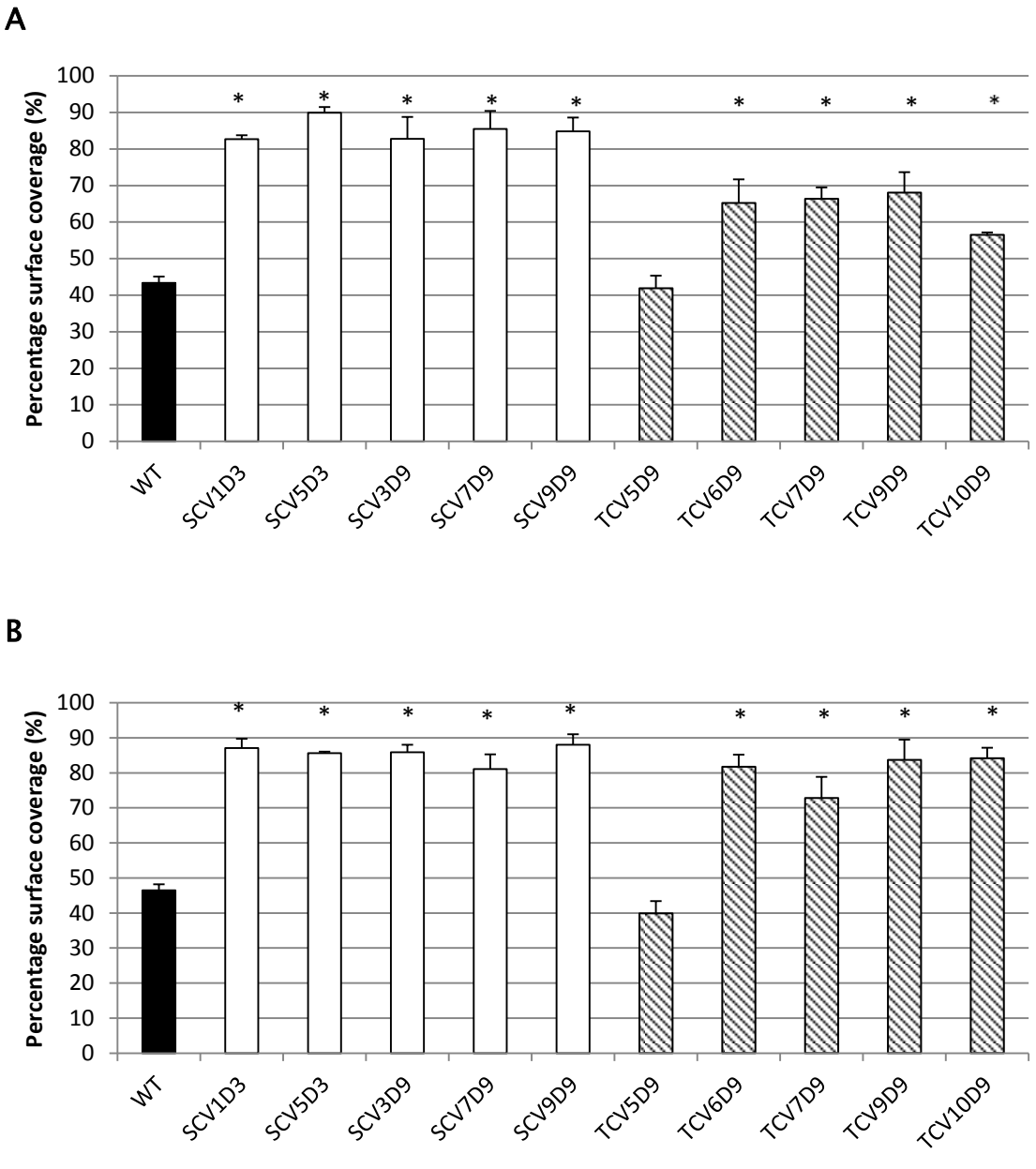


Figure 4-3: Percentage surface coverage of pneumococcal variant biofilms

Biofilm formation was observed using a Nikon Eclipse LV100 epi-fluorescent microscope and the *Bac*Light live/dead stain. Triplicate biofilms were cultured in MatTek culture plates for 3 (A) and 6 (B) days. A total of 10 fields of view were taken, per sample, per time point, at 400x magnification with the Image-Pro 6.2 software and analysed using ImageJ software. A 2-sample t-test was used to compare each variant to the WT. Asterisks represent significant differences in percentage coverage of biofilm compared to the WT parent strain (where $p = <0.05$). Error bars represent standard error.

To verify these findings and assess the difference in biofilm depth, biofilm-derived variants (SCV, MCV, TCV) harvested from different biofilm replicates of day 9 biofilms and the wild type 22F parent strain were visualised on the Leica TCS SP5 confocal laser scanning microscope (Figure 4-4 A-D). CLSM revealed that the wild type parent produced relatively flat biofilms with an average thickness of approximately 6 µm. The wild type parent produced small microcolony towers with maximum thickness ranging between 12-15 µm (Figure 4-4A). In contrast the small colony variant produced structured biofilms with an average thickness of approximately 13 µm. Microcolony towers were well-defined with maximum thickness ranging between 25-35 µm (Figure 4-4B). Statistically all datasets had normal distribution and equal variance. The 2-sample t-test revealed that SCV biofilms had significantly greater biomass ($\mu m^3/\mu m^2$) ($p = 0.031$), surface area (μm^2) ($p = 0.024$), average thickness (µm) ($p = 0.048$) and maximum thickness (µm) ($p = 0.005$) compared to the parent strain (Figure 4-5). MCVs produced slightly larger biofilms with an average thickness of approximately 10 µm thick and maximum thickness ranging between 15-20 µm (Figure 4-4C). TCVs produced biofilm comparable with the wild type strain (Figure 4-4D); microcolony towers were small with a maximum thickness of 13-14 µm and an average thickness of 8 µm. Biofilm produced by the MCVs and TCVs did not differ significantly from the parent strain (Figure 4-5).

To estimate the percentage of dead cells within the biofilms, the biomass of live cells (SYTO®9 only channels) was subtracted from the biomass of the total biofilm (merged SYTO®9 and propidium iodide channels); this value was then used to determine percentage dead in the total biofilm. The percentage of dead cells within the wild type and SCV biofilms were comparable at 11 % and 9 % respectively. The difference in the percentage of dead cells was not significant ($p = 0.431$). Interestingly the MCV and TCV biofilms had a higher percentage of dead cells with 25 % and 17 % respectively. The difference in the percentage of dead cells compared to the wild type was not significant ($p_{MCV} = 0.155$ and $p_{TCV} = 0.355$). This is likely due to the level of standard error between replicates.

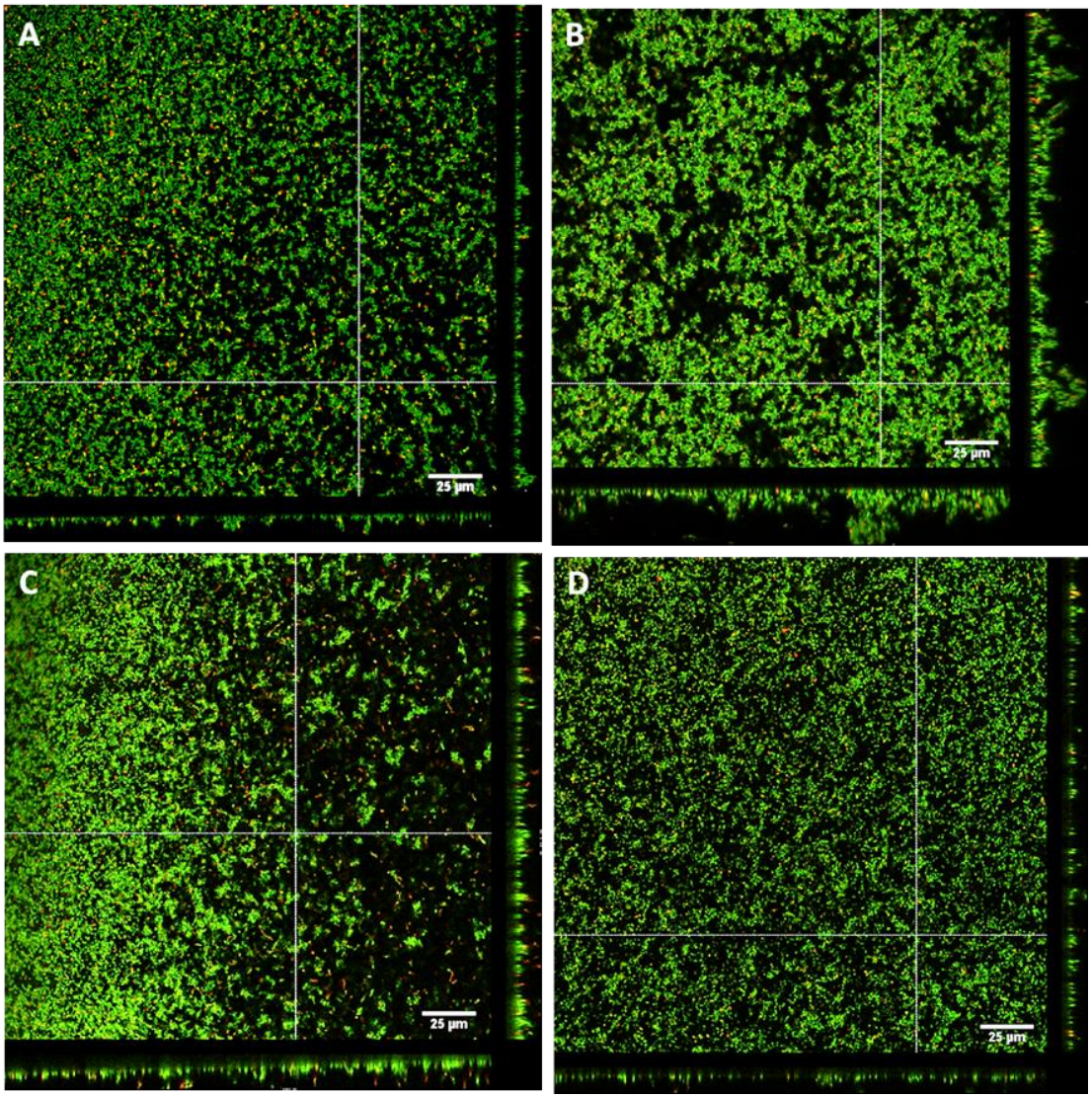


Figure 4-4: Day 3 biofilms of biofilm-derived colony variants

Day 3 biofilms of the wild type parent (A), small colony variant (B), medium colony variant (C), and typical colony variant (D). Biofilms were visualised on a Leica TCS SP5 confocal laser scanning microscope using *Bac*Light live/dead stain. Green cells indicate live cells and red cells indicate dead cells. Five fields of view were taken per triplicate biofilm to get an accurate indication of biofilm growth. Z-Scans were performed every 0.3 µm on each field of view. White bars represent 25 µm.

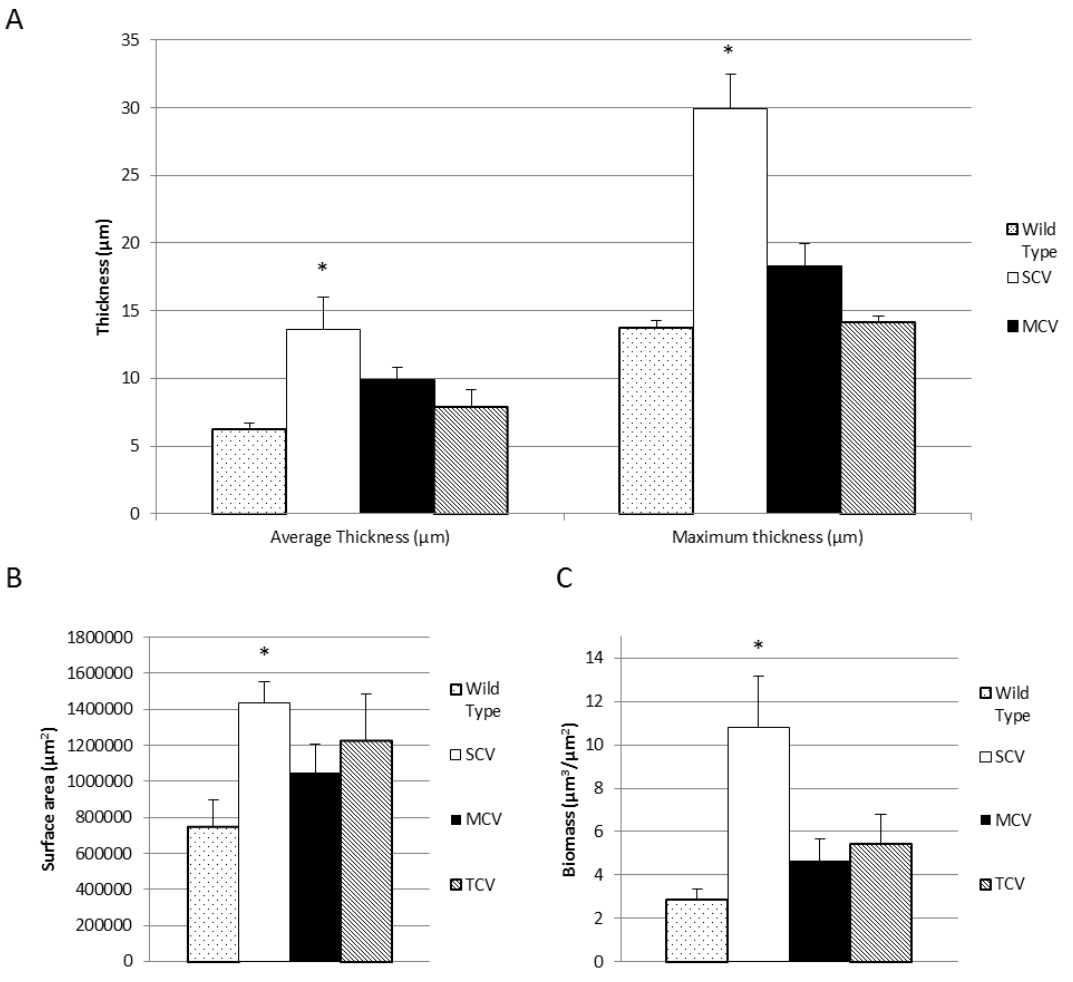


Figure 4-5: COMSTAT analysis of biofilm-derived colony variants

Biofilm formation was observed using CLSM and the *Bac*Light live/dead stain. Analysis of the confocal images was performed using COMSTAT 1 to generate data on A) average thickness maximum thickness, B) surface area and C) biomass. A 2-sample t-test was used to determine whether there was a significant difference between each variant and the WT parent. Asterisks represent significant differences compared to the WT parent strain. Error bars represent standard error.

4.3.6 Antibiotic susceptibility of the 22F biofilm-derived variants

Minimum inhibitory concentration (M.I.C) strips were used to test whether the biofilm-derived colony variants of serotype 22F had different antibiotic susceptibilities to four commonly used antibiotics [penicillin, erythromycin, tetracycline and ceftriaxone]. All SCVs had had increased sensitivity to penicillin, 4/5 SCVs also had an increased sensitivity to erythromycin and 3/5 SCVs had an increased sensitivity to tetracycline (Figure 4-6).

In contrast all TCVs tested had no difference in antibiotic sensitivity compared to the wild type strain. No difference in sensitivity to ceftriaxone was observed in all the variants.

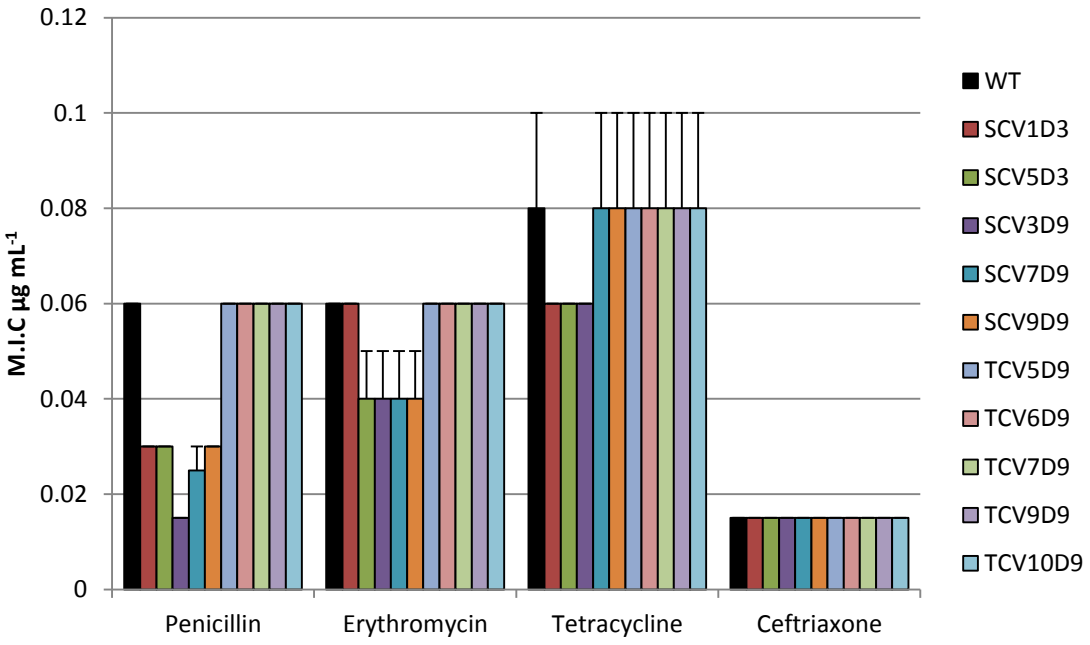


Figure 4-6: Minimum inhibitory concentration (M.I.C) antibiotic resistance assay

M.I.C strips were used to test whether the biofilm-derived colony variants of serotype 22F had different antibiotic susceptibilities to 4 commonly used antibiotics [penicillin, erythromycin, tetracycline and ceftriaxone]. After incubation for 18 hours at 37 ^{o}C in 5 % CO_2 zones of inhibition were measured according to the manufacturer's instructions. Error bars represent standard error.

4.3.7 Growth curves of the 22F biofilm-derived variants

To determine the growth profile of the each variant, variants were grown in BHI broth for a total of 10 hours. OD_{600} readings were taken hourly and CFU counts were taken every two hours. SCVs recorded an increased lag phase of approximately 2 hours and lower OD_{600} values (Figure 4-7) and CFU counts (Figure 4-8) by 8 and 10 hours compared to the wild type and TCVs.

Furthermore to confirm a difference in growth rate; CFU data from the exponential growth phase (between 4 and 8 hours of growth) of the SCVs was compared to the wild type. All samples recorded lower mean rates compared to the wild type strain. All datasets had a normal distribution and equal variance. A 2-sample t-test was used to compare all the SCVs to the wild type. The results showed that there was a significant difference in growth rate during the exponential growth phase between the SCVs and the wild type ($p = <0.05$) with the exception of SCV5D3 ($p = 0.08$) (Figure 4-9). The growth rate was calculated using the following formula:

$$\mu = ((\log10\ N - \log10\ N_0)\ 2.303) / (t - t_0))$$

Where:

N = final CFU count (time point 8 hours)

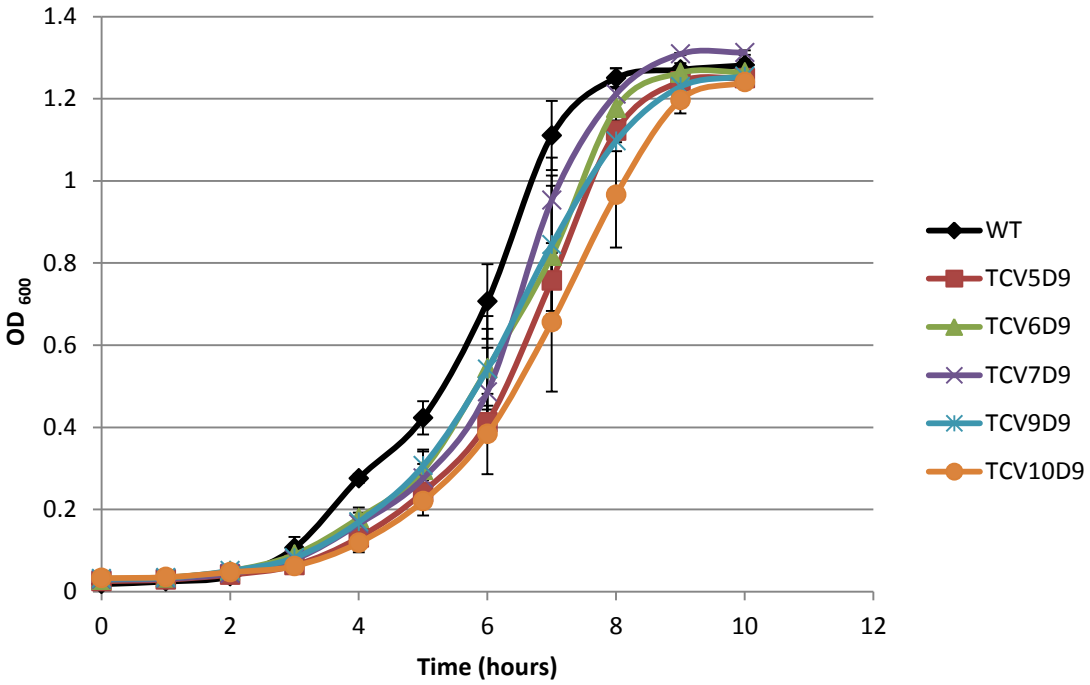
N_0 = initial CFU count (time point 4 hours)

t = initial time (time point 4 hours)

t_0 = final time (time point 8 hours)

μ = growth rate

A



B

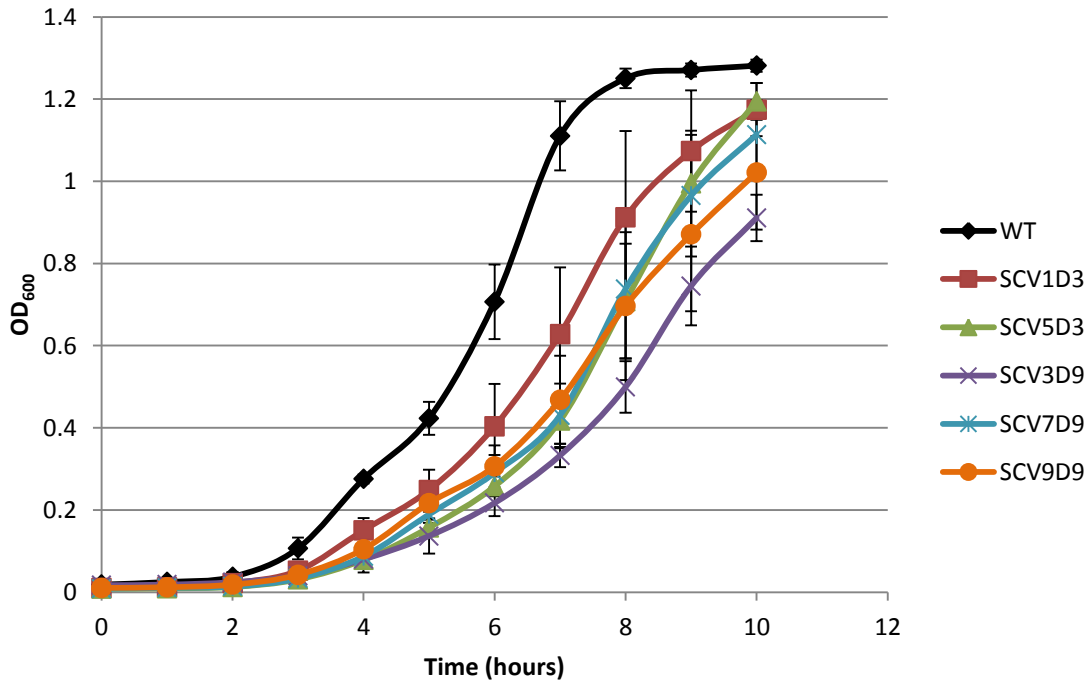


Figure 4-7: Optical density growth curve of biofilm-derived variants

The growth profiles were assessed for typical colony variants (A) and small colony variants (B) chosen to undergo whole genome sequencing. Variants were grown in BHI broth for a total of 10 hours. OD_{600} readings were taken every hour for 10 hours. Error bars represent standard error.

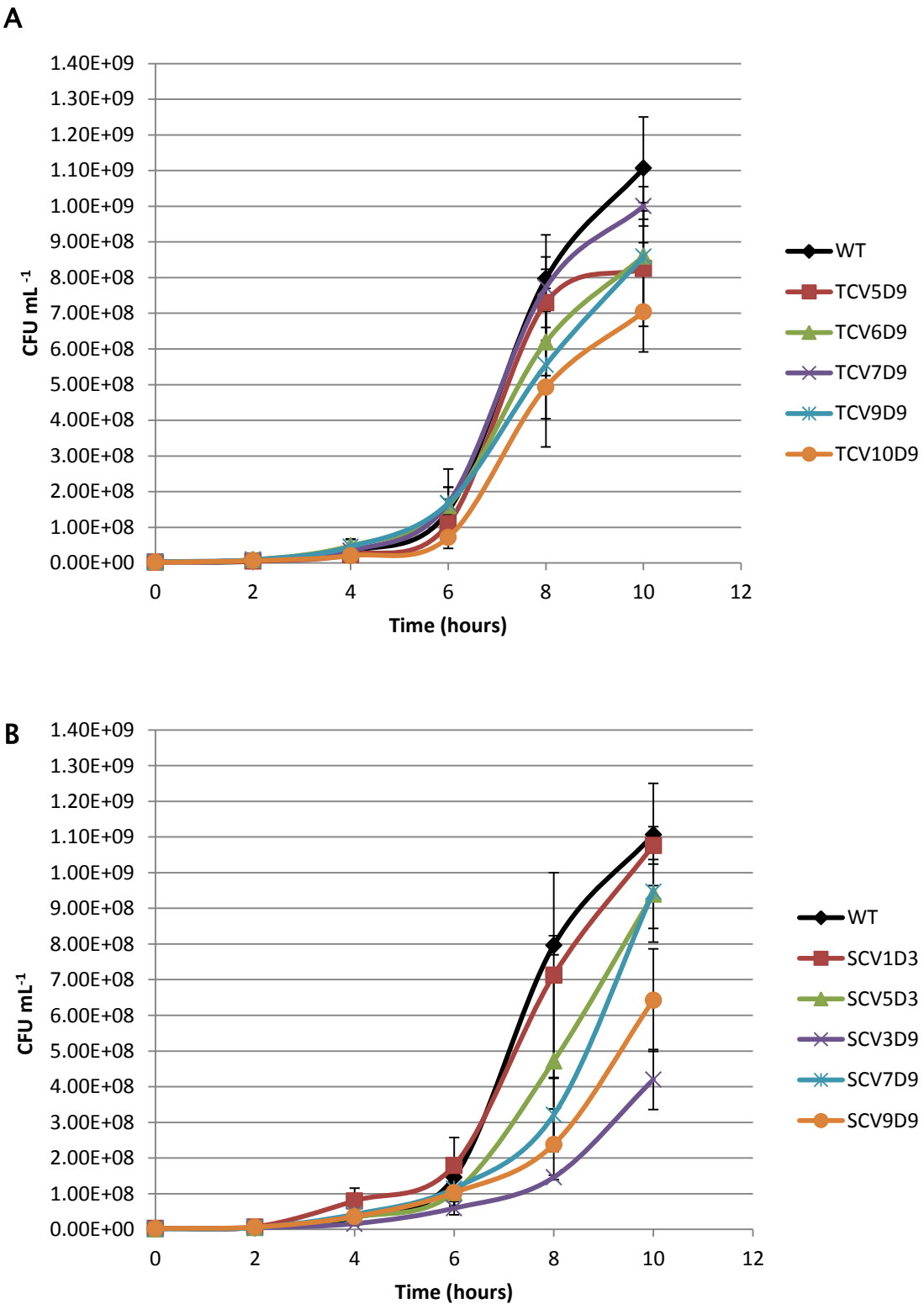


Figure 4-8: Colony Forming Unit growth curve of biofilm-derived variants

The growth profiles were assessed for typical colony variants (A) and small colony variants (B) chosen to undergo whole genome sequencing. Variants were grown in BHI broth for a total of 10 hours. CFU counts were taken every two hours. Error bars represent standard error.

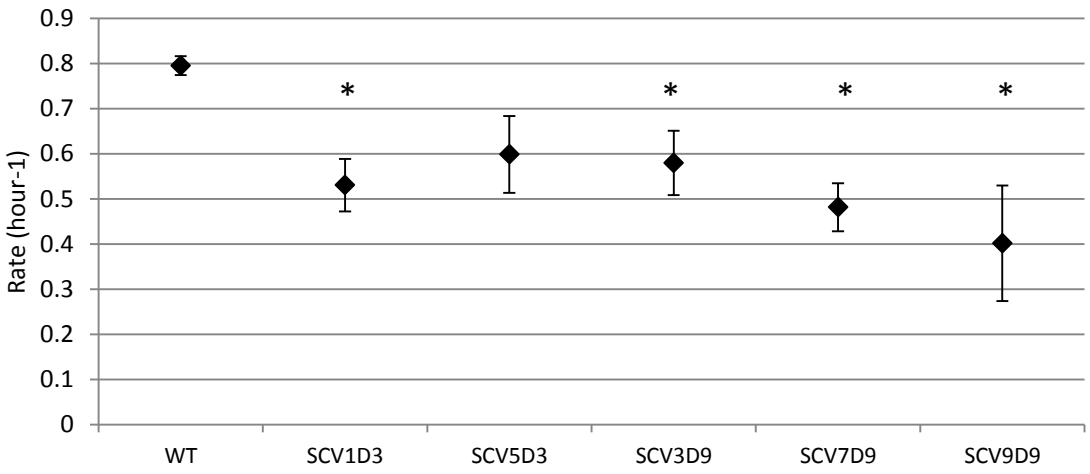


Figure 4-9: Growth rate of biofilm-derived small colony variants

The growth rate was assessed for all SCVs using the equation $\mu = ((\log10\ N - \log10\ N_0)\ 2.303) / (t - t_0))$ and CFU data from the exponential growth phase (between 4 and 8 hours of growth). A 2-sample t-test was used to determine whether there was a significant difference in growth rate between each variant and the WT parent. Asterisks denote SCVs with significant different growth rates compared to the WT (where $p =<0.05$). Error bars represent standard error.

4.3.8 Capsule staining of the 22F biofilm-derived variants

The capsule of *S. pneumoniae* acts as a major virulence factor and remains the primary target for vaccine design. Moreover studies have shown the capsule to be an important factor in colonization, survival, dissemination and attachment of pneumococci to host cells (Hammerschmidt *et al*., 2005). Stains-all solution was used to quantify the amount of capsule the biofilm-derived variants as described in section 2.14. A 2-sample t-test revealed a significant decrease ($p = <0.05$) in capsule staining in the small colony variants compared to both the typical colony variants and wild type parent strain (Figure 4-10). Despite the fact that the typical colony variants appear to have greater capsule production, this difference was not significant. All cultures were grown to approximately 10^8 CFU mL^{-1}, rejecting the possibility that cell number is responsible for the difference seen.

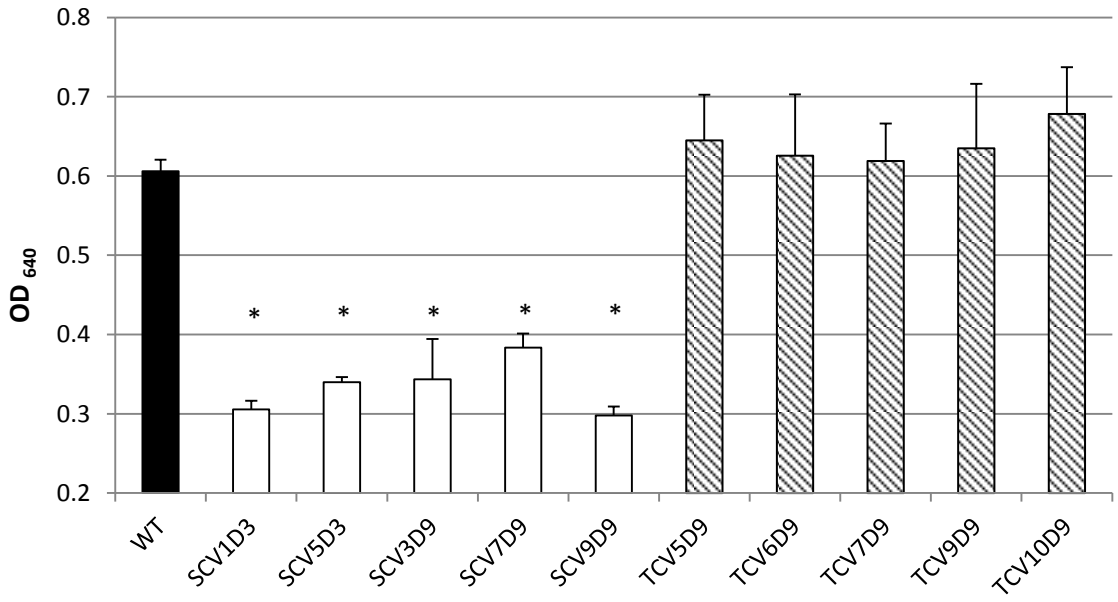


Figure 4-10: Capsule quantification of the biofilm-derived pneumococcal variants

Biofilm-derived pneumococcal variants were cultured in BHI to an optical density of approximately 10^8 CFU mL^{-1} and pelleted. Cells were re-suspended and stained with Stain-all solution to quantify the capsule. The absorbance was measured on a spectrophotometer at 640 nm and subtracted from the negative control (250 µL of sterile distilled water stained with 1 mL of Stains-all). A 2-sample t-test was used to determine whether there was a significant difference in capsule staining between each variant and the WT parent. Asterisks denote significantly different values from the WT (where $p = <0.05$). Error bars represent standard error.

4.3.9 Mutation frequency of the 22F biofilm-derived variants

Prior to WGS the rifampicin assay was performed as previous described (Lee *et al*., 2010) to indicate whether the biofilm-derived variants had an increased mutation frequency compared to the parent strain as a result of biofilm growth. The assay was performed in triplicate for each sample, using independent bacterial cultures. The mutation frequency values were reported as the proportion of rifampicin-resistant colonies (after 48 hours of incubation in a 5 % CO_2 atmosphere) compared to the total viable cell counts.

Table 4-4: Rifampicin mutation frequency of biofilm-derived pneumococcal variants

Sample	Mutation Frequency
WT	5.9×10^{-08}
SCV1D3	3.9×10^{-09}
SCV5D3	1.2×10^{-08}
SCV3D9	4.2×10^{-08}
SCV7D9*	1.1×10^{-07}
SCV9D9	1.4×10^{-08}
TCV5D9*	9.5×10^{-08}
TCV6D9	5.2×10^{-08}
TCV7D9	3.0×10^{-08}
TCV9D9	3.1×10^{-08}
TCV10D9	3.5×10^{-08}

*A mutator strain refers to an isolate with a mutation frequency $\geq 7.5 \times 10^{-8}$ in accordance with Morosini *et al.* (2003).

Table 4-4 represents the mean mutation frequency from colony variants harvested from a serotype 22F biofilm. Based on previous work with mutation frequency in *S. pneumoniae*, a mutator strain was defined as an isolate with a mutation frequency $\geq 7.5 \times 10^{-8}$ (Morosini *et al*., 2003). In this work two isolates would be considered mutator strains, SCV7D9 and TCV5D9 which had mean mutation frequencies of 1.1×10^{-07} and 9.5×10^{-08} respectively. Interestingly, all colony variants (with the exception of SCV7 and TCV5) had reduced mutation frequencies compared to the wild type parent strain. There appears to be no association between the variant morphology and the mutation frequency.

4.4 **Discussion of Results**

The aim of this chapter was to establish the level of phenotypic and genetic diversity that arises in pneumococci as a result of growth under biofilm conditions. Due to the stability of serotype 22F SCVs, phenotypic analysis has focused primarily on five serotype 22F small colony variants and five serotype 22F typical colony variants, selected to undergo WGS with the wild type strain (Table 4-3). Phenotyping was achieved by performing assays including the API Rapid ID 32 Strep assay, crystal violet initial attachment assay, antibiotic resistance assay and by assessing growth rate, capsule production and mutation frequency.

The API Rapid ID 32 assay was used to see if there is any change in phenotypic profile of the variants by comparing the phenotypic profile to the wild type serotype 22F strain. Interestingly, the phenotypic analysis revealed that 5 of the 10 SCVs tested were unable to metabolise one or all of the carbon substrates, including maltose, pullulan, D-trehalose, and D-saccharose. Of these substrates, trehalose and pullulan were the only substrates not metabolised in multiple SCVs. Trehalose is a disaccharide which broken down into glucose by the catabolic enzyme trehalose-6-phosphate hydrolase *tre*A (predicted from Gene Ontology). Pullulan is a polysaccharide which consists of three glucose monomers and is metabolised by pullulanases (Bongaerts *et al*., 2000), such as *spu*A (predicted from Gene Ontology). To date the metabolic pathways concerning trehalose and pullulan in *S. pneumoniae* are not well-defined; as such, the clinical significance of this finding remains unclear.

S. pneumoniae is a nutritionally fastidious facultative anaerobe. Carbohydrates are converted to pyruvate via glycolysis (see Figure 7-1) and are the only nutrients which support growth and cell division in *S. pneumoniae* (Hoskins *et al*., 2001), but the availability of a glucose source within the human nasopharynx is thought to be in low concentrations (Philips *et al*., 2003). There is a lack of research into the metabolic profile of the human airway which may support the growth of the pneumococcus during colonisation (King, 2010), however it is thought, that one major carbohydrate source comes from the modification of host glycan structures by pneumococcal glycosidases (King, 2010). Mucins are glycosylated proteins produced by epithelial cells in the lung, and *S. pneumoniae* has been shown to grow on mucin (Yesilkaya *et al*., 2008) indicating that the sugar structures within mucin may act as a source of carbohydrate for metabolism. In 2009, a study by Trappetti *et al.* into the effect of different sugars on biofilm growth in pneumococcal strains D39 and TIGR4, showed that of the 27 sugars tested, sialic acid was shown to significantly increase the attachment of bacteria to the

wells after 24 hours, the concentration of which is similar to that seen in human saliva. The same result was shown using soy and yeast based mediums suggesting that it was sialic acid that was responsible for the increase in cell attachment. In a 2012 study, Bidossi *et al.* revealed 32 carbon sources that *S. pneumoniae* may utilise including hyaluronic acid and glycans found on mucin and the human mucosal cell surface. Interestingly there was a notable difference in the generation time of the bacterium when cultured in different carbon source, with glucose being the most efficient carbon source, with generation time of 31 minutes, to hyaluronic acid, with a generation time of 131 minutes (Bidossi *et al.,* 2012). Taken together, these data suggest that the availability of carbon sources in the nasopharynx plays a crucial role in directing sessile growth of pathogens. As biofilm formation is thought to represent a survival strategy in a nutritionally limited environment (Domenech *et al*., 2012), the inability to metabolise all the carbon substrates on the API strip may suggest that the SCVs are "phenotypically programmed" to increase biofilm formation as seen in the EF and CLSM microscopy. Moreover, the inability to metabolise certain sugars may indicate mutations in metabolic enzymes or carbohydrate transporters which may account for the relatively small colony phenotype and reduced capsule. Indeed, a direct link between capsule biosynthesis and basic metabolism has been seen in serotype 3 (Hardy *et al*., 2000). Here, the authors showed that mutations in the glycolytic phosphoglucomutase gene, *pgm*, resulted in a 10 % reduction in capsule compared to the parent strain (Hardy *et al*., 2000). Further discussion regarding carbohydrate metabolism in biofilm-derived SCVs will be discussed in chapter 5.

In this work, the TCVs tested revealed no deficiency in metabolising the carbon substrates and the results obtained for the wild type 22F strain concurred with the *S. pneumoniae* profile from the table of identification in the API ID 32 strep handbook, suggesting the test is accurate and that the differences seen in the biofilm-derived variants are genuine. The use of a more comprehensive assay such as the Biolog^{TM} phenotypic microarray could also be used to fortify or refute these data. Due to the smaller colony phenotype, SCVs required longer growth periods prior to being inoculated into the API strip. This is because the API strip requires a cell density equivalent to McFarland no.4. As a result each SCV, with the exception of SCV1D3 and SCV2D3, had to be processed in duplicate to confirm to a satisfactory percentage (>90 %) that the organism were *S. pneumoniae*. This information is relevant because in routine identification within a clinical microbiology laboratory, small colony variants of *S. pneumoniae* isolated from a clinical sample may be misidentified in the population as an alternative *Streptococcal* species such as *S. mitis*, *Streptococcus oralis*, *Streptococcus parasanguinis* or *Streptococcus pseudopneumoniae*. The inability of the SCVs to metabolise

certain sugars may be due to mutations which have arisen during biofilm growth. Whole genome sequencing of the SCVs should help elucidate the nature and extent of genomic changes within SCVs.

The API Rapid ID 32 assay is a diagnostic assay used to identify *S. pneumoniae* from unknown cultures. The different profiles exhibited by the variants may have suggested that the variants themselves were not in fact *S. pneumoniae*. In order to confirm that all the TCVs and SCVs were *S. pneumoniae* all variants underwent PCR (Figure 4-1) for the 22F gene and the *cps*A capsule gene, specific to *S. pneumoniae* (Park *et al*., 2010). The presence of both the 22F gene and the *cps*A confirmed that all the variants were in fact *S. pneumoniae.* These data, combined with the fact that all the variants were confirmed to be Gram-positive cocci and exhibited alpha-haemolysis and susceptibility to optochin antibiotic meant that the variants could confidently be classed as *S. pneumoniae*.

In order to test whether the inability to metabolise certain sugars in the SCV isolates was a result of slow growth, the growth profiles of the variants was established. Data in Figure 4-7 and Figure 4-8 shows that all SCVs tested had an increased lag phase and lower OD reading compared to the wild type and TCVs. Further analysis revealed that 4/5 SCVs had significantly lower growth rates compared to the wild type. Despite samples SCV1D3, SCV3D9 and SCV9D9 being able to metabolise the full profile of carbon substrates in the API assay the growth rates associated with these phenotypes was still significantly less than the parent strain. These data are interesting as they suggest that the inability to metabolise the sugars in the API assay may not be only cause for the slow growth phenotype of the SCVs. Further work could be performed by comparing growth phenotypes in a range of media without the presence of key carbon substrates, to see if a more apparent difference in growth is present. Potentially mutations have arisen in this variant sub-population which reduces the growth rate. Whole genome sequencing data should help elucidate what mutations are responsible for these changes.

Previous work on SCVs showed the variants to have increased biofilm formation (Allegrucci and Sauer, 2007). This work has shown that all 22F SCVs tested have increased biofilm formation compared to the wild type strain. However, a significant difference was only observed by day 3; crystal violet initial attachment studies revealed four SCVs had no significant difference in attachment and one SCV (SCV1D3) had lower OD_{600} values, indicating reduced attachment. These data suggest that a difference in biofilm forming ability is present but not in the initial stages of cell attachment. CLSM revealed that SCVs produce structured

biofilms with significantly greater biomass, thickness and surface area compared to the parent strain. This is an interesting observation considering the SCVs also display a slow growth phenotype and may seem contradictory. The difference in capsule production may help explain this phenomenon; this work has shown that biofilm-derived SCVs have significantly reduced capsule staining compared to the parent strain, which is indicative of reduced capsule production. Previous work has shown that capsule production is important for biofilm formation and that down-regulation of the capsule is needed to form biofilms (Hall-Stoodley *et al*., 2008). A reduction in capsule seen in the SCVs therefore may in fact aid biofilm production as a reduced capsule has previously been shown to increase biofilm formation (Hall-Stoodley *et al*., 2008). In turn, may help facilitate long-term persistence within the host. CLSM showed TCVs to have a slight increase in surface area compared to wild type 22F strain. The difference was shown to be non-significant. The lack of significance means that the result is not consistent with the percentage coverage in the EF data. This discrepancy between CLSM and EF microscopy data in the TCVs can be accounted for by both the sensitivity of the microscopes used and the image analysis performed. CLSM uses lasers to excite SYTO®9 and propidium iodide at the wavelengths of 485 nm and 535 nm, respectively, to generate more accurate images whereas EF uses filter blocks. Additionally the use of z-scanning in the CLSM images generates a more robust image compared to EF microscopy.

Previous work has also revealed small colony variants which have increased resistance to antibiotics (Besier *et al*., 2008). This work however revealed that all SCVs had increased sensitivity to penicillin, which is likely due to increased penetration due to the putative reduction in capsule. In contrast all TCVs tested had no difference in antibiotic sensitivity compared to the wild type 22F strain. No association was also seen in the mutation frequency of the variants and the morphology of the variant. Indeed the parent strain had higher mutation frequency in all but two of the variants tested. This suggests that mutations that may have arisen in the biofilm variants do not affect the mutation frequency and that all variants should, in theory, have the same likelihood of developing mutations as a result of biofilm growth.

This chapter has shown that biofilm-derived small colony variants have distinct phenotypic diversity from the parent strain including a reduced growth rate, increased sensitivity to penicillin, altered metabolic profile and reduced capsule production coupled with increased biofilm production. This work remains to determine the level of genetic diversity which arises in the biofilm-derived variants and assess whether a mutations within the genome can account for the observed changes in phenotypic profile.

Chapter 5

Whole genome sequencing of pneumococcal biofilm-derived colony variants

5 Chapter 5: Whole genome sequencing of pneumococcal biofilm-derived colony variants

5.1 Introduction

Next generation sequencing (NGS) has revolutionised the way prokaryote genomes are sequenced; allowing detailed high-throughput data generation in a matter of days. Data from the phenotypic analysis of biofilm-derived phenotypic variants has revealed that the small colony phenotypes have a reduced growth rate, increased sensitivity to penicillin, altered metabolic profile and reduced capsule production coupled with increased biofilm production. To further understand the genetic mechanisms responsible for these phenotypic changes, a pilot investigation was performed as described in chapter 4. From this pilot study a total of five 22F SCVs, five 22F TCVs and the parent wild type 22F strain, which have been characterised phenotypically, underwent WGS using the Roche GS Junior^{TM} 454 and Illumina Miseq^{TM} bench top sequencers. WGS was employed to gain a comprehensive overview of possible mutations across the entire genome which can be attributed to growth under biofilm conditions. This work had three primary aims 1) Characterise colony variants genetically using next generation WGS to account for changes in the phenotypic profile. 2) Assess whether biofilm growth can result in mutations which change the serotype or multi-locus sequence type (ST) of the organism, and 3) Assess whether biofilm growth results in mutations within genes relevant to invasive disease and vaccine design. Mutations within vaccine relevant genes may change the epitope structure and prevent antibody binding and as such facilitate vaccine escape. This work will highlight whether growth under biofilm condition could be a mechanism for vaccine escape and furthermore assess for changes in serotype and MLST.

A total of 10 biofilm-derived variants (5 SCVs, and 5 TCVs from 3 replicate biofilms) and the parent wild type 22F strain underwent WGS as described in materials and methods section 2.16 and 2.17. Depiction of the two genomic analysis strategies can be seen in Figure 5-1.

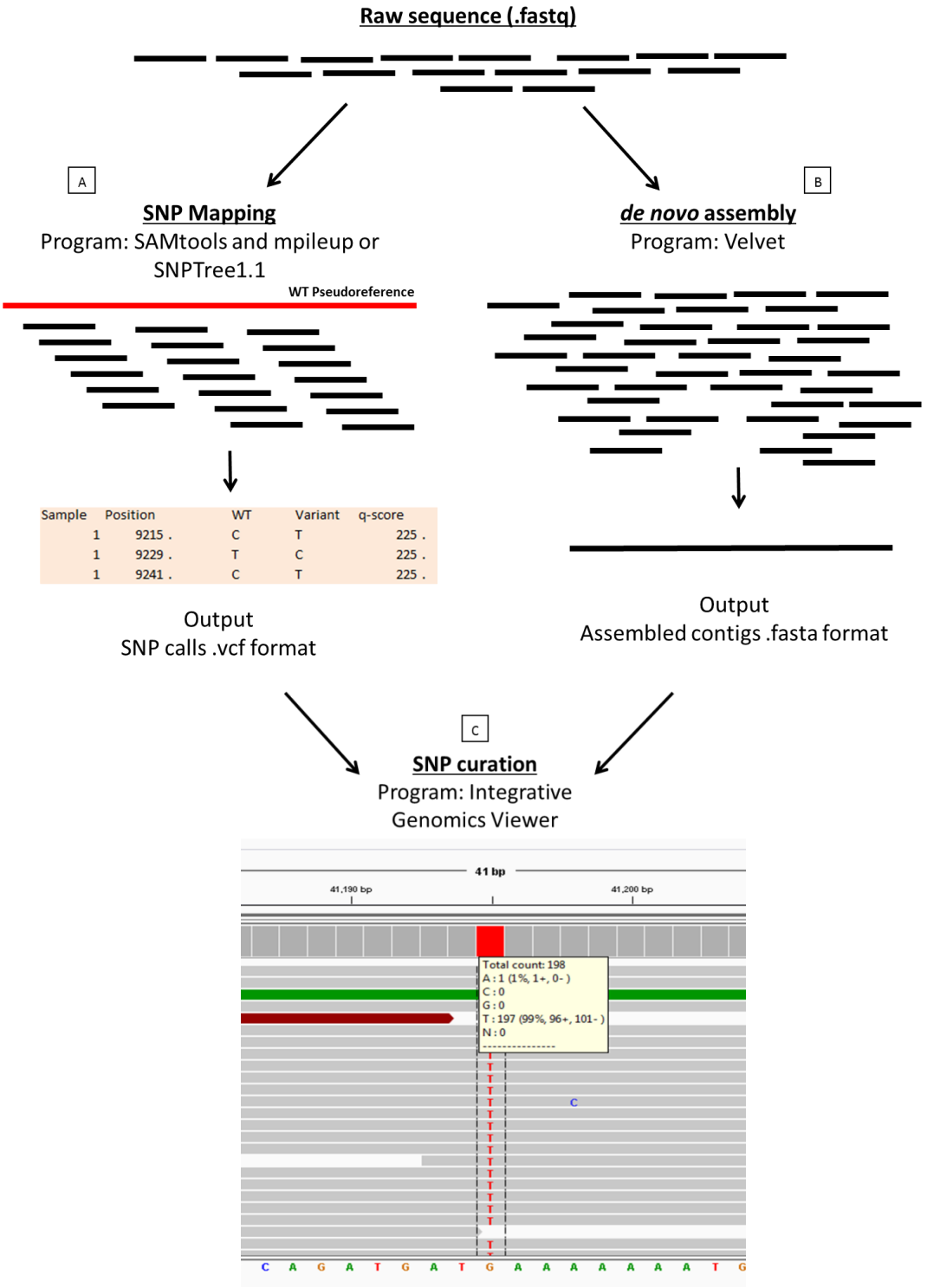


Figure 5-1: Annotation of the next generation sequencing genomic analysis

Using the raw sequence data from the biofilm-derived variants (.fastq files), three methods of analysis were performed. Sequences were mapped against the WT sequence to assess for SNP changes in the genome using the programs SAMtools and Mpileup or SNPTree1.1 which generates a .vcf output table of SNPs and their position in the genome (A). Raw sequence data were also assembled *de novo* using the program Velvet (B). Each SNP was manually curated by visualising the SNP positions in the *de novo* assemblies using Integrative Genomics Viewer (C) to determine coverage and base call quality.

5.2 Materials and Methods

See sections 2.1 to 2.5, 2.8, 2.13, 2.15-2.21 and 2.26 for more detail.

5.3 Results

5.3.1 SNP mapping of biofilm-derived variants

In order to identify SNPs in the biofilm-derived variants, sequence data for each isolate was mapped against the annotated parent strain reference as described in materials and methods section 2.19 to generate variant call files (.vcf) containing the list of SNPs in the biofilm-derived variants compared to the wild type parent strain.

Table 5-1 displays the confirmed mutations identified in this work. The WGS revealed confirmed SNPs in all but one sequenced isolate (TCV10). Notably all sequenced SCVs were shown to contain mutations within the DNA-directed RNA polymerase delta subunit (RpoE) (Table 5-1). These mutations occurred independently at different positions within the gene in separate biofilms and on different days. Furthermore, 4 out of 5 of the SCVs SNPs result in a premature stop codon. In the case of SCV9D9, initially the .vcf output revealed an A-C SNP at position 422795 however on further inspection of the *de novo* assembly there appeared to be a 264 bp deletion. The deletion was subsequently verified using PCR primers designed for the *rpo*E gene (Figure 5-2).

To further verify to quality of SNPs each SNP was manually curated using the Integrative Genomics Viewer (IGV) software (Robinson *et al*., 2011, Thorvaldsdottir *et al*., 2013) to determine coverage at each SNP base and the corresponding Phred scores of each base. Only SNPs which were present in the SNP mapping data and the *de novo* assembled contigs were classed as confident SNPs (Figure 5-1). As an additional metric, the percentage confidence of each SNP was calculated as the coverage of the identified variant divided by the total coverage at the variant position. Variants with >90% confidence were classed as high-confidence mutations. All reported SNPs had coverage greater than or equal to 20x coverage (Table 5-1) with high level Phred scores >Q20. Those SNPs which had the lowest coverage of 20x coverage (present in TCV6D9 and TCV7D9) had the highest percentage confidence of 100 % (Table 5-2), suggesting that these SNPs were indeed genuine. The discrepancy between the

identified G>A SNP at position Contig21: 23295 in TCV5D9 (two-component response regulator SA14-24) (Table 5-1) and the corresponding coverage of 32x T (Table 5-2) can be explained by the fact that the Two-component response regulator is the only mutated gene which is coded for on the negative DNA strand. Additional SNPs in mobile elements and BOX elements were observed however did not match the selection criteria for the sequences could not be verified in the *de novo* assemblies. SNPs were also detected within potential cell surface proteins such as Choline binding protein J, and the iron compound ABC uptake transporter permease protein PiuB. These SNPs were present in both TCV and SCVs, and resulted in a non-synonymous amino acid change.

Table 5-1: Confirmed SNPs in the biofilm-derived variants from the pilot experiment

Variant	Reference Genome Position	WT	Variant SNP	Gene Description	Amino Acid change	SNP Coverage [Ψ]
SCV1D3	423010	G	T	DNA-directed RNA polymerase delta subunit	E173Stop	49x
SCV5D3	422977	G	T	DNA-directed RNA polymerase delta subunit	E163Stop	69x
SCV3D9	422921	C	A	DNA-directed RNA polymerase delta subunit	S144Stop	44x
SCV3D9	328131	A	C	Choline binding protein J	Q118P	30x
SCV3D9	908199	T	G	Hydrolase, putative	L26V	27x
SCV7D9	422977	G	T	DNA-directed RNA polymerase delta subunit	E163Stop	243x
SCV9D9	422795	A	-	DNA-directed RNA polymerase delta subunit	Deletion*	n/a
SCV9D9	1620361	C	A	Iron compound ABC uptake transporter permease protein PiuB	A278D	47x
TCV5D9	1077117	G	A	Two-component response regulator SA14-24	T213T	32x
TCV6D9	1813063	C	T	N-acetylglucosamine-6-phosphate deacetylase	N112N	20x
TCV7D9	158492	C	A	Aminopeptidase YpdF (MP-, MA-, MS-, AP-, NP- specific)	L115I	20x
TCV9D9	1671732	G	T	Immunity protein, probable	S140S	46x

Asterisk (*) denotes a deletion event not detected by SNP mapping analysis. [Ψ] SNP coverage was confirmed manually by visualising the aligned variant .bam files in Integrative Genomics Viewer.

Table 5-2: Coverage of confirmed SNPs in the biofilm-derived variants from the pilot experiment

Variant	Reference Genome Position	WT	Variant SNP	SNP Coverage [Ψ]	A	G	C	T	N	Total Base Coverage	Confidence Percentage (%)
SCV1D3	423010	G	T	49x	0	0	1	49	0	50x	98
SCV5D3	422977	G	T	69x	0	0	0	69	0	69x	100
SCV3D9	422921	C	A	44x	44	0	0	0	0	44x	100
SCV3D9	328131	A	C	30x	0	0	30	0	1	31x	96.7
SCV3D9	908199	T	G	27x	0	27	0	0	1	28x	96.4
SCV7D9	422977	G	T	243x	0	0	0	243	0	243x	100
SCV9D9	1620361	C	A	47x	47	0	0	0	1	48x	97.9
TCV5D9	1077117	G	A	32x	0	0	0	32	0	32x	100
TCV6D9	1813063	C	T	20x	0	0	0	20	0	20x	100
TCV7D9	158492	C	A	20x	20	0	0	0	0	20x	100
TCV9D9	1671732	G	T	46x	0	0	0	46	1	47x	97.8

[Ψ] SNP coverage was confirmed manually by visualising the aligned variant .bam files in Integrative Genomics Viewer.

5.3.2 PCR of the *rpo*E gene of *S. pneumoniae*

Data from the WGS analysis of the biofilm-derived variants revealed that 4 of the 5 SCVs contained mutations within the DNA-directed RNA polymerase delta subunit which result in a premature stop codon. In the fifth SCV, SCV9D9, a 264 bp deletion was observed. The reason for this partial sequence of the *rpo*E gene may have been due to poor alignment in the Velvet assembly. A search for the 264 bp deleted region using The Basic Local Alignment Search Tool (BLAST) confirmed that the sequence was not assembled in an alternative region of the SCV9D9 genome. To confirm whether this deleted region was in fact a true deletion, PCR primers were designed for the *rpo*E gene. To date no literature has published PCR primers for the *rpo*E gene in *S. pneumoniae*. The *rpo*E gene is 588 bp long; in order to design primers with acceptable GC percentage and melting temperature, the primers spanned both the start and end of the 5' and 3' coding sequence. If the deletion was a false-positive all SCVs would generate a band at 609 bp, however in the case of a true deletion, a band should be present at 345 bp. As shown in Figure 5-2, the presence of a band at ~350 bp in the *rpo*E PCR confirmed the 264 bp deletion is SCV9D9. In order to determine whether this deletion was a one-time phenomenon, PCR of the *rpo*E gene region was performed on all SCVs harvested from both the pilot project and a duplicate, independent biofilm experiment of serotype 22F (Figure 5-3).

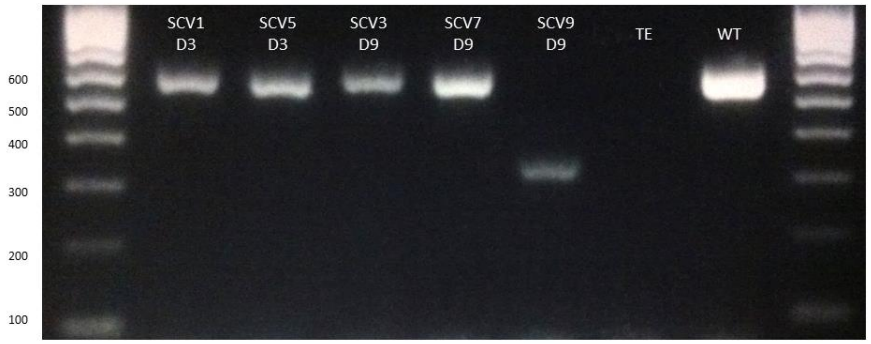
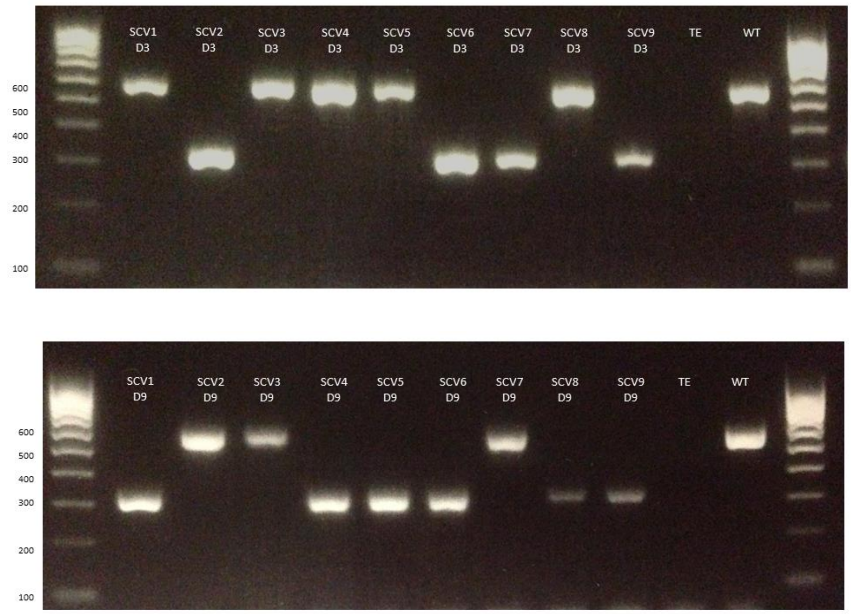


Figure 5-2: *rpo*E PCR of the biofilm-derived sequenced SCVs

PCR was run out on 2 % agarose gel for 1.5 hours at 125 V and stained with Gel Red. Bioline hyperladder^{TM} IV was used to size the bands. Full sequence = 609 bp, deletion sequence = 345 bp

Pilot project (Experiment 1)



Second experiment (Experiment 2 (E2))

Figure 5-3: *rpo*E PCR of all biofilm-derived SCVs

RpoE PCR of the biofilm-derived pneumococcal SCVs. PCR was run out on 2 % agarose gel for 1.5 hours at 125 V and stained with Gel Red. Bioline hyperladder^{TM} IV was used to size the bands. Full sequence = 609 bp, deletion sequence = 345 bp

In the pilot study (Experiment 1) 10 biofilm-derived SCV isolates were observed with a 264 bp deletion; 6 isolates in day 3 biofilms and 4 in day 9 biofilms (no SCVs were harvested at day 6) (Figure 5-3). These mutations occurred in separate triplicate biofilms and were observed in isolates harvested on different days. Whole genome sequencing revealed that the banding at ~345 bp was a result of a 264 bp deletion starting at position 298 bp in the *rpo*E gene resulting in a 324 bp gene product. Based on the sequence this deletion is likely due to a recombination event between two similar sequences (GACGACG(A/C)TGAA) at positions starting at 298 bp and 562 bp respectively within the *rpo*E gene (Figure 5-4).

In the second experiment (E2), three different types of deletion were observed (Figure 5-3). In day 3; 4 isolates (from the same biofilm replicate) displayed a band at ~450 bp. Whole genome sequencing revealed that this banding was a result of a 129 basepair deletion starting at basepair 268 of the *rpo*E gene, resulting in a 459 bp *rpo*E gene product. The deletion is likely due to a recombination event between two similar gene sequences (GACGAAGA) at positions 268 bp and 397 bp, respectively (Figure 5-5). In day 6 samples, 2 isolates were observed that displayed a band at ~345 bp as well as 1 isolate that displayed a band at ~450 bp (all three isolates were from the same biofilm replicate). Whole genome sequencing revealed that the banding at ~345 bp was a result of a 264 bp deletion starting at position 310bp resulting in a 324 bp gene product. This mutation was seen on separate days in separate biofilms in two independent biofilm experiments. In day 9 samples one isolate displayed no band (SCV7D9); this isolate did test positive for the 22F band and the *cps*A band confirming that the isolate was *S. pneumoniae* and that DNA was present in the sample (data not shown). WGS revealed that this banding was a result of a 62 bp deletion, including 51 bp at the 3' end starting from bp 538 resulting in a 537 bp gene product and 11 bp downstream of the *rpo*E coding region. Thus the absence of the band in the PCR was due to the inability of the 3' primer to bind. Based on the sequence data, there is no evidence that this deletion was a result of a recombination event (Figure 5-6).

5’-
TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATATTAG
AACTTCGTGGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAA
ACAGCGATATCCGCGAAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGG
ACAACAAATGGGGTCTTCGTTCATGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAA
T**GACGACGATGAA**GTAGCACCAAAAGCTAAGAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCC
ATTGACTACAATGCAGATGATCCAGAAGACGAAGATGCATACGAAGCAGATCCAGCTCTTTCATACGATGATGA
AAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGAAATCAACGAAATCGCTCCAGATGACTTG
GGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGAT**GACGACGCTGAA**ACCAGCGAGGAATAA
-3’

Figure 5-4: *rpo*E gene sequence for the sequenced variant SCV9D9

Sequence highlighted in blue represents the 264 bp deleted region. The sequence in bold and underlined indicates the regions of possible recombination.

5’-
TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATATTAG
AACTTCGTGGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAA
ACAGCGATATCCGCGAAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGG
ACAACAAATGGGGTCTTCGTTCATGGTATGGTGTGGACGAAATC**GACGAAGA**AATCATCGCTCTTGAAGAAAA
TGACGACGATGAAGTAGCACCAAAAGCTAAGAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCC
ATTGACTACAATGCAGATGATCCAGAA**GACGAAGA**TGCATACGAAGCAGATCCAGCTCTTTCATACGATGATG
AAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGAAATCAACGAAATCGCTCCAGATGACTT
GGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGACGCTGAAACCAGCGAGGAATA
A -3’

Figure 5-5: *rpo*E gene sequence for the sequenced variant SCV1D3E2

Sequence highlighted in blue represents the 129 bp deleted region. The sequence in bold and underlined indicates the regions of possible recombination.

5’-
TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATATTAG
AACTTCGTGGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCA
AACAGCGATATCCGCGAAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGG
GACAACAAATGGGGTCTTCGTTCATGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAA
ATGACGACGATGAAGTAGCACCAAAAGCTAAGAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGC
CATTGACTACAATGCAGATGATCCAGAAGACGAAGATGCATACGAAGCAGATCCAGCTCTTTCATACGATGATG
AAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGAAATCAACGAAATCGCTCCAGATGACTT
GGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGACGCTGAAACCAGCGAGGAATA
A·GAGTTAGCTAT -3’

Figure 5-6: *rpo*E gene sequence for the sequenced variant SCV7D9E2

Sequence highlighted in blue represents the 62 bp deleted region (51 in *rpo*E coding region) at the 3’ end starting from basepair 538 (537 bp remaining). The symbol · denotes where the *rpo*E coding region ends.

In addition to PCR of the *rpo*E gene, an additional seven SCVs from a second independent biofilm experiment were sequenced and assessed for SNP mutations. These SNPs were confirmed in three additional SCVs from days 3, 6 and 9 biofilms (Table 5-3). Coverage of the SNPs varied between 122-197x with >99 % confidence (Table 5-4). In addition an insertion of a Guanine nucleotide was observed in sample SCV6D6E2 resulting in a D177G amino acid change and subsequent frameshift coding resulting in a stop codon at amino acid position 188 in the RpoE protein. Sequence coverage of this insertion was high at 92x, as such is unlikely to be due to sequencer error (Table 5-4). In addition to the *rpo*E gene, SNPs were confirmed in the beta-galactosidase gene and the glycine/D-amino acid oxidases family gene, in variant SCV4D6E2 and *Streptococcal* histidine triad protein in variant SCV7D6E2. All sequenced SCVs from the second experiment were phenotyped based on API profile, growth rate and capsule quantification (See Appendix 4: Figure 8-2 to Figure 8-4 and Table 8-5). The SCV phenotype was consistent with SCVs from the pilot study (Figure 4-7 to Figure 4-10).

Table 5-3: Confirmed SNPs in the biofilm-derived variants from the second experiment

Variant	Reference Genome Position	WT	Variant SNP	Gene Description	Amino Acid change	SNP Coverage [Ψ]
SCV1D3E2	422765	A	-	DNA-directed RNA polymerase delta subunit	Deletion	n/a
SCV2D6E2	422799	G	-	DNA-directed RNA polymerase delta subunit	Deletion	n/a
SCV4D6E2	422947	G	T	DNA-directed RNA polymerase delta subunit	E153Stop	197x
SCV4D6E2	591574	G	T	Beta-galactosidase	A146S	129x
SCV4D6E2	1396435	C	T	Glycine/D-amino acid oxidases family	R214H	137x
SCV6D6E2	423020	-	G	DNA-directed RNA polymerase delta subunit	Insertion	92x
SCV5D9E2	423028	G	T	DNA-directed RNA polymerase delta subunit	E180Stop	146x
SCV6D9E2	422962	G	T	DNA-directed RNA polymerase delta subunit	E158Stop	122x
SCV7D9E2	423028	G	-	DNA-directed RNA polymerase delta subunit	Deletion	n/a
SCV7D9E2	1035274	G	A	*Streptococcal* histidine triad protein	Q111stop	106x

[Ψ] SNP coverage was confirmed manually by visualising the aligned variant .bam files in Integrative Genomics Viewer.

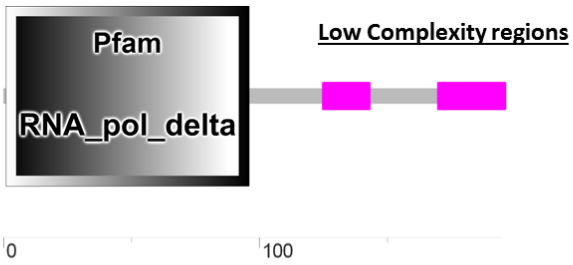
Table 5-4: Coverage of confirmed SNPs in the biofilm-derived variants from the second experiment

Variant	Reference Genome Position	WT	Variant SNP	SNP Coverage [Ψ]	A	G	C	T	N	Total Base Coverage	Confidence Percentage (%)
SCV4D6E2	422947	G	T	197x	1	0	0	197	0	198x	99.5
SCV4D6E2	591574	G	T	129x	1	0	1	129	0	131x	98.5
SCV4D6E2	1396435	C	T	137x	0	0	0	137	0	131	100
SCV6D6E2	423020	-	G	92x	-	92	-	-	-	98x	93.9
SCV5D9E2	423028	G	T	146x	0	0	0	146	0	146x	100
SCV6D9E2	422962	G	T	121x	0	1	0	120	0	122x	99.2
SCV7D9E2	1035274	G	A	106x	106	0	0	0	0	106x	100

[Ψ] SNP coverage was confirmed manually by visualising the aligned variant .bam files in Integrative Genomics Viewer.

5.3.3 Amino acid structure of RpoE

The functional effects of the truncated *rpo*E gene mutations were postulated based on the amino acid structure of the RpoE protein from the Ensembl SMART database (Schultz *et al*., 1998, Letunic *et al*., 2012). The RpoE protein is divided into 4 distinct regions of interest; the conserved RNA-polymerase delta domain at the N-terminus and three low complexity regions located towards the C-terminus (Figure 5-7). In the context of this work, the insertion and SNP mutations seen in the SCVs would affect just the C-terminal low complexity regions. The 129 bp and 264 bp deletions would affect both the low complexity regions and the N-terminal RNA-polymerase delta domain. All mutations in the SCVs would thus likely affect the binding of the delta subunit. This observation is relevant as there is evidence that low complexity regions have both a positional dependant role and a role important for protein-protein interaction (Coletta *et al*., 2010).



Wild Type amino acid sequence

MELEVFAGQEKSELSMIEVARAILELRGRDHEMHFSDLVNEIQNYLGTSNSDIRE
ALPLFYTELNFDGSFISLGDNKWGLRSWYGV**DEIDEEIIALEENDDDE**VAPKAKK
KRVNAFMDGDSDAI**DYNADDPEDEDAYEADPA**LSYDDENPDDEKNEVEAYDA
EINEIAP**DDLGEDVDLNEDDDEFSDDDAETSEE**

Figure 5-7: Graphical representation of the RpoE amino acid sequence from the Ensembl SMART database

The corresponding amino acid sequence highlighted in grey represents the conserved RNA-polymerase delta domain at the N-terminus; the regions highlighted in pink represent the C-terminal low complexity regions. An additional low complexity regions exists which spans the RNA-polymerase delta domain (in bold and underlined).

5.3.4 RpoE variation in pneumococcal carriage isolates

This work has shown that the *rpo*E gene is consistently targeted in pneumococci grown under biofilm conditions. To better understand the significance of *rpo*E variation in the wider population WGS data from 518 pneumococcal carriage isolates was used to identify different *rpo*E alleles and furthermore determine whether there is a link between the *rpo*E alleles seen and the serotype and or multi locus sequence type of the organisms. The 518 pneumococcal carriage isolates used in this work were isolated from an on-going nasopharyngeal carriage study in children aged 4 years and under at UHS. This carriage study was also the source of the serotype 22F parent isolate used in this thesis.

Twenty-two different alleles were identified in the 518 isolates over the first 5-year period. Taking into account multiple serotypes with the same *rpo*E allele, we see that the most common allele is *rpo*E1, which occurred in 25 % of isolates, followed by *rpo*E2 and *rpo*E4 (occurring at 15 % and 14 % respectively) (Figure 5-8). Sequences for the *rpo*E alleles can be seen in Table 8-6 in Appendix 5.

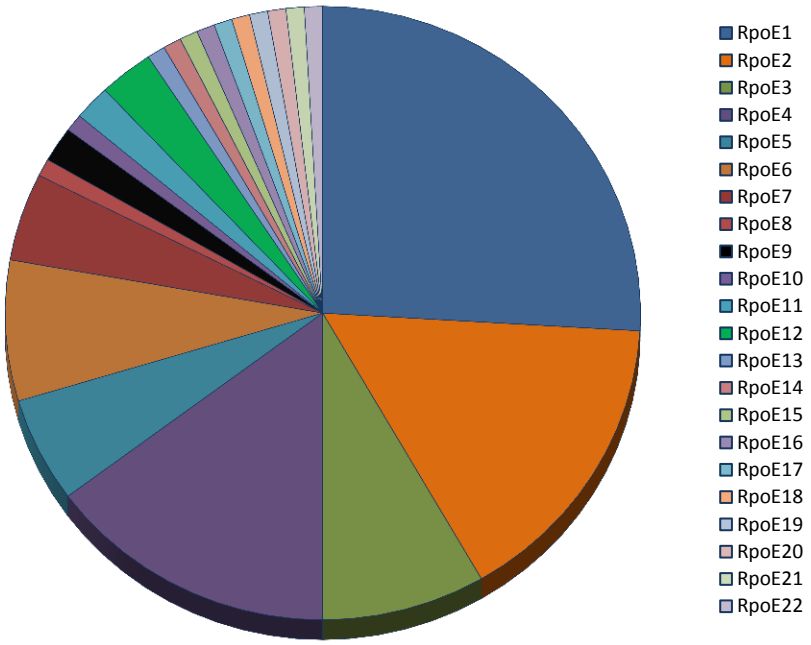


Figure 5-8: Percentage of *rpo*E alleles in pneumococcal carriage isolates

*rpo*E alleles were extracted from whole genome data from 518 pneumococcal isolates from a 5-year carriage study in children aged 4 years and under. The 22 different alleles were identified which occurred on 108 occasions. This graph represents the percentage of *rpo*E alleles over the 5 years, percentages are independent of serotype.

To determine whether there is a link between the *rpo*E alleles seen in the population and the serotype and or multi locus sequence type (ST) of the organisms, pneumococcal WGS data obtained from 518 isolates from the first five years of an on-going pneumococcal carriage study in children under 4 years-old, were organised into serotype and/or MLST and *rpo*E alleles and displayed graphically (Figure 5-9-Figure 5-10).

These data show that some serotypes have a tendency towards a small range of *rpo*E alleles, whereas other serotypes have more diverse *rpo*E alleles. This can be seen in serotypes 14 and 17F, which only contained allele *rpo*E2, and serotype 35F, which only contained *rpo*E7, and in serotype 23A, which only contained *rpo*E5 (Figure 5-9). In contrast, serogroup 15 contained a more diverse number of alleles including *rpo*E1, *rpo*E2, *rpo*E3, *rpo*E4, *rpo*E6, *rpo*E9, and *rpo*E11 (Figure 5-9). This observation can perhaps be explained by comparing the *rpo*E alleles to the MLST of the isolates (Figure 5-11); here we see that certain STs only have certain *rpo*E alleles. For example ST62 only contains *rpo*E2 alleles and ST1635 only contains *rpo*E7. In many cases the number of STs isolated is too small to draw any conclusion however looking at the ST that arose >10 times over the 5 year course we see that isolates with the sequence type ST36, ST62, ST65, ST138, ST162, ST176, ST438, ST439, and ST1635 only contained a single RpoE allele. These data suggest therefore that there is a link between the serotype/MLST of a pneumococcus and the *rpo*E allele in its genome. Due to the disparity in the number of serotypes isolated each year, a yearly comparison of *rpo*E alleles and serotype/MLST could not be performed. As this carriage study spans the introduction of the pneumococcal conjugate vaccine PCV7 and PCV13, it would be of interest to see if the diversity of rpoE alleles changes in response to the use of the vaccine. It is important to note that no alleles matched the mutations seen in the sequenced SCVs from the biofilm experiments.

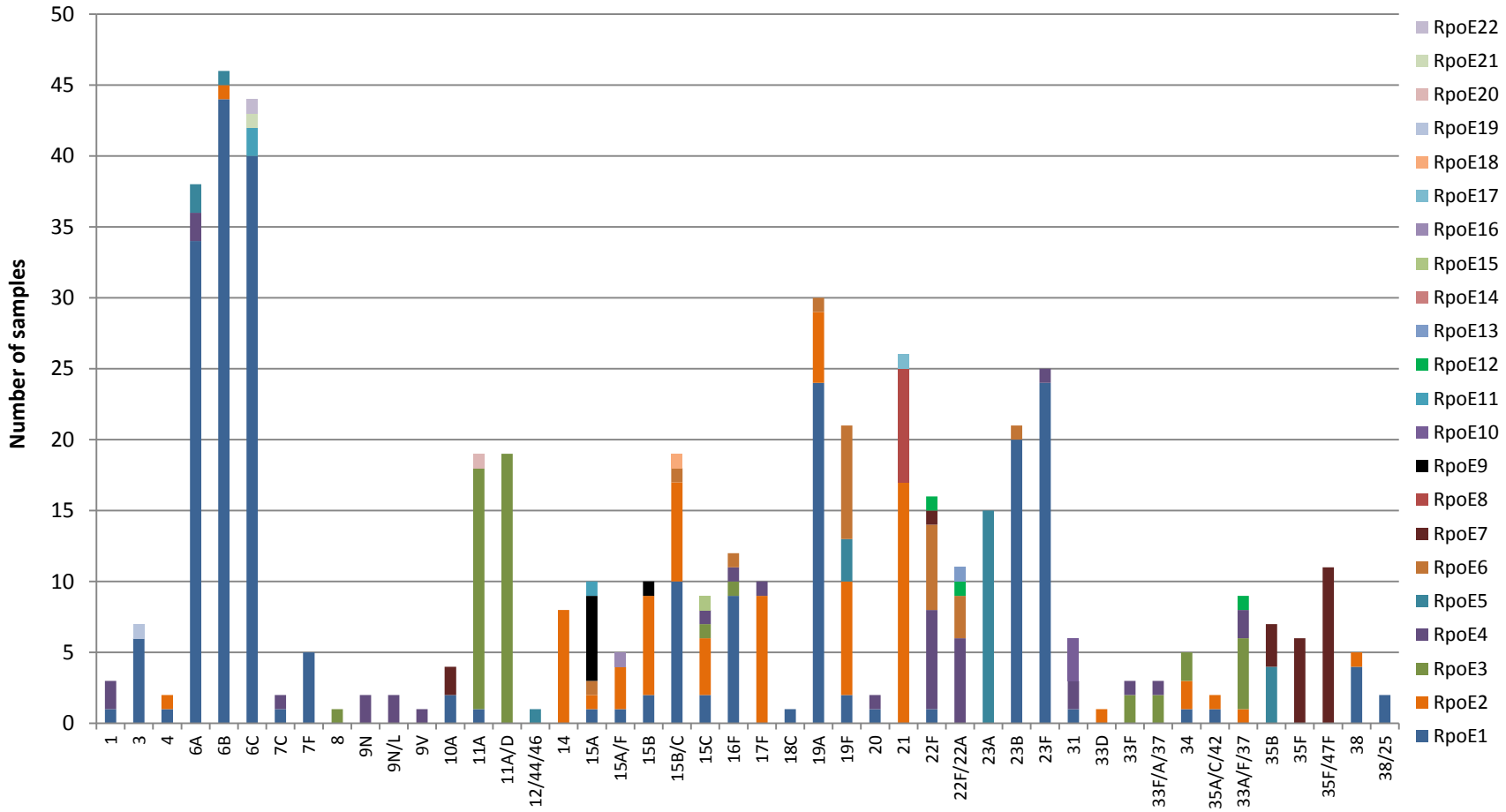


Figure 5-9: Number of rpoE alleles compared to serotype in pneumococcal carriage isolates

RpoE alleles were extracted from whole genome data from 518 pneumococcal isolates from a 5-year carriage study in children aged 4 years and under. This graph represents the distribution of the 22 identified *rpo*E alleles between 46 serotypes over the 5 years.

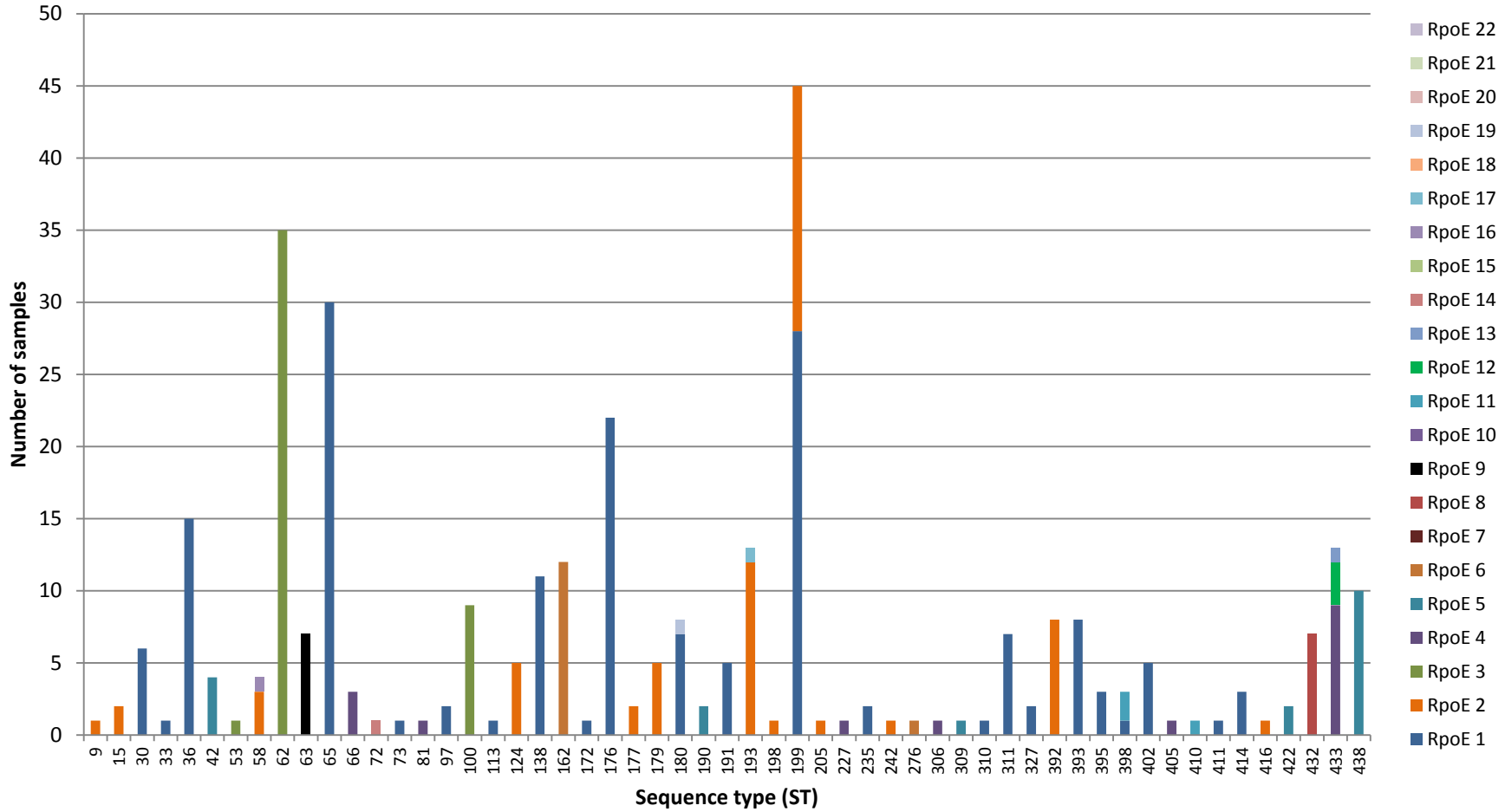


Figure 5-10: Number of *rpo*E alleles compared to multi locus sequence typing in pneumococcal carriage isolates

RpoE alleles were extracted from whole genome data from 518 pneumococcal isolates from a 5-year carriage study in children aged 4 years and under. The 22 identified *rpo*E alleles were distributed between 112 sequence types (STs). This graph represents the distribution of the 22 identified *rpo*E alleles between the first 55 STs over the 5 years.

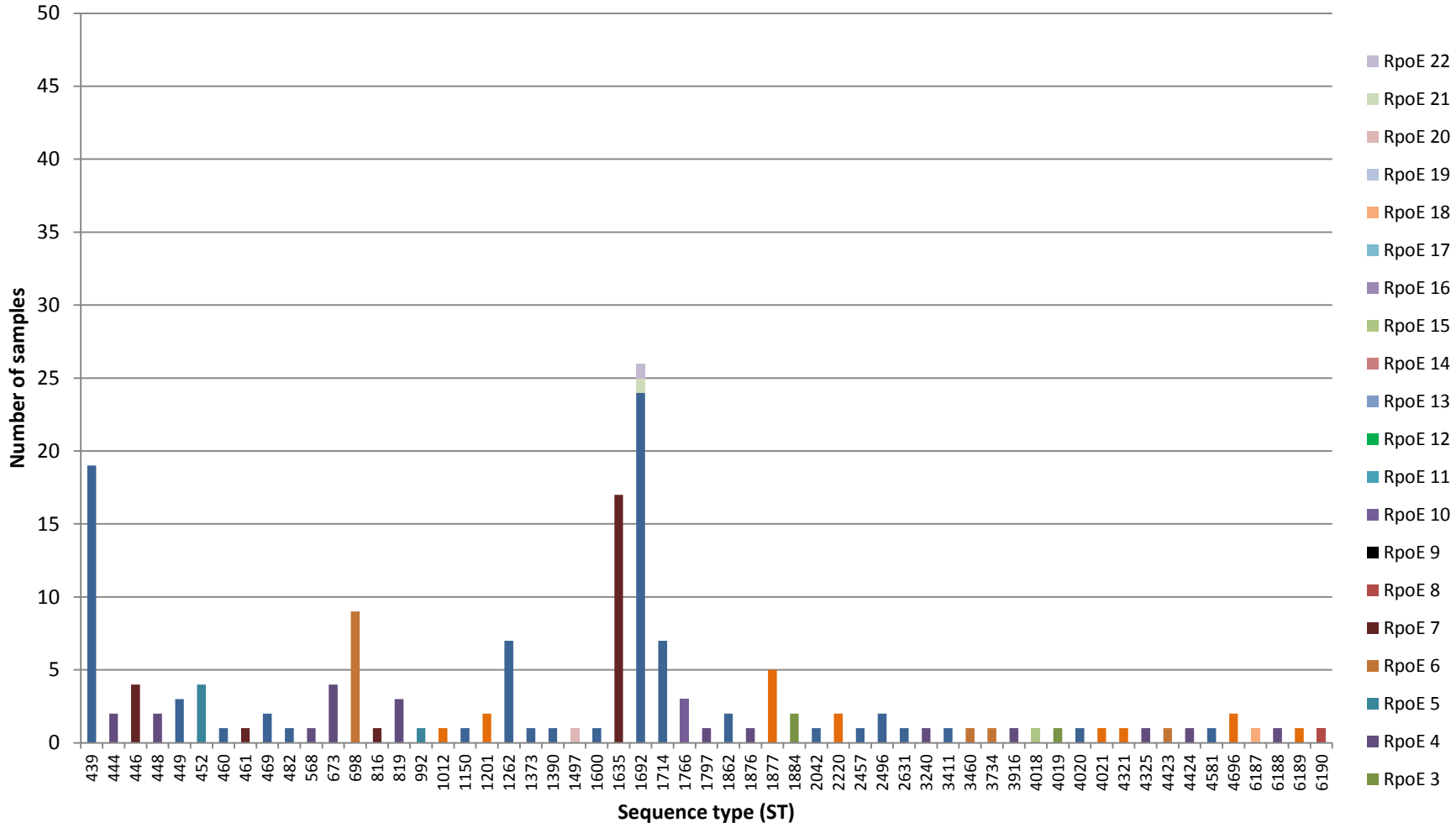


Figure 5-11: Number of *rpo*E alleles compared to multi locus sequence typing in pneumococcal carriage isolates

RpoE alleles were extracted from whole genome data from 518 pneumococcal isolates from a 5-year carriage study in children aged 4 years and under. The 22 identified *rpo*E alleles were distributed between 112 sequence types (STs). This graph represents the distribution of the 22 identified *rpo*E alleles between the next 57 STs over the 5 years.

5.3.5 Phylogenetic analysis of RpoE in carriage isolates

To assess how the *rpo*E alleles related to each other functionally, the amino acid (aa) sequence of the 22 RpoE alleles were assessed phylogenetically using the online resource phylogeny.fr as described in section 2.21. From this alignment we see that 12 out of the 22 alleles had no difference in amino acid sequence, suggesting that the protein is functionally conserved. For comparison, using the Dayhoff substitution model generated a similar phylogeny (data not shown).

What is evident from Figure 5-12 is that RpoE8 forms a separate clade suggesting a potential difference in function. RpoE8 is unique to serotype 21 and seen only in years 4 and 5 of the carriage study, suggesting that this may be a previously un-sampled clone present in increasing number the wider population, however the number of RpoE8 isolates is too small (n = 3 in year 4 and n = 5 in year 5) to draw any meaningful conclusion. Those alleles which have diverged more (rpoE 14, 15, 16, 17, 21, and 22) are also in low numbers within the population (n = 1). The exception is RpoE3 which was recorded in 50 isolates. This allele was seen in all 5 years and was seen in numerous serotypes including 8, 11A, 11A/D, 16F, 33, 33F/A/37 and 34. Interestingly, *rpo*E alleles 14, 15, 17, and 22 occurred after the implementation of PCV7 vaccine and *rpo*E alleles 13, 16 and 21 occurred after the implementation of both PCV7 and PCV13 vaccine. In particular, *rpo*E allele 21 was the only allele to have 585 bp as opposed to the full 588 bp product. These isolates is too small to draw any meaningful conclusion that these alleles have diverged as a direct result of vaccine implementation however future studies would be advised to record the prevalence of these alleles to see if they increase in the population.

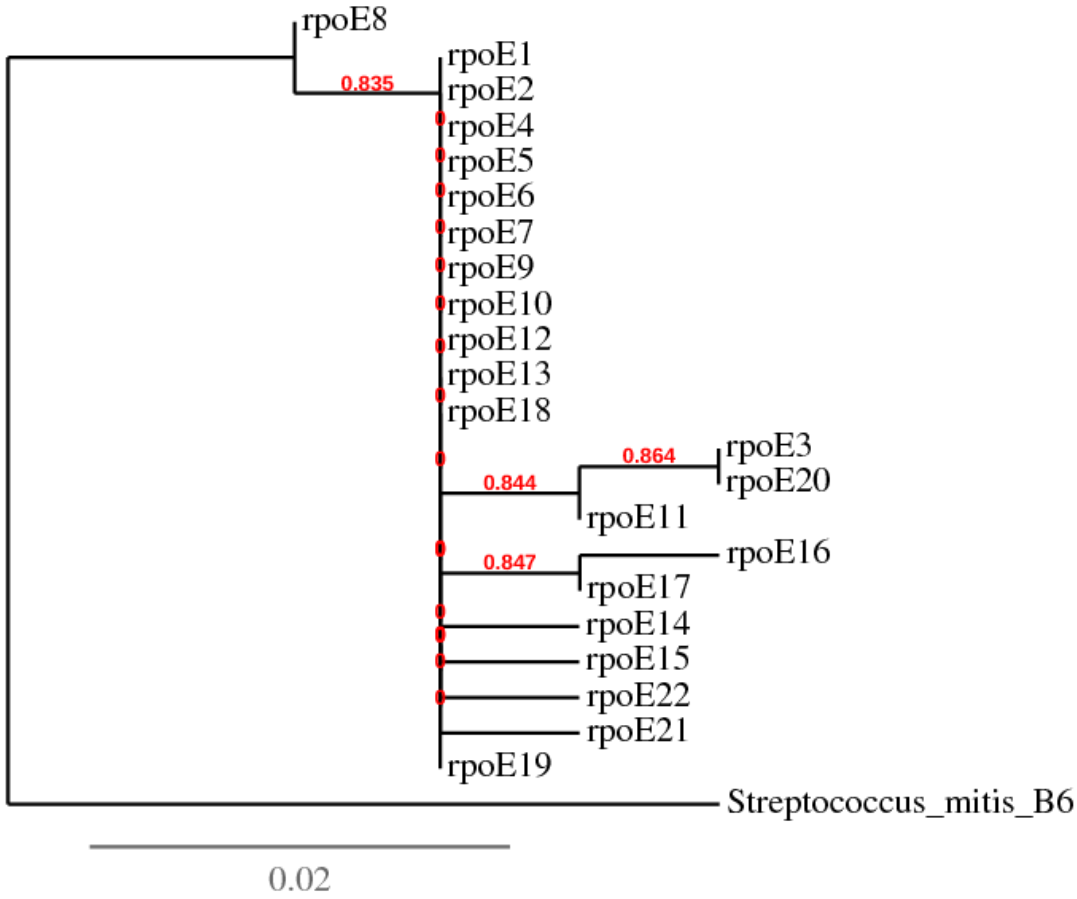


Figure 5-12: Phylogenetic tree of RpoE amino acid sequences in pneumococcal carriage isolates

Amino acid sequences were aligned using MUSCLE and curated using Gblocks. The phylogenetic tree was reconstructed using PhyML and maximum likelihood method using the Blosum62 substitution model. The tree was constructed using TreeDyn. Average likelihood ratio test (ALRT) values are displayed in red. Scale bar represents nucleotide substitutions per site.

Although no alleles matched the mutations seen in the sequenced SCVs from the biofilm experiments in the first 5 years of the carriage study, four isolates were discovered to have a smaller colony phenotype in later years from two independent carriage studies. Three samples (8081, 8112 and 8140) were isolated from year 8 (2013/14) of the on-going nasopharyngeal carriage study in children aged 4 years and under at UHS. The additional sample (sample 695) was isolated in 2012 from a separate carriage study (the SMART study). All samples displayed alpha-haemolysis and sensitivity to optochin antibiotic with the exception of sample 8112, indicating that this sample may not be a pneumococcus. DNA from all four samples was extracted for species conformation, characterisation of the *rpo*E gene sequence, serotyping and MLST profiling.

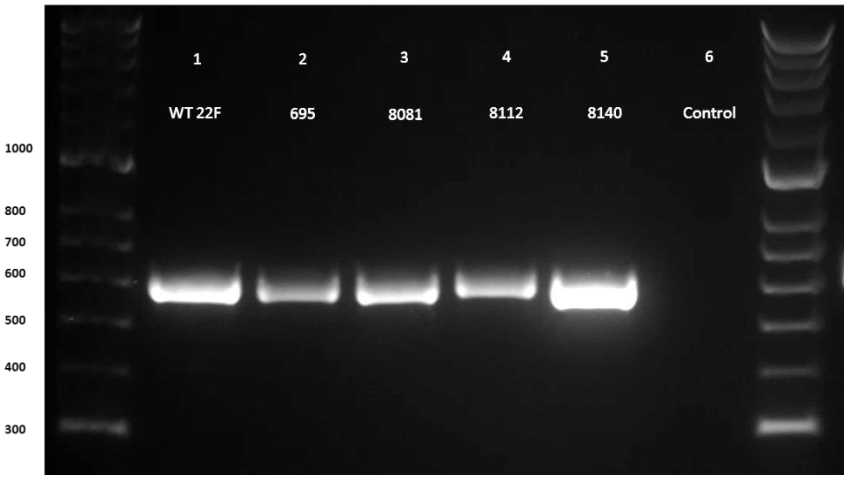


Figure 5-13: *rpo*E PCR of SCVs in carriage Isolates

PCR was run out on 1.5 % agarose gel for 45 minutes at 90 V and stained with Gel Red. Bioline hyperladder^{TM} II was used to size the bands. Full sequence = 609 bp

PCR of *rpo*E gene in the four SCV carriage isolates revealed no large scale deletions in the *rpo*E gene (Figure 5-13). Sequencing of the *rpo*E gene revealed that samples 695 and 8081 had a 3 basepair deletion at position 541-543 resulting in a 585 bp gene product as opposed to 588 bp. This deletion is consistent with *rpo*E allele 21 (RpoE21) seen in the first five years of carriage. This 3 basepair results in the lack of a Aspartic acid (D) residue at position 181 in the amino acid sequence. Serotyping of the four small colony variants revealed all were PCR-negative for the capsule genes (including *cps*A) suggesting the isolates were non-typeable. Only one isolate (8140) was confirmed as a pneumococcus via MLST (ST488). The remaining three isolates (695, 8081 and 8112) had sequence types not recorded on the MLST database. All isolates also underwent a BLAST search for sequence similarity. BLAST search revealed that samples 695 and 8081 shared closer similarity with *S. pseudopneumoniae* and 8112 shared similarity *S. mitis*. Factoring in the lack of capsule and lack of known ST, these data suggest isolates 695, 8081 and 8112 are not pneumococcus.

5.3.6 *in silico* serotyping and multi locus sequence typing of biofilm-derived variants

All biofilm-derived variant sequences were assembled *de novo* as described in section 2.18. The Velvet optimiser script was used to determine the K-mer value required to achieve approximately 20x coverage of the whole genome. This was achieved in all but two of the biofilm-derived isolates; however the level of coverage was acceptable for determining serotype and sequence type (ST). All biofilm-derived variants reported a sequence type of ST433 (*aro*E-1, *ddl*-17, *gdh*-1, *gki*-4 *rec*P-1, *spi*-18, *xpt*-58), suggesting that no mutations had arisen under biofilm growth which resulted in a change in allele sequence, this result was subsequently determined using the SRST2 script (http://katholt.github.io/srst2/) as described in section 2.18 (Table 5-5). Using an in-house *in silico* PCR script designed in conjunction with the Sanger Institute Cambridge, all sequenced isolates were confirmed as serotype 22F (Table 5-5).

Table 5-5: Serotype MLST and Velvet coverage level for WGS of all biofilm-derived variants

Isolate	Velvet coverage	Serotype	MLST
WT	-	22F	433
SCV1D3	23	22F	433
SCV5D3	28	22F	433
SCV3D9	24	22F	433
SCV7D9	122	22F	433
SCV9D9	18	22F	433
TCV5D9	20	22F	433
TCV6D9	20	22F	433
TCV7D9	17	22F	433
TCV9D9	20	22F	433
TCV10D9	21	22F	433

Isolate	Velvet coverage	Serotype	MLST
WT	-	22F	433
SCV1D3E2	76	22F	433
SCV2D6E2	84	22F	433
SCV4D6E2	94	22F	433
SCV6D6E2	52	22F	433
SCV5D9E2	80	22F	433
SCV6D9E2	53	22F	433
SCV7D9E2	87	22F	433

5.4 **Discussion of results**

The aim of this chapter was to investigate the level of genetic diversity that arises in pneumococcal colony morphology variants isolated from a biofilm. The primary aims of this work were to 1) Characterise the biofilm-derived colony variants genetically using next generation WGS to account for changes in the phenotypic profile. 2) Assess whether biofilm growth can result in mutations which change the serotype or multi-locus sequence type (ST) of the organism. And 3) Assess whether biofilm growth results in mutations within genes relevant to invasive disease and vaccine design. This work has shown that 1) changes in the phenotypic profile of SCVs are linked to mutations within the *rpo*E gene, 2) biofilm growth did not result in mutations which change the multi-locus sequence type or the serotype of the organism, and 3) SNPs were detected within potential vaccine targets which resulted in an amino acid change. For each aim the results will be discussed.

Using WGS numerous SNPs have been identified which are distributed throughout the genome of all variants sequenced. Of note, all SCVs sequenced had mutations within the DNA-directed RNA polymerase delta subunit (RpoE) ranging from premature stop codons to large scale deletions. These mutations occurred in multiple replicates of two independent biofilm experiments at different positions within the gene and on different days. In bacteria, RNA polymerase (RNAP) is the enzyme complex responsible for the transcription of RNA from genomic DNA. The core enzyme consists of five subunits αI, αII, β, β', and ω and has shown to be highly conserved in both eukaryotes and prokaryotes (Ebright, 2000). This core enzyme complex subsequently binds to ancillary subunits to initiate transcription (Gruber and Gross, 2003) including the DNA-directed RNA polymerase delta subunit (RpoE) in *Firmicutes* (Lopez de Saro *et al*., 1995). In Gram-positive bacteria RpoE is thought to bind to the core RNAP complex via the β and β'subunits (Weiss *et al*., 2014) and has been shown to be an important protein involved in environmental adaptation and virulence (Xue *et al*., 2010, Xue *et al*., 2011).

To date RpoE has not been studied in *S. pneumoniae* and thus the function of RpoE must be inferred from studies on closely related Gram-positive bacteria. Using *Bacillus subtilis* as a model organism *rpo*E has been attributed to increasing transcriptional specificity by increasing binding of RNAP to promoter sequences (Achberger and Whiteley, 1981, Achberger *et al*., 1982). Additionally *rpo*E has been attributed to increased efficiency of RNA synthesis and decreased affinity for nucleic acids due to enhanced recycling of RNA (Juang and Helmann, 1994, Lopez de Saro *et al*., 1999). Understanding of the RpoE function is primarily based on

gene deletion studies; *rpo*E deletion mutants of *B. subtilis* have been shown to be viable suggesting that RpoE is non-essential (Lopez de Saro *et al*., 1999), however phenotypically *rpo*E mutant strains have been shown to display an increased lag phase, altered cell morphology (Lopez de Saro *et al*., 1999) and altered biofilm architecture (Xue *et al*., 2010). In *S. mutans*, *rpo*E mutants were shown to exhibit reduced virulence with impaired growth and reduced tolerance to environmental stress (Xue *et al*., 2010). These data are consistent when a proteomic comparison was made between the *rpo*E mutant and the wild type 22F strain under various stress conditions (Xue *et al*., 2012). Conversely, the same group also showed *rpo*E mutants to exhibit traits akin to increased virulence including increased self-aggregation and co-aggregation in addition to increased carbohydrate metabolism and adhesion to human extracellular matrix proteins such as fibronectin (Xue *et al*., 2011). The slight disparity between studies was attributed to the differences in growth media and stress conditions employed (Xue *et al*., 2012). Recent work in *S. aureus* has shown down-regulation of virulence factor and up-regulation of mobile genetic elements in the *rpo*E mutants (Weiss *et al*., 2014) and is consistent with Xue *et al.* (2010). Taken together these studies show that RpoE provides an important role in the survival, adaptation and virulence of the organism. In all studies to date disruption of the *rpo*E gene has been mediated by directly removing or altering the sequence, mutations in this work are the first to provide evidence that mutations affecting the *rpo*E gene can occur spontaneously *in vitro*.

Structurally the RpoE protein is divided into 4 distinct regions; the conserved RNA-polymerase delta domain at the N-terminus and an unstructured C-terminus consisting of three low complexity regions (Lopez de Saro *et al*., 1995). Both the truncation mutations and the large scale deletions seen in the biofilm-derived SCVs would affect one or more of the C-terminal low complexity regions. Although little is known about the role of low complexity regions within proteins, there is evidence that these regions have both a positional dependant role and a role important for protein-protein interaction (Coletta *et al*., 2010). A hypothesis based on this information could be that mutations within the low complexity regions of the *rpo*E gene may change the transcriptomic/proteomic profile of the variants which may facilitate survival within a specific biofilm niche.

The increased lag phase and reduction in growth rate seen in the SCVs tested (Figure 4-7Figure 4-7 and Figure 8-2Figure 8-3) is consistent with previous RpoE studies in *B. subtilis* (Lopez de Saro *et al*., 1999). This slow growth may help this sub-population of SCVs in the biofilm context as fluctuations in growth between different sub-populations within the biofilm may prevent direct competition between neighbouring cells. In this work the mutations within

the *rpo*E gene do not appear to affect the viability of the organism; the slower growth phenotype, reduced capsule production and different metabolic profile is coupled with increased biofilm formation which may facilitate long term carriage within the host. The slower growth phenotype may help explain the differences seen in the SCV API profiles, including the inability to metabolise trehalose; the inability to metabolise certain sugars seen in the SCVs is consistent with the transcriptomic and proteomic profiles seen in *rpo*E mutants of *S. mutans* (Xue *et al*., 2010, Xue *et al*., 2012) suggesting that the *rpo*E mutations seen in the SCVs are likely the causative factor.

We observed multiple large scale deletions in the *rpo*E genes of biofilm-derived SCVs which were confirmed using PCR. Due to the technique in harvesting of the biofilm-derived isolates, the multiple observations of isolates with large scale deletions may be due, in some part, to clonal expansion of a single clone within each triplicate biofilm. Considering, however, that both the 129 bp deletion was observed on multiple days and the 264 bp deletion was observed in independent experiments on multiple days, suggests that these deletions are not a rare phenomenon in SCVs. Based on the sequence data, the 129 bp deletion and the 264 bp deletions are thought to have occurred due to recombination events and in both recombination events, the starting sequence was GAC·GTC. Triplet repeat sequences (TRS) such as this have previously been linked to recombination events (Jakupciak and Wells, 1999). The number of triplet repeat sequences within the *rpo*E gene may help explain why recombination has occurred multiple times in independent experiments. In all four deletion events a pattern in the first 8 bp of the deleted sequence can be seen as each deleted sequence commences with the sequence 5'-GA(C/A)GA(A/C)GA-3'. This pattern of sequence may represent a recognition site for the pneumococcal recombinase machinery which in turn results in the deletions seen. This pattern may be analogous to the *Chi* site seen in *E. coli*. The *Chi* site is a sequence of DNA (5'-GCTGGTGG-3') near which homologous recombination is likely to occur at a higher frequency and serves as a signal to activate the RecBCD helicase-exonuclease pathway (Prudhomme *et al*., 2002). The remaining SCVs had point mutations resulting in a premature stop codon; the majority of the point mutations were G-T, altering the original codon of GAA (Glutamic acid, E) to TAA (Stop). With the exception of SCV5D3 and SCV7D9 all SNPs occurred at different positions in the gene, the reason for the consistent mutation is likely due to the fact that the *rpo*E gene contains 29 GAA triplet sequences. This high proportion of GAA triplet sequences may increase the chance that a G-T SNP would result in a stop codon and thus a truncated product. Eight of the 12 SCVs sequenced only contained a mutation in the *rpo*E gene and with the exception of SCV5D3 and SCV7D9 all SNPs occurred at

different positions in the gene. This observation not only supports the hypothesis that mutations in the *rpo*E gene are directly linked with the SCV phenotype but also the consistent mutations in a single gene in multiple replicates of independent experiments is clear evidence of biofilm-mediated parallel evolutionary events in real-time. Such a phenomenon has been reported previously in *P. aeruginosa* (McElroy *et al*., 2014) biofilms but this is the first evidence of such an event in *S. pneumoniae*. Parallel evolution of specific genes has been shown to be clinically relevant in determining the pathogenesis of *Burkholderia dolosa* (Lieberman *et al*., 2011) and thus is an important field of research. Furthermore, recent work by Kim *et al.* (2014) has shown that spontaneous mucoid variants repeatedly arise in *Pseudomonas fluorescens* colonies. These variants were shown to spatially position themselves on the surface of the colony to optimise access to oxygen and nutrients. The authors proposed that the mucoid variants underwent strong selection in mixed colony experiments where rapid increase in the mucoid phenotype resulted in dominance over the wild type strain. Furthermore genome sequencing of over 500 mucoid variants revealed a striking example of parallel evolution at the *rsm*E locus with mutations that altered the competitiveness depending on the variant (Kim *et al*., 2014). Taken together, these data suggest that parallel evolution is a key driving force for diversification in surface-attached communities of bacteria, such as biofilms.

The reason for the *rpo*E gene being consistently targeted for mutation remains unclear but the answer is likely to be due to its function. RpoE has been linked to environmental adaptation and virulence (Xue *et al*., 2010, Xue *et al*., 2011), one could hypothesise that the mutations seen here benefit the bacteria to survive in the context of biofilm formation by altering the gene expression in favour of colonisation rather than virulence. This notion is supported by the fact that SCVs form biofilms with a greater biomass, thickness and surface area than the wild type 22F strain. Moreover, RpoE has been shown to be up-regulated in response to quorum sensing molecule autoinducer-2 in *Streptococcus mutans* (Sztajer *et al*., 2008). RpoE is responsible for transcription of genes involved in the metabolism of carbohydrates (Xue *et al*., 2010, Xue *et al*., 2011) and the mutations may help explain the diversity seen in the SCV API profiles.

Previously, the emergence of SCVs within a biofilm community has supported the notion that SCVs may have a competitive advantage within the biofilm niche (Singh *et al*., 2010). This work would support previous studies and furthermore the *rpo*E mutations that are present in SCVs may affect the carbon utilisation within the biofilm. RpoE has been shown to be important for rapid changes in gene expression in order to adapt to changing environmental conditions (Rabatinova *et al*., 2013). This may help explain why this gene is

consistently targeted within a sub-population of biofilm-derived isolates. As mentioned previously, RpoS has been shown to be expressed in response to environmental stress such as nutrient depletion in planktonic cultures and has been shown to be a relevant factor in biofilm development (Stoodley *et al*., 2002), this work furthers previous work by highlighting the importance of RpoE as a possible source of genetic variation facilitating survival within the biofilm environment. The large scale deletions seen in the SCVs may also suggest that these variants act as DNA donors for other cells within the biofilm; however this hypothesis is simply speculation at this point.

To determine whether there is a link between the *rpo*E alleles seen in the population and the serotype and or multi locus sequence type (ST) of the organisms, pneumococcal WGS data obtained from 518 isolates from an on-going pneumococcal carriage study in children under 4 years-old, were organised into serotype and/or MLST and *rpo*E alleles and displayed graphically (Figure 5-9-Figure 5-11). These data suggest that there is a pattern between the sequence type of the organism and the *rpo*E allele. For example ST62 only contains *rpo*E2 alleles and ST1635 only contains *rpo*E7. In many cases the number of STs isolated is too small to draw any conclusion; however looking at the ST that arose >10 times over the 5 year course we see that isolates with the sequence type ST36, ST62, ST65, ST138, ST162, ST176, ST438, ST439, and ST1635 only contained a single *rpo*E allele. This pattern may be due to the housekeeping genes used to determine ST (*aro*E, *gdh*, *gki*, *rec*P, *spi*, *xpt*, *ddl*) and the *rpo*E gene mutating at a similar rate. Alternatively this could be evidence that there are persistent clones cultured year after year within the carriage study population which, in itself, is an interesting observation. The *rpo*E gene allele could potentially be used to identify novel clones within the wider population. It is interesting to note that no alleles matched the mutations seen in the sequenced SCVs. This is not unexpected, due to the fact that the mutations seen were unique to small colony variants. It is important to note that during the isolation of the 518 carriage isolates, colony morphology was not documented. Colony morphology is not regularly reported during carriage studies as such the proportion of SCVs isolated during routine carriage studies remains unknown. When specifically looking to isolate SCVs, four were observed from two independent carriage studies. These samples were all shown to be non-typeable, indicating the lack of a capsule. Of the four SCV isolated only one isolate was confirmed as a pneumococcus (8140). Analysis of the *rpo*E gene in 8140 revealed no truncation mutation. Taken together this suggests that the SCV phenotype seen in 8140 was due to a lack of capsule, not mutations within the *rpo*E gene. This single sample is not large enough to draw any meaningful conclusions, to fully investigate whether isolates containing mutations within

the *rpo*E gene exist in the clinical setting a much larger study would need to be performed. The fact that 3 of the 4 SCVs isolated from carriage were misidentified as *S. pneumoniae* highlights the challenges of pneumococcal SCV-based research, however this is evidence that smaller pneumococcal colony phenotypes do exist in the clinical setting and that further work is needed to determine their clinical relevance. Due to a lack of research into pneumococcal SCVs, their true importance within the epidemiology of this bacterium remains unclear.

It is important to note that additional SNPs were identified in the variant call files (with a quality score >20) which were not present in the *de novo* assembled contigs. Additionally there were SNPs identified in both the .fasta and .vcf files that did not have sufficient coverage to confirm a confident SNP (<20x coverage). This highlights the need to validate SNPs manually instead of relying on automated quality scores, which may result in erroneous data. These SNPs were present in BOX elements, mobile genetic elements, acetate kinase, general stress protein Gls24 family IS1380-Spn1, and transposase. Re-sequencing using traditional Sanger sequencing may help elucidate whether these SNPs are genuine. An additional point to note is that we observed no mutations within the capsule genes of the biofilm-derived variants. This challenges previous studies by Allegrucci and Sauer (2007), who implicated a 7kb deletion in the *cps3DSU* operon to be responsible for the phenotype of the SCV variants of serotype 3 (Allegrucci and Sauer, 2007). All of the genomes sequenced in this work showed no change in the serotype or MLST profile compared to the parent strain (ST433). This was not unexpected as the housekeeping genes used to determine ST do not readily mutate due to their essential function within the cell. To investigate this further a larger sample number would need to be taken from multiple biofilm experiments, to see if mutations can arise which result in a change in ST.

Chapter 6

Proteomic analysis of biofilm-derived small colony variants

6 Chapter 6: Proteomic analysis of biofilm-derived small colony variants

6.1 Introduction

Using conventional phenotyping coupled with next generation WGS it has been observed that biofilm-derived small colony variants of *S. pneumoniae* exhibit increased biofilm formation, reduced capsule expression and slower growth phenotype which has been attributed to mutations within the DNA-directed RNA polymerase delta subunit (RpoE) ranging from truncation mutations to large scale deletions. RpoE has been linked to virulence and environmental adaptation (Xue *et al*., 2010, Xue *et al*., 2011) and is the only gene which is consistently mutated within the sequenced biofilm-derived variants.

The amino acid sequence of the RpoE protein is divided into 4 distinct regions; a conserved RNA-polymerase delta domain at the N-terminus and three low complexity regions located towards the C-terminus. All mutations seen in the biofilm-derived SCVs would affect one or more of the C-terminal low complexity regions and as such are likely to affect the function of this gene. Low complexity regions are thought to have both a positional dependant role and a role important for protein-protein interaction (Coletta *et al*., 2010). The *rpo*E gene is unique to Gram-positive bacteria and a definitive role for *rpo*E has yet to be discovered; the aim of this work was to extract the whole proteome of the all five sequenced biofilm-derived SCVs from the pilot experiment and the wild type 22F parent strain for UPLC/MS_E to determine what affect the *rpo*E mutations may have on the protein expression of the small colony variants. Based on previous literature in other *Streptococcus spp.* (Xue *et al*., 2010, Xue *et al*., 2011), it can be hypothesised that discrepancies in metabolic proteins will be seen between the SCVs and the parent strain.

6.2 Materials and Methods

See sections 2.1 to 2.5 and 2.23 to 2.26 for more detail.

6.3 Results

6.3.1 Optimisation of whole proteome extraction

In order to optimise the extraction of the whole proteome from the pneumococcus, protein yields were obtained using 100 mM Triethylammonium bicarbonate (TEAB) buffer with 6 different extraction reagents [(1)100 mM TEAB + 0.1 % RapiGest or (2) 1 M urea or (3) 4 M guanidine hydrochloride or (4) 1 M urea + 1 % deoxycholic acid or (5) 4 M guanidine hydrochloride + 1 % deoxycholic acid or (6) 4 M guanidine hydrochloride + 1 % deoxycholic acid + 0.1 % RapiGest].

Initial optimisation using 10 mL planktonic cultures revealed that the extraction buffers containing 100 mM TEAB + 4 M guanidine hydrochloride or 1 M urea yielded the highest protein concentrations of 0.076 mg mL^{-1} and 0.87 mg mL^{-1} respectively (Table 6-1). The extraction buffer containing TEAB + 1 M Urea + 1 % deoxycholic acid yielded the lowest value of 0.0094 mg mL^{-1}. In all cases the yields attained was not sufficient for UPLC/MS_E analysis (~0.5-1 mg mL^{-1}).

Table 6-1: Protein extraction buffers and the relative protein concentrations achieved and from 10 mL planktonic cultures

Extraction buffer (100 mM TEAB +)	Buffer number	Protein (mg mL^{-1})
0.1 % RapiGest	1	0.027
1 M urea	2	0.087
4 M guanidine hydrochloride	3	0.076
1 M urea + 1 % deoxycholic acid	4	0.0094
4 M guanidine hydrochloride + 1 % deoxycholic acid	5	0.021
4 M guanidine hydrochloride + 1 % deoxycholic acid + 0.1 % RapiGest	6	0.028

In order to increase the protein yield, extractions were repeated using larger 40 mL cultures and the following buffers [(2) 1 M Urea or (3) 4 M guanidine hydrochloride or (5) 4 M guanidine hydrochloride + 1 % deoxycholic acid or (6) 4 M guanidine hydrochloride + 1 % deoxycholic acid + 0.1 % RapiGest]. Due to its putative ability to extract membrane bound proteins (Lin *et al*., 2014), deoxycholic acid was an ideal component to include in the extraction buffer. Therefore, and despite the lower protein yields, extraction buffers containing 1 % deoxycholic acid were repeated.

Table 6-2: Protein extraction buffers and the relative protein concentrations achieved and from 40 mL planktonic cultures

Extraction buffer (100 mM TEAB +)	Buffer number	Protein (mg mL^{-1})
1 M urea	2	0.72
4 M guanidine hydrochloride	3	0.98
4 M guanidine hydrochloride + 1 % deoxycholic acid	5	0.71
4 M guanidine hydrochloride + 1 % deoxycholic acid + 0.1 % RapiGest	6	0.78

The repeated extractions using 40 mL planktonic cultures revealed that buffer 3 (containing 100 mM TEAB + 4 M guanidine hydrochloride) yielded the highest protein concentration of 0.98 mg mL^{-1} (Table 6-2). Using 40 mL cultures, all extraction buffers yielded protein concentrations sufficient for MS analysis (Table 6-2). To determine the purity of the

extracted proteins and to ensure that a wide spectrum of proteins was extracted, all samples were run on a one-dimensional (1-D) 12 % Bis-tris polyacrylamide gel (Figure 6-1). Extraction buffers 5 and 6 (which contained 1 % deoxycholic acid) both exhibited an unexpected banding pattern at >260 kDa suggesting impurities in the protein samples. A similar staining pattern can be seen in buffer 2 (1 M urea) but not in buffer 3 (4 M guanidine hydrochloride) suggesting that buffer 3 extracts a protein sample free from impurities (Figure 6-1). In addition, intense staining in lower molecular weight proteins in samples extracted with buffers 3, 5, 6 was observed which was not seen in buffer 2, suggesting that the 4 M guanidine hydrochloride may help increase the extraction of lower molecular weight proteins. Considering the buffer containing 100 mM TEAB + 4 M guanidine hydrochloride yielded the highest protein concentration of 0.98 mg mL^{-1} and the cleanest banding pattern, buffer 3 was used for all subsequent extractions for UPLC/MS_E analysis.

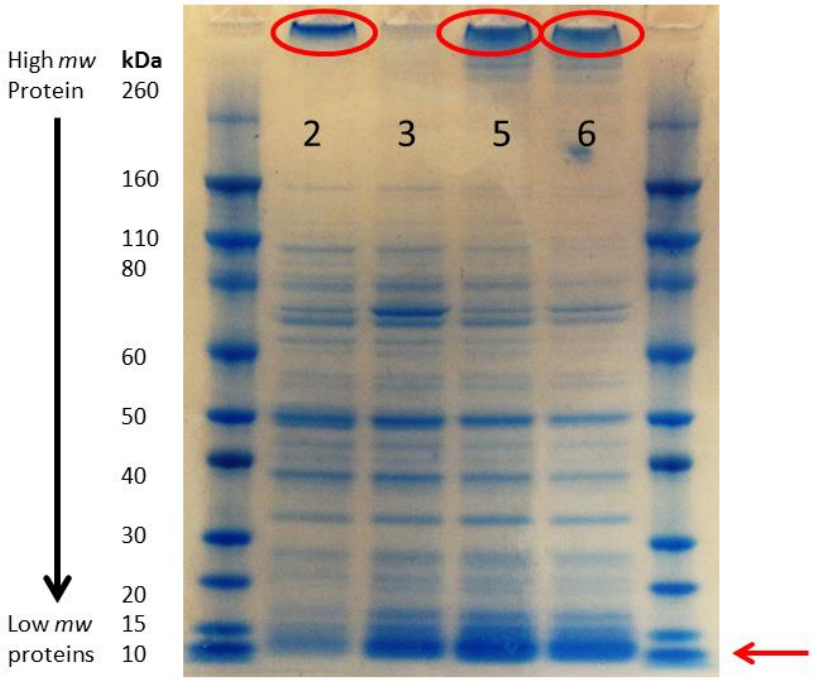


Figure 6-1: One dimensional gel of protein extraction buffers

15 µg of protein was loaded onto a standard 12 % Bis-tris Nu-PAGE pre-cast polyacrylamide gel in MOPS buffer with 10 µL Novex® Sharp Pre-stained Protein Standards to determine band weight. Banding was stained with SafeBLUE Protein Stain. Irregular banding in the high molecular weight proteins in buffers 2, 5 and 6 is highlighted (red circles). Increased staining in the lower molecular weight proteins ~ 10 kDa is also highlighted (red arrow).

6.3.2 Extraction of the variant proteomes

All proteins for the SCVs and wild type samples were extracted using buffer 3 containing 100 mM TEAB + 4 M guanidine hydrochloride. Each sample was extracted in triplicate from independent cultures. Using the Bradford's assay (Bradford, 1976) regression curve described in the materials and methods section 2.23, the following final concentrations were determined for the three biological replicates: all samples were normalised to 45 µg prior to mass spectrometry analysis (Table 6-3).

Table 6-3: Extracted protein concentrations from the biofilm-derived SCVs

Sample	Sample abbreviation	Replicate 1 (µg)	Replicate 2 (µg)	Replicate 3 (µg)
WT	WT	99.4	101.8	45.8
SCV1D3	S1	91.8	49.9	68.7
SCV5D3	S5	82.8	82.3	46.6
SCV3D9	S3	82.8	68.5	62.6
SCV7D9	S7	78.5	75.2	71.3
SCV9D9	S9	53.8	55.2	49.9

To determine purity of the extracted protein and to visualise the overall protein profile, all samples were run on a 1-D 12 % Bis-tris polyacrylamide gel. The protein profile appeared very similar between the variants; an interesting observation was the presence of a higher intensity band at the molecular weight of ~ 75 kDa in the SCV proteome compared to the WT sample (Figure 6-2). Similarly, the presence of a higher intensity band at the molecular weight of ~ 70 kDa in the WT sample compared to the SCV profiles was also observed in the 1-D gel. These observations suggested putative gene up-regulation and down-regulation in the SCVs. Finally the intensity of the banding at ~65 kDa appeared greater in the SCVs than in the WT suggesting increased expression. Importantly all extractions ran cleanly suggesting the samples are free from contaminant which may have affected the protein migration though the polyacrylamide gel. All samples were sent to the Centre for Proteomic Research at the University of Southampton for quantification via label-free UPLC/MS_E.

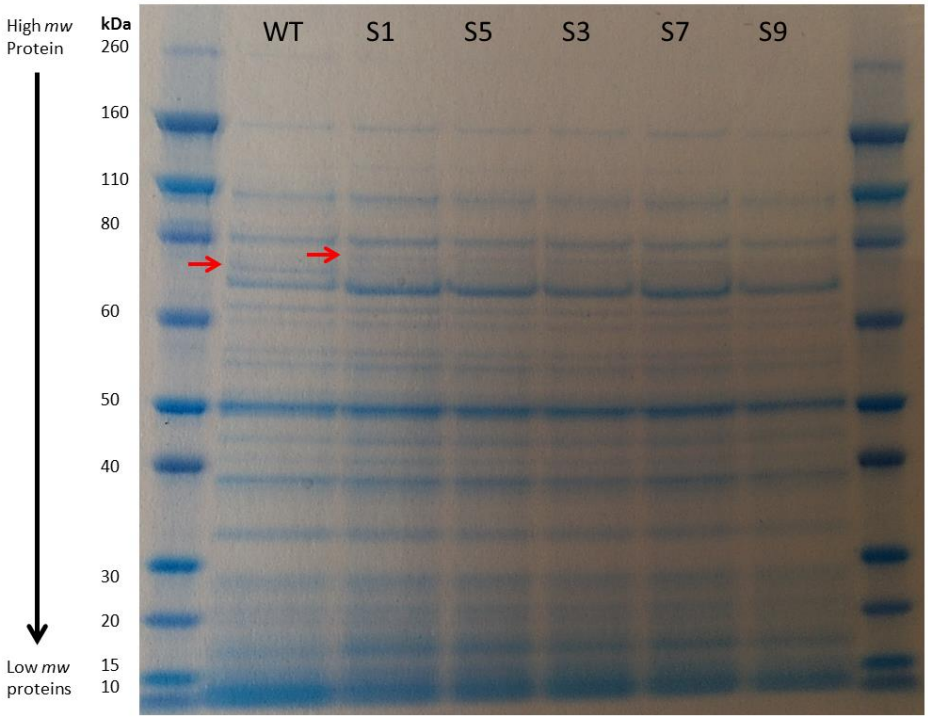


Figure 6-2: One dimensional gel of extracted protein from biofilm-derived SCVs

15 µg of protein was loaded onto a standard 12 % Bis-tris Nu-PAGE pre-cast polyacrylamide gel in MOPS buffer with 10 µL Novex® Sharp Pre-stained Protein Standards to determine band weight. Banding was stained with SafeBLUE Protein Stain. Additional band in WT at ~70 kDa and additional band in SCVs at ~75 kDa are highlighted with a red arrow.

6.3.3 UPLC/MS_E Mass spectrometry

Analysis of the proteins detected using label-free UPLC/MS_E was performed on the wild type 22F parent strain and two SCVs (SCV5D3 and SCV3D9). Three biological replicates of each sample tested were analysed. One-way ANOVA revealed no significant difference between the total detectable protein concentration between the wild type and the SCV datasets ($p = 0.305$). Prior to analysis all protein concentrations were normalised by dividing the protein concentration by the total protein concentration detected in each technical replicate. Proteins concentrations were subsequently filtered to exclude proteins that were present within in only one MS technical replicate in a given sample. Proteins were further filtered by excluding proteins that were present in only one biological replicate. Using these exclusion criteria, a total of 253 proteins and 222 proteins were detected in SCV5D3 and SCV3D9, respectively, compared to a total of 173 in the wild type parent strain.

Further analysis revealed a total of 26 up-regulated proteins (>1.5 fold increase in protein concentration) and 34 down-regulated proteins (>1.5 fold decrease in protein concentration) in SCV5D3 relative to the wild type parent. In comparison, a total of 17 up-regulated proteins and 5 down-regulated proteins were identified in SCV3D9. Fold changes were only considered significant when the average protein concentrations had non-intersecting 95 % confidence intervals (CIs) with the wild type. Protein accession numbers and descriptions of up and down-regulated proteins can be seen in Table 8-7-Table 8-8 for SCV5D3 and Table 8-9-Table 8-10 for SCV3D9 in Appendix 6.

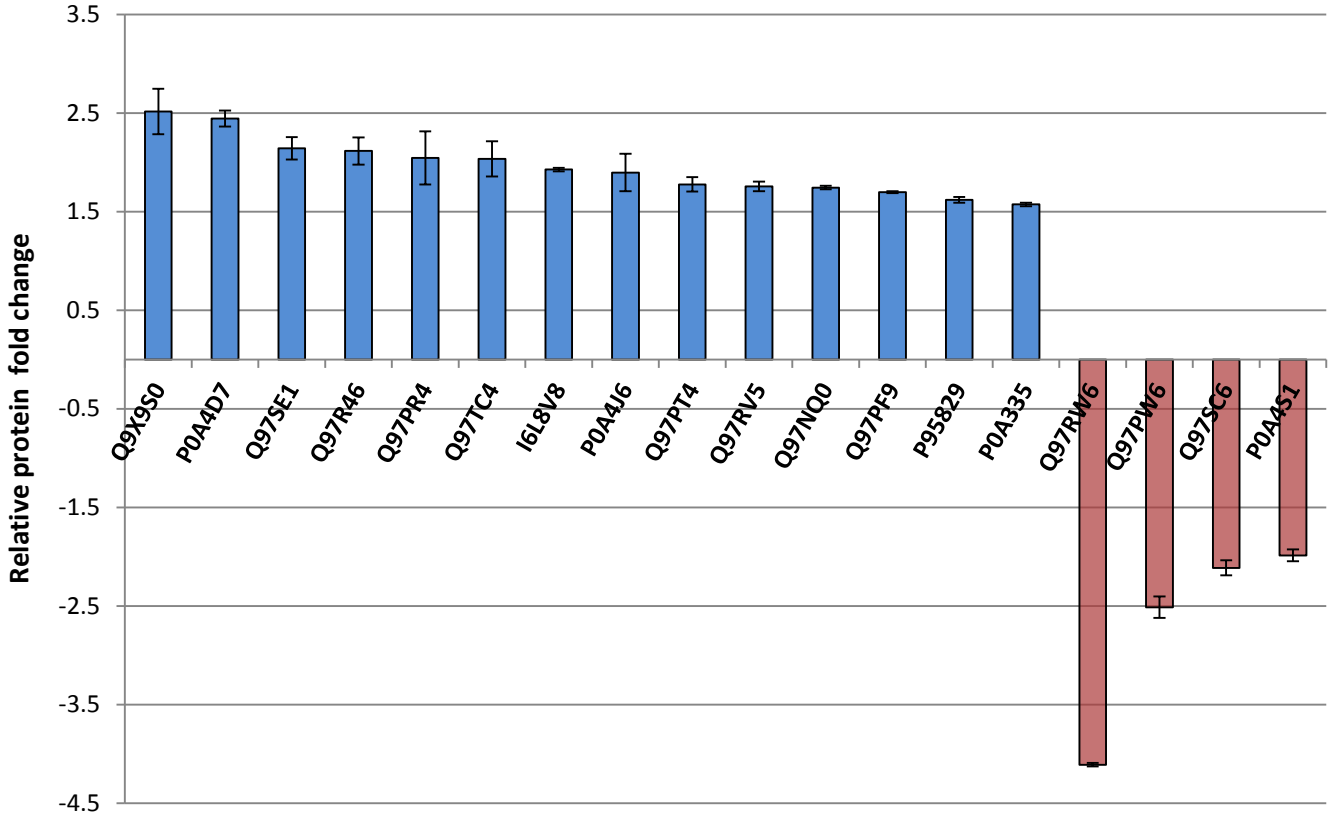
A total of 14 proteins were consistently up-regulated between the two SCVs tested as depicted in Figure 6-3. The highest fold increase was seen in the probable dihydroorotate dehydrogenase A (Q9X9S0) with an average fold increase of 2.51 x, followed by the probable ATP-dependent RNA helicase exp9 (P0A4D7) (2.44 x average fold increase), a putative uncharacterized protein (Q97SE1) (2.14 x fold increase), the bifunctional protein GlmU (Q97R46) (2.11 x average fold increase), the Cof family protein/peptidyl-prolyl cis-trans isomerase cyclophilin type (Q97PR4) (2.04 x average fold) and the hypoxanthine-guanine phosphoribosyltransferase (Q97TC4) (2.03 x average fold increase). Three proteins were consistently up-regulated between the two SCVs tested which have links to stress response, these included the manganese superoxide dismutase (P0A4J6) (1.90 x average fold increase), chaperone protein DnaK (P95829) (1.62 x average fold increase) and the 60 kDa chaperonin (P0A335) (1.57 x average fold increase).

In contrast a total only 4 proteins were consistently down-regulated between the two SCVs tested as depicted in Figure 6-3. The highest fold decrease was seen in a putative uncharacterized protein (Q97RW6) with an average fold increase of 4.14 x followed by glycine--tRNA ligase beta subunit (Q97PW6) (2.72 x average fold decrease), the formate acetyltransferase (Q97SC6) (2.26 x average fold decrease) and finally the fructose-bisphosphate aldolase (P0A4S1) (1.98 x average fold decrease).

Of note an increase in the ATP-dependent Clp protease (Q97SK0) was observed in both SCV5D3 (2.41 fold) and SCV3D9 (3.17 fold). The ATP-dependent Clp protease ATP-binding is 77 kDa in weight and likely accounts for increased intensity in the band at the molecular weight of ~ 75 kDa in the SCV profiles in the 1-D gel (Figure 6-2). Similarly a decrease in the glycine--tRNA ligase beta subunit (Q97PW6) was observed in both SCV5D3 (2.41 fold) and SCV3D9 (3.17 fold). The glycine--tRNA ligase beta subunit is 75 kDa in weight and likely accounts for increased intensity in the band at the molecular weight of ~ 70 kDa in the WT profile displayed in the 1-D gel (Figure 6-2).

The probable DNA-directed RNA polymerase subunit delta (P66717) (RpoE) was detected in both the SCVs and the wild type parent strain, however was not comparable due to the lack of detection in all biological replicates. Nonetheless its presence indicates the protein was expressed in SCVs, and suggests its regulation was unaffected by the mutations. To determine any known or predicted protein-protein interactions of the up and down-regulated proteins and to determine any direct association with RpoE, accession numbers were inputted into the STRING 9.1 online database (Jensen *et al*., 2009). Of the down-regulated proteins only fructose-bisphosphate aldolase was shown to have an interaction network with up-regulated protein 60 kDa chaperonin, chaperone protein DnaK, and the cell division protein FtsZ. In contrast, a total of 15 known or predicted interactions were observed in the up-regulated proteins, as depicted in Figure 6-4A. A clear interaction network was present among the 60 kDa chaperonin, chaperone protein DnaK, cell division protein FtsZ and manganese superoxide dismutase which had association scores ranging from 0.712 to 0.99 (Figure 6-4B). Of note the only observed association with *rpo*E was with the cell division protein FtsZ with an association score of 0.436 (Figure 6-4A&B).

A



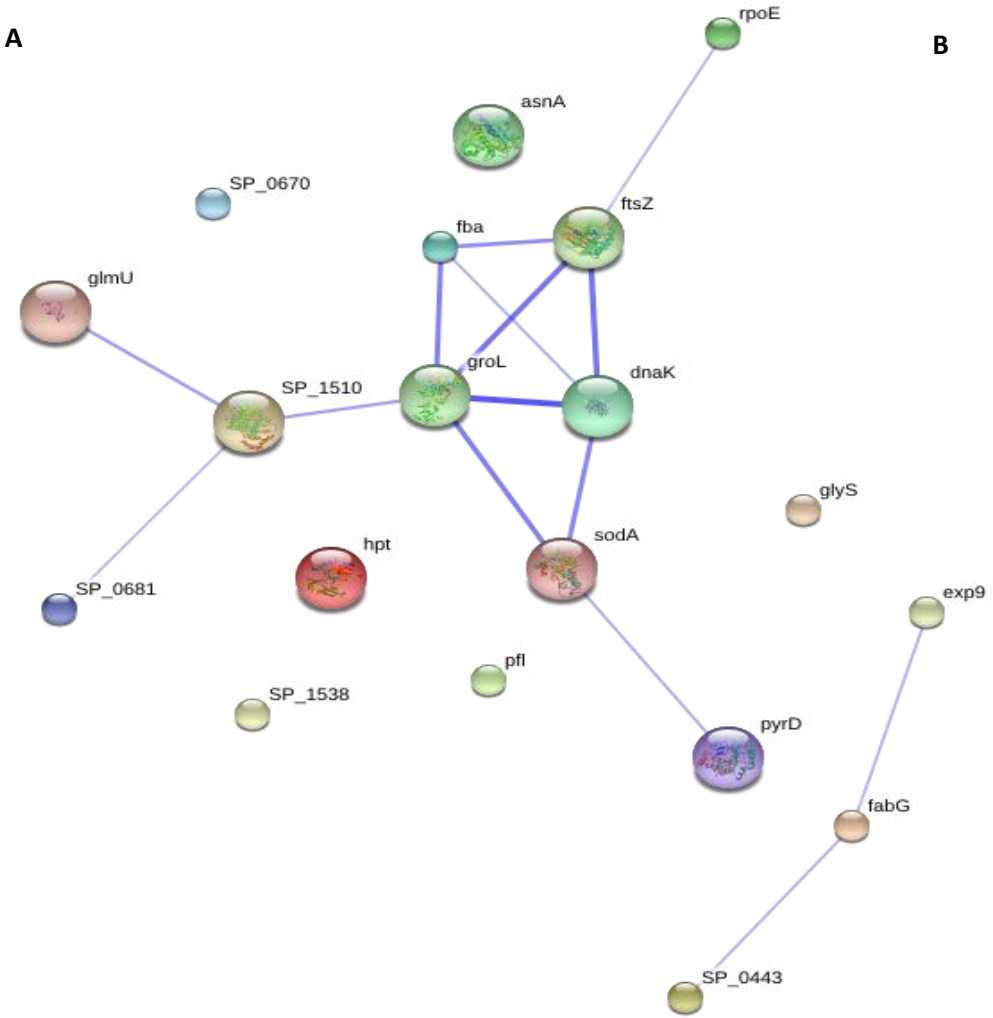
B

Protein ID	Protein Name	Description
Q9X9S0	pyrD	Probable dihydroorotate dehydrogenase A (fumarate)
P0A4D7	exp9	Probable ATP-dependent RNA helicase exp9
Q97SE1	SP_0443	Putative uncharacterized protein
Q97R46	glmU	Bifunctional protein GlmU
Q97PR4	SP_1538	Cof family protein/peptidyl-prolyl cis-trans isomerase_ cyclophilin type
Q97TC4	hpt	Hypoxanthine-guanine phosphoribosyltransferase
I6L8V8	fabG	3-oxoacyl-[acyl-carrier protein] reductase
P0A4J6	sodA	Superoxide dismutase [Mn]
Q97PT4	SP_1510	ATP synthase subunit alpha
Q97RV5	SP_0681	Elongation factor Tu family protein
Q97NQ0	asnA	Aspartate--ammonia ligase
Q97PF9	FtsZ	Cell division protein FtsZ
P95829	DnaK	Chaperone protein DnaK
P0A335	groL	60 kDa chaperonin
Q97RW6	SP_0670	Putative uncharacterized protein
Q97PW6	glyS	Glycine--tRNA ligase beta subunit
Q97SC6	pfl	Formate acetyltransferase
P0A4S1	fba	Fructose-bisphosphate aldolase

Figure 6-3: Protein fold change in biofilm-derived small colony variants relative to the WT 22F parent strain

Proteomes extracted from the wild type 22F parent strain and two independent SCVs were quantified via UPLC/MS_E mass spectrometry. A total of 14 proteins were consistently up-regulated (blue) and 4 were consistently down-regulated (red) in all SCVs tested (A) (>1.5 fold increase with non-intersecting 95 % confidence intervals) and corresponding protein descriptions (B). Error bars represent standard error between the two SCVs tested.

A



B

Protein 1	Description		Protein 2	Description	Combined association score
groL	60 kDa chaperonin	↔	**dnaK**	Chaperone protein DnaK	0.990
ftsZ	Cell division protein FtsZ	↔	**dnaK**	Chaperone protein DnaK	0.824
groL	60 kDa chaperonin	↔	**ftsZ**	Cell division protein FtsZ	0.817
groL	60 kDa chaperonin	↔	**sodA**	Superoxide dismutase [Mn]	0.761
groL	60 kDa chaperonin	↔	**fba**	Fructose bisphosphate aldolase	0.743
sodA	Superoxide dismutase [Mn]	↔	**dnaK**	Chaperone protein DnaK	0.712
ftsZ	Cell division protein FtsZ	↔	**fba**	Fructose bisphosphate aldolase	0.673
SP_1510	ATP synthase subunit alpha	↔	**glmU**	Bifunctional protein GlmU	0.564
groL	60 kDa chaperonin	↔	**SP_1510**	ATP synthase subunit alpha	0.559
sodA	Superoxide dismutase [Mn]	↔	**pyrD**	Probable dihydroorotate dehydrogenase A	0.476
dnaK	Chaperone protein DnaK	↔	**fba**	Fructose bisphosphate aldolase	0.463
fabG	3-oxoacyl-[acyl-carrier protein] reductase	↔	**SP_0443**	Putative uncharacterized protein	0.458
fabG	3-oxoacyl-[acyl-carrier protein] reductase	↔	**exp9**	Probable ATP-dependent RNA helicase exp9	0.456
rpoE	DNA-directed RNA-polymerase delta subunit	↔	**ftsZ**	Cell division protein FtsZ	0.436
SP_1510	ATP synthase subunit alpha	↔	**SP_0681**	Elongation factor Tu family protein	0.432

Figure 6-4: STRING network of *rpo*E and up-regulated protein in biofilm-derived SCVs

Protein accession numbers were inserted into the STRING 9.1 online database to determine the known and predicted protein-protein interactions with *rpo*E (A) and corresponding combined association score for each protein-protein interaction (B). Interactions are represented by the blue lines; strength of association is represented by the thickness of the line.

6.4 Discussion

In this chapter the protein extraction procedure for *S. pneumoniae* has been successfully optimised and used to extract proteomes for quantitative analysis to determine the effect of the *rpo*E mutations in the biofilm-derived SCVs.

Initially a 1-D gel (Figure 6-2) was used to display the full protein profile of the extracted SCV proteins. This showed that the protein profile of the SCVs appeared very similar to the wild type profile; an interesting observation was the presence of a band in the SCVs at the molecular weight of ~75 kDa which was not evident in the wild type sample, suggesting a gene up-regulation. Analysis of the UPLC/MS_E would suggest this banding to be due to a 2-3 fold increase in the ATP-dependent Clp protease. Furthermore the presence of a band at the molecular weight of ~ 70 kDa which was present in the wild type but less visible in the SCVs suggested gene down-regulation in the SCVs. Analysis of the UPLC/MS_E would suggest that this banding pattern is be due to 2-fold decrease of the glycine-tRNA ligase beta subunit.

The exclusion criteria used in the UPLC/MS_E analysis were devised to include the presence of proteins detected in at least two technical replicates and two biological replicates, these criteria reduced increasing bias into the analysis by excluding data unnecessarily. A total of 253 proteins and 222 proteins were detected in SCV5D3 and SCV3D9, respectively, using these criteria, compared to a total of 173 in the wild type parent strain. This increase of detectable proteins in the SCVs may suggest one of three hypotheses; firstly, the increase in proteins may be due to technical limitations of the UPLC/MS_E to detect proteins within the wild type samples, perhaps due to errors introduced during the preparation stage. Secondly, the increase in proteins may suggest an increase in unregulated gene expression in the SCVs as a direct consequence of the *rpo*E mutations. This hypothesis is not implausible considering *rpo*E has previously been linked to global gene up-regulation in *S. mutans* (Xue *et al*., 2010, Xue *et al*., 2011, Xue *et al*., 2012). Thirdly, the reduction in capsule of the SCVs may have aided protein extraction, including extraction of membrane bound proteins. Considering, however that the SCVs yielded lower protein concentrations in the first and second biological replicates, this latter hypothesis is unlikely.

Overall a total of 26 proteins were up-regulated and a total of 34 proteins were down-regulated in SCV5D3 relative to the wild type parent and a total of 17 proteins were up-regulated and 5 proteins were down-regulated in SCV3D9. This up and down regulation seen is

consistent with the normalisation of gene expression observed in *rpo*E mutants of *S. aureus* by Weiss *et al.* (2014). These authors showed that loss of the delta subunit in *S. aureus* strains resulted in down-regulation of virulence factor and up-regulation of mobile genetic elements. Considering the role of *rpo*E this is not unexpected and suggests that the mutations seen in the SCVs have a direct effect on the protein expression in the cell.

When comparing the proteomic dataset from SCV5D3 and SCV3D9, a total of 14 consistently up-regulated and 4 consistently down-regulated proteins were observed. The up-regulation of the 60 kDa chaperonin (P0A335), the chaperone protein DnaK (P95829) and the manganese superoxide dismutase (P0A4J6) is consistent with the increased ability of SCVs to form biofilm as all four proteins are induced under stress conditions (Yesilkaya *et al*., 2000, Kwon *et al*., 2003). SCV3D9 reported a 3-fold increase in the ATP-dependent Clp protease (Q97SK0), a 2-fold increase was also observed in SCV5D9 however due to intersecting 95 % CIs with the wild type, this increase could not be considered significant. Nonetheless the 1-D gel indicates this protein is up-regulated among all SCVs; pneumococcal Clp genes have been directly linked to pneumococcal biofilm formation (Allan *et al*., 2014) in addition to adherence and virulence (Chastanet *et al*., 2001, Tu *et al*., 2007, Zhang *et al*., 2009). Thus the ATP-dependent Clp protease up-regulation may be a key phenotype of biofilm-derived SCVs and help explain the increased biofilm formation seen in these variants. Consistent down-regulation of the fructose-bisphosphate aldolase (P0A4S1), formate acetyltransferase (Q97SC6) and the glycine-tRNA ligase beta subunit (Q97PW6) is consistent with the decrease in growth rate in SCVs as all three proteins are involved in pneumococcal metabolism. Furthermore fructose-bisphosphate aldolase has previously been shown to be down-regulated in pneumococcal biofilms along with three additional glycolytic proteins (Allan *et al*., 2014). This down-regulation was suggested to promote a non-invasive phenotype to increase pneumococcal persistence (Allan *et al*., 2014). Considering SCV3D9 was able to metabolise the full profile of the carbon substrates in the API Rapid ID 32 Strep assay, this down-regulation is unlikely to account for the metabolic differences seen in all SCV API profiles. However the down-regulation of 19 ribosomal proteins, coupled with the down-regulation of the PTS mannose-specific IIAB component (Q97SP2) and the fructose-bisphosphate aldolase (P0A4S1) seen in SCV5D3 may help explain the inability of this variant to metabolise the full carbon substrate profile in the API Rapid ID 32 Strep assay as the PTS mannose-specific IIAB component is important for the uptake and phosphorylation of specific carbohydrates from the extracellular environment (Hoskins *et al.,* 2001, Bidossi *et al.,* 2012) (see Figure 7-1) and

the fructose-bisphosphate aldolase plays an important role in the glycolysis pathway (Allan *et al.,* 2014).

Interestingly there are a number of similarities and disparities between these data and that observed by Xue *et al.* (Xue *et al.,* 2012) in their ΔRpoE *S. mutans* mutant. Notably this work observed up-regulation in the ATP-dependent Clp protease, 3-oxoacyl-[acyl-carrier-protein] synthase 2 and superoxide dismutase [Mn] whereas Xue *et al.* observed down-regulation in the ΔRpoE *S. mutans* mutant of -2.3-fold, -1.4-fold and -8.1-fold, respectively, when grown under various conditions. Furthermore this work observed decreases in the fructose-bisphosphate aldolase and alcohol dehydrogenase which was shown to be up-regulated 1.5 and 2.1 fold respectively, in the ΔRpoE *S. mutans* mutant. Similarities between these studies include the up-regulation of the chaperone protein DnaK and down-regulation of formate acetyltransferase (Xue *et al*., 2012). The disparity between datasets can be accounted for by the difference in growth conditions used, *Streptococcal* species variation and severity of the ΔRpoE *S. mutans* mutant. Nonetheless, these data both suggest that *rpo*E plays a vital role in the regulation of these proteins and has a clear role in stress response.

Furthermore, the chaperone protein DnaK (P95829), inosine-5'-monophosphate dehydrogenase (Q97N43), Cof family protein/peptidyl-prolyl cis-trans isomerase cyclophilin type (Q97PR4), elongation factor Tu family protein (Q97RV5) and the dihydroorotate dehydrogenase A (fumarate) (Q9X9S0) were up-regulated in one or both of the SCVs. All five proteins have been shown previously to be involved in biofilm maturation (Allegrucci *et al*., 2006) between 3 and 9 days. This observation may simply be an artefact of SCV5D3 and SCV3D9 being harvested from 3 and 9 day-old biofilms. Alternatively, it may account for the increased biofilm formation seen in the SCVs and further support the hypothesis that *rpo*E plays a vital role in biofilm formation. An additional up-regulated protein in SCV5D3 which merits discussion is the 2-fold increase in the protein RecA (P0A451). RecA is an important protein in the transformation machinery, facilitating the recognition of homologous regions within the bacterial host chromosome for the exchange of transformed DNA (Johnston *et al*., 2014). The increased expression of RecA, coupled with the reduced capsule expression (Figure 4-10) may suggest an increased transformation efficiency and higher propensity to take up exogenous DNA.

Method development in this chapter is also noteworthy of discussion. Initially 10 mL cultures of a 10^8 CFU mL^{-1} cell concentration were used to optimise protein extraction using 6 different extraction buffers [(1)100 mM TEAB + 0.1 % RapiGest or (2) 1 M urea or (3) 4 M

guanidine hydrochloride or (4) 1 M urea + 1 % deoxycholic acid or (5) 4 M guanidine hydrochloride + 1 % deoxycholic acid or (5) 4 M guanidine hydrochloride + 1 % deoxycholic acid+ 0.1 % RapiGest]. All reagents are all well-documented methods of protein extraction (Leon *et al*., 2013, Allan *et al*., 2014, Lin *et al*., 2014). The rationale behind the use of these reagents was based on literature research and previous optimisation within the Welcome Trust Research facility, UHS performed by Dr Ray Allan. In the mechanical lysis stage, it was noticed that the use of Qiagen Pathogen lysis tubes S yielded poor protein titres so lysis was repeated using MP lysis matrix B tubes, which yielded a higher protein concentration. The increased number of 0.1 mm silica spheres in the MP lysis matrix B tubes was most likely the cause of increased yield. Using the Bradford's assay regression curve for quantification, the extraction buffers containing 100 mM TEAB + 1 M urea OR 4 M guanidine hydrochloride were shown to achieve the best protein yield. Interestingly the extraction buffer comprising of 100 mM TEAB + 1 M urea + 1 % deoxycholic acid revealed a very low protein concentration suggesting that perhaps the urea and deoxycholic acid interact, preventing efficient extraction. In all cases the protein yield using 10 mL cultures was too low for UPLC/MS_E analysis (0.009-0.87 mg mL^{-1}).

In order to increase the protein yield, extractions were repeated as before but with 40 mL cultures of a 10^8 CFU mL^{-1} cell concentration. The extraction buffers used included (2) 100 mM TEAB + 1 M urea or (3) 4 M guanidine hydrochloride or (5) 4 M guanidine hydrochloride + 1 % deoxycholic acid or (6) 4 M guanidine hydrochloride + 1 % deoxycholic acid+ 0.1 % RapiGest]. Upon repeating the extractions, the protein yields were improved; ranging from 0.71-0.98 mg mL^{-1} and all four extraction buffers produced an adequate banding pattern. Interestingly, the two extraction buffers which contained 1 % deoxycholic acid displayed an irregular banding pattern at the high molecular weight markers (>260 kDa). This pattern was also present but not as distinct in the 100 mM TEAB + 1 M urea buffer. The irregular banding is likely an indication that the sample was not clean and contained some residual deoxycholic acid or urea which may have interacted with components within the stacking gel to prevent protein migration and thus create the banding pattern seen. It was also observed that the low molecular weight proteins were visually less abundant in the 100 mM TEAB + 1 M urea buffer. From these four buffers 100 mM TEAB + 4 M guanidine hydrochloride was selected as the optimal buffer to use; due to the fact that it yielded the greatest concentration of protein (0.98 mg mL^{-1}) and had the cleanest banding pattern. Protein samples were normalised to approximately 15 µg prior to being run on the 1-D gel, which allowed the

observations to be semi-quantitative. It should also be noted that all optimisation was performed only on the wild type 22F parent strain.

Overall this work has shown valuable insight into the effect of *rpo*E mutations on the proteomic phenotype of the SCVs, however it would benefit from the further study of the remaining 10 sequenced SCVs. This work has shown that SCVs exhibit up-regulation of a common sub-set of stress-inducible proteins which are part of an interaction network consisting of the 60 kDa chaperonin, chaperone protein DnaK, cell division protein FtsZ and manganese superoxide dismutase. These proteins can be directly linked to adherence and virulence of the pneumococcus, providing insight into the increased biofilm forming properties of the SCVs. This uniformity in expression, coupled with evidence of association with *rpo*E, suggests that the *rpo*E mutations may have had a direct result in their up-regulation. Furthermore, the change in expression of biofilm-related proteins such as ATP-dependent Clp protease and fructose-bisphosphate aldolase may suggest that biofilm-derived mutation within the *rpo*E gene re-programs the SCV proteome in favour of biofilm formation, colonisation and ultimately long-term persistence. To further verify the role of *rpo*E a knockout mutant and complementation mutant should be made to corroborate the current findings.

Chapter 7

Discussion and future work

7 Chapter 7: Discussion

S. pneumoniae is a commensal human pathogen and the causative agent of invasive pneumococcal disease worldwide (O'Brien *et al*., 2009). Carriage and persistence of the pneumococcus in the nasopharynx of humans is thought to be mediated by biofilm formation and the inherent diversification within the biofilm environment (Marks *et al*., 2012b). To date this work has shown that pneumococci grown under biofilm conditions exhibit colony morphology variation and that this variation can be used as a marker for phenotypic and genetic variation. Previously biofilm-derived pneumococcal biofilm studies have focused predominantly on serotype 3 (Waite *et al*., 2001, Allegrucci and Sauer, 2007, McEllistrem *et al*., 2007, Allegrucci and Sauer, 2008, Domenech *et al*., 2009); this work has extended previous work by showing that the morphological variation is common to multiple serotypes, not just serotype 3. The aim of this thesis was to understand the level of phenotypic and genetic diversity that arises in clinical isolates of *S. pneumoniae* when grown under biofilm conditions. The objectives of this thesis were to document the repertoire of genetic mutations in biofilm-derived colony morphology variants and relate these mutations to the phenotypic diversity seen. This project has helped understand the role of biofilms in driving phenotypic and genetic diversification in new variant lineages, shown that the mutations seen are recurring and predictable in nature, and that the mutations seen may enhance the fitness of the organism by increasing the ability to form biofilms thus facilitating long-term persistence within the host.

Serotype 22F was selected as a model serotype due to its ability to form good biofilms and the relative stability of the harvested biofilm-derived morphological variants. Serotype 14 did produce stable SCVs but due to the limited number of stable variants this serotype remains to be characterised phenotypically or genetically. In order to phenotypically characterise the biofilm-derived colony variants of serotype 22F assessment of the growth rate, metabolic profile, biofilm formation, antibiotic resistance, capsule production and mutation frequency was performed and compared to the wild type parent strain. We have shown that biofilm-derived small colony variants have a reduced growth rate, a reduced metabolic profile, less capsule production and increased biofilm formation compared to the parent strain.

To determine a causal link between the phenotypic diversity and genetic diversity of the biofilm-derived variants WGS, using NGS platforms, was employed. Sequence data for each isolate was mapped against the annotated parent strain reference in order to identify verified

mutations. In addition all sequences were assembled *de novo* and the assembled contigs were assessed for changes in serotype and MLST compared to the 22F parent sequence type (ST433).

Using next generation WGS this work has shown that all SCVs sequenced had mutations within the DNA-directed RNA polymerase delta subunit (RpoE) ranging from premature stop codons to large scale deletions. These mutations were not present in the parent population and were seen in SCVs harvested from multiple replicates from different aged biofilms from independent experiments. Moreover the mutations occurred at different positions within the *rpo*E gene indicating clear evidence of parallel evolution. SCVs were not observed after multiple days of planktonic culture and were only seen after three days of biofilm growth. Considering the fact that these mutations were seen at different positions within the *rpo*E gene, and that the original parent population originated from a single clone, we can confidently say that these adaptive mutations are a direct result of biofilm growth and that these mutations did not exist in the original parent population. Based on the amino acid structure of the RpoE protein both the truncation mutations and the large scale deletions seen in the biofilm-derived SCVs would affect one or more of the C-terminal low complexity regions or the conserved RNA-polymerase delta domain at the N-terminus. From this information we can hypothesise that the function of this gene will be affected. Indeed, UPLC/MS_E analysis of the SCVs revealed the consistent up-regulation of a common sub-set of stress-inducible proteins which are part of an interaction network consisting of the 60 kDa chaperonin, chaperone protein DnaK, cell division protein FtsZ and manganesse superoxide dismutase. This work highlights some of the first evidence of biofilm-mediated, real-time parallel evolutionary events in *S. pneumoniae* and suggests a strong positive selection for mutations in the *rpo*E, which may identify *rpo*E as an important gene for biofilm formation, carriage and pathogenicity of the pneumococcus.

RpoE is specific to Gram-positive bacteria and is thought to be responsible for transcription of genes involved in the metabolism of carbohydrates (Xue *et al*., 2011). The mutations in the *rpo*E gene may help explain the reduced metabolic profile seen in the API Rapid ID assay. The observed down-regulation of 19 ribosomal proteins, coupled with the down-regulation of the PTS mannose-specific IIAB component (Q97SP2) and the fructose-bisphosphate aldolase (P0A4S1) seen in UPLC/MS_E analysis of SCV5D3 may help explain the inability of SCVs to metabolise the full carbon substrate profile in the API Rapid ID 32 Strep assay. The PTS is responsible for the detection, translocation and phosphorylation of carbohydrates in *Streptococcal* species (Abranches *et al.,* 2003, Tong *et al.,* 2011). In a 2012 study, Bidossi *et al.* showed that of the 30 identified pneumococcal sugar transporters in

26 pneumococcal genomes, the PTS contributes 15-20 % of the genome, with 21 PTSs identified (Bidossi *et al.,* 2012). Figure 7-1 depicts the carbohydrate transport system and glycolysis pathway within *S. pneumoniae* (Hoskins *et al*., 2001). As shown, the PTS provides a crucial role in facilitating the movement of key carbohydrates across the membrane. Carbohydrates are converted to pyruvate via glycolysis and are the only nutrients which support growth and cell division in *S. pneumoniae* (Hoskins *et al*., 2001). A reduction in carbohydrate metabolism may also help explain the reduction in capsule as formation of the capsule requires the synthesis of polysaccharides (Hardy *et al.,* 2000, Yother, 2011), and reduced metabolic profile would likely result in less energy available for capsule synthesis. As such *rpo*E mutation may seem disadvantageous to the organism, however, a reduction in capsule production has been linked to increased biofilm formation, and using CLSM we have shown that SCVs form biofilms with greater thickness, biomass and surface area than the wild type parent strain. Therefore a hypothesis based on this information could be that mutations within the low complexity regions of the *rpo*E gene may change the transcriptomic/proteomic profile of the variants which may promote colonisation in the form of biofilm formation and thus increase survival within the host.

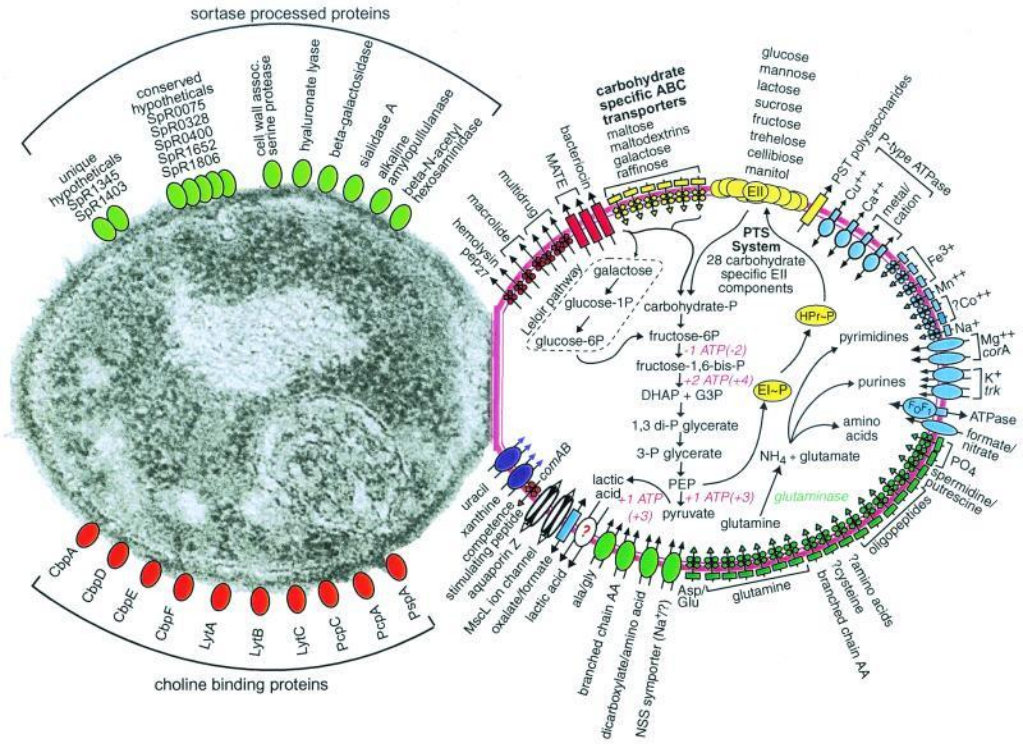


Figure 7-1: *S. pneumoniae* substrate transport systems

Figure depicts the substrate transport systems and glycolysis pathway in *S. pneumoniae*. Transporters include: red = multidrug and peptide exporters, yellow = carbohydrates, blue = cations, green = anions and amino acids, purple = nucleosides, purines, and pyrimidines, white = other. Rectangles = permeases, cylinders = porins, ovals = ATPases. Figure was reproduced from Hoskins *et al.* (2001). Additional permission was not required from the American Society for Microbiology for the republishing of this image in accordance with Copyright Clearance Center's RightsLink® service.

Additional regions of interest within the pneumococci genome included the capsule genes, and major pneumococcal virulence factors, including pneumolysin, pneumococcal surface protein A, neurominidase NanA and NanB, and choline binding protein A; we identified no mutations in these genes, suggesting that these genes are not readily mutated under biofilm conditions. The use of WGS meant that a comprehensive overview of all possible mutations across the entire genome was possible. Previous studies have focused only on the capsule genes thus limiting their findings to these select loci (Allegrucci and Sauer, 2007). WGS analysis also showed that the sequence type (ST) of the variants did not change from the ST of the parent strain (ST433). This was not unexpected as the housekeeping genes used to determine ST do not readily mutate due to their essential function within the cell. To investigate this further a larger sample number would need to be taken from multiple biofilm experiments, to see if mutations can arise which result in a change in ST. An additional experiment of interest would be to culture biofilms of multiple sequence types to see whether there is crossover of different ST alleles which may result in a novel MLST profile. Regions within pneumococcal phage genomes were also relevant considering the work by Webb *et al*. (2004). In 2004, Webb *et al*. showed that exposure of filamentous bacteriophage Pf4 to *P. aeruginosa* results in a SCV due to phase variation, highlighting the fact that the diversity seen from biofilm cultures may be a result of mutations in the phage genome rather than mutations in the host genome (Webb *et al*., 2004). In this work, the pneumococcal phage hyaluronidase gene was also seen to contain SNPs in the initial analysis however the quality of the SNPs were poor and these could not be verified in the *de novo* assemblies or alignment mapping, and were deemed as false-positive SNPs. Further work would be required to confirm the role phage plays in the generating diversity in pneumococcal biofilms.

To test the clinical significance of the biofilm-derived SCV *rpo*E mutations in the wider pneumococcal population whole genome data from 518 pneumococcal carriage isolates was examined to identify different *rpo*E alleles to see if the observed biofilm-derived *rpo*E truncation mutations were present. Twenty-two (22) different alleles were identified in the 518 isolates over the 5-year period; however no alleles matched the mutations seen in the sequenced SCVs. This was not unexpected, due to the fact that the mutations seen were unique to SCVs, and colony morphology was not recorded when the 518 isolates were isolated and sequenced. SCVs are not regularly isolated when culturing isolates from routine carriage studies due to the likelihood of misidentification as seen in the four SCVs isolated from carriage studies in this work. Due to the lack of research into SCVs of the pneumococcus, the epidemiology of this phenotype within this organism remains unclear. A link between the *rpo*E

gene and MLST may exist which means that the *rpo*E allele could potentially be used to identify novel clones within the wider population, however this is simply speculation at this point and further study is needed.

7.1 Limitations and Future work

This work has assessed the phenotypic diversity, genomic diversity and the proteomic diversity which exists between the biofilm-derived SCV phenotype and the wild type parent. It has delivered novel insight into the level of diversity which arises as a result of biofilm growth and shown evidence of clear phenotype-genotype relationships. Nonetheless limitations do exist within this thesis. Three notable limitations in this work include; 1) the fact that the genomic analysis was limited to biofilm-derived variants of a single serotype, 22F, 2) the relatively small number of isolates sequenced, 3) the fact that proteomic analysis was only achievable on two SCVs within the given timeframe and funding restraints. All three limitations are easily resolvable; firstly, biofilm-derived variants from days 1-9 of serotypes 1, 3 and 14 are currently stored at – 80 ^{o}C and await genomic analysis when funding can be obtained. This task is essential to determine whether the *rpo*E-targeted mutations are a phenomenon unique to serotype 22F or whether this level of diversity and evolution is seen in multiple serotypes. Secondly, duplicate biofilm experiments with triplicate replicates were performed for each serotype, in accordance with previous literature (Hall-Stoodley *et al*., 2008). To increase the sample size additional experiments can be performed as described in this thesis, in addition a more comprehensive selection of variants should be analysed; ideally a minimum of 10 isolates of each variant (SCV, MCV and TCV) from different replicates of each day, and from multiple independent biofilm experiments could be genome sequenced and phenotyped in full to help strengthen these findings. Finally all sequenced SCVs were phenotyped based on colony size, Gram staining, API profile, growth rate, capsule quantification, MLST and serotyping in addition to *rpo*E gene sequence. Proteomics-based phenotyping remains to be performed on 10 of the sequenced SCVs and would be advisable to achieve a more comprehensive understanding of the effect the additional *rpo*E mutations have on the protein expression. This would be achievable with additional funding. Nonetheless this work has shown that biofilm-derived *rpo*E-targeted mutations do occur in multiple biofilms from independent experiments and that the SCV phenotype is linked to slow growth rate, changes in metabolism, reduced capsule formation and increased biofilm formation. Moreover, label free UPLC/MS_E of the rpoE

variants revealed up-regulation of oxidative stress and virulence-related proteins suggesting *rpo*E may be important to colonisation and virulence of the pneumococcus.

The rationale for using a proteomic approach after the genomic sequencing was to assess whether the mutations that arise in biofilm-derived cells may alter protein targets for vaccine design. In hindsight, an alternative approach of analysis would have been to assess the transcriptomic diversity using DNA-microarrays to see what genes may be affected by the *rpo*E mutations prior to analysing the proteome. Validation of transcriptomic profiles could be confirmed via qRT-PCR. Proteins were extracted at late exponential phase for two reasons, firstly *rpo*E expression has been shown to correspond with both early and late exponential phase, secondly late exponential phase cultures delivered the required cell density to extract sufficient protein for mass spectrometry. This work would benefit from quantifying the expression profile of the *rpo*E gene in different environmental conditions to better understand its role clinically. This could be achieved through the use of quantitative real time-PCR; messenger RNA could be extracted at designated time periods and compared to the expression of the pneumococcal housekeeping gene *gyr*A as an internal control. The relative fold change can be calculated from $2^{\Delta\Delta Ct}$, to determine relative expression as described previously (Song *et al*., 2009).

In addition competition studies comparing the wild type and SCV strain in both planktonic and biofilm setting will be useful to assess the relative fitness of the SCV phenotype. Distinction of the SCV from the wild type could be achieved by the introduction of a green fluorescent protein-containing plasmid or other selectable marker. Further phenotypic analysis of the additional sequenced SCVs would be beneficial to confirm the SCV phenotype; attachment studies on *in vitro* cell lines and infectivity studies in mice may help to see if the SCV phenotype has increased colonisation.

One aspect of this project, yet to be investigated, is the level of variation that can occur as a result of growth in multi-serotype and multi-species biofilms. Many chronic infections play host not to one single pathogen but too many, for example *S. pneumoniae* has been implicated as a major component of polymicrobial infections and is often associated with *M. catarrhalis* and *H. influenzae* in otitis media (Hall-Stoodley *et al*., 2006). Moreover multiple serotypes of *S. pneumoniae* from single patients have been observed (Meats *et al*., 2003) in carriage, indicating that serotype-serotype interactions may play a vital role in shaping the biofilm environment. As previously mentioned, penicillin-sensitive pneumococcal strains have shown to be protected from amoxicillin and benzylpenicillin antibiotic when co-cultured with

β-lactamase producing *M. catarrhalis* in a continuous biofilm system (Budhani and Struthers, 1998). Recent work from Cope *et al.* (2011), looking at the chronic rhinosinusitis (CRS) model of infection, suggests that co-culture of *S. pneumoniae* with non-typeable *H. influenzae* (NTHi) may actually be synergistic as the NTHi type IV pili, encoded by a gene essential for NTHi biofilm formation, was only expressed in co-culture and that the *S. pneumoniae* pyruvate oxidase gene was also up-regulated in co-culture. In contrast the inhibition of *H. influenzae* by hydrogen peroxide produced by the *S. pneumoniae spx*B gene (Pericone *et al*., 2000) is evidence for the concept that *S. pneumoniae* acts as a protector against *H. influenzae* and *S. aureus* during nasopharyngeal carriage. These data suggest that mechanisms involved in co-culture biofilms is complex and that further work is needed to better understand the interactions involved in polymicrobial biofilms. Initial work into multi-serotype biofilms of *S. pneumoniae* suggested potential antagonistic interaction between the various pneumococci (data not shown), however further repeats are needed before any conclusion to be drawn.

This project also remains to investigate whether growth under biofilm conditions can alter the antigenic profile of the streptococci. *S. pneumoniae* is a highly diverse organism, comprising of over 90 different serotypes. These serotypes have arisen via numerous mutational events and it has been shown that even a single point mutation can alter the serotype of an organism (Sheppard *et al*., 2010). Genetic diversity acquired during the biofilm mode of growth may change immunogenic targets on the bacteria's surface making vaccines ineffective. Such mutations were identified Sheppard *et al.* in 2010; the authors identified four *S. pneumoniae* isolates capable of expressing serotypes 6A and 6B. This phenomenon was due to a point mutation in WciP (Sheppard *et al*., 2010). Although this paper made no reference to the role of biofilms in generating this distinctive mutation, bacteria grown under biofilm conditions have shown to exhibit an increase in mutation frequency (Driffield *et al*., 2008) making point mutations such as that seen in Sheppard *et al* (2010) highly plausible. Thus biofilms may provide a stable environment for serotype switching (Spratt and Greenwood, 2000). Research has shown that capsule production is down-regulated in pneumococcal biofilm development (Moscoso *et al*., 2006, Hall-Stoodley *et al*., 2008) suggesting it does not have a vital role in biofilm development. The unnecessary need for the capsule's expression could lead to the hypothesis that mutations within capsule genes could occur without detrimental effects during the biofilm mode of growth and that upon release from the biofilm, when capsule expression is up-regulated, these mutations may result in a change in the capsule phenotype.

This work aims to inform vaccine design by highlighting the relevance of biofilms in generating diversity which may alter the antigenic property of the pneumococcus and thus result in vaccine escape variants, such as those already reported (Golubchik *et al*., 2012). Pneumococcal vaccine design still relies on epidemiological data to track the dynamic change in serotype distribution and highlight serotype most prone to cause invasive disease. The relevance of this research, in contrast to previous research, is that this work focuses on genetic diversity in the context of biofilms. At present pneumococcal vaccines are developed to prevent invasive diseases, such as meningitis and pneumonia, this work will contribute to the development of a vaccine for chronic infections such as pneumococcal carriage and otitis media and in doing so may prevent invasive disease by reducing the initial carriage of the pneumococcus. In order to understand the nature and extent of the genetic change that can occur within biofilms of *S. pneumoniae* and assess the implications of this variation for the ability to vaccinate against this pathogen; *in vivo* infection assay should be performed.

Future work could involve the use of a murine model of infection to see if the *rpo*E mutations and the increased ability to form biofilms seen in the SCVs will increase the chance of infectivity. Previous work has shown that cells harvested from biofilms are attenuated for invasive disease but not for nasopharyngeal colonization (Sanchez *et al*., 2011b). Based on data in this thesis we would hypothesise a similar result for the *rpo*E SCV mutants, however as the isolates used in Sanchez *et al,* (2011b) were not sequenced; testing the role of *rpo*E in infectivity remains an exciting prospect.

7.2 **Conclusion**

This project aimed to characterise the biofilm forming properties of four clinically relevant strains of *S. pneumoniae* and furthermore assess the level of genetic and phenotypic variation that arises as a direct result of biofilm formation using colony morphology variation as a marker for genetic variation. This has been achieved and it can be confidently stated that diversification within pneumococcal biofilms is driven, in part, by parallel evolution which results in sub-populations of colonies which are morphologically, phenotypically and genetically distinct from the parent strain. This work supports, whilst simultaneously extends, previous literature regarding biofilm-derived variation and has set the foundations for future work; in particular examining the role of RpoE in directing genetic and phenotypic variation in pneumococci biofilms which, in turn, might provide a valuable target for future pneumococcal vaccination or drug development.

8 Appendices

8.1 Appendix 1: API Rapid ID 32 Strep reference information

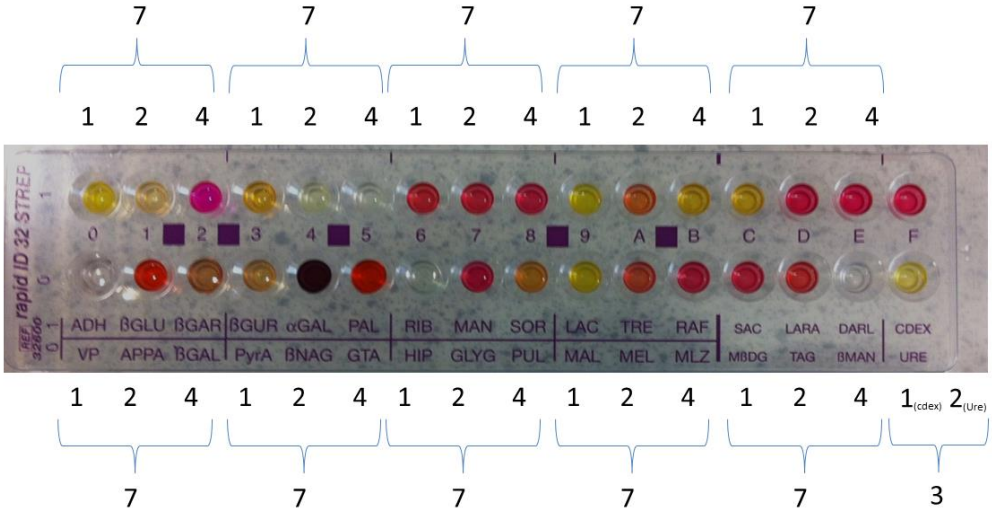


Figure 8-1: API ID 32 Strep scoring system

The chemical reactions in the API strip are divided into blocks of three with each well designated a score of 1, 2, or 4 respectively. A positive reaction for all three reactions in a block would score a maximum of 7 (1 +2 +4 = 7). Negative reactions score a zero. Each block is scored to generate an 11-digit code. This code is then entered into the online database, *APIweb* (https://apiweb.biomerieux.com) and the code identifies the organism.

Table 8-1: The 32 enzymatic tests contained within the API Rapid ID 32 Strep assay

Test	Reaction/enzyme
ADH	Arginine Dihydrolase
BGLU	β Glucosidase
BGAR	β Galactosidase
BGUR	β Giucuronidase
aGAL	α Galactosidase
PAL	Alkaline Phosphatase
RIB	Ribose
MAN	Manitol
SOR	Sorbitol
LAC	Lactose
TRE	Trehalose
RAF	Raffinose
SAC	Saccharose
LARA	L-arabinose
DARL	D-Arabitol
CDEX	Cyclodextrin
VP	Voges Proskauer
APPA	Alanyl-Phenylalanyl-Proline Arylamidase
BGAL	β Galactosidase
PyrA	Pyroglutamic acid Arylamidase
BNAG	N-Acetyl-B-Glucosaminidase
GTA	Glycyl-Tryptophan Arylamidase
HIP	Hydrolysis of Hippurate
GLYG	Glycogen
PUL	Pullulan
MAL	Maltose
MEL	Melibiose
MLZ	Melezitose
MBDG	Methyl- βD Gluco-pyranoside
TAG	Tagatose
BMAN	β Mannosidase
URE	Urease

8.2 Appendix 2: Whole genome sequence assembly metrics

Table 8-2: Assembly Metrics of WGS sequences

	WT	SCV1D3E1	SCV5D3E1	SCV3D9E1	SCV7D9E1	SCV9D9E1
Number of contigs	147	446	177	642	156	527
Total size of contigs	2076001	2054277	2067588	2051357	2062388	2067031
Longest contig	122005	57712	159272	49030	155002	56956
N50 contig length	70963	10317	48039	7604	44794	11160

	SCV4D6E2	SCV2D6E2	SCV6D6E2	SCV5D9E2	SCV6D9E2	SCV7D9E2
Number of contigs	126	152	116	130	151	141
Total size of contigs	2066668	2058841	2068240	2068728	2068292	2066742
Longest contig	261439	102900	253995	158746	118385	167883
N50 contig length	60856	53520	61264	64214	71697	53591

	TCV5	TCV6	TCV7	TCV9	TCV10
Number of contigs	711	809	790	343	426
Total size of contigs	2062443	2058673	2054444	2067867	2066182
Longest contig	35837	30501	43228	77301	68258
N50 contig length	7732	6910	8132	15051	11825

Average	WT	SCV	TCV
Number of contigs	147	243.08333	615.8
Total size of contigs	2076001	2063301.1	2061921.8
Longest contig	122005	139540	51025
N50 contig length	70963	44357.167	9930

8.3 Appendix 3: *in silico* PCR primers for pneumococcal serotyping

Table 8-3: *in silico* PCR primers for pneumococcal serotyping

Serogroup/type	primers
srt1	srt1 CTCTATAGAATGGAGTATATAAACTATGGTTA CCAAAGAAAATACTAACATTATCACAATATTGGC 400 200
srt2	srt2 TATCCCAGTTCAATATTTCTCCACTACACC ACACAAAATATAGGCAGAGAGAGACTACT 400 200
srt3	srt3 ATGGTGTGATTTCTCCTAGATTGGAAAGTAG CTTCTCCAATTGCTTACCAAGTGCAATAACG 500 300
srt4	srt4 CTGTTACTTGTTCTGGACTCTCGATAATTGG GCCCACTCCTGTTAAAATCCTACCCGCATTG 600 300
srt5	srt5 ATACCTACACAACTTCTGATTATGCCTTTGTG GCTCGATAAACATAATCAATATTTGAAAAAGTATG 500 300
srg6A_6B_6C_6D	srg6A_6B_6C_6D AATTTGTATTTTATTCATGCCTATATCTGG TTAGCGGAGATAATTTAAAATGATGACTA 350 200
srg6C_6D	srg6C_6D CATTTTAGTGAAGTTGGCGGTGGAGTT AGCTTCGAAGCCCATACTCTTCAATTA 900 600
srts7C_7B_40	srts7C_7B_40 CTATCTCAGTCATCTATTGTTAAAGTTTACGACGGGA GAACATAGATGTTGAGACATCTTTTGTAATTTC 400 200
srg7F_7A	srg7F_7A TCCAAACTATTACAGTGGGAATTACGG ATAGGAATTGAGATTGCCAAAGCGAC 700 500
srt8	srt8 GAAGAAACGAAACTGTCAGAGCATTTACAT CTATAGATACTAGTAGAGCTGTTCTAGTCT 300 100
srg9N_9L	srg9N_9L GAACTGAATAAGTCAGATTTAATCAGC ACCAAGATCTGACGGGCTAATCAAT 600 400
srg9V_9A	srg9V_9A GGGTTCAAAGTCAGACAGTGAATCTTAA CCATGAATGAAATCAACATTGTCAGTAGC 900 700
srt10A	srt10A GGTGTAGATTTACCATTAGTGTCGGCAGAC GAATTTCTTCTTTAAGATTCGGATATTTCTC 700 500
srts10F_10C_33C	srts10F_10C_33C GGAGTTTATCGGTAGTGCTCATTTTAGCA CTAACAAATTCGCAACACGAGGCAACA 350 200
srg11A_11D	srg11A_11D GGACATGTTCAGGTGATTTCCCAATATAGTG GATTATGAGTGTAATTTATTCCAACTTCTCCC 600 400
srts12F_12A_44_46	srts12F_12A_44_46 GCAACAAACGGCGTGAAAGTAGTTG CAAGATGAATATCACTACCAATAACAAAAC 500 300
srt13	srt13 TACTAAGGTAATCTCTGGAAATCGAAAGG CTCATGCATTTTATTAACCGCTTTTTGTTC 800 600
srt14	srt14 GAAATGTTACTTGGCGCAGGTGTCAGAATT GCCAATACTTCTTAGTCTCTCAGATGAAT 300 100
srg15A_15F	srg15A_15F ATTAGTACAGCTGCTGGAATATCTCTTC GATCTAGTGAACGTACTATTCCAAAC 600 400
srg15B_15C	srg15B_15C TTGGAATTTTTTAATTAGTGGCTTACCTA CATCCGCTTATTAATTGAAGTAATCTGAACC 700 400
srt16F	srt16F GAATTTTTCAGGCGTGGGTGTTAAAAG CAGCATATAGCACCGCTAAGCAAATA 800 600
srt17F	srt17F TTCGTGATGATAATTCCAATGATCAAACAAGAG GATGTAACAAATTTGTAGCGACTAAGGTCTGC 800 600
srg18A_18B_18C_18F	srg18A_18B_18C_18F CTTAATAGCTCTCATTATTCTTTTTTTAAGCC TTATCTGTAAACCATATCAGCATCTGAAAC 700 500
srt19A	srt19A GAGAGATTCATAATCTTGCACTTAGCCA CATAATAGCTACAAATGACTCATCGCC 700 500
srt19F	srt19F GTTAAGATTGCTGATCGATTAATTGATATCC GTAATATGTCTTTAGGGCGTTTATGGCGATAG 400 250
srt20	srt20 GAGCAAGAGTTTTTCACCTGACAGCGAGAAG CTAAATTCCTGTAATTTAGCTAAAACTCTTATC 600 450
srt21	srt21 CTATGGTTATTTCAACTCAATCGTCACC GGCAAACTCAGACATAGTATAGCATAG 300 100
srg22F_22A	srg22F_22A GAGTATAGCCAGATTATGGCAGTTTTATTGTC CTCCAGCACTTGCGCTGGAAACAACAGACAAC 800 600
srt23A	srt23A TATTCTAGCAAGTGACGAAGATGCG

	CCAACATGCTTAAAAACGCTGCTTTAC 850 650
srt23B	srt23B CCACAATTAGCGCTATATTCATTCAATCG GTCCACGCTGAATAAAATGAAGCTCCG 300 100
srt23F	srt23F GTAACAGTTGCTGTAGAGGGAATTGGCTTTTC CACAACACCTAACACTCGATGGCTATATGATTC 500 300
srg24A_24B_24F	srg24A_24B_24F GCTCCCTGCTATTGTAATCTTTAAAGAG GTGTCTTTTATTGACTTTATCATAGGTCGG 200 50
srt31	srt31 GGAAGTTTTCAAGGATATGATAGTGGTGGTGC CCGAATAATATATTCAATATATTCCTACTC 800 600
srts33F_33A_37	srts33F_33A_37 GAAGGCAATCAATGTGATTGTGTCGCG CTTCAAAATGAAGATTATAGTACCCTTCTAC 500 300
srt34	srt34 GCTTTTGTAAGAGGAGATTATTTTCACCCAAC CAATCCGACTAAGTCTTCAGTAAAAAACTTTAC 500 300
srts35A_35C_42	srts35A_35C_42 ATTACGACTCCTTATGTGACGCGCATA CCAATCCCAAGATATATGCAACTAGGTT 400 200
srt35B	srt35B GATAAGTCTGTTGTGGAGACTTAAAAAGAATG CTTTCCAGATAATTACAGGTATTCCTGAAGCAAG 800 600
srt35F_47F	srt35F_47F GAACATAGTCGCTATTGTATTTTATTTAAAGCAA GACTAGGAGCATTATTCCTAGAGCGAGTAAACC 700 500
srts38_25F_25A	srts38_25F_25A CGTTCTTTTATCTCACTGTATAGTATCTTTATG ATGTTTGAATTAAAGCTAACGTAACAATCC 700 500
srt39	srt39 TCATTGTATTAACCCTATGCTTTATTGGTG GAGTATCTCCATTGTATTGAAATCTACCAA 200 50
***cps*A**	cpsA GCAGTACAGCAGTTTGTTGGACTGACC GAATATTTTCATTATCAGTCCCAGTC 300 100

8.4 Appendix 4: Phenotypic profiling for serotype 22F variants from second biofilm experiment

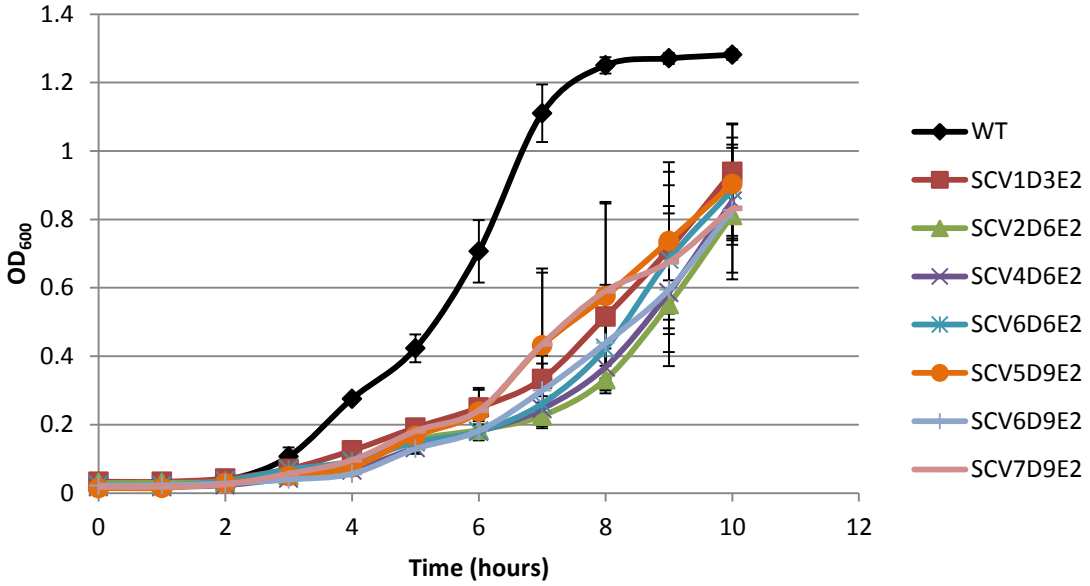


Figure 8-2: Optical density growth curve of biofilm-derived small colony variants from the second experiment

The growth profiles were assessed for all variants chosen to undergo whole genome sequencing. Variants were grown in brain heart infusion media for a total of 10 hours. OD_{600} readings were taken every hour for 10 hours. Error bars represent standard error.

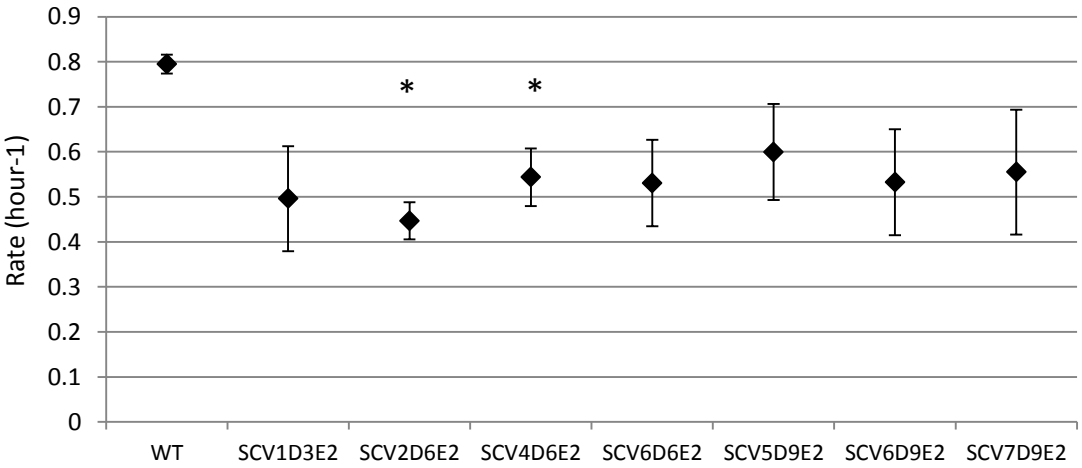


Figure 8-3: Growth rate of biofilm-derived small colony variants from the second experiment

The growth rate was assessed for all SCVs using the equation $\mu = ((\log10\ N - \log10\ N_0)\ 2.303) / (t - t_0)$ and CFU data from the exponential growth phase (between 4 and 8 hours of growth). A 2-sample t-test was used to compare each the SCVs to the WT. Asterisks denote SCVs with significant different growth rates compared to the WT (where $p = <0.05$). Error bars represent standard error.

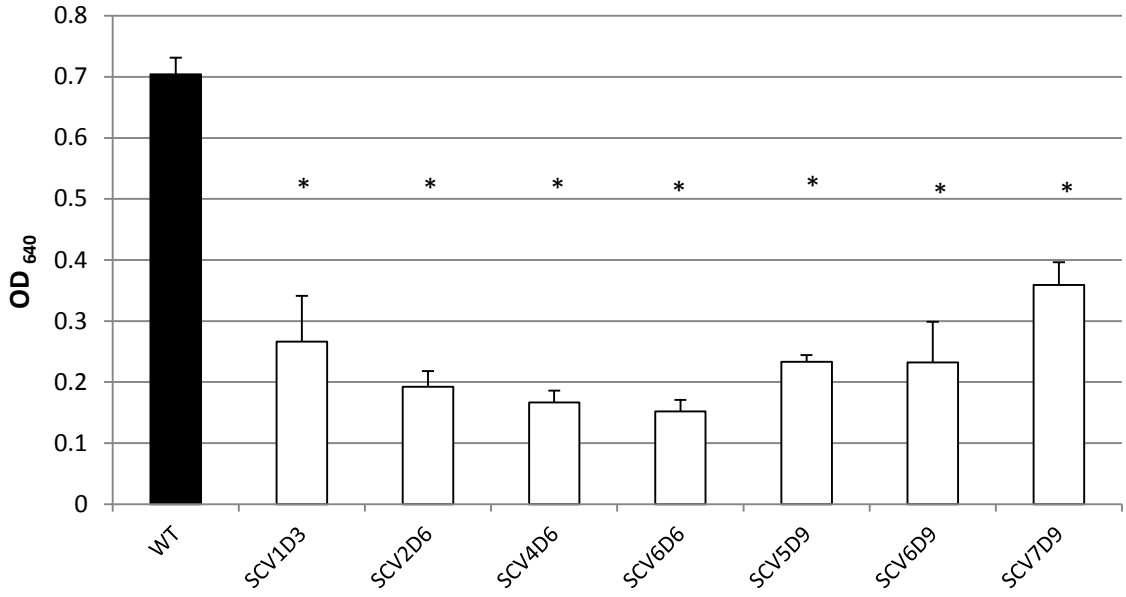


Figure 8-4: Capsule quantification of biofilm-derived small colony variants from the second experiment

Biofilm-derived pneumococcal variants were cultured in BHI to an optical density of approximately 10^8 CFU mL^{-1} and pelleted. Cells were re-suspended and stained with Stain-all solution to quantify the capsule. The absorbance was measured on a spectrophotometer at 640 nm and subtracted from the negative control (250 µL of sterile distilled water stained with 1 mL of Stains-all). A 2-sample t-test was used to compare each SCV to the WT. Asterisks denote significantly different values from the WT (where p = <0.05). Error bars represent standard error.

Table 8-4: API Rapid ID 32 Strep assay profile for the biofilm-derived small colony variants from second experiment [Ψ]

	ADH	BGLU	BGAR	BGUR	aGAL	PAL	RIB	MAN	SOR	LAC	TRE	RAF	SAC	LARA	DARL	CDEX	VP	APPA	BGAL	PyrA	BNAG	GTA	HIP	GLYG	PUL	MAL	MEL	MLZ	MBDG	TAG	BMAN	URE
SCV1 D3E2	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	-	+	-	-	-	-	-	-
SCV2 D6E2	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	-	+	-	-	-	-	-	-
SCV4 D6E2	-	-	+	-	+	-	-	-	-	+	+	+	-	-	-	-	-	+	-	-	+	+	-	-	-	-	-	-	-	-	-	-
SCV6 D6E2	-	-	+	-	+	-	-	-	-	+	+	+	-	-	-	-	-	+	-	-	+	+	-	-	-	-	-	-	-	-	-	-
SCV5 D9E2	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	-
SCV6 D9E2	-	-	+	-	+	-	-	-	-	+	-	+	-	-	-	-	-	+	-	-	+	+	-	-	-	-	-	-	-	-	-	-
SCV7 D9E2	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	-	+	-	-	-	-	-	-
WT	-	-	+	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	-	-

[Ψ] Positive symbols indicate active metabolism, negative symbols indicate no metabolism. Results were recorded visually using the table of identification in the manufacturer's protocol booklet. Results which differ from the WT profile are coloured red.

Table 8-5: API*web* identification of the biofilm-derived variants from second experiment using the Rapid ID 32 Strep assay tests

Isolate	*APIweb* percentage similarity to *S. pneumonie* (%)[Ψ]	*APIweb* percentage similarity to *S. oralis* (%)	*APIweb* percentage similarity to *S. parasanguinis* (%)
WT 22F	99.9	0.1	-
SCV1D3E2	99.8	-	0.1
SCV2D6E2	99.8	-	0.1
SCV4D6E2	99.9	0.1	-
SCV6D6E2	99.9	0.1	-
SCV5D9E2	99.1	0.7	-
SCV6D9E2	98.5	-	0.4
SCV7D9E2	99.8	-	0.1

[Ψ] Where 99.9 % similarity was not achieved alternative species similarity was recorded. Those isolates highlighted in bold were selected for whole genome sequencing.

8.5 Appendix 5: *rpo*E alleles from 5-year pneumococcal carriage study

Table 8-6: All *rpo*E alleles from 5-year pneumococcal carriage study

Allele ID	Sequence	bp	Isolates
***rpo*E1**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATCTTGGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTTTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGACGC TGAAACCAGCGAGGAATAA	588	246
***rpo*E2**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATCTTGGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTTTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGATGC TGAAACCAGCGAGGAATAA	588	77
***rpo*E3**	TTGAAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATCTTGGAACTTCGTG GTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCGA AGCTTTGCCTTTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCAT GGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAACACCAAAAGCTAAG AAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGCA TACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGAA ATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGACGACGACGCT GAAACCAGCGAGGAATAA	588	50
***rpo*E4 (WT 22F)**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATATTAGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGACGC TGAAACCAGCGAGGAATAA	588	35
***rpo*E5**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATCTTGGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTTTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGACGC TGAAACCAGCGAGGAATAA	588	26
***rpo*E6**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATCTTAGAACTTCGTG GTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCGA AGCTTTGCCTTTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCAT GGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAAG AAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGCA TACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGAA ATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGATGCT GAAACCAGCGAGGAATAA	588	25
***rpo*E7**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATATTAGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTTTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGACGC TGAAACCAGCGAGGAATAA	588	25
***rpo*E8**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATCTTGGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTTTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGATGC TGAAACCATCGAGGAATAA	588	8
***rpo*E9**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATCTTGGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTTTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTTAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGACGC TGAAACCAGCGAGGAATAA	588	7

***rpo*E10**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATATTAGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGATAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGACGC TGAAACCAGCGAGGAATAA	588	3
***rpo*E11**	TTGAAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATCTTGGAACTTCGTG GTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCGA AGCTTTGCCTTTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCAT GGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAAG AAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGCA TACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGAA ATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGACGCT GAAACCAGCGAGGAATAA	588	3
***rpo*E12**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATATTAGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTATAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGACGC TGAAACCAGCGAGGAATAA	588	3
***rpo*E13**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATATTAGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAGGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGACGC TGAAACCAGCGAGGAATAA	588	1
***rpo*E14**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATATTAGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAGCGAAGACGACGACGAGTTTTCAGATGACGACG CTGAAACCAGCGAGGAATAA	588	1
***rpo*E15**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATCTTGGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTTTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAGTGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGATGC TGAAACCAGCGAGGAATAA	588	1
***rpo*E16**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATCTTAGAACCTCGT GGTCGCGATTATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTTTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGATGC TGAAACCAGCGAGGAATAA	588	1
***rpo*E17**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATCTTAGAACCTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTTTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTTAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGACGC TGAAACCAGCGAGGAATAA	588	1
***rpo*E18**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATCTTAGAACTTCGTG GTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCGA AGCTTTGCCTTTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCAT GGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAAG AAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGCA TACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGAA ATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGACGCT GAAACCAGCGAGGAATAA	588	1
***rpo*E19**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATCTTGGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTTTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGATGACGC TGAAACCAGCGAGGAATAA	588	1

***rpo*E20**	TTGAAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATCTTGGAACTTCGTG GTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCGA AGCTTTGCCTTTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCAT GGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAACACCAAAAGCTAAG AAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGCA TACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGAA ATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGACGCT GAAACCAGCGAGGAATAA	588	1
***rpo*E21**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATCTTGGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTTTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGACGCTGA AACCAGCGAGGAATAA	585	1
***rpo*E22**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATCTTGGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTTTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTTTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGACGC TGAAACCAGCGAGGAATAA	588	1
***rpo*E23 SCV1D3**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATATTAGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGATAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGACGC TGAAACCAGCGAGGAATAA	588	0
***rpo*E24 SCV3D9**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATATTAGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTAATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGACGC TGAAACCAGCGAGGAATAA	588	0
***rpo*E25 SCV5D3**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATATTAGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCATA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGACGC TGAAACCAGCGAGGAATAA	588	0
***rpo*E26 SCV7D9**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATATTAGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCATA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGACGC TGAAACCAGCGAGGAATAA	588	0
***rpo*E27 SCV9D9**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATATTAGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGCTGAAACCAGCGAGGAATAA	324	0
***rpo*E28 SCV1D3 E2**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATATTAGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGATGCATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGAT GAAAAAAATGAAGTGGAAGCTTATGATGCAGAAATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAA CGAAGACGACGACGAGTTTTCAGATGACGACGCTGAAACCAGCGAGGAATAA	459	0
***rpo*E29 SCV2D6 E2**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATATTAGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGCTGAAACCAGCGAGGAATAA	324	0
***rpo*E30 SCV4D6 E2**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATATTAGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATTAAAAAAATGAAGTGGAAGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGACGC TGAAACCAGCGAGGAATAA	588	0
***rpo*E31 SCV6D6 E2**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATATTAGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA	589	0

	TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGAC GCTGAAACCAGCGAGGAATAA		
***rpo*E32 SCV5D9 E2**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATATTAGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACTAAGACGACGACGAGTTTTCAGATGACGACGC TGAAACCAGCGAGGAATAA	588	0
***rpo*E33 SCV6D9 E2**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATATTAGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGTAAGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGACGC TGAAACCAGCGAGGAATAA	588	0
***rpo*E34 SCV7D9 E2**	TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTGAGGTAGCGCGTGCTATATTAGAACTTCGT GGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGATATCCGCG AAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTTGACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCA TGGTATGGTGTGGACGAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACCAAAAGCTAA GAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGC ATACGAAGCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAAGCTTATGATGCAGA AATCAACGAAATCGCTCCAGATGACTTGGGAGAAGATGTGGATCTCAAC	537	0

8.6 Appendix 6: UPLC/MS_E mass spectrometry protein accession numbers and gene ontology

Table 8-7: Protein accession number and description of up-regulated proteins in biofilm-derived SCV5D3

Protein Accession Number	Protein Description	Gene name	Gene Ontology terms (Biological process)	Fold increase (x)
Q9F2F0	Peptide deformylase	*def*	Protein biosynthesis, hydrolase, translation	3.27
Q9X9S0	Probable dihydroorotate dehydrogenase A (fumarate)	*pyr*DA	Pyrimidine biosynthesis, UMP biosynthetic process, pyrimidine nucleobase biosynthetic process	2.98
Q97SF6	Enoyl-CoA hydratase/isomerase family protein	SP_0415	Isomerase	2.84
P0A4D7	Probable ATP-dependent RNA helicase exp9	*exp*9	ATP-dependent helicase activity, RNA binding	2.60
Q97PR4	Cof family protein/peptidyl-prolyl cis-trans isomerase_ cyclophilin type	SP_1538	Isomerase, protein folding	2.58
Q97TC4	Hypoxanthine-guanine phosphoribosyltransferase	*hpt*	Purine salvage	2.39
Q97R46	Bifunctional protein GlmU	*glm*U	UDP-N-acetylglucosamine biosynthetic process, lipid A/lipopolysaccharide/peptidoglycan biosynthesis	2.39
Q97SE1	Putative uncharacterized protein	SP_0443	Glycerol metabolic process	2.37
P0A4J6	Superoxide dismutase [Mn]	*sod*A	Superoxide metabolic process, oxidative stress response	2.27
P0A451	Protein RecA	*rec*A	DNA recombination, SOS response, DNA repair	2.17
Q97PK2	UDP-glucose 4-epimerase	SP_1607	Isomerase, galactose metabolic process	2.00
I6L8V8	3-oxoacyl-[acyl-carrier protein] reductase	*fab*G	Fatty acid biosynthetic process	1.96
P58313	UTP--glucose-1-phosphate uridylyltransferase	*cap*4C	UDP-glucose metabolic process, capsule polysaccharide biosynthetic process	1.89
Q97R51	Foldase protein PrsA	*prs*A	Protein folding and secretion	1.85
Q97NQ0	Aspartate--ammonia ligase	*asn*A	L-asparagine biosynthetic process, tRNA aminoacylation for protein translation	1.78
P65755	Probable manganese-dependent inorganic	*ppa*C	Catalytic activity	1.77

	pyrophosphatase			
I6L8U5	3-oxoacyl-[acyl-carrier-protein] synthase 2	*fab*F	Fatty acid biosynthetic process	1.76
Q97PF9	Cell division protein FtsZ	*fts*Z	Cell division/cytokinesis by binary fission, protein polymerization	1.72
I6L8Q7	Enoyl-(Acyl-carrier-protein) reductase	*fab*K	Oxidoreductase	1.70
P95829	Chaperone protein DnaK	*dna*K	Stress response, protein folding	1.68
Q97RV5	Elongation factor Tu family protein	SP_0681	Protein biosynthesis	1.66
Q97PT4	ATP synthase subunit alpha	SP_1510	ATP hydrolysis coupled proton transport, ATP synthesis	1.63
P18791	Oligopeptide-binding protein AmiA	*ami*A	Peptide/protein transport	1.60
Q97RQ3	Uracil phosphoribosyltransferase	*upp*	Uracil salvage	1.55
P0A335	60 kDa chaperonin	*gro*L	Protein refolding	1.54

Table 8-8: Protein accession number and description of down-regulated proteins in biofilm-derived SCV5D3

Protein Accession Number	Protein Description	Gene name	Gene Ontology terms (Biological process)	Fold decrease (x)
Q97RW6	Putative uncharacterized protein	SP_0670	Unknown function	-4.07
Q97SU4	50S ribosomal protein L30	*rpm*D	Translation	-3.79
P61360	50S ribosomal protein L33 type 3	*rpm*G3	Translation	-3.72
P66524	30S ribosomal protein S21	*rps*U	Translation	-3.69
Q97T34	UPF0356 protein SP_0122	SP_0122	Protein binding	-3.25
P0A483	50S ribosomal protein L29	*rpm*C	Translation	-3.09
P66135	50S ribosomal protein L27	*rpm*A	Translation	-3.01
P0A4B3	30S ribosomal protein S17	*rps*Q	Translation	-2.78
P66200	50S ribosomal protein L31 type B	*rpm*E2	Translation	-2.66
Q97PI9	30S ribosomal protein S15	*rps*O	Translation	-2.60
Q97S86	Glutamine synthetase type I	*gln*A	glutamine biosynthetic process, nitrogen fixation	-2.45
Q97PW6	Glycine--tRNA ligase beta subunit	*gly*S	Protein biosynthesis, glycyl-tRNA aminoacylation	-2.30
P60629	50S ribosomal protein L24	*rpl*X	Translation	-2.24
P0A4B5	30S ribosomal protein S19	*rps*S	Translation	-2.20
Q97N37	Serine protease	SP_2239	Protease, Serine-type endopeptidase activity	-2.17
Q97NB7	50S ribosomal protein L32	*rpm*F	Translation	-2.17
Q97SP1	Alcohol dehydrogenase_ zinc-containing	SP_0285	Alcohol dehydrogenase (NAD) activity, zinc binding	-2.11
P0A2W0	Acyl carrier protein	*acp*P	Fatty acid biosynthesis/metabolism	-2.06
P66112	50S ribosomal protein L20	*rpl*T	Translation	-2.06
Q97QU4	50S ribosomal protein L21	*rpl*U	Translation	-2.03
Q97SP2	PTS mannose-specific IIAB component	*man*L	Phosphoenolpyruvate-dependent sugar phosphotransferase system	-2.00

Q97SV3	50S ribosomal protein L23	*rplW*	Translation	-1.98
Q97QT6	DNA-binding protein HU	*hup*	Chromosome condensation	-1.96
Q97SC6	Formate acetyltransferase	*pfl*	Acyltransferase, carbohydrate metabolic process	-1.96
Q97NV3	10 kDa chaperonin	*groS*	Protein folding	-1.96
Q97RK7	Aminopeptidase N	*pepN*	Aminopeptidase activity, metallopeptidase activity, zinc binding	-1.95
P66339	30S ribosomal protein S10	*rpsJ*	Translation	-1.87
P0A4S1	Fructose-bisphosphate aldolase	*fba*	Fructose 1,6-bisphosphate metabolic process, glycolytic process	-1.86
Q97SU6	50S ribosomal protein L18	*rplR*	Translation	-1.75
P66392	30S ribosomal protein S13	*rpsM*	Translation	-1.67
P66730	DNA-directed RNA polymerase subunit omega	*rpoZ*	Transcription	-1.65
Q97ST6	50S ribosomal protein L17	*rplQ*	Translation	-1.56
P0A473	50S ribosomal protein L14	*rplN*	Translation	-1.51

Table 8-9: Protein accession number and description of up-regulated proteins in biofilm-derived SCV3D9

Protein Accession Number	Protein Description	Gene name	Gene Ontology terms (Biological process)Gene Ontology terms (Biological process)	Fold increase (x)
Q97SK0	Putative ATP-dependent Clp protease_ ATP-binding subunit	SP_0338	ATP-binding, nucleotide binding, peptidase activity	3.17
P0A4D7	Probable ATP-dependent RNA helicase exp9	*exp*9	ATP-dependent helicase activity, RNA binding	2.28
Q9X9S0	Probable dihydroorotate dehydrogenase A (fumarate)	*pyr*DA	Pyrimidine biosynthesis, UMP biosynthetic process, pyrimidine nucleobase biosynthetic process	2.05
Q97PT4	ATP synthase subunit alpha	SP_1510	ATP hydrolysis coupled proton transport, ATP synthesis	1.92
Q97SE1	Putative uncharacterized protein	SP_0443	Glycerol metabolic process	1.91
I6L8V8	3-oxoacyl-[acyl-carrier protein] reductase	*fab*G	Fatty acid biosynthetic process	1.89
Q97RV5	Elongation factor Tu family protein	SP_0681	Protein biosynthesis	1.85
P63413	Acetate kinase	*ack*A	Acetyl-CoA biosynthetic process, organic acid metabolic process	1.84
Q97R46	Bifunctional protein GlmU	*glm*U	UDP-N-acetylglucosamine biosynthetic process, lipid A/lipopolysaccharide/peptidoglycan biosynthesis	1.84
Q97NQ0	Aspartate--ammonia ligase	*asn*A	L-asparagine biosynthetic process, tRNA aminoacylation for protein translation	1.71
Q97TC4	Hypoxanthine-guanine phosphoribosyltransferase	*hpt*	Purine salvage	1.68
Q97PF9	Cell division protein FtsZ	*fts*Z	Cell division/cytokinesis by binary fission, protein polymerization	1.68
P0A335	60 kDa chaperonin	*gro*L	Protein refolding	1.61
P95829	Chaperone protein DnaK	*dna*K	Stress response, protein folding	1.56
Q97N43	Inosine-5'-monophosphate dehydrogenase	*gua*B	Purine biosynthesis, GMP biosynthetic process	1.52
P0A4J6	Superoxide dismutase [Mn]	*sod*A	Superoxide metabolic process, oxidative stress response	1.52
Q97PR4	Cof family protein/peptidyl-prolyl cis-trans isomerase_ cyclophilin type	SP_1538	Isomerase, protein folding	1.50

Table 8-10: Protein accession number and description of down-regulated proteins in biofilm-derived SCV3D9

Protein Accession Number	Protein Description	Gene name	Gene Ontology terms (Biological process)	Fold decrease (x)
Q97RW6	Putative uncharacterized protein	SP_0670	Unknown function	-4.14
Q97PW6	Glycine--tRNA ligase beta subunit	*gly*S	Protein biosynthesis, glycyl-tRNA aminoacylation	-2.73
Q97P40	Putative general stress protein 24	SP_1804	Unknown function	-2.45
Q97SC6	Formate acetyltransferase	*pfl*	Acyltransferase, carbohydrate metabolic process	-2.27
P0A4S1	Fructose-bisphosphate aldolase	*fba*	Fructose 1,6-bisphosphate metabolic process, glycolytic process	-2.11

9 References

ABRANCHES, J., CHEN, Y. Y. & BURNE, R. A. 2003. Characterization of *Streptococcus mutans* strains deficient in EIIAB Man of the sugar phosphotransferase system. *Appl Environ Microbiol,* 69**,** 4760-9.

ACHBERGER, E. C., HILTON, M. D. & WHITELEY, H. R. 1982. The effect of the delta subunit on the interaction of *Bacillus subtilis* RNA polymerase with bases in a SP82 early gene promoter. *Nucleic Acids Res,* 10**,** 2893-910.

ACHBERGER, E. C. & WHITELEY, H. R. 1981. The Role of the Delta-peptide of the *Bacillus subtilis* RNA-polymerase in promoter selection. *Journal of Biological Chemistry,* 256**,** 7424-7432.

AHL, J., MELANDER, E., ODENHOLT, I., TVETMAN, L., THORNBLAD, T., RIESBECK, K. & RINGBERG, H. 2014. Risk factors for pneumococcal carriage in day care centers: a retrospective study during a 10-year period. *Pediatr Infect Dis J,* 33**,** 536-8.

ALLAN, R. N., SKIPP, P., JEFFERIES, J., CLARKE, S. C., FAUST, S. N., HALL-STOODLEY, L. & WEBB, J. 2014. Pronounced metabolic changes in adaptation to biofilm growth by *Streptococcus pneumoniae*. *PLoS One,* 9**,** e107015.

ALLEGRUCCI, M., HU, F. Z., SHEN, K., HAYES, J., EHRLICH, G. D., POST, J. C. & SAUER, K. 2006. Phenotypic characterization of *Streptococcus pneumoniae* biofilm development. *J Bacteriol,* 188**,** 2325-35.

ALLEGRUCCI, M. & SAUER, K. 2007. Characterization of colony morphology variants isolated from *Streptococcus pneumoniae* biofilms. *Journal of Bacteriology,* 189**,** 2030-2038.

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. *Journal of Bacteriology,* 190**,** 6330-6339.

ALLESEN-HOLM, M., BARKEN, K. B., YANG, L., KLAUSEN, M., WEBB, J. S., KJELLEBERG, S., MOLIN, S., GIVSKOV, M. & TOLKER-NIELSEN, T. 2006. A characterization of DNA release in *Pseudomonas aeruginosa* cultures and biofilms. *Mol Microbiol,* 59**,** 1114-28.

ARAI, J., HOTOMI, M., HOLLINGSHEAD, S. K., UENO, Y., BRILES, D. E. & YAMANAKA, N. 2011. *Streptococcus pneumoniae* isolates from middle ear fluid and nasopharynx of children with acute otitis media exhibit phase variation. *J Clin Microbiol,* 49**,** 1646-9.

ARTIMO, P., JONNALAGEDDA, M., ARNOLD, K., BARATIN, D., CSARDI, G., DE CASTRO, E., DUVAUD, S., FLEGEL, V., FORTIER, A., GASTEIGER, E., GROSDIDIER, A., HERNANDEZ, C., IOANNIDIS, V., KUZNETSOV, D., LIECHTI, R., MORETTI, S., MOSTAGUIR, K., REDASCHI, N., ROSSIER, G., XENARIOS, I. & STOCKINGER, H. 2012. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research,* 40**,** W597-W603.

AVERY, O. T., MACLEOD, C. M. & MCCARTY, M. 1944. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. *Journal of Experimental Medicine,* 79**,** 137-58.

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 Genetics,* 2**,** 262-269.

BESIER, S., ZANDER, J., KAHL, B. C., KRAICZY, P., BRADE, V. & WICHELHAUS, T. A. 2008. The thymidine-dependent small-colony-variant phenotype is associated with hyermutability and antibiotic resistance in clinical *Staphylococcus aureus* isolates. *Antimicrobial Agents and Chemotherapy,* 52**,** 2183-2189.

BETTS, A., KALTZ, O. & HOCHBERG, M. E. 2014. Contrasted coevolutionary dynamics between a bacterial pathogen and its bacteriophages. *Proc Natl Acad Sci U S A*.

BIDOSSI, A., MULAS, L., DECOROSI, F., COLOMBA, L., RICCI, S., POZZI, G., DEUTSCHER, J., VITI, C. & OGGIONI, M. R. 2012. A functional genomics approach to establish the complement of carbohydrate transporters in *Streptococcus pneumoniae*. *PLoS One,* 7**,** e33320.

BOGAERT, D., DE GROOT, R. & HERMANS, P. W. 2004. *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. *Lancet Infect Dis,* 4**,** 144-54.

BOIKOS, C. & QUACH, C. 2013. Risk of invasive pneumococcal disease in children and adults with asthma: a systematic review. *Vaccine,* 31**,** 4820-6.

BOLES, B. R. & SINGH, P. K. 2008. Endogenous oxidative stress produces diversity and adaptability in biofilm communities. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 12503-12508.

BOLES, B. R., THOENDEL, M. & SINGH, P. K. 2004. Self-generated diversity produces "insurance effects" in biofilm communities. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 16630-16635.

BONGAERTS, R. J., HEINZ, H. P., HADDING, U. & ZYSK, G. 2000. Antigenicity, expression, and molecular characterization of surface-located pullulanase of *Streptococcus pneumoniae*. *Infect Immun,* 68**,** 7141-3.

BRADFORD, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem,* 72**,** 248-54.

BRANDA, S. S., VIK, S., FRIEDMAN, L. & KOLTER, R. 2005. Biofilms: the matrix revisited. *Trends Microbiol,* 13**,** 20-6.

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**,** 270-83.

BUDHANI, R. K. & STRUTHERS, J. K. 1997. The use of Sorbarod biofilms to study the antimicrobial susceptibility of a strain of *Streptococcus pneumoniae*. *Journal of Antimicrobial Chemotherapy,* 40**,** 601-602.

BUDHANI, R. K. & STRUTHERS, J. K. 1998. Interaction of *Streptococcus pneumoniae* and *Moraxella catarrhalis*: Investigation of the indirect pathogenic role of beta-lactamase-producing *moraxellae* by use of a continuous-culture biofilm system. *Antimicrobial Agents and Chemotherapy,* 42**,** 2521-2526.

CAMILLI, R., PANTOSTI, A. & BALDASSARRI, L. 2011. Contribution of serotype and genetic background to biofilm formation by *Streptococcus pneumoniae*. *Eur J Clin Microbiol Infect Dis,* 30**,** 97-102.

CASTRESANA, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution,* 17**,** 540-552.

CHAN, E. Y. 2009. Next-generation sequencing methods: impact of sequencing accuracy on SNP discovery. *Methods Mol Biol,* 578**,** 95-111.

CHASTANET, A., PRUDHOMME, M., CLAVERYS, J. P. & MSADEK, T. 2001. Regulation of *Streptococcus pneumoniae clp* genes and their role in competence development and stress survival. *Journal of Bacteriology,* 183**,** 7295-7307.

CHEVENET, F., BRUN, C., BANULS, A. L., JACQ, B. & CHRISTEN, R. 2006. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *Bmc Bioinformatics,* 7.

CHIA, N., WOESE, C. R. & GOLDENFELD, N. 2008. A collective mechanism for phase variation in biofilms. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 14597-14602.

CIOFU, O., RIIS, B., PRESSLER, T., POULSEN, H. E. & HOIBY, N. 2005. Occurrence of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis patients is associated with the oxidative stress caused by chronic lung inflammation. *Antimicrobial Agents and Chemotherapy,* 49**,** 2276-2282.

CIRZ, R. T., GINGLES, N. & ROMESBERG, F. E. 2006. Side effects may include evolution. *Nature Medicine,* 12**,** 890-891.

CLARKE, S. C., JEFFERIES, J. M. C., SMITH, A. J., MCMENAMIN, J., MITCHELL, T. J. & EDWARDS, G. F. S. 2006. Pneumococci causing invasive disease in children prior to the introduction of pneumococcal conjugate vaccine in Scotland. *Journal of Medical Microbiology,* 55**,** 1079-1084.

CLAVERYS, J. P., MARTIN, B. & HAVARSTEIN, L. S. 2007. Competence-induced fratricide in streptococci. *Molecular Microbiology,* 64**,** 1423-1433.

COCHRAN, W. L., SUH, S. J., MCFETERS, G. A. & STEWART, P. S. 2000. Role of RpoS and AlgT in *Pseudomonas aeruginosa* biofilm resistance to hydrogen peroxide and monochloramine. *J Appl Microbiol,* 88**,** 546-53.

COFFEY, T. J., ENRIGHT, M. C., DANIELS, M., MORONA, J. K., MORONA, R., HRYNIEWICZ, W., PATON, J. C. & SPRATT, B. G. 1998. Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of *Streptococcus pneumoniae*. *Molecular Microbiology,* 27**,** 73-83.

COLETTA, A., PINNEY, J. W., SOLIS, D. Y. W., MARSH, J., PETTIFER, S. R. & ATTWOOD, T. K. 2010. Low-complexity regions within protein sequences have position-dependent roles. *Bmc Systems Biology,* 4.

CONIBEAR, T. C., COLLINS, S. L. & WEBB, J. S. 2009. Role of mutation in Pseudomonas aeruginosa biofilm development. *PLoS One,* 4**,** e6289.

COPE, E. K., GOLDSTEIN-DARUECH, N., KOFONOW, J. M., CHRISTENSEN, L., MCDERMOTT, B., MONROY, F., PALMER, J. N., CHIU, A. G., SHIRTLIFF, M. E., COHEN, N. A. & LEID, J. G. 2011. Regulation of virulence gene expression resulting from Streptococcus pneumoniae and nontypeable Haemophilus influenzae interactions in chronic disease. *PLoS One,* 6**,** e28523.

COSTERTON, J. W. 2001. Cystic fibrosis pathogenesis and the role of biofilms in persistent infection. *Trends Microbiol,* 9**,** 50-2.

COSTERTON, J. W., LEWANDOWSKI, Z., CALDWELL, D. E., KORBER, D. R. & LAPPINSCOTT, H. M. 1995. Microbial Biofilms. *Annual Review of Microbiology,* 49**,** 711-745.

COSTERTON, J. W., STEWART, P. S. & GREENBERG, E. P. 1999. Bacterial biofilms: a common cause of persistent infections. *Science,* 284**,** 1318-22.

CRAMTON, S. E., GERKE, C., SCHNELL, N. F., NICHOLS, W. W. & GOTZ, F. 1999. The intercellular adhesion (ica) locus is present in Staphylococcus aureus and is required for biofilm formation. *Infection and Immunity,* 67**,** 5427-5433.

CRAMTON, S. E., ULRICH, M., GOTZ, F. & DORING, G. 2001. Anaerobic conditions induce expression of polysaccharide intercellular adhesin in Staphylococcus aureus and Staphylococcus epidermidis. *Infection and Immunity,* 69**,** 4079-4085.

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.

CRUICKSHANK, H. C., JEFFERIES, J. M. & CLARKE, S. C. 2014. Lifestyle risk factors for invasive pneumococcal disease: a systematic review. *BMJ Open,* 4**,** e005224.

CUCARELLA, C., SOLANO, C., VALLE, J., AMORENA, B., LASA, I. & PENADES, J. R. 2001. Bap, a *Staphylococcus aureus* surface protein involved in biofilm formation. *Journal of Bacteriology,* 183**,** 2888-2896.

CUNDELL, D. R., WEISER, J. N., SHEN, J., YOUNG, A. & TUOMANEN, E. I. 1995. Relationship between Colonial Morphology and Adherence of *Streptococcus pneumoniae*. *Infection and Immunity,* 63**,** 757-761.

DARLING, A. E., MAU, B. & PERNA, N. T. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One,* 5**,** e11147.

DASGUPTA, T., DE KIEVIT, T. R., MASOUD, H., ALTMAN, E., RICHARDS, J. C., SADOVSKAYA, I., SPEERT, D. P. & LAM, J. S. 1994. Characterization of lipopolysaccharide-deficient mutants of *Pseudomonas aeruginosa* derived from serotypes O3, O5, and O6. *Infect Immun,* 62**,** 809-17.

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. *Journal of Immunology,* 173**,** 471-477.

DEBEER, D., STOODLEY, P. & LEWANDOWSKI, Z. 1996. Liquid flow and mass transport in heterogeneous biofilms. *Water Research,* 30**,** 2761-2765.

DEBEER, D., STOODLEY, P., ROE, F. & LEWANDOWSKI, Z. 1994. Effects of Biofilm Structures on Oxygen Distribution and Mass-Transport. *Biotechnology and Bioengineering,* 43**,** 1131-1138.

DEREEPER, A., GUIGNON, V., BLANC, G., AUDIC, S., BUFFET, S., CHEVENET, F., DUFAYARD, J. F., GUINDON, S., LEFORT, V., LESCOT, M., CLAVERIE, J. M. & GASCUEL, O. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res,* 36**,** W465-9.

DERETIC, V., SCHURR, M. J., BOUCHER, J. C. & MARTIN, D. W. 1994. Conversion of *Pseudomonas aeruginosa* to mucoidy in cystic fibrosis: environmental stress and regulation of bacterial virulence by alternative sigma factors. *J Bacteriol,* 176**,** 2773-80.

DEZIEL, E., COMEAU, Y. & VILLEMUR, R. 2001. Initiation of biofilm formation by *Pseudomonas aeruginosa* 57RP correlates with emergence of hyperpiliated and highly adherent phenotypic variants deficient in swimming, swarming, and twitching motilities. *J Bacteriol,* 183**,** 1195-204.

DI GUILMI, A. M. & DESSEN, A. 2002. New approaches towards the identification of antibiotic and vaccine targets in *Streptococcus pneumoniae*. *EMBO Rep,* 3**,** 728-34.

DOMENECH, M., ARAUJO, L., GARCIA, E. & MOSCOSO, M. 2013a. In vitro biofilm formation by *Streptococcus pneumoniae* as a predictor of post-vaccination emerging serotypes colonizing the human nasopharynx. *Environ Microbiol*.

DOMENECH, M., GARCIA, E. & MOSCOSO, M. 2009. Versatility of the capsular genes during biofilm formation by *Streptococcus pneumoniae*. *Environmental Microbiology,* 11**,** 2542-2555.

DOMENECH, M., GARCIA, E. & MOSCOSO, M. 2012. Biofilm formation in *Streptococcus pneumoniae*. *Microb Biotechnol,* 5**,** 455-65.

DOMENECH, M., RAMOS-SEVILLANO, E., GARICA, E., MOSCOSO, M. & YUSTE, J. 2013b. Biofilm Formation Avoids Complement Immunity and Phagocytosis of *Streptococcus pneumoniae*. *Infection and Immunity,* 81**,** 2606-2615.

DONLAN, R. M. & COSTERTON, J. W. 2002. Biofilms: Survival mechanisms of clinically relevant microorganisms. *Clinical Microbiology Reviews,* 15**,** 167-+.

DONLAN, R. M., PIEDE, J. A., HEYES, C. D., SANII, L., MURGA, R., EDMONDS, P., EL-SAYED, I. & EL-SAYED, M. A. 2004. Model system for growing and quantifying *Streptococcus pneumoniae* biofilms in situ and in real time. *Applied and Environmental Microbiology,* 70**,** 4980-4988.

DRAGO, L., DE VECCHI, E., TORRETTA, S., MATTINA, R., MARCHISIO, P. & PIGNATARO, L. 2012. Biofilm formation by bacteria isolated from upper respiratory tract before and after adenotonsillectomy. *APMIS,* 120**,** 410-6.

DRENKARD, E. & AUSUBEL, F. M. 2002. *Pseudomonas* biofilm formation and antibiotic resistance are linked to phenotypic variation. *Nature,* 416**,** 740-743.

DRIFFIELD, K., MILLER, K., BOSTOCK, J. M., O'NEILL, A. J. & CHOPRA, I. 2008. Increased mutability of *Pseudomonas aeruginosa* in biofilms. *Journal of Antimicrobial Chemotherapy,* 61**,** 1053-1056.

EARL, D., BRADNAM, K., ST JOHN, J., DARLING, A., LIN, D., FASS, J., YU, H. O., BUFFALO, V., ZERBINO, D. R., DIEKHANS, M., NGUYEN, N., ARIYARATNE, P. N., SUNG, W. K., NING, Z., HAIMEL, M., SIMPSON, J. T., FONSECA, N. A., BIROL, I., DOCKING, T. R., HO, I. Y., ROKHSAR, D. S., CHIKHI, R., LAVENIER, D., CHAPUIS, G., NAQUIN, D., MAILLET, N., SCHATZ, M. C., KELLEY, D. R., PHILLIPPY, A. M., KOREN, S., YANG, S. P., WU, W., CHOU, W. C., SRIVASTAVA, A., SHAW, T. I., RUBY, J. G., SKEWES-COX, P., BETEGON, M., DIMON, M. T., SOLOVYEV, V., SELEDTSOV, I., KOSAREV, P., VOROBYEV, D., RAMIREZ-GONZALEZ, R., LEGGETT, R., MACLEAN, D., XIA, F., LUO, R., LI, Z., XIE, Y., LIU, B., GNERRE, S., MACCALLUM, I., PRZYBYLSKI, D., RIBEIRO, F. J., YIN, S., SHARPE, T., HALL, G., KERSEY, P. J., DURBIN, R., JACKMAN, S. D., CHAPMAN, J. A., HUANG, X., DERISI, J. L., CACCAMO, M., LI, Y., JAFFE, D. B., GREEN, R. E., HAUSSLER, D., KORF, I. & PATEN, B. 2011. Assemblathon 1: a competitive assessment of *de novo* short read assembly methods. *Genome Res,* 21**,** 2224-41.

EBRIGHT, R. H. 2000. RNA polymerase: Structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *Journal of Molecular Biology,* 304**,** 687-698.

EHRLICH, G. D., HU, F. Z., SHEN, K., STOODLEY, P. & POST, J. C. 2005. Bacterial plurality as a general mechanism driving persistence in chronic infections. *Clinical Orthopaedics and Related Research***,** 20-24.

ENRIGHT, M. C. & SPRATT, B. G. 1998. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. *Microbiology-Uk,* 144**,** 3049-3060.

FOSTER, P. L. 2007. Stress-induced mutagenesis in bacteria. *Critical Reviews in Biochemistry and Molecular Biology,* 42**,** 373-397.

GARCIA-CASTILLO, M., MOROSINI, M. I., VALVERDE, A., ALMARAZ, F., BAQUERO, F., CANTON, R. & DEL CAMPO, R. 2007. Differences in biofilm development and antibiotic susceptibility among *Streptococcus pneumoniae* isolates from cystic fibrosis samples and blood cultures. *Journal of Antimicrobial Chemotherapy,* 59**,** 301-304.

GIAO, M. S., AZEVEDO, N. F., WILKS, S. A., VIEIRA, M. J. & KEEVIL, C. W. 2010. Effect of Chlorine on Incorporation of *Helicobacter pylori* into Drinking Water Biofilms. *Applied and Environmental Microbiology,* 76**,** 1669-1673.

GIAO, M. S. & KEEVIL, C. W. 2014. *Listeria monocytogenes* Can Form Biofilms in Tap Water and Enter Into the Viable but Non-Cultivable State. *Microbial Ecology,* 67**,** 603-611.

GLADSTONE, R. A., JEFFERIES, J. M., FAUST, S. N. & CLARKE, S. C. 2011. Continued control of pneumococcal disease in the UK - the impact of vaccination. *J Med Microbiol,* 60**,** 1-8.

GOLUBCHIK, T., BRUEGGEMANN, A. B., STREET, T., GERTZ, R. E., JR., SPENCER, C. C., HO, T., GIANNOULATOU, E., LINK-GELLES, R., HARDING, R. M., BEALL, B., PETO, T. E., MOORE, M. R., DONNELLY, P., CROOK, D. W. & BOWDEN, R. 2012. Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event. *Nat Genet,* 44**,** 352-5.

GOTOH, H., KASARANENI, N., DEVINENI, N., DALLO, S. F. & WEITAO, T. 2010. SOS involvement in stress-inducible biofilm formation. *Biofouling,* 26**,** 603-11.

GOTZ, F. 2002. *Staphylococcus* and biofilms. *Molecular Microbiology,* 43**,** 1367-1378.

GRUBER, T. M. & GROSS, C. A. 2003. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annual Review of Microbiology,* 57**,** 441-466.

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**,** 202-11.

HALL-STOODLEY, L., NISTICO, L., SAMBANTHAMOORTHY, K., DICE, B., NGUYEN, D., MERSHON, W. J., JOHNSON, C., HU, F. Z., STOODLEY, P., EHRLICH, G. D. & POST, J. C. 2008. Characterization of biofilm matrix, degradation by DNase treatment and evidence of

capsule downregulation in *Streptococcus pneumoniae* clinical isolates. *Bmc Microbiology,* 8. 8:173

HALL-STOODLEY, L. & STOODLEY, P. 2009. Evolving concepts in biofilm infections. *Cellular Microbiology,* 11**,** 1034-1043.

HAMMERSCHMIDT, S., BETHE, G., REMANE, P. H. & CHHATWAL, G. S. 1999. Identification of pneumococcal surface protein a as a lactoferrin-binding protein of *Streptococcus pneumoniae*. *Infection and Immunity,* 67**,** 1683-1687.

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. *Infection and Immunity,* 73**,** 4653-67.

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**,** 1854-63.

HARRO, J. M., PETERS, B. M., O'MAY, G. A., ARCHER, N., KERNS, P., PRABHAKARA, R. & SHIRTLIFF, M. E. 2010. Vaccine development in *Staphylococcus aureus*: taking the biofilm phenotype into consideration. *Fems Immunology and Medical Microbiology,* 59**,** 306-323.

HAUSDORFF, W. P., DAGAN, R., BECKERS, F. & SCHUERMAN, L. 2009. Estimating the direct impact of new conjugate vaccines against invasive pneumococcal disease. *Vaccine,* 27**,** 7257-7269.

HAUSNER, M. & WUERTZ, S. 1999. High rates of conjugation in bacterial biofilms as determined by quantitative in situ analysis. *Applied and Environmental Microbiology,* 65**,** 3710-3713.

HAUSSLER, S. 2004. Biofilm formation by the small colony variant phenotype of *Pseudomonas aeruginosa*. *Environmental Microbiology,* 6**,** 546-551.

HAUSSLER, S., ZIEGLER, I., LOTTEL, A., VON GOTZ, F., ROHDE, M., WEHMHOHNER, D., SARAVANAMUTHU, S., TUMMLER, B. & STEINMETZ, I. 2003. Highly adherent small-colony variants of *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Journal of Medical Microbiology,* 52**,** 295-301.

HAVARSTEIN, L. S., MARTIN, B., JOHNSBORG, O., GRANADEL, C. & CLAVERYS, J. P. 2006. New insights into the pneumococcal fratricide: relationship to clumping and identification of a novel immunity factor. *Molecular Microbiology,* 59**,** 1297-1307.

HENDERSON, I. R., OWEN, P. & NATARO, J. P. 1999. Molecular switches - the ON and OFF of bacterial phase variation. *Molecular Microbiology,* 33**,** 919-932.

HENGGEARONIS, R. 1993. Survival of Hunger and Stress - the Role of Rpos in Early Stationary Phase Gene-Regulation in *Escherichia coli*. *Cell,* 72**,** 165-168.

HEYDORN, A., NIELSEN, A. T., HENTZER, M., STERNBERG, C., GIVSKOV, M., ERSBOLL, B. K. & MOLIN, S. 2000. Quantification of biofilm structures by the novel computer program COMSTAT. *Microbiology,* 146 (Pt 10)**,** 2395-407.

HILLER, N. L., AHMED, A., POWELL, E., MARTIN, D. P., EUTSEY, R., EARL, J., JANTO, B., BOISSY, R. J., HOGG, J., BARBADORA, K., SAMPATH, R., LONERGAN, S., POST, J. C., HU, F. Z. & EHRLICH, G. D. 2010. Generation of Genic Diversity among *Streptococcus pneumoniae* Strains via Horizontal Gene Transfer during a Chronic Polyclonal Pediatric Infection. *Plos Pathogens,* 6.

HOFFMAN, L. R., DEZIEL, E., D'ARGENIO, D. A., LEPINE, F., EMERSON, J., MCNAMARA, S., GIBSON, R. L., RAMSEY, B. W. & MILLER, S. I. 2006. Selection for Staphylococcus aureus small-colony variants due to growth in the presence of Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 19890-19895.

HOIBY, N., CIOFU, O. & BJARNSHOLT, T. 2010. *Pseudomonas aeruginosa* biofilms in cystic fibrosis. *Future Microbiol,* 5**,** 1663-74.

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**,** 5709-17.

HYAMS, C., YUSTE, J., BAX, K., CAMBERLEIN, E., WEISER, J. N. & BROWN, J. S. 2010. Streptococcus pneumoniae resistance to complement-mediated immunity is dependent on the capsular serotype. *Infection and Immunity,* 78**,** 716-25.

INOUYE, M., CONWAY, T. C., ZOBEL, J. & HOLT, K. E. 2012. Short read sequence typing (SRST): multi-locus sequence types from short reads. *BMC Genomics,* 13**,** 338.

JACOBS, M. R., GOOD, C. E., BAJAKSOUZIAN, S. & WINDAU, A. R. 2008. Emergence of *Streptococcus pneumoniae* Serotypes 19A, 6C, and 22F and Serogroup 15 in Cleveland, Ohio, in Relation to Introduction of the Protein-Conjugated Pneumococcal Vaccine. *Clinical Infectious Diseases,* 47**,** 1388-1395.

JANSEN, A. G., SANDERS, E. A., A, V. D. E., AM, V. A. N. L., HOES, A. W. & HAK, E. 2008. Invasive pneumococcal and meningococcal disease: association with influenza virus and respiratory syncytial virus activity? *Epidemiol Infect,* 136**,** 1448-54.

JEFFERIES, J. M., CLARKE, S. C., WEBB, J. S. & KRAAIJEVELD, A. R. 2011. Risk of red queen dynamics in pneumococcal vaccine strategy. *Trends Microbiol,* 19**,** 377-81.

JEFFERIES, J. M., SMITH, A. J., EDWARDS, G. F. S., MCMENAMIN, J., MITCHELL, T. J. & CLARKE, S. C. 2010. Temporal Analysis of Invasive Pneumococcal Clones from Scotland Illustrates Fluctuations in Diversity of Serotype and Genotype in the Absence of Pneumococcal Conjugate Vaccine. *Journal of Clinical Microbiology,* 48**,** 87-96.

JENSEN, L. J., KUHN, M., STARK, M., CHAFFRON, S., CREEVEY, C., MULLER, J., DOERKS, T., JULIEN, P., ROTH, A., SIMONOVIC, M., BORK, P. & VON MERING, C. 2009. STRING 8-a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Research,* 37**,** D412-D416.

JOHNSON, H. L., DELORIA-KNOLL, M., LEVINE, O. S., STOSZEK, S. K., FREIMANIS HANCE, L., REITHINGER, R., MUENZ, L. R. & O'BRIEN, K. L. 2010. Systematic evaluation of serotypes causing invasive pneumococcal disease among children under five: the pneumococcal global serotype project. *PLoS Med,* 7.

JOHNSTON, C., CAMPO, N., BERGE, M. J., POLARD, P. & CLAVERYS, J. P. 2014. *Streptococcus pneumoniae*, le transformiste. *Trends Microbiol,* 22**,** 113-119.

JUANG, Y. L. & HELMANN, J. D. 1994. The delta subunit of *Bacillus subtilis* RNA polymerase. An allosteric effector of the initiation and core-recycling phases of transcription. *J Mol Biol,* 239**,** 1-14.

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. *Nature Reviews Microbiology,* 6**,** 288-301.

KAHL, B., HERRMANN, M., EVERDING, A. S., KOCH, H. G., BECKER, K., HARMS, E., PROCTOR, R. A. & PETERS, G. 1998. Persistent infection with small colony variant strains of *Staphylococcus aureus* in patients with cystic fibrosis. *Journal of Infectious Diseases,* 177**,** 1023-1029.

KECK, T., LEIACKER, R., RIECHELMANN, H. & RETTINGER, G. 2000. Temperature profile in the nasal cavity. *Laryngoscope,* 110**,** 651-4.

KIM, J. H., KIM, C. H., HACKER, J., ZIEBUHR, W., LEE, B. K. & CHO, S. H. 2008. Molecular characterization of regulatory genes associated with biofilm variation in a *Staphylococcus aureus* strain. *Journal of Microbiology and Biotechnology,* 18**,** 28-34.

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*. *Journal of Infectious Diseases,* 177**,** 368-377.

KIM, W., RACIMO, F., SCHLUTER, J., LEVY, S. B. & FOSTER, K. R. 2014. Importance of positioning for microbial evolution. *Proc Natl Acad Sci U S A,* 111**,** E1639-47.

KING, S. J. 2010. Pneumococcal modification of host sugars: a major contributor to colonization of the human airway? *Mol Oral Microbiol,* 25**,** 15-24.

KIRISITS, M. J., PROST, L., STARKEY, M. & PARSEK, M. R. 2005. Characterization of colony morphology variants isolated from Pseudomonas aeruginosa biofilms. *Appl Environ Microbiol,* 71**,** 4809-21.

KIRKHAM, L. A. S., JEFFERIES, J. M. C., KERR, A. R., JING, Y., CLARKE, S. C., SMITH, A. & MITCHELL, T. J. 2006. Identification of invasive serotype 1 pneumococcal isolates that express nonhemolytic pneumolysin. *Journal of Clinical Microbiology,* 44**,** 151-159.

KLAUSEN, M., AAES-JORGENSEN, A., MOLIN, S. & TOLKER-NIELSEN, T. 2003a. Involvement of bacterial migration in the development of complex multicellular structures in *Pseudomonas aeruginosa* biofilms. *Molecular Microbiology,* 50**,** 61-68.

KLAUSEN, M., HEYDORN, A., RAGAS, P., LAMBERTSEN, L., AAES-JORGENSEN, A., MOLIN, S. & TOLKER-NIELSEN, T. 2003b. Biofilm formation by *Pseudomonas aeruginosa* wild type, flagella and type IV pili mutants. *Molecular Microbiology,* 48**,** 1511-1524.

KOS, M. I., STENZ, L., FRANCOIS, P., GUYOT, J. P. & SCHRENZEL, J. 2009. Immuno-detection of Staphylococcus aureus Biofilm on a Cochlear Implant. *Infection,* 37**,** 450-454.

KUROLA, P., TAPIAINEN, T., SEVANDER, J., KAIJALAINEN, T., LEINONEN, M., UHARI, M. & SAUKKORIIPI, A. 2011. Effect of xylitol and other carbon sources on *Streptococcus pneumoniae* biofilm formation and gene expression in vitro. *Apmis,* 119**,** 135-142.

KWON, H. Y., KIM, S. W., CHOI, M. H., OGUNNIYI, A. D., PATON, J. C., PARK, S. H., PYO, S. N. & RHEE, D. K. 2003. Effect of heat shock and mutations in ClpL and ClpP on virulence gene expression in *Streptococcus pneumoniae*. *Infect Immun,* 71**,** 3757-65.

LARSEN, M. V., COSENTINO, S., RASMUSSEN, S., FRIIS, C., HASMAN, H., MARVIG, R. L., JELSBAK, L., SICHERITZ-PONTEN, T., USSERY, D. W., AARESTRUP, F. M. & LUND, O. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol,* 50**,** 1355-61.

LAUDERDALE, K. J., BOLES, B. R., CHEUNG, A. L. & HORSWILL, A. R. 2009. Interconnections between Sigma B, agr, and Proteolytic Activity in *Staphylococcus aureus* Biofilm Maturation. *Infection and Immunity,* 77**,** 1623-1635.

LEE, J. Y., SONG, J. H. & KO, K. S. 2010. Recombination rates of Streptococcus pneumoniae isolates with both erm(B) and mef(A) genes. *Fems Microbiology Letters,* 309**,** 163-169.

LEEKITCHAROENPHON, P., KAAS, R. S., THOMSEN, M. C., FRIIS, C., RASMUSSEN, S. & AARESTRUP, F. M. 2012. snpTree--a web-server to identify and construct SNP trees from whole genome sequence data. *BMC Genomics,* 13 Suppl 7**,** S6.

LEON, I. R., SCHWAMMLE, V., JENSEN, O. N. & SPRENGER, R. R. 2013. Quantitative assessment of in-solution digestion efficiency identifies optimal protocols for unbiased protein analysis. *Mol Cell Proteomics,* 12**,** 2992-3005.

LETUNIC, I., DOERKS, T. & BORK, P. 2012. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res,* 40**,** D302-5.

LEWIS, K. 2010. Persister cells. *Annu Rev Microbiol,* 64**,** 357-72.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & PROC, G. P. D. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-2079.

LI, Y. H., LAU, P. C. Y., LEE, J. H., ELLEN, R. P. & CVITKOVITCH, D. G. 2001. Natural genetic transformation of *Streptococcus mutans* growing in biofilms. *Journal of Bacteriology,* 183**,** 897-908.

LIEBERMAN, T. D., MICHEL, J. B., AINGARAN, M., POTTER-BYNOE, G., ROUX, D., DAVIS, M. R., SKURNIK, D., LEIBY, N., LIPUMA, J. J., GOLDBERG, J. B., MCADAM, A. J., PRIEBE, G. P. & KISHONY, R. 2011. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nature Genetics,* 43**,** 1275-U148.

LIN, Y., LIN, H., LIU, Z., WANG, K. & YAN, Y. 2014. Improvement of a sample preparation method assisted by sodium deoxycholate for mass-spectrometry-based shotgun membrane proteomics. *J Sep Sci*.

LIZCANO, A., CHIN, T., SAUER, K., TUOMANEN, E. I. & ORIHUELA, C. J. 2010. Early biofilm formation on microtiter plates is not correlated with the invasive disease potential of *Streptococcus pneumoniae*. *Microbial Pathogenesis,* 48**,** 124-130.

LOMAN, N. J., CONSTANTINIDOU, C., CHAN, J. Z., HALACHEV, M., SERGEANT, M., PENN, C. W., ROBINSON, E. R. & PALLEN, M. J. 2012a. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat Rev Microbiol,* 10**,** 599-606.

LOMAN, N. J., MISRA, R. V., DALLMAN, T. J., CONSTANTINIDOU, C., GHARBIA, S. E., WAIN, J. & PALLEN, M. J. 2012b. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol,* 30**,** 434-9.

LOPEZ DE SARO, F. J., WOODY, A. Y. & HELMANN, J. D. 1995. Structural analysis of the *Bacillus subtilis* delta factor: a protein polyanion which displaces RNA from RNA polymerase. *J Mol Biol,* 252**,** 189-202.

LOPEZ DE SARO, F. J., YOSHIKAWA, N. & HELMANN, J. D. 1999. Expression, abundance, and RNA polymerase binding properties of the delta factor of *Bacillus subtilis*. *J Biol Chem,* 274**,** 15953-8.

LUNTER, G. & GOODSON, M. 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res,* 21**,** 936-9.

MANDSBERG, L. F., CIOFU, O., KIRKBY, N., CHRISTIANSEN, L. E., POULSEN, H. E. & HOIBY, N. 2009. Antibiotic resistance in *Pseudomonas aeruginosa* strains with increased mutation frequency due to inactivation of the DNA oxidative repair system. *Antimicrob Agents Chemother,* 53**,** 2483-91.

MARGOLIS, E., YATES, A. & LEVIN, B. R. 2010. The ecology of nasal colonization of *Streptococcus pneumoniae, Haemophilus influenzae* and *Staphylococcus aureus:* the role of competition and interactions with host's immune response. *Bmc Microbiology,* 10**,** -.

MARKS, L. R., DAVIDSON, B. A., KNIGHT, P. R. & HAKANSSON, A. P. 2013a. Interkingdom Signaling Induces *Streptococcus pneumoniae* Biofilm Dispersion and Transition from Asymptomatic Colonization to Disease. *MBio,* 4**,** (4): e00438-13. .

MARKS, L. R., PARAMESWARAN, G. I. & HAKANSSON, A. P. 2012a. Pneumococcal interactions with epithelial cells are crucial for optimal biofilm formation and colonization in vitro and in vivo. *Infect Immun,* 80**,** 2744-60.

MARKS, L. R., REDDINGER, R. M. & HAKANSSON, A. P. 2012b. High levels of genetic recombination during nasopharyngeal carriage and biofilm formation in *Streptococcus pneumoniae*. *MBio,* 3**,** (5) e00200-12.

MARKS, L. R., REDDINGER, R. M. & HAKANSSON, A. P. 2013b. Biofilm Formation enhances Fomite Survival of *S. pneumoniae* and S. pyogenes. *Infect Immun*.

MARSHALL, K. C., STOUT, R. & MITCHELL, R. 1971. Mechanism of Initial Events in Sorption of Marine Bacteria to Surfaces. *Journal of General Microbiology,* 68**,** 337-&.

MARTINEZ-SOLANO, L., MACIA, M. D., FAJARDO, A., OLIVER, A. & MARTINEZ, J. L. 2008. Chronic *Pseudomonas aeruginosa* Infection in Chronic Obstructive Pulmonary Disease. *Clinical Infectious Diseases,* 47**,** 1526-1533.

MATHEE, K., CIOFU, O., STERNBERG, C., LINDUM, P. W., CAMPBELL, J. I. A., JENSEN, P., JOHNSEN, A. H., GIVSKOV, M., OHMAN, D. E., MOLIN, S., HOIBY, N. & KHARAZMI, A. 1999. Mucoid conversion of *Pseudomonas aeruginosa* by hydrogen peroxide: a

mechanism for virulence activation in the cystic fibrosis lung. *Microbiology-Sgm,* 145**,** 1349-1357.

MAVROIDI, A., AANENSEN, D. M., GODOY, D., SKOVSTED, I. C., KALTOFT, M. S., REEVES, P. R., BENTLEY, S. D. & SPRATT, B. G. 2007. Genetic relatedness of the *Streptococcus pneumoniae* capsular Biosynthetic loci. *Journal of Bacteriology,* 189**,** 7841-7855.

MCELLISTREM, M. C., RANSFORD, J. V. & KHAN, S. A. 2007. Characterization of in vitro biofilm-associated pneumococcal phase variants of a clinically relevant serotype 3 clone. *Journal of Clinical Microbiology,* 45**,** 97-101.

MCELROY, K. E., HUI, J. G., WOO, J. K., LUK, A. W., WEBB, J. S., KJELLEBERG, S., RICE, S. A. & THOMAS, T. 2014. Strain-specific parallel evolution drives short-term diversification during *Pseudomonas aeruginosa* biofilm formation. *Proc Natl Acad Sci U S A,* 111**,** E1419-27.

MCKENNA, A., HANNA, M., BANKS, E., SIVACHENKO, A., CIBULSKIS, K., KERNYTSKY, A., GARIMELLA, K., ALTSHULER, D., GABRIEL, S., DALY, M. & DEPRISTO, M. A. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res,* 20**,** 1297-303.

MCKENZIE, G. J., HARRIS, R. S., LEE, P. L. & ROSENBERG, S. M. 2000. The SOS response regulates adaptive mutation. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 6646-6651.

MEATS, E., BRUEGGEMANN, A. B., ENRIGHT, M. C., SLEEMAN, K., GRIFFITHS, D. T., CROOK, D. W. & SPRATT, B. G. 2003. Stability of serotypes during nasopharyngeal carriage of *Streptococcus pneumoniae*. *Journal of Clinical Microbiology,* 41**,** 386-392.

MELEGARO, A., EDMUNDS, W. J., PEBODY, R., MILLER, E. & GEORGE, R. 2006. The current burden of pneumococcal disease in England and Wales. *J Infect,* 52**,** 37-48.

MELTER, O. & RADOJEVIC, B. 2010. Small colony variants of Staphylococcus aureus - review. *Folia Microbiologica,* 55**,** 548-558.

MICHAELS, M. L. & MILLER, J. H. 1992. The Go system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-Dihydro-8-Oxoguanine). *Journal of Bacteriology,* 174**,** 6321-6325.

MILES, A. A., MISRA, S. S. & IRWIN, J. O. 1938. The estimation of the bactericidal power of the blood. *J Hyg (Lond),* 38**,** 732-49.

MITCHELL, A. M. & MITCHELL, T. J. 2010. *Streptococcus pneumoniae*: virulence factors and variation. *Clinical Microbiology and Infection,* 16**,** 411-418.

MITCHELL, G., BROUILLETTE, E., SEGUIN, D. L., ASSELIN, A. E., JACOB, C. L. & MALOUIN, F. 2010a. A role for sigma factor B in the emergence of Staphylococcus aureus small-colony variants and elevated biofilm production resulting from an exposure to aminoglycosides. *Microbial Pathogenesis,* 48**,** 18-27.

MITCHELL, G., SEGUIN, D. L., ASSELIN, A. E., DEZIEL, E., CANTIN, A. M., FROST, E. H., MICHAUD, S. & MALOUIN, F. 2010b. Staphylococcus aureus sigma B-dependent emergence of small-colony variants and biofilm production following exposure to Pseudomonas aeruginosa 4-hydroxy-2-heptylquinoline-N-oxide. *Bmc Microbiology,* 10.

MITCHELL, T. J. 2000. Virulence factors and the pathogenesis of disease caused by Streptococcus pneumoniae. *Res Microbiol,* 151**,** 413-9.

MOROSINI, M. I., BAQUERO, M. R., SANCHEZ-ROMERO, J. M., NEGRI, M. C., GALAN, J. C., DEL CAMPO, R., PEREZ-DIAZ, J. C. & BAQUERO, F. 2003. Frequency of mutation to rifampin resistance in Streptococcus pneumoniae clinical strains: hexA and hexB polymorphisms do not account for hypermutation. *Antimicrob Agents Chemother,* 47**,** 1464-7.

MOSCOSO, M., GARCIA, E. & LOPEZ, R. 2006. Biofilm formation by *Streptococcus pneumoniae:* Role of choline, extracellular DNA, and capsular polysaccharide in microbial accretion. *Journal of Bacteriology,* 188**,** 7785-7795.

MURPHY, T. F., BAKALETZ, L. O. & SMEESTERS, P. R. 2009. Microbial Interactions in the Respiratory Tract. *Pediatric Infectious Disease Journal,* 28**,** S121-S126.

NISTICO, L., KREFT, R., GIESEKE, A., COTICCHIA, J. M., BURROWS, A., KHAMPANG, P., LIU, Y., KERSCHNER, J. E., POST, J. C., LONERGAN, S., SAMPATH, R., HU, F. Z., EHRLICH, G. D., STOODLEY, P. & HALL-STOODLEY, L. 2011. Adenoid reservoir for pathogenic biofilm bacteria. *J Clin Microbiol,* 49**,** 1411-20.

NOLLMANN, M., GILBERT, R., MITCHELL, T., SFERRAZZA, M. & BYRON, O. 2004. The role of cholesterol in the activity of pneumolysin, a bacterial protein toxin. *Biophys J,* 86**,** 3141-51.

NORMARK, B. H., KALIN, M., ORTQVIST, A., AKERLUND, T., LILJEQUIST, B. O., HEDLUND, J., SVENSON, S. B., ZHOU, J., SPRATT, B. G., NORMARK, S. & KALLENIUS, G. 2001. Dynamics of penicillin-susceptible clones in invasive pneumococcal disease. *Journal of Infectious Diseases,* 184**,** 861-869.

NOSANCHUK, J. D. & CASADEVALL, A. 2003. The contribution of melanin to microbial pathogenesis. *Cellular Microbiology,* 5**,** 203-223.

O'BRIEN, K. L., WOLFSON, L. J., WATT, J. P., HENKLE, E., DELORIA-KNOLL, M., MCCALL, N., LEE, E., MULHOLLAND, K., LEVINE, O. S., CHERIAN, T. & DIS, H. P. G. B. 2009. Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. *Lancet,* 374**,** 893-902.

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**,** 1196-210.

OLIVER, A. & MENA, A. 2010. Bacterial hypermutation in cystic fibrosis, not only for antibiotic resistance. *Clinical Microbiology and Infection,* 16**,** 798-808.

OLIVER, M. B., VAN DER LINDEN, M. P., KUNTZEL, S. A., SAAD, J. S. & NAHM, M. H. 2013. Discovery of *Streptococcus pneumoniae* serotype 6 variants with glycosyltransferases synthesizing two differing repeating units. *J Biol Chem,* 288**,** 25976-85.

OVERBEEK, R., OLSON, R., PUSCH, G. D., OLSEN, G. J., DAVIS, J. J., DISZ, T., EDWARDS, R. A., GERDES, S., PARRELLO, B., SHUKLA, M., VONSTEIN, V., WATTAM, A. R., XIA, F. F. & STEVENS, R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research,* 42**,** D206-D214.

OVERWEG, K., PERICONE, C. D., VERHOEF, G. G. C., WEISER, J. N., MEIRING, H. D., DE JONG, A., DE GROOT, R. & HERMANS, P. W. M. 2000. Differential protein expression in phenotypic variants of *Streptococcus pneumoniae*. *Infection and Immunity,* 68**,** 4604-4610.

PAI, R., GERTZ, R. E. & BEALL, B. 2006. Sequential multiplex PCR approach for determining capsular serotypes of *Streptococcus pneumoniae* isolates. *Journal of Clinical Microbiology,* 44**,** 124-131.

PARK, H. K., LEE, S. J., YOON, J. W., SHIN, J. W., SHIN, H. S., KOOK, J. K., MYUNG, S. C. & KIM, W. 2010. Identification of the cpsA gene as a specific marker for the discrimination of Streptococcus pneumoniae from viridans group streptococci. *J Med Microbiol,* 59**,** 1146-52.

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. *Infection and Immunity,* 77**,** 3722-30.

PARSEK, M. R., VAL, D. L., HANZELKA, B. L., CRONAN, J. E., JR. & GREENBERG, E. P. 1999. Acyl homoserine-lactone quorum-sensing signal generation. *Proc Natl Acad Sci U S A,* 96**,** 4360-5.

PERICONE, C. D., OVERWEG, K., HERMANS, P. W. M. & WEISER, J. N. 2000. Inhibitory and bactericidal effects of hydrogen peroxide production by *Streptococcus pneumoniae* on other inhabitants of the upper respiratory tract. *Infection and Immunity,* 68**,** 3990-3997.

PETERSEN, F. C., PECHARKI, D. & SCHEIE, A. A. 2004. Biofilm mode of growth of *Streptococcus intermedius* favored by a competence-stimulating signaling peptide. *Journal of Bacteriology,* 186**,** 6327-6331.

PETERSON, S. N., SUNG, C. K., CLINE, R., DESAI, B. V., SNESRUD, E. C., LUO, P., WALLING, J., LI, H. Y., 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. *Molecular Microbiology,* 51**,** 1051-1070.

PHILIPS, B. J., MEGUER, J. X., REDMAN, J. & BAKER, E. H. 2003. Factors determining the appearance of glucose in upper and lower respiratory tract secretions. *Intensive Care Med,* 29**,** 2204-10.

POTERA, C. 1999. Microbiology - Forging a link between biofilms and disease. *Science,* 283**,** 1837-+.

PROCTOR, R. A., VON EIFF, C., KAHL, B. C., BECKER, K., MCNAMARA, P., HERRMANN, M. & PETERS, G. 2006. Small colony variants: a pathogenic form of bacteria that facilitates persistent and recurrent infections. *Nat Rev Microbiol,* 4**,** 295-305.

PRUDHOMME, M., LIBANTE, V. & CLAVERYS, J. P. 2002. Homologous recombination at the border: Insertion-deletions and the trapping of foreign DNA in *Streptococcus pneumoniae*. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 2100-2105.

PRUNIER, A. L., MALBRUNY, B., LAURANS, M., BROUARD, J., DUHAMEL, J. F. & LECLERCQ, R. 2003. High rate of macrolide resistance in *Staphylococcus aureus* strains from patients with cystic fibrosis reveals high proportions of hypermutable strains. *Journal of Infectious Diseases,* 187**,** 1709-1716.

QIN, L., KIDA, Y., IMAMURA, Y., KUWANO, K. & WATANABE, H. 2013. Impaired capsular polysaccharide is relevant to enhanced biofilm formation and lower virulence in *Streptococcus pneumoniae*. *J Infect Chemother,* 19**,** 261-71.

RABATINOVA, A., SANDEROVA, H., JIRAT MATEJCKOVA, J., KORELUSOVA, J., SOJKA, L., BARVIK, I., PAPOUSKOVA, V., SKLENAR, V., ZIDEK, L. & KRASNY, L. 2013. The delta subunit of RNA polymerase is required for rapid changes in gene expression and competitive fitness of the cell. *J Bacteriol,* 195**,** 2603-11.

REID, S. D., HONG, W. Z., DEW, K. E., WINN, D. R., PANG, B., WATT, J., GLOVER, D. T., HOLLINGSHEAD, S. K. & SWORDS, W. E. 2009. *Streptococcus pneumoniae* Forms Surface-Attached Communities in the Middle Ear of Experimentally Infected Chinchillas. *Journal of Infectious Diseases,* 199**,** 786-794.

ROBINSON, J. T., THORVALDSDOTTIR, H., WINCKLER, W., GUTTMAN, M., LANDER, E. S., GETZ, G. & MESIROV, J. P. 2011. Integrative genomics viewer. *Nat Biotechnol,* 29**,** 24-6.

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**,** 819-29.

ROTH, J. R., KUGELBERG, E., REAMS, A. B., KOFOID, E. & ANDERSSON, D. I. 2006. Origin of mutations under selection: The adaptive mutation controversy. *Annual Review of Microbiology,* 60**,** 477-501.

RUKKE, H. V., HEGNA, I. K. & PETERSEN, F. C. 2012. Identification of a functional capsule locus in Streptococcus mitis. *Mol Oral Microbiol,* 27**,** 95-108.

RYDER, C., BYRD, M. & WOZNIAK, D. J. 2007. Role of polysaccharides in *Pseudomonas aeruginosa* biofilm development. *Curr Opin Microbiol,* 10**,** 644-8.

SADOWSKA, B., BONAR, A., VON EIFF, C., PROCTOR, R. A., CHMIELA, M., RUDNICKA, W. & ROZALSKA, B. 2002. Characteristics of *Staphylococcus aureus*, isolated from airways of cystic fibrosis patients, and their small colony variants. *Fems Immunology and Medical Microbiology,* 32**,** 191-197.

SALO, J., SEVANDER, J. J., TAPIAINEN, T., IKAHEIMO, I., POKKA, T., KOSKELA, M. & UHARI, M. 2009. Biofilm formation by Escherichia coli isolated from patients with urinary tract infections. *Clinical Nephrology,* 71**,** 501-507.

SANCHEZ, C. J., HURTGEN, B. J., LIZCANO, A., SHIVSHANKAR, P., COLE, G. T. & ORIHUELA, C. J. 2011a. Biofilm and planktonic pneumococci demonstrate disparate immunoreactivity to human convalescent sera. *BMC Microbiol,* 11**,** 245.

SANCHEZ, C. J., KUMAR, N., LIZCANO, A., SHIVSHANKAR, P., DUNNING HOTOPP, J. C., JORGENSEN, J. H., TETTELIN, H. & ORIHUELA, C. J. 2011b. *Streptococcus pneumoniae* in biofilms are unable to cause invasive disease due to altered virulence determinant production. *PLoS One,* 6**,** e28738.

SANCHEZ, C. J., SHIVSHANKAR, P., STOL, K., TRAKHTENBROIT, S., SULLAM, P. M., SAUER, K., HERMANS, P. W. & ORIHUELA, C. J. 2010. The pneumococcal serine-rich repeat protein is an intra-species bacterial adhesin that promotes bacterial aggregation in vivo and in biofilms. *PLoS Pathog,* 6**,** e1001044.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A,* 74**,** 5463-7.

SAUER, K., CAMPER, A. K., EHRLICH, G. D., COSTERTON, J. W. & DAVIES, D. G. 2002. *Pseudomonas aeruginosa* displays multiple phenotypes during development as a biofilm. *Journal of Bacteriology,* 184**,** 1140-1154.

SAUER, K., CULLEN, M. C., RICKARD, A. H., ZEEF, L. A. H., DAVIES, D. G. & GILBERT, P. 2004. Characterization of nutrient-induced dispersion in *Pseudomonas aeruginosa* PAO1 biofilm. *Journal of Bacteriology,* 186**,** 7312-7326.

SCHNEIDER, M., MUHLEMANN, K., DROZ, S., COUZINET, S., CASAULTA, C. & ZIMMERLI, S. 2008. Clinical characteristics associated with isolation of small-colony variants of Staphylococcus aureus and Pseudomonas aeruginosa from respiratory secretions of patients with cystic fibrosis. *Journal of Clinical Microbiology,* 46**,** 1832-4.

SCHULTZ, J., MILPETZ, F., BORK, P. & PONTING, C. P. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A,* 95**,** 5857-64.

SCHUTZE, G. E., MASON, E. O., BARSON, W. J., KIM, K. S., WALD, E. R., GIVNER, L. B., TAN, T. Q., BRADLEY, J. S., YOGEV, R. & KAPLAN, S. L. 2002. Invasive pneumococcal infections in children with asplenia. *Pediatric Infectious Disease Journal,* 21**,** 278-282.

SERRANO, I., MELO-CRISTINO, J. & RAMIREZ, M. 2006. Heterogeneity of pneumococcal phase variants in invasive human infections. *Bmc Microbiology,* 6.

SHELDON, J. R., YIM, M. S., SALIBA, J. H., CHUNG, W. H., WONG, K. Y. & LEUNG, K. T. 2012. Role of rpoS in *Escherichia coli* O157:H7 strain H32 biofilm development and survival. *Appl Environ Microbiol,* 78**,** 8331-9.

SHENDURE, J. A., PORRECA, G. J. & CHURCH, G. M. 2008. Overview of DNA sequencing strategies. *Curr Protoc Mol Biol,* Chapter 7**,** Unit 7 1.

SHEPPARD, C. L., PICHON, B., GEORGE, R. C. & HALL, L. M. C. 2010. *Streptococcus pneumoniae* Isolates Expressing a Capsule with Epitopes of Both Serotypes 6A and 6B. *Clinical and Vaccine Immunology,* 17**,** 1820-1822.

SIMELL, B., AURANEN, K., KAYHTY, H., GOLDBLATT, D., DAGAN, R., O'BRIEN, K. L. & PNEUMOCOCCAL CARRIAGE, G. 2012. The fundamental link between pneumococcal carriage and disease. *Expert Rev Vaccines,* 11**,** 841-55.

SINGH, R., RAY, P., DAS, A. & SHARMA, M. 2010. Enhanced production of exopolysaccharide matrix and biofilm by a menadione-auxotrophic Staphylococcus aureus small-colony variant. *J Med Microbiol,* 59**,** 521-7.

SKINNER, J. M., INDRAWATI, L., CANNON, J., BLUE, J., WINTERS, M., MACNAIR, J., PUJAR, N., MANGER, W., ZHANG, Y., ANTONELLO, J., SHIVER, J., CAULFIELD, M. & HEINRICHS, J. H. 2011. Pre-clinical evaluation of a 15-valent pneumococcal conjugate vaccine (PCV15-CRM197) in an infant-rhesus monkey immunogenicity model. *Vaccine,* 29**,** 8870-6.

SLEEMAN, K., KNOX, K., GEORGE, R., MILLER, E., WAIGHT, P., GRIFFITHS, D., EFSTRATIOU, A., BROUGHTON, K., MAYON-WHITE, R. T., MOXON, E. R., CROOK, D. W., SERV, P. H. L. & G, O. P. S. 2001. Invasive pneumococcal disease in England and Wales: Vaccination implications. *Journal of Infectious Diseases,* 183**,** 239-246.

SONG, X. M., CONNOR, W., HOKAMP, K., BABIUK, L. A. & POTTER, A. A. 2009. The growth phase-dependent regulation of the pilus locus genes by two-component system TCS08 in *Streptococcus pneumoniae*. *Microb Pathog,* 46**,** 28-35.

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. *Journal of Clinical Investigation,* 116**,** 2828-2828.

SORENSEN, U. B. S., BLOM, J., BIRCHANDERSEN, A. & HENRICHSEN, J. 1988. Ultrastructural-Localization of Capsules, Cell-Wall Polysaccharide, Cell-Wall Proteins, and F-Antigen in Pneumococci. *Infection and Immunity,* 56**,** 1890-1896.

SOTO, S. M., SMITHSON, A., MARTINEZ, J. A., HORCAJADA, J. P., MENSA, J. & VILA, J. 2007. Biofilm formation in uropathogenic Escherichia coli strains: Relationship with prostatitis, urovirulence factors and antimicrobial resistance. *Journal of Urology,* 177**,** 365-368.

SPRATT, B. G. & GREENWOOD, B. M. 2000. Prevention of pneumococcal disease by vaccination: does serotype replacement matter? *Lancet,* 356**,** 1210-1211.

STEIN, K. E. 1992. Thymus-Independent and Thymus-Dependent Responses to Polysaccharide Antigens. *Journal of Infectious Diseases,* 165**,** S49-S52.

STERNBERG, C., CHRISTENSEN, B. B., JOHANSEN, T., NIELSEN, A. T., ANDERSEN, J. B., GIVSKOV, M. & MOLIN, S. 1999. Distribution of bacterial growth activity in flow-chamber biofilms. *Applied and Environmental Microbiology,* 65**,** 4108-4117.

STEWART, P. S. & COSTERTON, J. W. 2001. Antibiotic resistance of bacteria in biofilms. *Lancet,* 358**,** 135-138.

STOODLEY, P., HALL-STOODLEY, L. & LAPPIN-SCOTT, H. M. 2001. Detachment, surface migration, and other dynamic behavior in bacterial biofilms revealed by digital time-lapse imaging. *Microbial Growth in Biofilms, Pt B,* 337**,** 306-318.

STOODLEY, P., SAUER, K., DAVIES, D. G. & COSTERTON, J. W. 2002. Biofilms as complex differentiated communities. *Annual Review of Microbiology,* 56**,** 187-209.

SUNTHARALINGAM, P. & CVITKOVITCH, D. G. 2005. Quorum sensing in *Streptococcal* biofilm formation. *Trends in Microbiology,* 13**,** 3-6.

SZTAJER, H., LEMME, A., VILCHEZ, R., SCHULZ, S., GEFFERS, R., YIP, C. Y., LEVESQUE, C. M., CVITKOVITCH, D. G. & WAGNER-DOBLER, I. 2008. Autoinducer-2-regulated genes in *Streptococcus mutans* UA159 and global metabolic effect of the *lux*S mutation. *J Bacteriol,* 190**,** 401-15.

TAPIAINEN, T., KUJALA, T., KAIJALAINEN, T., IKAHEIMO, I., SAUKKORIIPI, A., RENKO, M., SALO, J., LEINONEN, M. & UHARI, M. 2010. Biofilm formation by *Streptococcus pneumoniae* isolates from paediatric patients. *Apmis,* 118**,** 255-260.

THORVALDSDOTTIR, H., ROBINSON, J. T. & MESIROV, J. P. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform,* 14**,** 178-92.

TOCHEVA, A. S., JEFFERIES, J. M., RUBERY, H., BENNETT, J., AFIMEKE, G., GARLAND, J., CHRISTODOULIDES, M., FAUST, S. N. & CLARKE, S. C. 2011. Declining serotype coverage of new pneumococcal conjugate vaccines relating to the carriage of *Streptococcus pneumoniae* in young children. *Vaccine,* 29**,** 4400-4.

TOCHEVA, A. S., JEFFERIES, J. M. C., CHRISTODOULIDES, M., FAUST, S. N. & CLARKE, S. C. 2010. Increase in Serotype 6C Pneumococcal Carriage, United Kingdom. *Emerging Infectious Diseases,* 16**,** 154-155.

TONG, H., ZENG, L. & BURNE, R. A. 2011. The EIIABMan phosphotransferase system permease regulates carbohydrate catabolite repression in *Streptococcus gordonii*. *Appl Environ Microbiol,* 77**,** 1957-65.

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., 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. *Journal of Infectious Diseases,* 199**,** 1497-1505.

TRAPPETTI, C., POTTER, A. J., PATON, A. W., OGGIONI, M. R. & PATON, J. C. 2011b. LuxS mediates iron-dependent biofilm formation, competence, and fratricide in *Streptococcus pneumoniae*. *Infect Immun,* 79**,** 4550-8.

TRAPPETTI, C., VAN DER MATEN, E., AMIN, Z., POTTER, A. J., CHEN, A. Y., VAN MOURIK, P. M., LAWRENCE, A. J., PATON, A. W. & PATON, J. C. 2013. Site of isolation determines biofilm formation and virulence phenotypes of Streptococcus pneumoniae serotype 3 clinical isolates. *Infect Immun,* 81**,** 505-13.

TREVORS, J. T. 2011. Viable but non-culturable (VBNC) bacteria: Gene expression in planktonic and biofilm cells. *J Microbiol Methods,* 86**,** 266-73.

TU, L. N., JEONG, H. Y., KWON, H. Y., OGUNNIYI, A. D., PATON, J. C., PYO, S. N. & RHEE, D. K. 2007. Modulation of adherence, invasion, and tumor necrosis factor alpha secretion during the early stages of infection by Strepococcus pneumoniae ClpL. *Infection and Immunity,* 75**,** 2996-3005.

VIDAL, J. E., HOWERY, K. E., LUDEWICK, H. P., NAVA, P. & KLUGMAN, K. P. 2013. Quorum-Sensing Systems LuxS/Autoinducer 2 and Com Regulate *Streptococcus pneumoniae* Biofilms in a Bioreactor with Living Cultures of Human Respiratory Cells. *Infection and Immunity,* 81**,** 1341-1353.

VON MOLLENDORF, C., COHEN, C., DE GOUVEIA, L., NAIDOO, N., MEIRING, S., QUAN, V., LINDANI, S., MOORE, D. P., REUBENSON, G., MOSHE, M., ELEY, B., HALLBAUER, U. M., FINLAYSON, H., MADHI, S. A., CONKLIN, L., ZELL, E. R., KLUGMAN, K. P., WHITNEY, C. G., VON GOTTBERG, A. & FOR THE SOUTH AFRICAN, I. P. D. C.-C. S. G. 2014. Risk Factors for Invasive Pneumococcal Disease among Children less Than 5 Years of Age in a High HIV-Prevalence Setting, South Africa, 2010 to 2012. *Pediatr Infect Dis J*.

VUONG, C., VOYICH, J. M., FISCHER, E. R., BRAUGHTON, K. R., WHITNEY, A. R., DELEO, F. R. & OTTO, M. 2004. Polysaccharide intercellular adhesin (PIA) protects Staphylococcus epidermidis against major components of the human innate immune system. *Cellular Microbiology,* 6**,** 269-275.

WAITE, R. D., STRUTHERS, J. K. & DOWSON, C. G. 2001. *Spontaneous sequence* duplication within an open reading frame of the pneumococcal type 3 capsule locus causes high-frequency phase variation. *Molecular Microbiology,* 42**,** 1223-1232.

WATSON, M. E., BURNS, J. L. & SMITH, A. L. 2004. Hypermutable *Haemophilus influenzae* with mutations in *mut*S are found in cystic fibrosis sputum. *Microbiology-Sgm,* 150**,** 2947-2958.

WEBB, J. S., GIVSKOV, M. & KJELLEBERG, S. 2003a. Bacterial biofilms: prokaryotic adventures in multicellularity. *Current Opinion in Microbiology,* 6**,** 578-585.

WEBB, J. S., LAU, M. & KJELLEBERG, S. 2004. Bacteriophage and phenotypic variation in *Pseudomonas aeruginosa* biofilm development. *Journal of Bacteriology,* 186**,** 8066-8073.

WEBB, J. S., THOMPSON, L. S., JAMES, S., CHARLTON, T., TOLKER-NIELSEN, T., KOCH, B., GIVSKOV, M. & KJELLEBERG, S. 2003b. Cell death in *Pseudomonas aeruginosa* biofilm development. *Journal of Bacteriology,* 185**,** 4585-4592.

WEIMER, K. E., ARMBRUSTER, C. E., JUNEAU, R. A., HONG, W., PANG, B. & SWORDS, W. E. 2010. Coinfection with *Haemophilus influenzae* promotes pneumococcal biofilm formation during experimental otitis media and impedes the progression of pneumococcal disease. *J Infect Dis,* 202**,** 1068-75.

WEISS, A., IBARRA, J. A., PAOLETTI, J., CARROLL, R. K. & SHAW, L. N. 2014. The delta Subunit of RNA Polymerase Guides Promoter Selectivity and Virulence in *Staphylococcus aureus*. *Infect Immun,* 82**,** 1424-35.

WHO 2007. Pneumococcal conjugate vaccine for childhood immunization – WHO position paper. *Weekly Epidemiological Record,* 82**,** 93–104.

XUE, X., LI, J., WANG, W., SZTAJER, H. & WAGNER-DOBLER, I. 2012. The global impact of the delta subunit RpoE of the RNA polymerase on the proteome of *Streptococcus mutans*. *Microbiology,* 158**,** 191-206.

XUE, X., SZTAJER, H., BUDDRUHS, N., PETERSEN, J., ROHDE, M., TALAY, S. R. & WAGNER-DOBLER, I. 2011. Lack of the delta subunit of RNA polymerase increases virulence related traits of *Streptococcus mutans*. *PLoS One,* 6**,** e20075.

XUE, X., TOMASCH, J., SZTAJER, H. & WAGNER-DOBLER, I. 2010. The delta subunit of RNA polymerase, RpoE, is a global modulator of *Streptococcus mutans* environmental adaptation. *J Bacteriol,* 192**,** 5081-92.

YACHI, S. & LOREAU, M. 1999. Biodiversity and ecosystem productivity in a fluctuating environment: The insurance hypothesis. *Proceedings of the National Academy of Sciences of the United States of America,* 96**,** 1463-1468.

YADAV, M. K., CHAE, S. W. & SONG, J. J. 2012. *In vitro Streptococcus pneumoniae* Biofilm Formation and *in vivo* Middle Ear Mucosal Biofilm in a Rat Model of Acute Otitis Induced by *S. pneumoniae*. *Clinical and Experimental Otorhinolaryngology,* 5**,** 139-144.

YANG, J. W., EVANS, B. A. & ROZEN, D. E. 2010. Signal diffusion and the mitigation of social exploitation in pneumococcal competence signalling. *Proceedings of the Royal Society B-Biological Sciences,* 277**,** 2991-2999.

YESILKAYA, H., KADIOGLU, A., GINGLES, N., ALEXANDER, J. E., MITCHELL, T. J. & ANDREW, P. W. 2000. Role of manganese-containing superoxide dismutase in oxidative stress and virulence of *Streptococcus pneumoniae*. *Infect Immun,* 68**,** 2819-26.

YESILKAYA, H., MANCO, S., KADIOGLU, A., TERRA, V. S. & ANDREW, P. W. 2008. The ability to utilize mucin affects the regulation of virulence gene expression in *Streptococcus pneumoniae*. *FEMS Microbiol Lett,* 278**,** 231-5.

YOTHER, J. 2011. Capsules of *Streptococcus pneumoniae* and other bacteria: paradigms for polysaccharide biosynthesis and regulation. *Annual Review of Microbiology, Vol 65,* 65**,** 563-581.

ZHANG, Q., XU, S. X., WANG, H., XU, W. C., ZHANG, X. M., WU, K. F., LIU, L. & YIN, Y. B. 2009. Contribution of ClpE to virulence of *Streptococcus pneumoniae*. *Canadian Journal of Microbiology,* 55**,** 1187-1194.

ZIEBUHR, W., KRIMMER, V., RACHID, S., LOSSNER, I., GOTZ, F. & HACKER, J. 1999. A novel mechanism of phase variation of virulence in *Staphylococcus epidermidis*: evidence for control of the polysaccharide intercellular adhesin synthesis by alternating insertion and excision of the insertion sequence element IS256. *Molecular Microbiology,* 32**,** 345-356.