Investigating the role of Rgg Transcriptional Regulators in *Streptococcus pneumoniae*

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by

Bushra Shlla

Department of Infection, Immunity and Inflammation
University of Leicester

April 2018

Statement of originality

This accompanying thesis submitted for the degree of PhD entitled "Investigating the role of Rgg Transcriptional Regulators in *Streptococcus pneumoniae*" is based on work conducted by the author in the Department of Infection, Immunity and Inflammation of the University of Leicester during the period between September 2014 and September 2017.

All the work recorded in this thesis is original unless otherwise acknowledged in the text or by references.

None of this work has been submitted for another degree in this or any other University.

Signed …….. Date ……..……

Investigating the role of Rgg Transcriptional Regulators in *Streptococcus pneumoniae*

Bushra Shlla

Abstract

Rgg is a group of transcriptional regulators found in Gram positive bacteria, and they play diverse roles including in metabolism and virulence in a range of bacteria. However, their roles in the important human pathogen *Streptococcus pneumoniae* are not known in detail. To assess the functional importance of Rgg homologs in *S. pneumoniae*, targeted mutation was used for the deletion of two rgg genes, SPD_0999 and SPD_1518 (*rgg0999* and *rgg1518*). *In vitro* analysis showed that under microaerobic conditions, the mutants were attenuated on mannose and galactose compared to the wild type, whereas their growth profiles were similar to that of wild type in BHI (Brain Heart Infusion) and CDM (Chemically Defined Medium) containing glucose. *In vivo* analysis of Rgg mutants showed that both SPD_0999 and SPD_1518 are required for pneumococcal virulence. All mice infected intranasaly with either Rgg mutant survived significantly longer and had less bacteria in their blood compared to the wild type. Moreover, microarray analysis using cultures grown on galactose and mannose showed that the largest number of differentially expressed genes in rgg mutant strains was seen on galactose, especially those encoding for essential physiological functions and virulence related genes such as the capsular locus. Electrophoretic mobility shift assay, and transcriptional *lacZ*-fusions are the most appropriate techniques used for investigating the regulating role of genes in transcriptional regulation. In this study, the results show that Rgg1518 has a direct ability to control the transcription of capsule locus genes (SPD_0315-SPD_0328) and the data also show that the protein Rgg1518 acts as an activator for capsule locus genes in presence of galactose. Therefore, so far the available results show that Rggs encoded by SPD_0999 or SPD_1518 play a major role in pneumococcal virulence.

Acknowledgements & dedications

Firstly, I would like to express my sincere gratitude to my supervisor Dr Hasan Yesilkaya for the continuous support of my PhD study and related research, for His patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my PhD study.

I would also like to acknowledge Prof. Peter Andrew as the second supervisor, and I am gratefully indebted to his very valuable and insightful comments and encouragement, but also for the hard questions which encouraged me to widen my research from various perspectives.

I would also like to thank my colleagues Anfal Motib, Hasan Kahya, Firas Younis, Hastyar Najmuldeen and the members of lab 125 for their useful discussions.

My sincere thanks also goes to my sponsor The Iraqi Ministry of Higher Education and Scientific Research (MoHESR) for funding my scholarship. I also acknowledge the role of the Iraqi cultural attache and University of Mosul for generous support.

Finally, I must express my very profound gratitude to my brother Ahmed Alshuwaili and sister Riffa Shlla for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Abbreviations

BHI	Brain Heart infusion	dH2O	Distilled water
BAB	Blood Agar Base	v/v	Volume per volume
CDM	Chemically defined medium	w/v	Weight per volume
LB	Lysogeny Broth	x g	Gravity force
LA	Lysogeny Agar	ml	Millilitre
CFU	Colony forming units	µl	Microliter
PBS	Phosphate Buffered Saline	g	Gram
EDTA	Ethylenediaminetetraacetic acid	OD	Optical density
SDS	Sodium dodecyl sulphate	dNTP	Deoxynucleotide triphosphate
µg	Microgram	PCR	Polymerase chain reaction
µM	Micromolar	THB	Todd-Hewitt Broth
DNA	Deoxyribonucleic acid	ONPG	2-Nitrophenyl β-Dgalactopyranoside
bp	Base Pare	IPTG	Isopropyl β-D-1-thiogalactopyranoside
kb	Kilobase	TEMED	Tetramethylethylenediamine
RNA	Ribonucleic acid	kDa	Kilodalton
EMSA	Electrophoretic mobility shift assay	His	Histidine
QS	Quorum sensing	TCS	Two component system
PFL	Pyruvate formate lyase	RNA	Ribonucleic acid
CSP	Competence stimulating peptide	PCR	Polymerase chain reaction
Gal	galactose	Man	Mannose
GAS	Group A Streptococci	LAB	Lactic acid bacteria

Table of Contents

Chapter 1. Introduction

1.1. General characteristics

Streptococcus pneumoniae, also known as the pneumococcus, was first discovered and identified by George Miller Sternberg in the United States and by Louis Pasteur in France (Austrian, 1999). The pneumococcus is a Gram-positive, catalase-negative, aerotolerant anaerobic bacterium (Hoskins *et al.,* 2001). It has a spherical shape, normally presents in the form of pairs. The optimal temperature for growth is between 35-37°C. The use of 5% CO_2 or a candle-jar increases pneumococcal growth rate, and it is usually cultured on blood agar plates (BAP) or on a chocolate agar plates (CAP). *S. pneumoniae* is sensitive to optochin (ethylhydrocupreine) therefore it is used for pneumococcal identification (Stevens and Kaplan, 2000). Colonies of *S. pneumoniae* on blood agar arise as small colonies (Arbique *et al.,* 2004), it displays α-hemolysis when cultured on blood agar, and appears as green due to incomplete breakdown of heme in erythrocytes. The microbe does not form spores and is immotile (Iyer and Camilli, 2007). Pneumococcus lacks catalase and does not have M protein, which is a virulence factor and protects the pathogen against phagocytosis via preventing opsonisation by C3b (Bisno, 1991; Fischetti, 2016). Sugars are considered to be the main source of energy for biosynthesis and growth (Bisno, 1991). The pneumococcus relies on fermentation and can ferment up to 32 different sugars (Iyer and Camilli, 2007; Langvas-Nielsen A, 1944). For this reason, *In vivo* fitness is suggested to be due to the efficient utilisation of host-derived sugars (Paixão *et al.,* 2015).

The circular genome of *S. pneumoniae* is AT-rich (60.3%), has between 2 - 2.1 million base pairs, and contains a small plasmid (Tettelin *et al.,* 2001). The identification of 90 different capsule types is the basis of antigenic serotyping of the microbe. Various types of capsular (polysaccharide) antigens, derived from the highly-prevalent strains are

included in anti-pneumococcal vaccines (Henrichsen, 1995). Therefore, the protection conferred by these vaccines is not universal against all serotypes.

1.2. Epidemiological impact of pneumococcal infection on human health

Streptococcus pneumoniae is one of the most common causes of serious diseases worldwide with high morbidity and mortality. Pneumococcal diseases have become a major concern because they are responsible for deaths of approximately 3-5 million people per year due to pneumonia, meningitis, otitis media, and bacteremia (WHO, 2012; van Deursen *et al.*, 2016). The pneumococcus is responsible for over 75% of pneumonia cases (Musher and Thorner, 2014). This dangerous pathogen is responsible for cause meningitis, bacteremia and otitis media among elder people over the age of 65 and children under the age of 5 years (Sleeman *et al.,* 2001). These high rates are a serious public health concerns.

Pneumococcal infections can be seen in all age groups, however, most infections occur among children under 5 years old, elderly individuals, and those who suffer from immune deficiencies such as human immunodeficiency virus infection (HIV). According to epidemiological studies, the risk of respiratory infections is increasing, specifically those caused by pneumococcus (Marongiu *et al.*, 2016). In the United States in 2009, it was reported that approximately 14 invasive pneumococcal disease (IPD) cases occurred per 100,000 in adults and children (CDC, 2010a), and 10% of these cases, result in mortality (CDC, 2010b). The community acquired pneumonia (CAP) cause up to 40% mortality among elderly people in Europe (Jansen *et al.*, 2009). Most pneumococcal deaths occur among children less than 5 years of age living in developing countries (WHO, 2012). It has been reported that 67% of the estimated 14.5 million pneumococcal infections are recorded in Africa or Asia among children each year and

around one million children under five died due to pneumonia in 2013 (O'Brien *et al.*, 2009; WHO, 2014). Moreover, depending on the age groups, 30-60% of survivors suffer long-term sequelae including hearing loss, neurological deficits, and neuropsychological impairments (Koedel *et al.*, 2002).

The progression of pneumococcal pneumonia to meningitis and bacteremia leads to high mortality rates among infected people. Some studies have reported data indicating that pneumococcal pneumonia causes 60-87% of pneumococcal bacteremia in the USA and its development may lead to severe complications such as respiratory failure, meningitis, and pleural effusion (Carbon *et al.*, 2006). In the USA, about 900,000 cases of pneumococcal pneumonia are recorded each year and 5-7% of these lead to death (Huang *et al.*, 2011). Young children below 2 years and elder people over 65 years are the more prone to developing bacteremia than other age groups (Carbon *et al.*, 2006). It has been estimated that 3,700 deaths occurred in the USA due to pneumococcal meningitis and bacteraemia in 2013 (CDC, 2015). In fact, pneumococcus was responsible for about 130,000 cases of pneumonia, 3300 cases of meningitis and 6 million cases of otitis media and, 14% of these cases led to death despite the presence of a well-established healthcare (CDC, 2005b). All recent studies have shown that pneumococcus is still a serious and persistent threat for public health around the world.

Pneumococcal infections place a heavy financial burden on national health systems. It has been estimated that in 2004, the direct medical costs for pneumococcal infections amounted to a total of $3.5 billion in the USA. Although pneumonia comprised only 22% of previous infections, pneumonia accounted for 72% of pneumococcal costs in the USA. This cost was significantly higher in comparison with that spent on acute otitis media which was only 16% of total costs (Huang, 2011). Another evaluation has

estimated that, the direct costs of pneumococcal diseases among persons aged over 50 years in the USA to have been $3.7 billion in 2007 (Weycker *et al.,* 2010). Furthermore, general rates and costs of treatment of invasive pneumococcal disease (IPD) are higher among persons with certain chronic and immune-compromising conditions in comparison with those without any such conditions.

These data reveal that pneumococcal infections are a serious public health issue in the world and the increase in the incidence of antibiotic resistant clones and shortcomings of the existing vaccines means that we urgently need to study the biology of this pathogen in order to identify novel therapeutics and vaccine targets.

1.3. Treatment and prevention (antibiotics and vaccines)

β-lactam antibiotics are used for the treatment of pneumococcal infections. The β-lactam family includes a wide spectrum of antibiotics such as benzylpenicillin (Penicillin G), cephalosporin C, ampicillin, and ceftriaxone, which are known to be bactericidal and inhibit cell wall synthesis (Normark and Normark, 2002; Charpentier and Tuomanen, 2000). Penicillin has been widely used for the treatment of pneumococcal infections, and was very effective against most pneumococcal serotypes, which resulted in reducing the mortality and disease rates of pneumococcal infections (Morton and Swartz, 2004). However, pneumococcal resistance to the β-lactam antibiotics has risen significantly over the years because the pneumococcus acquired the genetic ability to resist antibiotics (Hansman and Bullen, 1967), and due to the excessive usage of penicillin and other β-lactam antibiotics (Hansman *et al.,* 1971), the microbe has produced modified penicillin-binding proteins which reduces binding to β-lactam antibiotics (Dowson *et al.,* 1994; Dahesh *et al.,* 2008). Over the recent years, resistance to penicillin has been increasing significantly. This has been suggested to be

the main reason for increasing rates of infections (Hanna-Wakim *et al.,* 2012). According to the Clinical and Laboratory Standards Institute (CLSI), 46% of *S. pneumoniae* isolates were reported to be resistant to penicillin between 2005 to 2010 in Lebanon (Hanna-Wakim *et al.,* 2012; Araj *et al.,* 1999). Similarly, reports from Saudi Arabia showed that 44% of pneumococcus isolates were penicillin-resistant (Al-Mazrou *et al.,* 2005). In the USA the reports showed that 6-14% of invasive pneumococcal disease cases caused by pneumococcal penicillin-resistant strains in the period between 1998 and 2008 (Hampton *et al.,* 2011). In Europe, 2200 isolates collected from eight countries were used to investigate pneumococcal resistance to a wide spectrum of antibiotics including penicillin. This study found that 25% of the isolates were specifically resistant to penicillin (Reinert *et al.,* 2005). On the other hand, the available reports showed that the resistance to penicillin varies in different regions. For example, in the USA and Europe, the rates of the resistance to penicillin are lower than the rates reported in the Middle East very likely due to the drug prescription rules (Hampton *et al.,* 2011).

Pneumococcal resistance to penicillin is also apparent with other antibiotics used for the treatment of pneumococcal infections such as erythromycin and tetracycline. In Lebanon, studies reported an increase in erythromycin resistance, from 11.5% in 1997 to 29.3% in 2012 (Hanna-Wakim *et al.,* 2012). In Southern Europe and USA erythromycin resistance rates were 37% and 24.9% in 2009, respectively (Li˜nares *et al.,* 2010; CDC, 2010). In addition, Felmingham *et al.* (2004) reported that the pneumococcal resistance to tetracycline has reached 25% in Western Russia and the USA. Pneumococcal resistance to most antibiotics is believed to have risen due to adaptation of the microbe to various environmental changes (Richter and Musher,

2017). This serious problem should be overcome through studies on pneumococcal biology to identify targets to design novel therapies and effective antibiotics.

The pneumococcal antibiotic resistance have become a universal concern. Therefore, new effective vaccines are required to reduce infections. Currently, there are two different types of pneumococcal vaccines that are in use to protect adults from pneumococcal infection. The first one includes polysaccharides from 23 pneumococcal serotypes (PPV23), and the second includes protein-conjugated polysaccharides from 13 serotypes (PCV13) (Fedson *et al.,* 2014). In the 1970s, the first pneumococcal polysaccharide vaccine, 14-valent plain polysaccharide vaccine (PPV14), was licensed for use against pneumococcal pneumonia (PP) and invasive pneumococcal disease in South Africa. However, in 1983 the PPV14 was replaced by PPV23 due to the highest rate of incidence of pneumococcal pneumonia and invasive pneumococcal disease in the elderly over 60 years (Falkenhorst *et al.,* 2017). The PPV23 contains capsular polysaccharides (CPS) of 23 serotypes, including 1, 2, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19A, 20, 22F, 23F and 33F, covering approximately 90% of invasive pneumococcal infections (Ho and Lin, 2005). Surveillance studies reported that PPV23 showed high efficacy in older children and adults but showed a poor response in children aged under two years (Long *et al.*, 2012).

The fact that PPV23 had limitations in its efficacy for children under 2 years and in the elderly was attributed to inclusion of only 23 of the 92 known serotypes (Long *et al.,* 2012). Therefore, in order to augment the efficacy of PPV, the polysaccharide antigens were conjugated with immunogenic carrier proteins diphtheria toxin (Abraham-Van, 2004). In 2000, 7-valent conjugate vaccine (PCV7) was introduced to protect against a wide range of serotypes: 4, 6B, 9V, 14, 18C, 19F, and 23F, which were most

widespread in the United States and Europe. In 2009, PCV7 was developed to PCV10 by adding serotypes 1, 5, and 7F. In 2010, PCV10 was further developed to PCV13 by adding serotypes 3, 6A, and 19A (CDC, 2010). The current observational studies found that the 13-valent vaccine was reported to be immunogenically effective, safe and well-tolerated when given with other vaccines to infants at 2, 3, 4 and 12 months of age (Martinón-Torres *et al.,* 2015). Advisory Committee on Immunization Practices (ACIP) recommends giving PCV13 to all children up to age 59 months not previously vaccinated with PCV7 (CDC, 2010). However, it must be noted that the protection provided by the current pneumococcal vaccines is still serotype specific.

There is no doubt that the increase in rates of invasive pneumococcal disease has been a global concern. The high mortality worldwide caused by pneumonia and emergence of drug-resistant strains and the limitations of the existing vaccines are considered to be real challenges for scientists. Therefore, more biological studies on this microorganism are needed in order to discover effective drug and vaccines targets to reduce the burden of disease.

1.4. Pneumococcal pathogenesis and virulence determinants

*Streptococcus pneumonia*e is a part of the normal flora of the respiratory tract of humans. After colonisation by one of the 91 different serotypes, a newly acquired strain eliminates other competing pneumococcal serotypes and continues to colonise for weeks or months, usually without any negative repercussions (Weinberger *et al.,* 2008). Pneumococcal colonisation of the nasopharynx is considered to be the most important step to initiate the disease. From this niche the pneumococcus can spread to the alveolar space, middle ear, lungs and then to bloodstream or brain, causing otitis media,

pneumonia, septicemia, and meningitis, respectively (McCuller and Tuomanen, 2001) (Figure 1.1).

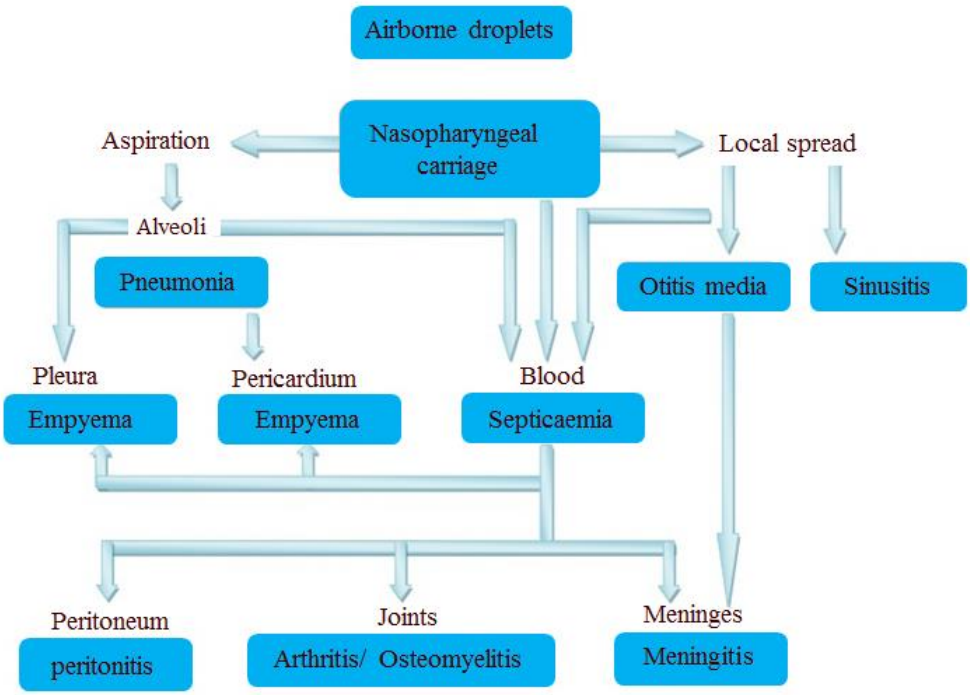


Figure 1.1. An illustration showing the regions in the host that can be infected by *S. pneumoniae*. The progression of the infection represented by light blue arrows. The pneumococcal infections are typically preceded by nasopharyngeal colonies, after this the infection may reach middle ear, causing otitis media and sinusitis, which might also result in meningitis. The infection can spread to the lungs causing pneumonia. Commonly, the microbe may also penetrate into the bloodstream, from where they can disseminate to the whole body (adapted from Eze *et al.*, 2013; Bogaert *et al.*, 2004).

This carrier state permits the organism to maintain its continuous existence within human populations and induces the host innate immune system (Lysenko *et al.*, 2005). The pneumococcus has strategies to enable itself to evade the defense system of its host as other pathogenic bacteria have (Kadioglu *et al.*, 2008). The non-invasive phenotype uses various surface adhesion factors, and immune evasion strategies. In contrast, the invasive phenotype relies on the expression of different virulence factors that affect the ability of *S. pneumoniae* to cause infection and the possibility of person-to-person spread by coughing (Wani *et al.*, 1996; Kadioglu *et al.*, 2008). Non- invasive pneumococcal strains of low virulence can stay circulating in human populations

relying on a persistent carriage of the pneumococcus (Zhang *et al.*, 2007). There are many known virulence factors of the pneumococcus that each contributes individually to the pathogenesis of *S. pneumoniae* (Figure 1.2).

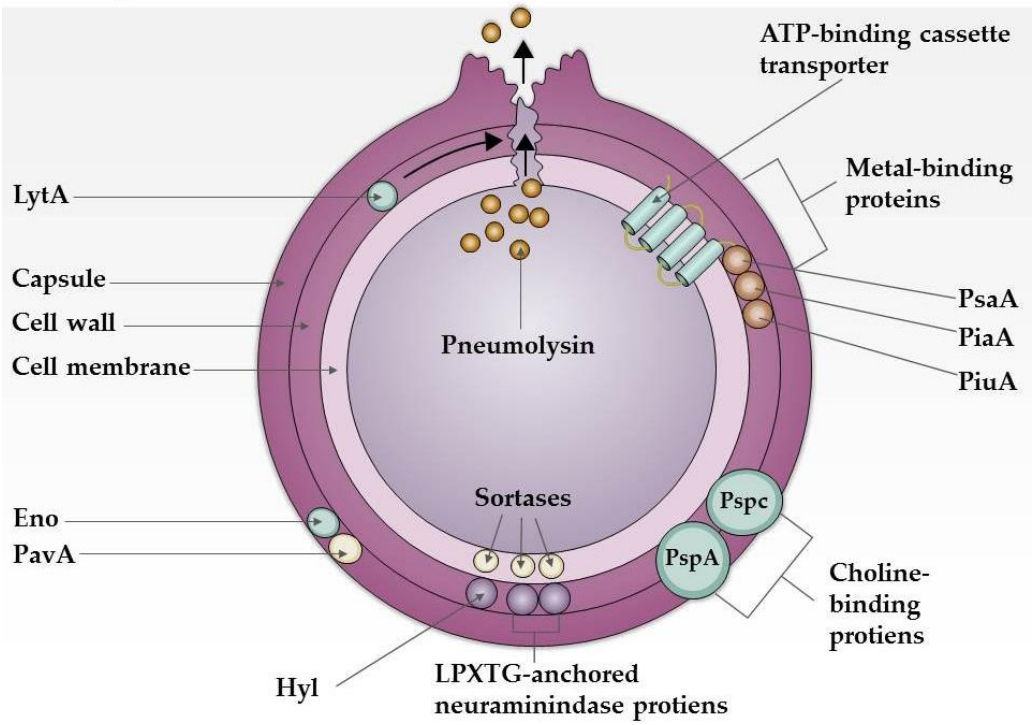


Figure 1.2. Schematic representation of pneumococcal virulence factors. Metal-binding proteins: PsaA pneumococcal surface adhesion A, PiaA pneumococcal iron acquisition A and PiuA pneumococcal iron uptake A; choline-binding proteins: PspC and PspA pneumococcal surface protein C and A respectively; LPXTG-anchored proteins: NanA: neuraminidase A; Hyl: hyaluromate lyase; PavA: pneumococcal adhesion and virulence A; Eno: enolase; LytA: autolysin A. (from Kadioglu *et al.*, 2008).

The polysaccharide capsule (CPS) is considered as the main virulence factor of the pneumococcus because the encapsulated strains are more virulent than unencapsulated pneumococci (Jedrzejas, 2001). Pneumococcal capsule prevents the pathogen from opsonisation by iC3b (Nelson *et al.*, 2007). The formation of the capsule has a crucial role in colonisation, as capsule prevents removal by mucus, limits autolysis, and

reduces the influence of antibiotics (Nelson *et al.,* 2007). There is a strong relationship between the thickness of capsule and the virulence of *S. pneumoniae* among certain CPS serotypes (MacLeod and Krauss, 1950). In addition, the capacity of pneumococci to cause disease varies with differences in CPS serotypes to which those pneumococci belong (Austrian, 1981). These differences in capsular structure may account for differential pneumococcal ability to resist phagocytosis, as well as differences in their ability to trigger a humoral immune response (Hostetter, 1986). Molecular epidemiological analysis showed that the properties specific to clonal types determine the potential of *S. pneumoniae* to cause invasive pneumococcal disease and involvement of capsular serotype in virulence (Sandgren *et al.,* 2005). The transition of *S. pneumoniae* from nasopharyngeal colonisation to invasive disease is considered to be an evolutionary process between the pneumococcus and its host. Therefore, the expression of the main pneumococcal virulence determinants allows pneumococci to adapt to its altered microenvironment, which leads to increase in the severity of the disease. The greater expression of capsule is required for systemic virulence, and the capsular thickness also has an effect on pneumococcal surface structures, such as adhesions (Talbot, 1996). The survival of *S. pneumoniae* in various host environments is directly linked to tight regulation of CPS production at the transcriptional, translational and post-translational level (Jedrzejas *et al.,* 2011). All pneumococcal serotypes have the *cps* locus responsible for capsule synthesis specifically the first four genes (CpsA-D) are common in these serotypes, and these four genes CpsA-D encode for proteins that have an effect on the level of CPS expression (Morona *et al.,* 2000). Despite its importance in pneumococcal colonisation and virulence, the genetic basis of capsule locus regulation is not known in detail.

Thiol-activated toxins pneumolysin (PLY) is another important virulence factor belongs to the cholesterol-dependent cytolysin family that is produced by certain Gram-positive bacteria (Kirkham *et al.,* 2006; Johnston, 1991). Benton *et al.* (1997) demonstrated that PLY is a 53-kDa protein synthesised by almost all pneumococcal clinical isolates, and it is detectable during the late exponential phase of growth. PLY is intracellular protein but recently it has also been shown to be cell wall-associated (Price and Camilli, 2009). Benton *et al.* (1995) suggested that during the first period of bacteremia, PLY plays a crucial role in virulence by preventing the generation of inflammation-based immunity, which results in increasing the exponential net growth of pneumococcus. The interactions between PLY and alveolar epithelial cells or pulmonary endothelial cells result in alveolar oedema and haemorrhage. These interactions cause an alveolar flooding providing nutrients for pneumococcal growth, and therefore allow pneumococcus to penetrate into the pulmonary interstitium and then into the bloodstream (Rubins and Janoff, 1998). The multiple functions of pneumolysin, especially in the early period of pneumococcal infection, are essential to the pneumococcal colonisation (Lawrence *et al.,* 2015).

Several pneumococcal enzymes are also implicated in virulence and pathogenesis such as autolysin, neuraminidase, and hyaluronidase (Jedrzejas, 2001). The autolysin is located on the cell wall of the pneumococcus, and is one of the several enzymes that can degrade bacterial peptidoglycan causing cell lysis. There are several types of pneumococcal autolysins, designated as autolysin A (LytA), autolysin B (LytB) and autolysin C, (LytC) (Garcia *et al.,* 1999; Hendriksen *et al.,* 2009). LytA is functionally the main pneumococcal autolysin contributes to virulence through its role in release of pneumolysin (Martner *et al.,* 2008). Previous studies reported that autolysin contributes

to pneumococcal colonisation and adherence as well as the modulation of meningeal inflammation (Eze *et al.,* 2013; Mitchell, 2000).

Neuraminidase is an enzyme, also known as sialidase, it cleaves N-acetylneuraminic acid from mucin, glycolipids, glycoproteins, and oligosaccharides on host cell surfaces (Rosenfeld *et al.,* 1992; Berry and Paton, 2000). The pneumococcus produces at least three distinct neuraminidase activities: NanA, NanB and NanC. All pneumococcal strains produce NanA, and most also produce NanB. However, only about 50% of strains produce NanC (Pettigrew *et al.,* 2006). Rosenfeld *et al.* (1992) suggested that neuraminidase could promote colonisation by diminishing the mucus viscosity, or by letting cell surface receptors to be directly exposed to *S. pneumoniae* (Berry *et al.,* 1996). NanA differs from NanB and NanC because it contains a C-terminal sequence LPXTG anchorage motif, which can bind to peptidoglycan of the cell wall (Camara *et al.,* 1994). NanA plays an important role in biofilm formation (Berry *et al.,* 1996; Oggioni *et al.,* 2006). Biofilms are a thick layer of surface-adherent microbes embedded in an exopolysaccharide matrix (Cvitkovitch *et al.,* 2003). The ability of pneumococcus to form a biofilm on host mucosal surfaces is clearly being recognised as a critical step in the pathogenesis of pneumococcal disease (Trappetti *et al.,* 2017). Manco *et al.* (2006) showed that NanA and NanB are essential for pneumococci to survive in the respiratory tract and bloodstream.

Hyaluronidase is one of the pneumococcal virulence factors that may aid *S. pneumoniae* to spread and colonise by degrading hexosaminidic linkages of hyaluronic acid, which is a component of the extracellular matrix (Kostyukova *et al.,* 1995; Berry and Paton, 2000). There is a strong relationship between the activity of hyaluronidase and the ability to induce meningitis (Berry and Paton, 2000). Moreover, hyaluronidase may be

responsible for pulmonary inflammation during pneumococcal infection by interaction with proinflammatory cytokines and chemokines (Irwin *et al.,* 1994; Mitchell and Mitchell, 2010).

The pneumococcal cell wall contains surface proteins that are likely to have an important role in virulence. These proteins adhere to the cell wall of the pathogen either by the attachment motif (LPXTG) or by choline-binding proteins (Cbps), which include pneumococcal surface protein A and C (PspA and PspC), via non-covalent interactions (Chhatwal, 2002; Mitchell, 2000). PspA is a lactoferrin-binding that has protein and has a notable effect on innate immunity (Shaper *et al.,* 2004). It is also reported to be responsible for inhibiting the complement system in addition to preventing adhesion of C3 onto the pneumococcal surface (Ren *et al.,* 2003; Kerr *et al.,* 2006). PspC interacts with human polymeric immunoglobin receptor (hpIgR) and this interaction leads to increased colonisation of the nasopharyngeal cells in the early stage of pathogenesis (Eze *et al.,* 2013). Metal-binding proteins include the pneumococcal surface adhesion (PsaA) which is a substrate-binding lipoprotein and plays a role in conferring pneumococcus a resistance to oxidative stress and also plays an active role in adhesion of pneumococci onto the nasopharynx of the host (Mitchell and Mitchell, 2010; Briles *et al.,* 2000). Additionally, pneumococcal iron acquisition A (PiaA) and pneumococcal iron uptake A (PiuA) are suggested to be involved in the pneumococcal virulence and colonisation (Jomaa *et al.,* 2006). Further to the above virulence factors, there are two other surface proteins, known as pneumococcal adhesion and virulence A (PavA) and enolase (Eno), which have roles in surface adhesions (Holmes *et al.,* 2001). PavA and Eno are reported to bind to the extracellular-matrix components fibronectin and plasminogen, respectively (Bergmann *et al.,* 2001). Finally, pili are hair-like structures

that are located on the surface of *S. pneumoniae* and have recently been identified in many, but not all, strains of *S. pneumoniae*. They play an active role in colonisation of the upper respiratory tract and contribute to TNF-α response by the immune system during the pneumococcal invasion (Hirst *et al.,* 2002).

1.5. Involvement of nutrient metabolism in virulence

To understand how *S. pneumoniae* colonises asymptomatically and causes pneumococcal invasive diseases, it is important to understand the way in which the pathogen utilises host sugars as a carbon source for nutrition. In addition, it is also important to reveal how the microbe regulates sugar metabolism, which will be investigated in this study. There is a strong relationship between pneumococcal virulence and efficient sugar acquisition and catabolism (Yesilkaya *et al.,* 2009). It is also well known that *S. pneumoniae* is only isolated from live tissues and there is no evidence for its existence in the environment (Tettelin *et al.,* 2001, Marion *et al.,* 2011). The importance of sugar metabolism is reflected in the pneumococcal genome as one third of the pneumococcal genes encode proteins involved in sugar metabolism, including sugar cleavage, transport, and metabolism (Terra *et al.,* 2010; Marion *et al.,* 2012). The pneumococcus is usually present in the multi-sugar environment of the host, and it can utilise at least 32 different sugars such as glucosides, galactosides and polysaccharides (Bidossi *et al.,* 2012). Thus, permit pneumococcus to utilise the available sugar in different host tissues (Bidossi *et al.,* 2012; Terra *et al.,* 2015). The pneumococcus relies on human host glycoconjugates for nutritional requirements due to the low concentrations of free carbohydrates in the respiratory tract (King, 2010; Philips *et al.,* 2003). For example, free glucose is scarce and below the pneumococcal requirements, in contrast to galactose which is abundant, in the respiratory tract (Philips

et al., 2003). Therefore, the pneumococcus utilises glycoproteins as the main source of carbon for nutritional requirements in the respiratory tract, and galactose plays an important role for *In vivo* survival of pneumococcus.

The pneumococcus is a pathogen that strictly relies on fermentation to obtain energy requirements via glycolysis as the pneumococcal genome lacks the complete set of genes required for respiration (Tettelin *et al.,* 2001; Hoskins *et al.,* 2001). Pneumococcal host-derived glycan utilisation plays an essential role for survival and virulence. Mucin is a high-molecular-weight glycoprotein and a very rich source of galactose, N-acetylgalactosamines, N-acetylglucosamine, mannose, fucose, sulfated sugars, and sialic acid (Terra *et al.,* 2010). During infection of the respiratory tract, *S. pneumoniae* shows a tendency to rely on mucin (Bogaert *et al.,* 2004) as the main source for its nutrition requirements due to the high content of carbohydrates in mucin which makes up to 85% of total dry weight (Rose and Voynow, 2006). Furthermore, mucin is the main component of the mucus that forms a considerable part of the human innate immune system (Terra *et al.,* 2010; Wiggins *et al.,* 2001). Pneumococcal mucin utilisation plays a key role in colonisation of the nasopharynx and invasion of the lungs and inner ear (Manco *et al.,* 2006; Tong *et al.,* 2000).

The importance of pyruvate metabolism has been emphasised in several studies (Yesilkaya *et al.,* 2009, Gaspar *et al.,* 2014). Although the pneumococcus is limited in its ability to utilise sugars only via glycolysis, once pyruvate is formed, there is metabolic flexibility as pyruvate can be used by several competing enzymes such as pyruvate formate lyase (PFL), lactate dehydrogenase (LDH) and pyruvate oxidase (SpxB) (Yesilkaya *et al.,* 2009, Gaspar *et al.,* 2014). The pyruvate formate lyase (PFL) is involved in pneumococcal galactose metabolism and survival. PFL converts pyruvate

to acetyl-CoA under microaerobic and anaerobic conditions (Yesilkaya *et al.,* 2009). Subsequently, acetyl-CoA is converted to ethanol and acetate by acetaldehyde-CoA/alcohol dehydrogenase (ADHE) or alcohol dehydrogenase (ADHA), and phosphotransacetylase (PTA) and acetate kinase (ACKA), respectively. The formation of acetate from acetyl-CoA leads to generation of an ATP molecule. (Yesilkaya *et al.,* 2009). The deletion of genes responsible for active PFL synthesis (SPD_0235 and SPD_0420) leads to reduction in ATP synthesis (Yesilkaya *et al.,* 2009). In addition, mutation of *ldh* attenuated the pathogen's ability to cause bacteremia and also the mutant failed to cause pneumonia when inoculated intravenously (Gaspar *et al.* 2014). Pyruvate oxidase (SpxB) converts pyruvate to acetyl phosphate and CO_2. SpxB is important for efficiently sensing and adapting to various environmental conditions (Spellerberg *et al.,* 1996). The mutation of *spxB* led to a reduction in virulence due to decrease in acetyl phosphate levels. For example, the deletion of *spxB* led to a reduction in ability to compete with other respiratory colonisers, decrease in adhesion proteins, and capsule production (Regev-Yochay *et al.,* 2007; Spellerberg *et al.,* 1996; Carvalho *et al.,* 2013).

Some of other pneumococcal enzymes worth mentioning here due to their importance for pneumococcal nutrient release *In vivo*. The pneumococcus encodes at least 10 extracellular glycosidases (sugar-cleaving enzymes). These glycosidases can modify many glycoproteins including N- and O-linked glycoproteins (King, 2010). O-glycosidase (Eng) and β-galactosidase C (BgaC) have ability to cleave O-linked glycans (Terra *et al.,* 2010, Jeong *et al.,* 2009), while N-linked glycans can be cleaved by β-galactosidase A (BgaA), neuraminidase A (NanA), neuraminidase B (NanB), neuraminidase C (NanC) and N-acetylglucosaminidase (StrH) (Burnaugh *et al.,* 2008; King *et al.,* 2006). This cleavage is processed in sequential pattern and requires the

initial removal of terminal sialic acid by NanA (King *et al.,* 2006). The sequential cleavage of sugars can be affected by any modifications in glycoproteins, such as acetylation (Kahya *et al.,* 2017).

1.6. Pneumococcal sugar transporters

Sugar transport systems take up sugars for catabolism after they are released by the glycosidases. One-third of the transporters in the pneumococcal genome are specific for sugar uptake. The multiplicity of transporters emphasises the importance of sugar metabolism (Terra *et al.,* 2015; Chimalapati *et al.,* 2012).

In *S. pneumoniae*, three different types of carbohydrate uptake systems have been identified, including the primary active transporters (ABC transporters), secondary active transporters (Figure 1.3) and the phosphoenolpyruvate (PEP): carbohydrate phosphotransferase system (PTS) (Figure 1.4). ABC systems utilise the energy of ATP hydrolysis to a large variety of biological activities. The ABC transporters take up carbohydrate, which is intracellularly hydrolysed to ADP and inorganic phosphate (Pi). These systems consist of the substrate-binding proteins (SBP), which are localised outside the cell wall (Davidson *et al.*, 2008; Schneider, 2001). The ABC systems also include two transmembranes (TM), which provide a pathway across the membrane for the entry of the substrate through the SBP. In addition to transmembranes, the ABC systems also contain two nucleotide-binding sites (NBDs) associated with the inner face of the membrane, and provide energy for activating the transport via ATP hydrolysis (Davidson *et al.*, 2008). The secondary active transporters couple the translocation of the carbohydrate to an electrochemical gradient (Davidson *et al.,* 2008). It consists of three units, (i) the uniport systems stimulate the translocation of one solute across cytoplasmic membrane; (ii) symport systems couple the transport of two or more

solutes in the same direction and (iii) antiport systems couple the transport of one solute in to the movement of another solute in the reverse direction (Konings *et al.*, 1994).

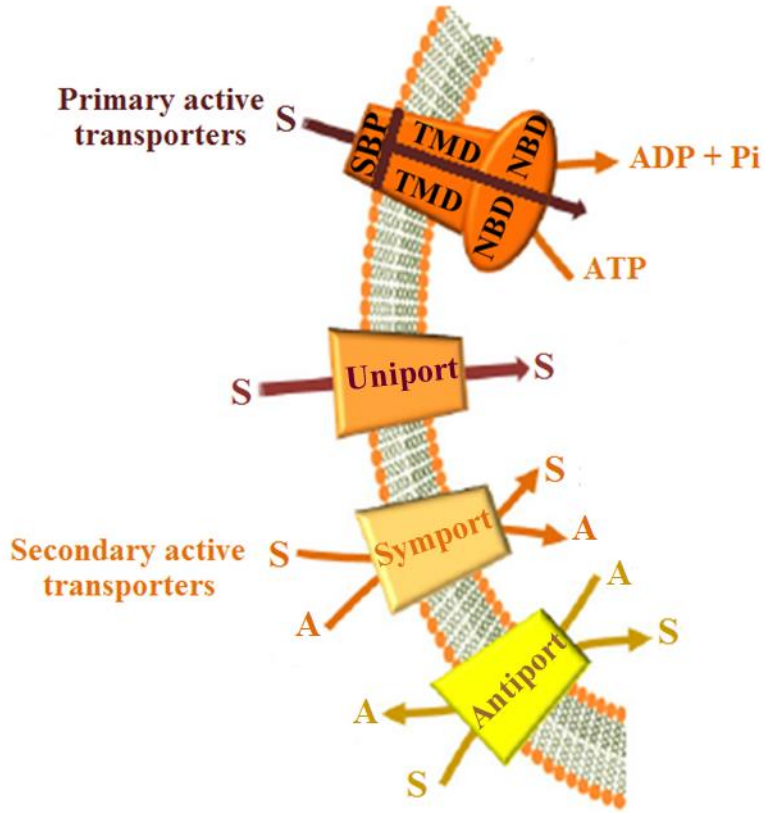


Figure 1.3. Schematic illustration represents the primary and secondary active transporters in *S. pneumoniae*. ATP hydrolysis for primary active transporters; translocation mechanisms of secondary active transporters are classified to uniport, symport and antiport systems. (A) indicates a generic solute and (S) indicates a generic sugar. The figure is designed based on (adapted from Scalise *et al.*, 2013; Davidson *et al.*, 2008).

The pneumococcal genome contains 15-20 phosphoenolpyruvate-dependent phosphotransferase systems (PEP: PTS) and in total 21 PTS systems have been identified in analysed strains (Bidossi *et al.*, 2012). This system is involved in the transport and phosphorylation of a wide range of carbohydrates (Postma *et al.*, 1993). PTS systems consist of histidine phospho carrier proteins (HPr) and enzymes II (Ells)

(Figure 1.5). HPr is a cytoplasmic protein with high solubility and involved in phosphorylation of all PTS carbohydrates used by the pathogen, while Ell is carbohydrate specific, and has three domains: A, B and C (Price *et al.*, 2012). Once His-15 phosphorylated by the PEP and enzyme I, the HPr protein serves as a phospho carrier for sugar PTSs. After phosphorylation, the HPr-His-P also participates in regulation of carbohydrate metabolism through transforming phosphoryl group to a sugar-specific EIIA (Price *et al.*, 2012). The phosphoryl group is ultimately transferred from EIIA to the carbohydrate bound to EIIBC resulting in sugar-6P (Wright and Axelsson, 2012).

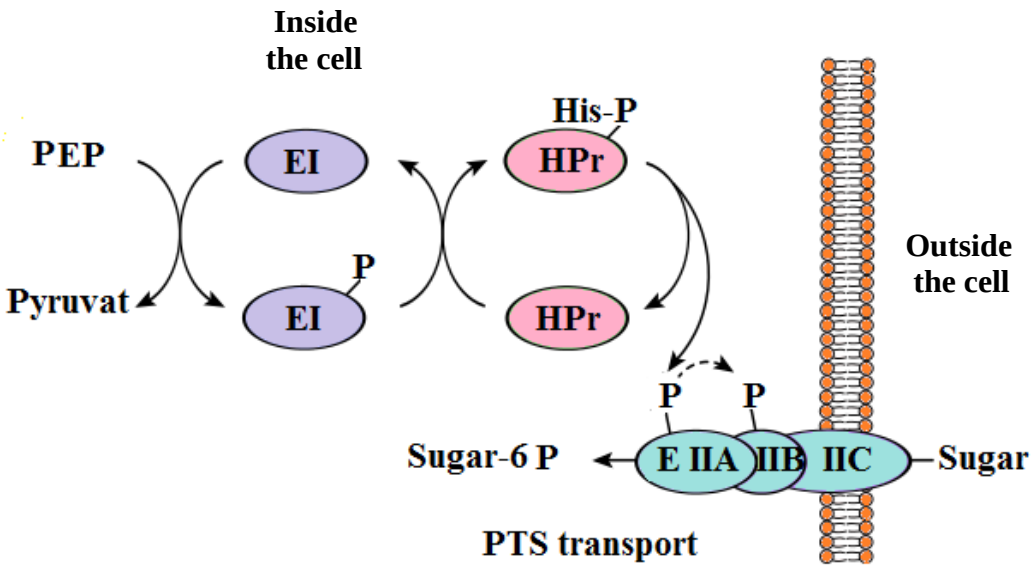


Figure 1.4. Schematic illustration represents the phosphoenolpyruvate-dependent phosphotransferase system in *S. pneumoniae*. (PEP) phosphoenolpyruvate; (EI) enzyme I; (PTS) phosphotransferase system; (Hpr) histidine phosphocarrier protein; (EIIA) enzyme II A; (EIIBC) enzyme II BC. His-15 phosphorylated by the PEP and enzyme I, the HPr protein serves as a phosphocarrier for sugar PTSs. The HPr-His-P regulates carbohydrate metabolism through transforming phosphoryl group to a sugar-specific EIIA resulting in sugar-6P. The figure is designed based on Clore and Venditti (2013) and Wright and Axelsson (2012).

1.7. Pneumococcal metabolism of galactose

Galactose is a very important sugar for pneumococcal biology in the respiratory tract where the microbe colonises and disseminates to cause serious diseases, because it forms a large percentage of mucin sugar content, up to 46%, in the form of galactose and N-acetylgalactosamines (Terra *et al.,* 2010). The pneumococcus also has the ability to cleave galactose from host glycans through the activity of β-galactosidase encoded by at least two genes, *bgaA* and *bgaC* (Terra *et al.,* 2010; King *et al.,* 2010). The presence of two genes for β-galactosidase activity reflects the evolutionary importance of galactose for the pneumococcus. Inactivation of either *bgaA* or *bgaC* was shown to adversely affect pneumococcal attachment (Terra *et al.,* 2010; King *et al.,* 2010). Due to abundance of galactose in mucosal surfaces, and its fundamental impact on the pneumococcal biology, a detailed understanding of pneumococcal galactose utilisation is warranted.

The pneumococcus is able to catabolise galactose, however, its catabolism differs from that of glucose. Glucose is the most preferred sugar utilised by lactic acid bacteria (LAB) for energy requirements, and is converted to pyruvate via the Embden-Meyerhof-Parnas (EMP) pathway or glycolysis. Glucose is either internalised and simultaneously phosphorylated by the PTS systems and then fed into glycolysis or imported by the enzyme glucose permease (GlcU) and phosphorylated by glucose kinase. After phosphorylation, glucose is converted to glucose-6- phosphate (G6P) (Figure 1.5A). G6P is converted to fructose-6- phosphate (F6P) by phosphoglucose isomerase (PGI), which is then phosphorylated by phosphofructokinase (PFK) to fructose-1,6-bisphosphate (FBP) (Postma *et al.,* 1993). The resulting sugar phosphate compounds enter into the glycolytic pathway in which these compounds are cleaved by fructose-1,6-bisphosphate aldolase (FBP) into dihydroxyacetone phosphate (DHAP)

and glyceraldehyde-3-phosphate (GAP) (Figure 1.5B). GAP (and DHAP via GAP) is simultaneously converted to pyruvate through a sequential metabolic process involving substrate-level phosphorylation. GAP is converted to 1,3-bisphosphoglycerate (1,3PG) by GAP dehydrogenase (GAPH), which is then converted to 3-phosphoglycerate (3PG) by phosphoglycerate kinase (PGK). The 3PG is converted to 2-phosphoglycerate (2GP) by phosphoglucrate mutase (PMG). Enolase converts 2GP to phosphenolpyruvate (PEP). Finally the phosphenolpyruvate is converted to pyruvate by phosphoglycerate kinase (PGK), generating 2 ATP (Figure 1.5C) (Wright and Axelsson, 2012).

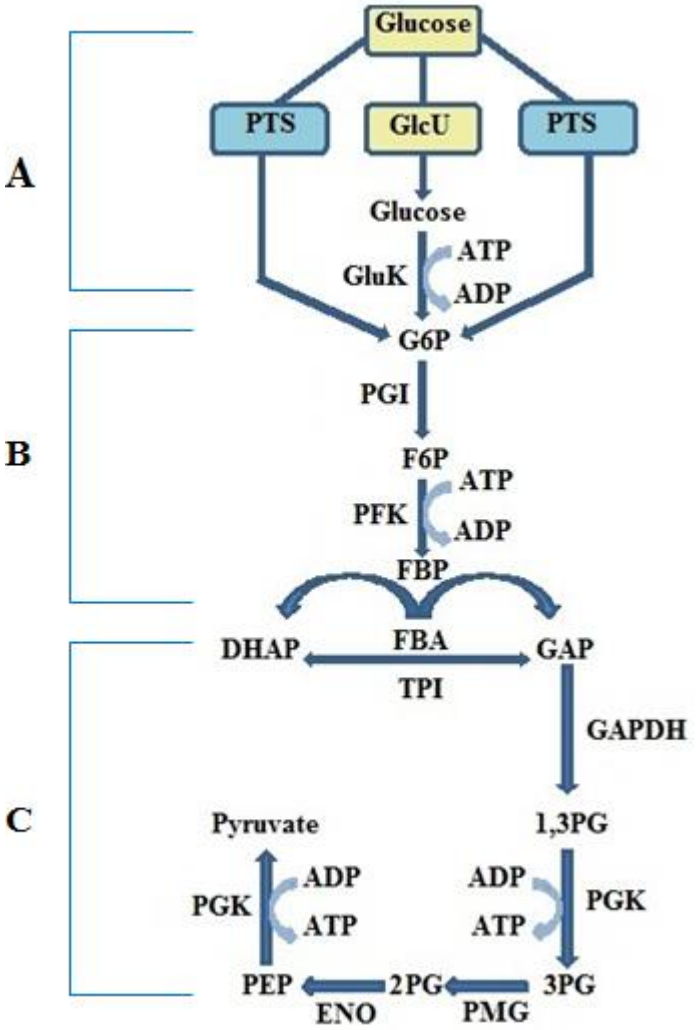


Figure 1.5. Catabolism of glucose via the EMP pathway in lactic acid bacteria. (A) Glucose is either internalised and phosphorylated by PTS systems and then fed into glycolysis or imported by the enzyme glucose permease (GlcU) and phosphorylated by glucose kinase. (B) Glucose-6-phosphate (G6P) is converted to fructose-1,6-bisphosphate (FBP) in sequential pattern. (C) GAP is converted to pyruvate through a sequential metabolic process involving substrate-level phosphorylation. Enzymes involved in the reactions: GluK, glucose kinase; PGI, phosphoglucose isomerase; PFK, phosphofructokinase; FBA, fructose-bisphosphate aldolase; TPI, triosephosphate isomerase; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; PGK, phosphoglycerate kinase; PMG, phosphoglycerate mutase; ENO, enolase; PK, pyruvate kinase. Abbreviations: G6P, Glucose-6-phosphate; F6P, fructose-6-phosphate; FBP, fructose-1,6-bisphosphate; DHAP, dihydroxyacetone phosphate; GAP, glyceraldehyde-3-phosphate; 1,3PG, 1,3-bisphosphoglycerate; 3PG, 3-phosphoglycerate; 2PG, 2-phosphoglycerate; and PEP, phosphoenolpyruvate (adapted from Price *et al.*, 2012; Wright and Axelsson 2012).

On the other hand, the catabolism of galactose differs from that of glucose in lactic acid bacteria (LAB) (Figure 1.6). Once galactose is utilised, it can be metabolised either by the Leloir or tagatose pathways (Paixão *et al.,* 2015). It is known that both pathways are active in *S. pneumoniae*. In the Leloir pathway, after transport, the precise nature of the transporter is currently unknown due to the functional redundancy among the pneumococcal transporters (Bidossi *et al.,* 2012). Galactose is taken up by the specific galactose permease (GalP) and converted to α-galactose by galactose mutarotase (GalM) (Price *et al.,* 2012). The α-galactose is phosphorylated by galactokinase (GalK) to form α-galactose-1-phosphate (α-Gal1P) from which α-glucose-1-phosphate (α-G1P) is formed by galactose-1-phosphate uridylyltransferase/UDP-glucose 4-epimerase (GalT/GalE). From this, glucose-6-phosphate can then be formed by phosphoglucomutase (PGM), which is then fed into glycolysis (Price *et al.*, 2012; Paixão *et al*., 2015). On the other hand, in the tagatose pathway, galactose is converted to tagatose-6-phosphate (T6P) by galactose-6-phosphate isomerase (Gal6P) and subsequently to tagatose 1,6-disphosphate (TDP) by tagatosephosphate kinase (TPK). TDP is converted by tagatose-dis-phosphate aldolase (TDA) to DHAP and GAP, which can then be fed into glycolysis (Price *et al.*, 2012; Paixão *et al*., 2015).

The Leloir and tagatose pathways have specific genes that encode for active enzymes in each pathway. The genes that encode for the enzymes for catabolism in Leloir pathway include *galM*, *galK*, *galT*, *galE*, and *pgm*, encoding for aldose 1-epimerase, galactokinase, galactose-1-phosphate uridylyltransferase, UDP-glucose 4-epimerase, and phosphoglucomutase, respectively, (Paixão *et al.,* 2015). The Leloir pathway genes are dispersed in the pneumococcal genome (Afzal *et al.,* 2014). In the tagatose pathway, *lacAB*, *lacC*, and *lacD*, code for galactose 6-phosphate isomerase, tagatose 6-phosphate

kinase, and tagatose 1,6-diphosphate aldolase, respectively, (Neves *et al.*, 2010), and the tagatose pathway are organised in one operon in genome (Grossiord *et al.*, 1998).

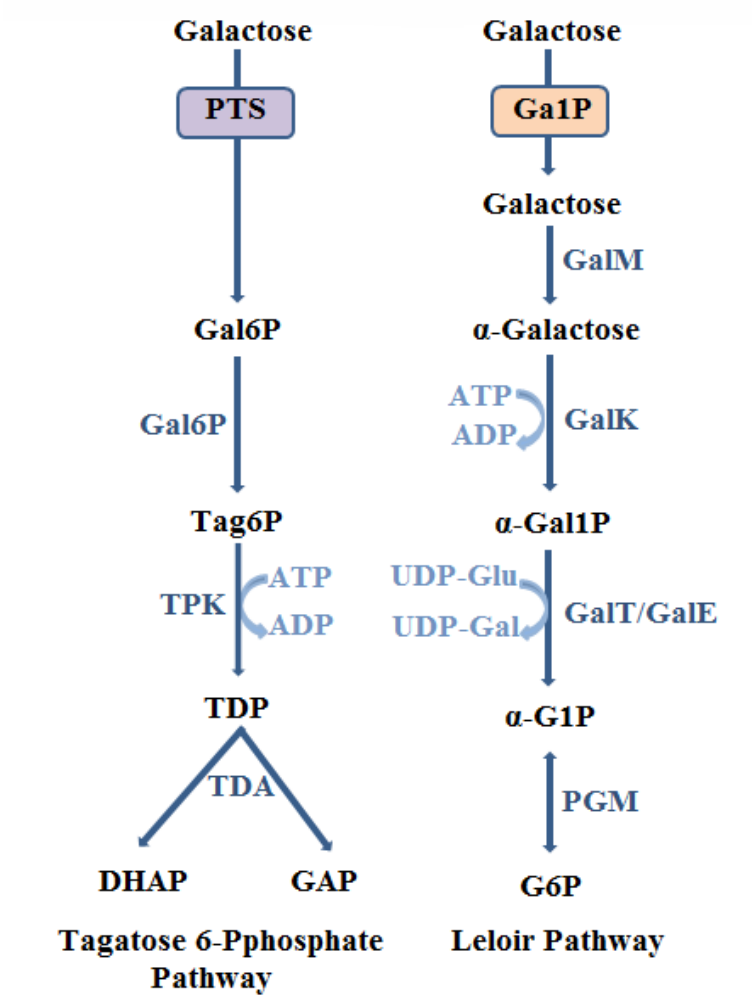


Figure 1.6. Catabolism of galactose via the tagatose and Leloir pathways in lactic acid bacteria. Enzymes involved in the reactions: Gal6P, galactose-6-phosphate isomerase; TPK, tagatose 6-phosphate kinase; TDA, tagatose-dis-phosphate aldolase; GalM, aldose 1-epimerase; GalK, galactokinase; GalT, galactose-1-phosphate uridylyltransferase; GalE, UDP-glucose 4epimerase; PGM, phosphoglucomutase. Abbreviations: Gal6P, galactose-6-phosphate; T6P, tagatose-6-phosphate; TDP, tagatose 1,6-disphosphate; DHAP, dihydroxyacetone phosphate; GAP, glyceraldehyde-3-phosphate; α-Gal1P, α-galactose-1-phosphate; α-G1P, α-glucose-1-phosphate; G6P, glucose-6-phosphate, Glu, glucose; Gal, galactose (adapted from Price *et al.*, 2012; Wright and Axelsson 2012; Paixão *et al.*, 2015).

A recent study by Paixão *et al.* (2015) showed that both pathways are important for galactose utilisation. Inactivation of either Leloir or tagatose pathways resulted in an extended lag phase (18 h), or total elimination of the pneumococcal growth in CDM supplemented with galactose as sole carbon source, respectively. Moreover, the same authors also demonstrated elegantly that *galK* and *lacD* mutation singly or in the same genetic background resulted in significant attenuation in pneumococcal virulence and colonisation, while the mutation of key genes of mannose or N-acetylglucosamine did not affect the colonisation and virulence properties of *S. pneumoniae*. This result clearly underlies the importance of galactose utilisation for *In vivo* survival of the microbe. In addition to pneumococcal energetics, the galactose catabolism generates important mediators of fatty acid and capsule synthesis (Paixão *et al.,* 2015). For example, acetyl-CoA, which can be produced through PFL activity, is an important precursor of fatty acid biosynthesis (Yesilkaya *et al.,* 2009). In addition, galactose plays an essential role in formation of essential virulence determinant the polysaccharide capsule. Eighteen different sugars are involved in the formation of 90 pneumococcal capsular serotypes, and interestingly, nearly all structurally characterised capsule types contain either galactose or the sugars synthesised from the intermediates formed during galactose catabolism (Carvalho *et al.,* 2011; Paixão *et al.,* 2015). For example, α-G1P, UDP-Gal, UDP-Glc produced through the Leloir pathway are required for the synthesis of capsule sugars rhamnose, galactofuranose, and glucuronic acid, respectively (Jansson *et al.*, 1988; Bonofiglio *et al.,* 2005). Hence, it is important to study the regulation of galactose metabolism, which provides insight into pneumococcal survival *In vivo*.

1.8. Transcriptional regulatory systems

The coordinated expression of pneumococcal genes *In vivo* is crucial. *S. pneumoniae* colonises initially the nasopharynx and has the ability to spread to the lower respiratory tract to cause various diseases including pneumonia, bacteraemia, otitis media, and meningitis. In different tissues, the pneumococcus encounters different temperature ranges, sugar composition, and oxygen concentrations. As presented above, galactose has a significant role in pneumococcal metabolism and virulence. However, the genetic mechanisms important for modulation of galactose metabolism are not known in detail. The pneumococcus has a wide spectrum of transcriptional regulators, which allow the microbe to adapt differing environmental conditions.

Environmental adaptation in Gram-positive bacteria is mediated mainly either by two-component systems (TCS) (Standish *et al.*, 2007) or stand-alone transcription regulators (McIver, 2009). Two-component systems (TCS) act as a stimuli-response system, and consists of a membrane-bound sensor protein histidine kinase (HK) and a cytoplasmic cognate response regulator (RR), both regulate gene expression in response to changes in environmental conditions (Hoch, 2000). The kinase domain of the HK sensor protein is activated and able to autophosphorylate the conserved histidine residue upon receiving a specific external stimulus. After that, the phosphate group is moved by the HK to a conserved aspartate residue in its cognate RR (Figure 1.7). This leads to changes in the RR conformation and the RR interacts with its target genes regulating gene expression or protein function. Generally, the RR has a DNA-binding feature and is responsible for regulating gene expression. Two-component systems are widely found in most bacterial species and can modify diverse cellular responses including osmoregulation, chemotaxis, sporulation, photosynthesis and pathogenicity (Hoch, 2000; Stock *et al.*, 2000).

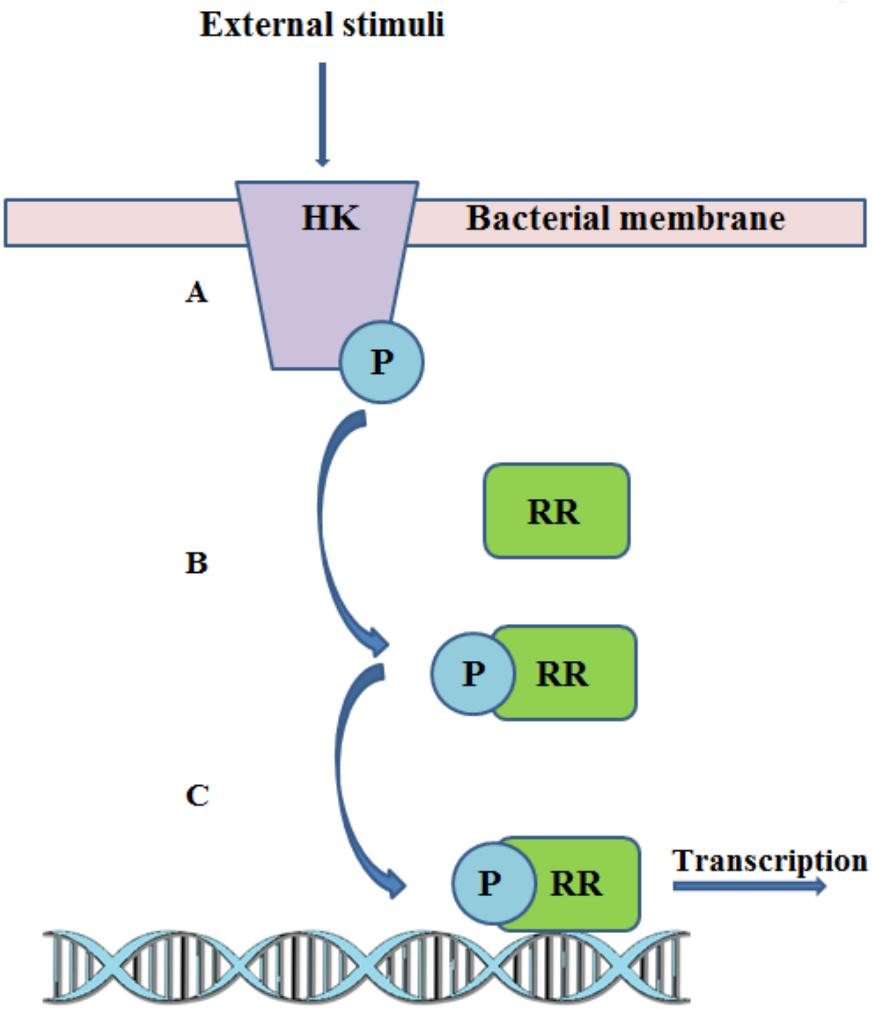


Figure 1.7. A diagram showing bacterial two-component signal transduction. (A) Following exposure to an environmental stimulus, cellular ATP is used by HK to autophosphorylate a conserved histidine residue. (B) The process includes transportation of the phosphoryl group from the HK to an aspartate residue on a cognate RR protein, found in the cytoplasm. (C) The phosphorylated RR protein is able to control gene regulation by binding to the promoter region upstream of its target genes (adapted from Paterson *et al.*, 2006).

In *S. pneumoniae*, there are 13 two-component regulatory systems (TCS) in addition to one identified orphan response regulator (Paterson *et al.*, 2006; Lange *et al.*, 1999; Throup *et al.*, 2000) (Table 1.1). These systems have various roles, from oxidative stress resistance, competence, to virulence. The CiaR/CiaH system is the first TCS identified in *S. pneumoniae* (Guenzi *et al.*, 1994). The CiaRH two component system

was suggested to have a key role in regulating 5 non-coding sRNAs that take part in control of competence and autolysis resistance. (Halfmann *et al.,* 2007; Laux *et al.,* 2015). It was reported that CiaRH is involved in the pneumococcal colonisation in the host tissues and found to have a role in the regulation of many genes, including those involved in temperature adaptation, competence development, autolysis, and bacteriocin production (Ibrahim *et al.,* 2004; Halfmann *et al.,* 2011). The ComDE system is another pneumococcal two-component signal-transducing system which was received considerable attention and widely studied (Claverys *et al.,* 2002). Natural competence in *S. pneumoniae* is a transient physiological state, which is initiated by a secreted pheromone known as the competence-stimulating peptide (CSP). The CSP stimulates competence when its concentration in the medium reaches a threshold level. Competence is also linked to stress response, and environmental adaptation (Claverys *et al.,* 2002; Steinmoen *et al.,* 2002).

Table 1.1. Pneumococcal TCSs and their regulatory functions.

TCSs	Alternative names	Regulatory function
01	480	virulence
02	*vic*, *micAB*, *yycFG*, 492	competence, virulence, redox/energy sensing
03	474	?
04	*pnpRS*, 481	virulence, resistance to oxidative stress, phosphate sensing
05	*ciaRH*, 494	competence, virulence, autolysis, bacteriocins production, resistance to β-lactam antibiotics,
06	478	virulence, adhesion
07	539	virulence
08	484	virulence, cellobios metabolism
09	*zmpSR*, 488	competence, virulence, carbohydrate intake
10	*vncRS*, 491	autolysis
11	479	?
12	*comDE*, 498	virulence, competence, quorum sensing system, autolysis
13	*blpRH*, 486	virulence, quorum sensing system, the production of bacteriocins
Orphan RR	*ritR*, 489	virulence, iron intake

1.9. Quorum sensing systems

Quorum sensing (QS) systems allow bacteria to communicate with each other to coordinate their behavior through the production and detection of signaling molecules (Parashar *et al.,* 2015). QS systems are essentially TCS but their operation differs from that of traditional TCS described above. The QS systems of Gram-negative bacteria use small molecules (homoserine lactones), while Gram-positive bacteria utilise processed peptides. In Gram positive QS systems, pheromone or the signal molecule is a peptide. This pheromone molecule is produced as pre-peptide, and exported outside the cell, when it reaches to sufficient concentration, it is imported back into the bacterial cell via ABC transporters and binds to its cytoplasmic receptor protein to orchestrate the expression of target genes including, the gene coding for the pheromone (Fleuchot *et al.,* 2011). The examples of this family of QS include RRNPP (Rgg, Rap, NprR, PlcR, and PrgX) family QS systems and they are discussed in detail below.

In certain QS systems, the pheromone does not interact with the cytoplasmic receptor but interacts with the sensor histidine kinase. The pneumococcal competence stimulating peptide belongs to the double glycine (Gly–Gly) peptide family. The characteristic of this group is that they have a cleavable N-terminal domain with a double glycine motif at the cleavage site (Havarstein *et al*., 1995a). The pro-peptide is processed and released by an ATP-binding cassette (ABC) transporter, and it is cleaved right after the double glycine (Havarstein *et al*., 1995b). The secreted peptides, then, bind the input domain of surface-exposed histidine kinases from a two component system (TCS) (Pestova *et al*., 1996; de Saizieu *et al*., 2000).

One of the widely distributed QS systems in Gram positive bacteria belong to the RRNPP family proteins. They are found in bacilli, streptococci, or enterococci. The genes that often encode for the RRNPP (Rgg, Rap, NprR, PlcR, and PrgX) family cytoplasmic proteins and their signaling pro-peptides are located in the chromosomes of bacteria or in the plasmid (Grenha *et al.,* 2013; Wilkening *et al.,* 2017). They are known to contain C-termenal tetratricopeptide repeats (TPR), which play a role in the peptide interaction (Declerck *et al.,* 2007; Parashar *et al.,* 2015).

Core and Perego (2003) demonstrated that there are 11 Rap phosphatases in *B. subtilis* which have six tetratricopeptide repeat (TPR) domains, which promote protein-protein interactions. Rap phosphatases in *B. subtilis* play a role in the sporulation (Jiang *et al.,* 2000). The Rap proteins are approximately 375 amino acids in length and they appear to have greater than 25% identity among them (Pottathil and Lazazzera, 2003). It was shown that RapA, RapB, and RapE act as negative regulators of the phosphorelay signal transduction system that initiates sporulation by enhancing dephosphorylation of the response regulator SpoOF, whereas RapC controls competence development via modulating the activity of response regulator and transcriptional factor ComA (Pottathil and Lazazzera, 2003). The Phr peptide, required for activation of Rap phosphatases, has a signal sequence for export. The secreted peptide is processed by proteolysis to mature signalling peptide that can be imported by the oligopeptide permease (Opp) (Perego, 1997; Perego *et al.,* 1991). At low cell density, Rap binds to the response regulator for inactivation but at high cell density, the mature Phr peptide binds to Rap to prevent its binding to response SpoOF regulator, which controls gene expression (Rocha-Estrada *et al.,* 2010).

The sex pheromone receptor (PrgX) in *E. faecalis* regulates antibiotic resistance plasmid pCF10 and it contributes to the regulation of conjugative transfer gene expression encoded in the plasmid. In this quorum sensing system, signalling happens between two types of cells, one with the plasmid (donor) and the other without plasmid (recipient) to allow the plasmid donor cells to regulate the expression of conjugation in response to the density of the recipient cell. By using this system *E. faecalis* mediates the conjugation process of the plasmid between donor and recipient cells (Kozlowicz *et al.,* 2006). The transcriptional regulator PrgX binds to two different signaling peptides, cCF10 (present in the bacterial chromosome) which acts as a specific inducer of pCF10 conjugation genes, and the inhibitor iCF10 (present in the plasmid pCF10 in the donor cells) which functions as repressor of pCF10 conjugation gene expression (Kozlowicz *et al.,* 2006). Both donor and recipient cells synthesize cCF10, and most of cCF10 from the donor cell is sequestered by PrgY in its membrane, but the free peptide and iCF10 are transported to the cytoplasm via oligo peptide permease. Then PrgX dimer specifically binds to pCF10 at two promoters in plasmid pCF10. cCF10 competes with iCF10 for binding to PrgX. When cCF10 binds to PrgX, it induces the gene expression while iCF10 binding causes inhibition of conjugation (Rocha-Estrada *et al.,* 2010).

Another quorum sensing system of RRNPP family is NprR, which modulates the neutral protease in *B. subtilis*. There are two alleles of this gene in different *B. subtilis* strains, *nprR1* and *nprR2*. *Bacilli* possessing *nprR1* produces 20 to 50 times less neutral proteases than others carrying *nprR2* (Toma *et al.,* 1986; Uehara *et al.,* 1979). Neutral protease is one of the essential enzymes for textile industry, and for detergents (Koetje *et al.,* 2003; Rocha-Estrada *et al.,* 2010).

PlcR regulators are also a member of RRNPP family quorum sensing systems. PlcR was described initially in *B. thuringiensis* as a positive transcriptional regulator of phosphatidylinositol specific phospholipase C gene (Lereclus *et al.,* 1996). This regulator uses PapR signaling peptide, which is located about 70 bp downstream from *plcR* (Gominet *et al.,* 2001). *papR* encodes for a 48-aa peptide which is excreted and remain intracellular and then transported to the extracellular milieu through Opp system. When the processed PapR is imported back, it binds with PlcR. Then 'activated' PlcR binds to *plcR* box, located in the promoter regions of PlcR regulated genes, controlling their transcription (Gominet *et al.,* 2001; Rocha-Estrada *et al.,* 2010).

In *B. cereus,* PlcR modulates most of extracellular virulence factors. It positively regulates 45 genes coding to produce extracellular proteins such as enterotoxins, haemolysins phospholipases and proteases (Agaisse *et al.,* 1999; Gohar *et al.,* 2008). In addition, PlcR regulates biofilm formation and virulence in several species of *Bacillus*. Recently a PlcR/PapR homolog, known as TprA/PhrA, was studied in *S. pneumoniae*. It was found that this QS system regulates bacteriocin production loci, and is involved in mucin utilisation, neuraminidase expression, and galactose metabolism (Hoover *et al.,* 2015; Motib *et al.,* 2017). In Leicester, it was shown by Motib *et al*. (2017) that this QS system is essential for virulence and colonisation in a mouse model of virulence and colonisation. In addition, the authors demonstrated that the system could be modulated using linear molecularly imprinted polymers, indicating that it can be used as a drug target.

Rgg family transcriptional regulators are also part of RRNPP family regulators. As the hypothesis of this thesis is on the functional characterization of Rggs, I will discuss this family of regulators in detail in section 1.10.

1.10. Rgg family regulators

Rgg-like regulators, also known as GadR, MutR, RopB, first identified in *S. gordonii* as the regulator of the glucosyltransferase gene, are a family of transcription factors that are widely disseminated in low GC Gram-positive bacteria including in *S. pyogenes* (Fleuchot *et al.,* 2011), *L. monocytogenes* (Neely *et al.,* 2003) and *L. lactis* (Sanders *et al.,* 1998). Rgg-like protein has a preserved helix-turn-helix (HTH) motif and several invariant residues in the amino terminus that are essential for binding to the promoter regions of Rgg-regulated genes. Members of Rgg family have distinctive helix-turn-helix motifs (HTH) in their N-terminal domains and C-terminal region (Fleuchot *et al.,* 2011). N-terminal domains are responsible for interactions with target promoter regions while C-terminal region comprised of tetratricopeptide repeats (TPR) which are responsible for peptide binding (Lasarre *et al.,* 2013; Rocha-Estrada *et al.,* 2010). Rgg proteins can act either as activators or repressors of gene expression (Samen *et al.,* 2006). Different strains of same microbial species can contain multiple homologs of *rgg* genes. In *S. pyogenes*, there are four rgg-like genes, *comR*, *ropB*, *rgg2*, and *rgg3* (Chang *et al.,* 2011). Rgg2 and Rgg3 are the most similar of the four *S. pyogenes Rgg* paralogs. In fact, Rgg2 and Rgg3 bind to the peptide pheromones Shp2 (short hydrophobic peptide) and Shp3, respectively (Parashar *et al.,* 2015). There are five homologs of *rgg* genes in *S. pneumoniae* D39 strain (SPD_0144, SPD_0939, SPD_0999, SPD_1518, and SPD_1952). *S. thermophilus* strain LMD-9 has six homologs of rgg genes (Ibrahim *et al.,* 2007a).

Certain Rggs, but not all, are associated with genes coding for short hydrophobic peptides (*shp*). It has been shown in other streptococci that Rgg-Shp pairs are the individual components of quorum sensing systems. Rgg pheromones are classified in

two groups, short hydrophobic peptides (Shps) and the peptides involved in competence pathway (XIPs) (Parashar *et al.,* 2015). The study of Shp amino acid sequences showed that there are three SHP groups. The Shps containing a conserved glutamate and aspartate belong to groups I and II, respectively, (Fleuchot *et al.,* 2011). The studies revealed that group III has seven peptides with amino acid sequences, one is in *S. thermophilus* strain LMD-9 and the others are from six strains of *S. pneumoniae* (Fleuchot *et al.*, 2011).

The Rgg called ComR, associated with the ComS (XIP) family of peptides, controls the triggering of competence in various species (Ibrahim *et al.,* 2007b; Fleuchot *et al.,* 2011). In *S. mutans*, *comS* encodes XIP for *sigX* inducing peptide. ComR is the receptor for XIP, and it is a member of the Rgg transcriptional regulators family (Cook and Federle, 2014). In *S. pyogenes*, both Rgg-XIP and Rgg-Shp regulate various processes such as biofilm formation and induction of a cryptic competence regulon (Chang *et al.,* 2011; Mashburn-Warren *et al.,* 2012). The XIP peptides are different from bacteriocin-like CSP peptides in that XIP do not contain a Gly-Gly motif in the leader sequence (von-Heijne, 1986; Cook and Federle, 2014).

Gram-positive bacteria produce short peptides, also known as pheromones, involved in the signaling and regulation of gene expression in quorum sensing systems (Miller and Bassler, 2001). The Rgg regulators interact directly with their pheromone peptide. This pheromone molecule is produced as pre-peptide, processed by pheromone-specific peptidase (Eep), The peptides are released outside the cells and processed to mature peptides that are internalised through oligopeptide permease (Opp) to bind with their cognate Rggs to orchestrate the expression of target genes including, the gene coding for the pheromone (Figure 1.8) (Chang *et al.,* 2011; Cook *et al.,* 2013; Fleuchot *et al.,*

2011). Associated *rgg* genes are present exclusively in the streptococci genus, including *S. thermophilus*, *S. pneumoniae*, *S. agalactiae*, *S. pyogenes* and *S. mutans* (Kawamura *et al.,* 1995). Almost all streptococcal genomes have one copy of *shp*, while others contain multiple *rgg* homologs, for instance, *S. thermophilus* strain LMD-9 has six copies. The mutant of *S. thermophilus* LMD-9ΔSTER_1357 led to significantly decrease in transcription of the gene coding for Pep1357C in addition to inactivate the synthesis of a short hydrophobic peptide, a transcriptional regulator, or the oligopeptide transport system. This confirms that the transcription of the Pep1357C-encoding gene is controlled by quorum-sensing system (Ibrahim *et al.*, 2007a). *S. pneumoniae* has two *Shp* homologues (*Shp0144* and *Shp0939*), these Shp peptides activate or repress the transcription of target genes to affect bacterial phenotype such as biofilm formation (Zhi *et al*., 2018).

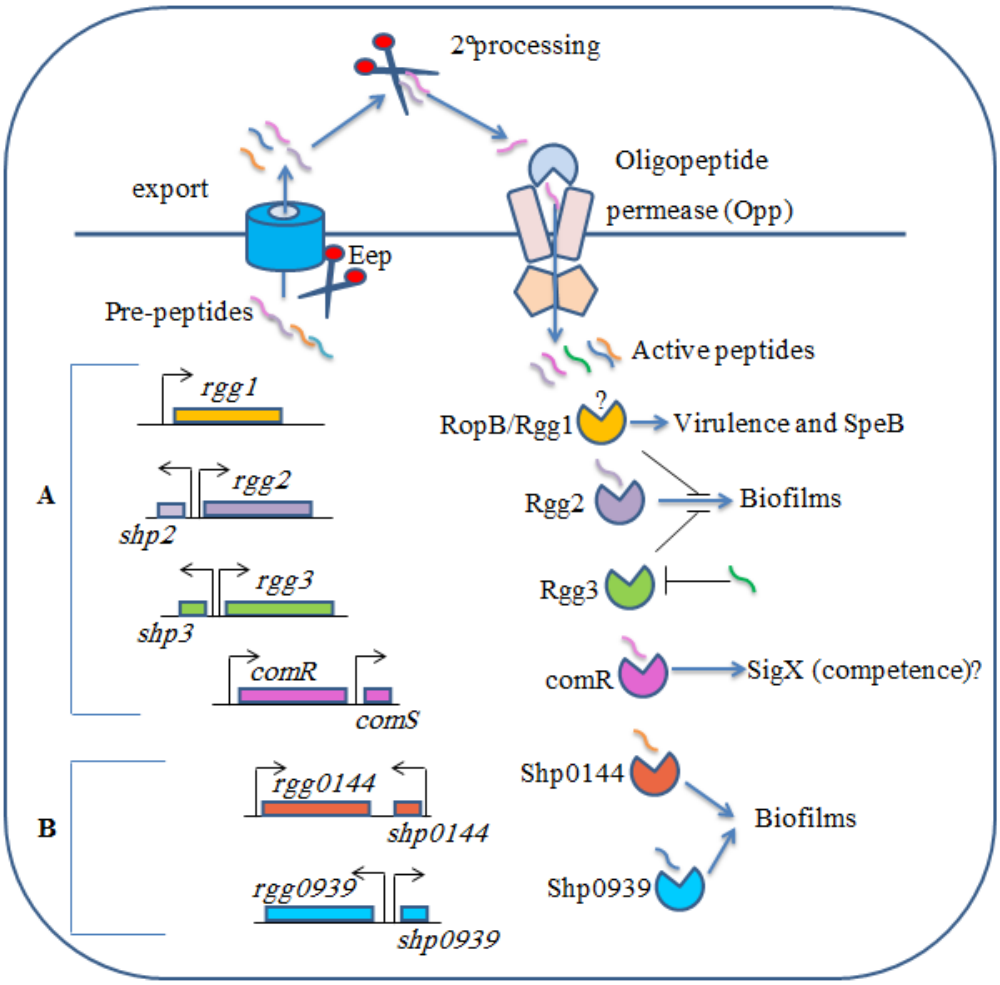


Figure 1.8. A diagram showing proposed model of Rgg-dependent quorum sensing systems in Streptococci. Pre-peptides, encoded by *shp*, are processed by Eep and secreted outside the cell. Secondary proposal processing, produces a mature Shp pheromone that is imported back by Opp. (A) Rgg-dependent quorum sensing in *S. pyogenes*, Shp3 interact with Rgg3, disrupting transcriptional repression of *shp* promoters, and Shp2 interact with Rgg2 to induce transcription of *shp* promoters. The Rgg called ComR, associated with the ComS of peptides, controls the triggering of competence. (B) Proposed model of Rgg-dependent quorum sensing in *S. pneumoniae.* (adapted from Cook *et al*., 2013; Zhi *et al*., 2018).

Rgg proteins play diverse roles in many bacteria: In *S. gordonii*, Rgg acts as a positive regulator of extracellular glucosyltransferase expression. The extracellular glucosyltransferase activity is essential for colonisation (Sanders *et al.*, 1998); (ii) in *L. lactis*, the Rgg-like regulator GadR is necessary for glutamate-dependent acid tolerance

(Sanders *et al.,* 1998); (iii) in *S. mutans*, the Rgg-like protein MutR controls the expression of the mutacin antibiotic (Qi *et al.,* 1999); and (iv) in *S. pyogenes*, Rgg (also known as RopB) is required for the expression of excreted proteins like SpeB (Neely *et al.,* 2003). In *S. agalactiae,* the Shp-associated Rgg is also known RovS, and it has been found to have a role in virulence due to its specificity to host liver and spleen cells using a murine model (Perez-Pascual *et al.,* 2015). A wide range of the Rgg regulators control the transcription of their adjacent genes (Fleuchot *et al.,* 2011) whilst Rggs from *S. pyogenes* NZ131, *S. agalactiae* NEM316 and *S. suis* SS2 act as global regulators because of their control over various genes involved in diverse biological functions (Samen *et al.,* 2006).

Rgg regulators have an important role in microbial biofilm formation. In *S. mutans*, CSP plays a critical role in the development of biofilms. It uses an autolysis pathway to provide nutrients and extracellular DNA for biofilm formation (Perry *et al.,* 2009). In *S. pyogenes* strain NZ131 uses both Rgg2/3 signaling pathway to regulate biofilm formation (Chang *et al.,* 2011). The deletion of both *rgg2* and *rgg3* (*Δrgg2 Δrgg3)* resulted in low biofilm production while the single deletion of *rgg3* led to increase in biofilm mass comparing to the wild-type strain. Additionally, the double mutant *Δrgg2ΔropB* displayed reduction of biofilm formation (Chang *et al.,* 2011).

It was shown that Rggs are global transcriptional regulators and they play important roles in virulence, and mediate non-glucose carbohydrate metabolism in various Gram-positive bacteria (Zheng *et al.,* 2011). Rgg of *S. pyogenes* (serotype M49) modulates the expression of virulence factors and metabolic enzymes that may be crucial for the pathogen's adaptation to different metabolic substrates. As expected, deletion of *rgg* led to significant differences in catabolic substrate preference during growth in complex

and defined media compared to the wild type strain in *S. pyogenes*. It was reported that inactivation of *rgg* abrogates *S. suis* serotype 2 (SS2) virulence significantly in a piglet infection model, and decreased the utilisation of non-glucose carbohydrate, including lactose and maltose (Zheng *et al.*, 2011).

Rggs have been considered as new therapeutic targets. It was proposed that by interfering with their allosteric regulation, bacterial behavior can be modulated to discover antiinfectives. To allow structure based drug design, Parashar *et al.* (2015) characterised the X-ray crystal structures of a *Streptococcus dysgalactiae* Rgg protein alone and in complex with a tight-binding signaling antagonist, the cyclic undecapeptide cyclosporin A. In future, Rgg structure can be utilised to develop antiinfectives targeting Rggs.

1.11. Rggs in *Streptococcus pneumoniae*

In the genome of the well-studied pneumococcal strain, D39, five putative proteins (SPD_0144, SPD_0939, SPD_0999, SPD_1518, and SPD_1952) have significant homology (between 23 to 28% identity and 42 to 48% similarity at the amino acid sequence level) to what is considered as the Rgg prototype; the Rgg from *S. gordonii* (SGO0496) (Figure 1.9) Furthermore, as can be seen in Figure 1.9, within the first 157 amino acids of all of the pneumococcal Rgg homologs there is a predicted helix-turn-helix motif and the three conserved amino acids typical of Rggs ; G8, R15 and W153 of the *S. gordonii* Rgg (SGO0496).

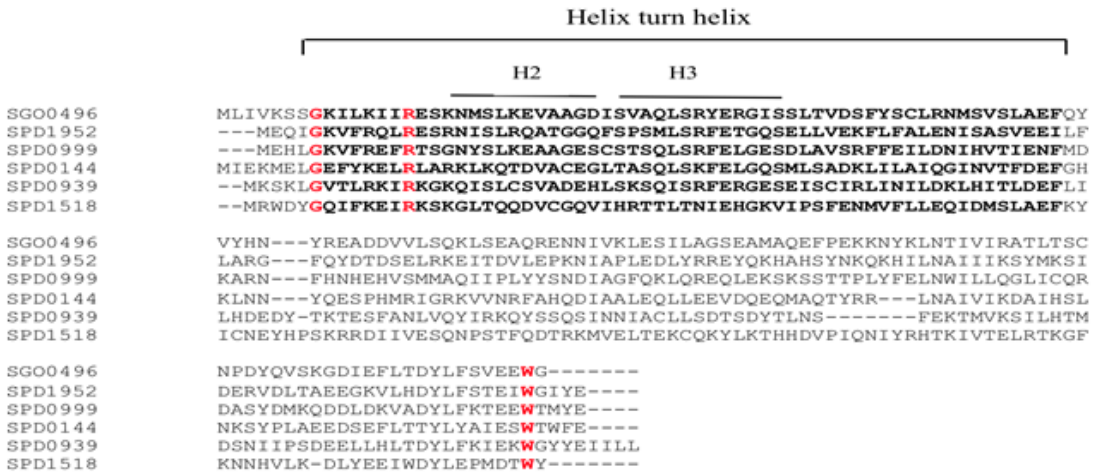


Figure 1.9. Alignment of deduced amino acid sequences of *S. gordonii* SGO0496 and putative pneumococcal Rgg proteins (SPD_1952, SPD_0999, SPD_0144, SPD_0939 and SPD_1518). The primary amino acid sequence of the putative pneumococcal Rgg have been aligned to the *S. gordonii* Rgg. Multiple alignments were performed using ClustalW. The putative helix-turn-helix motif indicated in bold typeface. including helix 2 (H2) and helix 3 (H3) which are predicted to make contact with target DNA. The conserved residues known to be invariant in most Rgg family member are shown in red Accession number: SGO0469 (AAA26968.1), SPD_1952 (ABJ54225.1), SPD_0999 (ABJ55078.1), SPD_0144 (ABJ54913.1), SPD_0939 (ABJ54684.1) and SPD_1518 (ABJ55270.1).

Recently, it was reported that the Rgg/Shp0939 system in *S. pneumoniae* D39 plays a significant role in surface polysaccharide synthesis (Junges *et al.*, 2017). The Rgg/Shp0939 system induces the transcription of *shp* and capsule gene locus, and over expression of the Rgg system increases capsule size and reduces the biofilm formation on lung epithelial cells. In addition, it has been shown that the Rgg family and its cognate short hydrophobic peptide modulates virulence peptide 1 (vp1) in *S. pneumoniae* (Cuevas *et al.*, 2017). *vp1* is a highly expressed Gly-Gly peptide encoding gene in chinchilla middle ear effusions. VP1 enhances the thickness and biomass of biofilm grown on chinchilla middle ear epithelial cells, therefore, it is considered as a novel regulatory peptide for biofilm formation and pneumococcal pathogenesis (Cuevas *et al.*, 2017). In Leicester, it was found by Zhi *et al*. (2018) that Rgg0144/Shp0144

(SPD_0144 locus) and Rgg0939/Shp0939 (SPD_0939 locus) operate as QS systems, are induced by mannose and galactose, and play major roles in colonisation and virulence. Rgg0144 is required for induction of Shp0144, and Rgg0939 is required for induction of Shp0939 in the presence of mannose and galactose. Both Rgg0144 and Rgg0939 acted as a repressor of the *cps* locus on mannose. The results showed that both Rgg0144 and Rgg0939 interacted directly with the putative promoter region of *cps* also shown in Figure 1.10. The regulatory interaction between all four Rggs (*rgg0999*, *rgg1518*, *rgg0144* and *rgg0939*) will be explained in the chapter four.

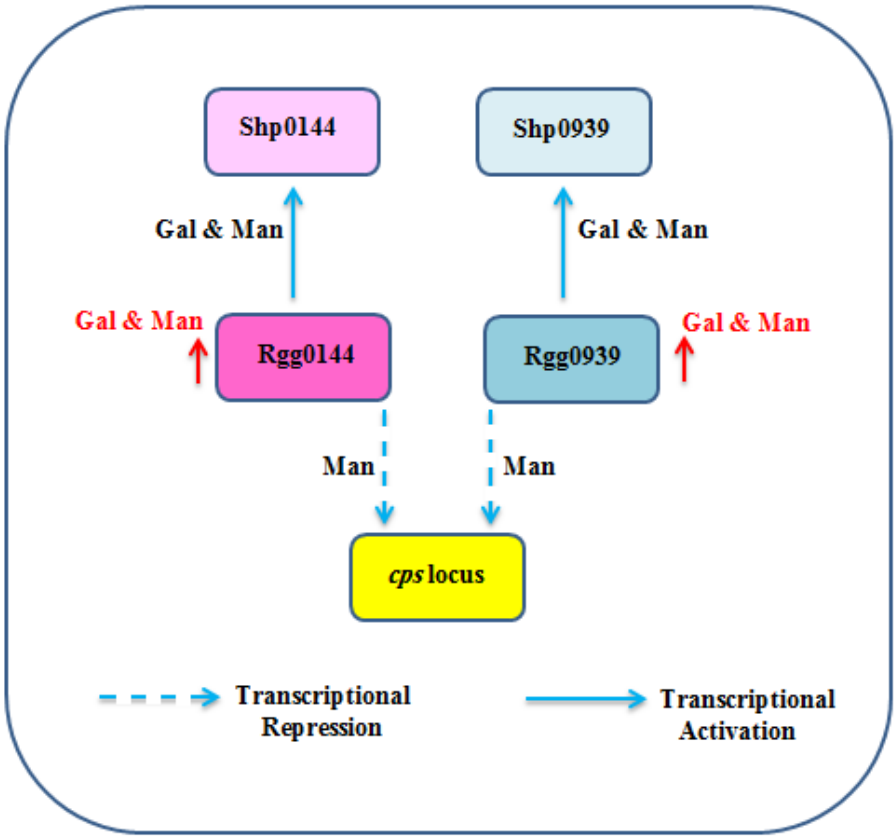


Figure 1.10. Schematic representation of proposed roles for the Rgg0144 and Rgg0939 transcriptional regulators in *S. pneumoniae*. Rgg0144 and Rgg0939 induced by mannose and galactose. Rgg0144 and Rgg0939 activate Shp0144 and Shp0939, respectively. Both these Rggs repress transcription of *cps* locus in the presence of mannose. These two transcriptional regulators interacted directly with the putative promoter region of *cps*. Gal: galactose, Man: mannose. Red arrow indicates increase in expression (adapted from Zhi *et al.*, 2018).

In addition to Rgg/Shp systems, there are also other Rgg homologs for which no pheromone has been identified. These ones are known as stand-alone response regulators (RRs), as they do not have a cognate sensor protein and they regulate gene expression by sensing the changes in intracellular conditions (Kreikemeyer *et al.*, 2003). Stand alone regulators have important regulatory functions in bacteria. McIver, (2009) reported that stand-alone regulators used by GAS are responsible for controlling virulence gene regulons with help of sensory components that have not yet fully understood. The multiple-virulence gene regulator (Mga) was identified as the first stand-alone regulatory system in GAS; it enables bacteria to adapt to environmental changes in the host and it is required for growth (Hondorp and McIver, 2007). The gene encoding Mga has been detected in all sequenced GAS genomes, displaying two divergent alleles (*mga1* and *mga2*) that are linked to infection regions (Bessen *et al.*, 2005). Moreover, Mga plays an essential role in biofilm formation, growth in blood, resistance to phagocytosis, and optimal virulence (Hondorp and McIver, 2007; Hondorp *et al.*, 2013).

Another standalone regulator worth mentioning is catabolite control protein A (CcpA). In some Gram-positive pathogenic bacteria, the virulence factors are regulated by a regulatory process known as carbon catabolic repression CCR (Iyer *et al.*, 2005). There are several common regulators of CCR but the main operative regulator is CcpA (Chauvaux, 1996). In most *Streptococcus* spp, the CcpA has been shown to play an active role in the regulation of catabolic operons and catabolite repression and the CcpA is needed for biofilm formation in *S. mutans* (Wen and Burne, 2002). Giammarinaro *et al.* (2002) showed that CcpA contributes to the virulence in *S. pneumoniae* D39.

Additionally, CodY is another well-studied stand-alone regulator. It is a global nutritional repressor and conserved in low-GC bacteria. CodY plays a regulatory role over operons involved in amino acid metabolism, essentially branched-chain amino acid (BCAA), in addition to proteolysis and peptide uptake and degradation (Hajaj *et al.*, 2017). CodY also plays an important role in regulation of virulence factor synthesis in some Gram-positive bacteria, for example *Bacillus cereus*, *Listeria monocytogenes*, *Clostridium difficile*, *S. aureus*, *S. pyogenes*, and *S. pneumoniae* (Somerville and Proctor, 2009). In *S. pneumoniae*, the inactivation of *codY* clearly leads to significant decrease in adherence to nasopharyngeal cells and minimizing the synthesis of oligopeptide permeases (Hendriksen *et al.*, 2008).

1.12. Aim of the study

The *S. pneumoniae* strain D39 genome contains five *rgg* homologues. The aim of this study was to test the hypothesis that the stand-alone Rggs, SPD_0999 (Rgg999) and SPD_1518 (Rgg1518), play an important role in pneumococcal biology. I selected SPD_0999 and SPD_1518 because their function in *S. pneumoniae* has not been studied yet. The involvement of Rggs in pneumococcal sugar metabolism was investigated by analysis of isogenic mutants in chemically defined medium (CDM) supplemented with different carbon sources and reporter assays. The regulon for each Rgg was determined on mannose and galactose by microarray analysis and the regulatory role among *rgg* genes (*rgg0999*, *rgg1518*, *rgg0144* and *rgg0939*) was investigated. Rgg1518 involvement in capsule synthesis was also evaluated by reporter assays and electromobility shift assays (EMSA). Finally, a mouse model of pneumococcal infection was used to determine the impact of these transcriptional regulators on pneumococcal virulence and nasopharyngeal colonisation.

Chapter 2. Materials & Methods

2.1. Chemicals and biological materials

All chemical in this study were obtained from Sigma, Fisher and New England Biolabs unless otherwise stated.

2.2. The strains of bacteria and plasmids

The *S. pneumoniae* and *Escherichia coli* strains and constructed plasmids used in this study are listed in the Table 2.1, 2.2 and 2.3, respectively.

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

***S. pneumoniae* strains**	**purpose of using**	**Source**
D39	Serotype 2 strain	Laboratory stock
Δrgg0999	D39, SPD0999, Spec resistant	This study
Δrgg1518	D39, SPD1518, Spec resistant	This study
Δrgg0999/1518	SPD0999, Spec resistant and SPD1518, Kan resistant	This study
*Δrgg0999*c	D39, *Rgg0999*, *Δrgg0999*, Spec & Kan resistant	This study
*Δrgg1518*c	D39, *Rgg1518* + *Δrgg1518*, Spec & Kan resistant	This study
Empty pPP2-*lacZ*-wt	D39, pPP2-*lacZ*, Tetra resistant	This study
P*rgg0999-lacZ*-wt	D39, P*rgg0999-lacZ*, Tetra resistant	This study
P*rgg0999-lacZ-*Δ*rgg0999*	Δ*rgg0999*, P*rgg0999-lacZ*, Tetra resistant	This study
P*rgg0999-lacZ-*Δ*rgg1518*	Δ*rgg1518*, P*rgg0999-lacZ*, Tetra resistant	This study
P*rgg0999-lacZ-*Δ*rgg0144*	Δ*rgg0144*, P*rgg0999-lacZ*, Tetra resistant	This study
P*rgg0999-lacZ-*Δ*rgg0939*	Δ*rgg0939*, P*rgg0999-lacZ*, Tetra resistant	This study
P*rgg1518-lacZ*-wt	D39, P*rgg1518-lacZ*, Tetra resistant	This study
P*rgg1518-lacZ-*Δ*rgg0999*	Δ*rgg0999*, P*rgg1518-lacZ*, Tetra resistant	This study
Prgg*1518-lacZ-*Δ*rgg1518*	Δ*rgg1518*, P*rgg1518-lacZ*, Tetra resistant	This study
P*rgg1518-lacZ-*Δ*rgg0144*	Δ*rgg0144*, P*rgg1518-lacZ*, Tetra resistant	This study
P*rgg1518-lacZ-*Δ*rgg0939*	Δ*rgg0939*, P*rgg1518-lacZ*, Tetra resistant	This study
P*cps-lacZ*-wt	D39, P*cps-lacZ*, Tetra resistant	This study
P*cps- lacZ-*Δ*rgg1518*	Δ*rgg1518*, P*cps-lacZ*, Tetra resistant	This study

Table 2.2. *Escherichia coli* strains dependent in this study.

***Escherichia coli* strains**	**purpose of using**	**Source**
One Shot® TOP10	Plasmid propagation	Invitrogen, UK
One Shot® TOP10-*rgg0999*	*Rgg0999*-pCEP Plasmid, Kan resistant	This study
One Shot® TOP10-*rgg1518*	*Rgg1518*-pCEP Plasmid, Kan resistant	This study
One Shot® TOP10-P*rgg0999*	P*rgg0999*-pPP2 plasmid, Amp resistant	This study
One Shot® TOP10-P*rgg1518*	P*rgg1518*-pPP2 plasmid, Amp resistant	This study
E. coli DH5 α- *rgg1518*	Cloning *Rgg1518*	PROTEX, University of Leicester
BL21 (DE3) pLysS	Protein expression	Agilent Tech, USA
BL21 (DE3) pLysS-*rgg1518*	pLEICS-01-*rgg1518*, Amp resistant	This study

Table 2.3. Plasmids employed in this study.

Plasmids	**purpose of using**	**Source**
pCEP	Genetic complementation, Kan resistant	Guiral *et al.*, 2006
pPP2	*lacZ* transcriptional report, Amp &Tetra resistant	Halfman *et al.*, 2007
pLEICS-01	protein expression, Amp resistant	PROTEX, UK

2.3. The growth conditions of bacteria

Brain Heart infusion (BHI) broth medium (Oxoid) or blood Agar Base (BAB) medium (Oxoid) supplemented with 5% defibrinated host blood were used to grow *S.pneumoniae* in the presence of 5% CO_2 to aid growth at 37ºC. All media used in this process were prepared according to manufacturer's instructions, and the autoclave was used to sterilise the media at 121°C at 15 pound / square inch for 15 min. When required, different antibiotics such as spectinomycin 100 μg/ml, kanamycin 200 μg/ml and tetracycline 15 μg/ml were used for supplementing the growth media. Pneumococcal strains were also grown microaerobically in chemically defined medium (CDM) (Kloosterman *et al*., 2006a). CDM contains basal solution (Table 2.4), amino

acids and other CDM components (Table 2.5), and vitamins (Table 2.6). Overall CDM composition used for growth of pneumococcal strains is represented in Table 2.7.

Table 2.4. Composition of basal solution*.

Components	**g/l**
Na_2-β-glycerophosphate	26
Na-Acetate	1.0
$(NH_4)_3$ citrate	0.6
KH_2PO_4	1.0
Cysteine-HCl	0.4

*pH 6, autoclaved and kept at -20ºC.

Table 2.5. Composition of amino acids and other CDM components.

Components	**g/l**
Amino acids*	
Alanine	3
Arginine	1.55
Asparagine	4.4
Aspartate	5.25
Glutamate	6.25
Glutamine	4.9
Glycine	2.2
Histidine	1.9
Isoleucine	2.65
Leucine	5.7
Lysine	5.5
Methionine	1.55
Phenylalanine	3.45
Proline	8.45
Serine	4.25
Threonine	2.8
Tryptophane	0.65
Valine	4.05
Micronutrients**	
$MgCl_2$	0.2
$CaCl_2$	0.038
$ZnSO_4$	0.005
Nitrogenous bases***	
Adenine	0.01
Uracil	0.01
Xanthine	0.01
Guanine	0.01
Pyruvate	
Sodium Pyruvate	10

*pH 6.5 kept at room temperature. ** It was dissolved in distilled water and kept at 4ºC. *** It was dissolved in 0.1 M NaOH kept in 4ºC.

Table 2.6. Composition of vitamins*.

Components	g/l
Na-p-Aminobenzoate	0.005
D-Biotin[2]	0.0025
Folic Acid[2]	0.001
Nicotinic Acid	0.001
Ca (D^+) Pantothenate	0.001
Pyridoxamine-HCl	0.0025
Pyridoxine-HCl	0.002
Riboflavin	0.001
Thiamine-HCl	0.001
DL-6,8-Thioctic Acid Amide	0.0015
Vitamin B_{12}	0.001

* Riboflavin was dissolved in about 2/4 of total volume and warming up to 70ºC. Other vitamins are not added until the solution has been cooled to 30-40ºC. D-Biotin and folic acid are dissolved in NaOH 2M then pH of solution is set at 6.5 and stored at 4ºC.

Table 2.7. CDM Composition utilized for growth of pneumococcal strains.

Medium composition	Volume ml
Basal solution	870
Amino acids	80
Sugar	40
Vitamins	10
Micronutrients	10
Nitrogenous bases	10
Pyruvate	1
Choline-HCl	4

The CDM was supplemented either with galactose, mannose or glucose Table 2.8. Pneumococci were initially grown to mid-exponential phase in 10 ml BHI. All cultures were spun down by centrifuging at 3400g (Sorvall legend T, Thermo Scientific) for 10 min at room temperature and the supernatant was then discarded. The pellets were subsequently resuspended in 1 ml of CDM and 200 µl of resuspended pellet was transferred into 30 ml universal tubes containing 10 ml of CDM supplemented with 55 mM of the selected sugar to get initial OD600 nm ~ 0.05.

Table 2.8. Sugars used to supplement the CDM.

Components	Stock g/l	ml/l CDM	Molarity
Galactose	250	40	55mM
Mannose	500	20	55mM
Glucose	500	20	55mM

The strains of *Escherichia coli* were grown for purpose of plasmid transformation and propagation. Lysogeny broth medium (LB) or Lysogeny agar (LA) plates consist of LB with 1.5 % w/v bacteriological agar were used for growing *E. coli*. Broth *E. coli* cultures were incubated on shaking incubator (220 g) at 37ºC. When required, the growth medium was supplemented with different antibiotics. Antibiotics were prepared in distilled water (dH_2O) or 50% (v/v) ethanol and filtered by using a 0.2 µm syringe filter, divided into stocks of 500 µl, and stored at -20ºC (Table 2.9).

Table 2.9. Antibiotic solutions used in this sudy.

Antibiotic	Stock solution	concentration*
Spectinomycin	100 mg/ml in H_2O	100 µg/ml
Kanamycin	100 mg/ml in H_2O	250 µg/ml
Ampicillin	100 mg/ml in H_2O	100 µg/ml
Tetracycline	15 mg/ml in 50% (v/v) ethanol	3 µg/ml

*Working concentration of kanamycin, for *S. pneumoniae*, 100 µg/ml; for *E. coli*, 250 µg/ml.

2.4. Viable counts using colony-forming units

According to the Miles and Misra method, the bacterial colony forming units (CFU) were calculated (Miles *et al*., 1938). In sterile condition, 20 µl of bacterial culture was mixed with 180 µl of phosphate buffered saline (PBS), pH 7.0, in a 96 wells U-bottom microplate, and was serially diluted. 40 µl of each dilution was platted out on blood agar base (BAB) plates; thereafter, the plates were dried and put in a candle jar, incubated overnight at 37°C. The next morning, viable colonies were counted in each section that

ranged from 30 to 300 colonies. The following formula was used to calculate bacterial colony forming units per ml.

CFU/ml = colonies number x dilution factor x (1000/40)

2.5. Extraction of pneumococcal chromosomal DNA

Pneumococcal DNA was extracted according to the method of Saito and Miura (1963). Firstly, *S. pneumoniae* was grown overnight at 37ºC in 10 ml BHI broth. Supernatant was discarded after centrifugation at 3500g for 10 min, and the pellet was resuspended in 400 µl TE buffer (1 M Tris-HCl and 500 mM EDTA, pH 8.0) containing 25% w/v sucrose, 60 µl of 500 mM EDTA, 40 µl of 10% w/v sodium dodecyl sulfate (1g of SDS in 10 ml nH_2O) and 2 µl of proteinase K (12.5 mg/ml). The mixture was incubated at 37ºC for 1-2 hours to get a clear lysate. The mixture then was centrifuged at 13000g (microfuge, Sigma) for 5min. The supernatant was gently mixed with an equal volume of liquidified phenol in Tris buffer pH 7.6. After centrifuging at 13000g for 5 min, the upper aqueous phase was carefully transferred to a new microcentrifuge tube without disturbing the white protein layer, and phenol: chloroform: isoamyl alcohol (25: 24: 1, v/v, respectively) (Invitrogen, UK) was added. After centrifugation, the upper aqueous phase was mixed with 5 volumes of 100% (v/v) ethanol and 0.1(v/v) volume of sodium acetate (3M pH 5.2). The pellet was washed with 70% ethanol (v/v), centrifuged at 13000g for 5 min. Next dried, resuspended in 250 µl TE buffer. DNA concentration was measured at OD260 nm using a NanoDropTM spectrophotometer (Thermo Scientific, UK). Pneumococcal genomic DNA was stored at -20ºC until use.

2.6. Agarose gel electrophoresis preparation

Agarose gel electrophoresis was used to determine the size and the quality of PCR products (Meyers *et al*., 1976). Gel electrophoresis was done according to the method of Sambrook *et al*., 1989. 1% w/v agarose (Bioline) was mixed with TAE buffer (40 mM Tris-acetate and 1 mM EDTA, pH 8.0). The mixture was heated, and after the agarose had cooled to 60°C, 5 µl (10 mg/ml) ethidium bromide for DNA staining was added to the mixture. The samples of DNA were mixed with loading dye (New England Biolabs (NEB, UK), and were then filled into wells and size marker was also included to estimate the approximate concentration or the size of DNA sample (1 kb or 100 bp ladder) (NEB, UK). The gel was run at 80-90 volts for approximately 60 min. DNA fragments were visualized using a long wave UV transilluminator.

2.7. Polymerase Chain Reaction (PCR)

Polymerase chain reaction was performed by using a thermal cycler (Biometra, Germany). PCR amplification was carried out using PrimeSTAR HS premix (Clontech, USA) and HotStarTaq Plus Master Mix (Qiagen,UK). PrimeSTAR HS premix, is an extremely active enzyme and has efficient proofreading activity, was used to amplify DNA for cloning and doing insertion-deletion mutations. The PCR reaction contained DNA template with concentration 20 ng/µl, gene specific forward and reverse primer mix (1 pmol each/reaction), DNase-RNase free water and PrimeSTAR HS premix or HotStarTaq Plus Master Mix (Table 2.10).

Table 2.10. General PrimeSTAR and HotStar PCR reaction.

Components	PrimeSTAR reaction µl	HotStar reaction µl
DNA template	2	2
Primer mix	2	2
DNase-RNase free water	21	6
Enzyme	25	10

The DNA sample was amplified using the following conditions for PrimeSTAR HS premix: initial denaturation at 105°C for 10 seconds, 30 cycles amplification (denaturation at 95ºC for 10 sec, annealing at 55°C for 45 sec, extension at 72°C for 1 min/1000 bp), 1 cycle of final extension at 72°C for 2 min, and hold at 4°C. For HotStarTaq: initial denaturation at 95°C for 10 min, 30 cycles amplification (denaturation at 95ºC for 45 sec, annealing at 55°C for 45 sec, extension at 72°C for 1 min/1000 bp), 1 cycle of final extension at 72°C for 2 min and hold at 4°C.

2.8. DNA purification using Wizard® SV Gel and PCR Clean-Up System kit

The Wizard® SV Gel and PCR Clean-Up System (Promega, UK) was used to purify and concentrate DNA from PCR reactions. 1 µl membrane binding solution (4.5 M guanidine isothiocyanate) was added per 1 µl of PCR product. To achieve DNA binding, the mixture was transferred into a SV minicolumn. Next, the SV minicolumn was inserted into a collection tube and incubated at room temperature for 1 minute. The SV minicolumn was centrifuged at 16000g for 1 min. Flowthrough was discarded and the SV minicolumn was washed with 700 µl membrane wash solution (10 mM potassium acetate pH 5.0, 80% ethanol and 16.7 µM EDTA pH 8.0), and centrifuged again at 16000g for 1 min. Flowthrough was discarded, and the SV minicolumn was reinserted into the collection tube, this, in turn, was emptied. The column assembly recentrifuged for 1 min to allow evaporation of any residual ethanol. For elution, the SV minicolumn was carefully transferred into a clean 1.5 ml microcentrifuge tube. Finally, 35 µl DNase-RNase free water was added to the SV minicolumn, incubated at room temperature for 1 min and centrifuged at 16000g for 2 min. The SV minicolumn was discarded, and DNA was stored at –20ºC.

2.9. Extraction and purification of DNA fragments from agarose gel

The target DNA fragments were extracted and purified by following the instructions of the Wizard® SV Gel and PCR Clean-Up System (Promega, UK). Following gel electrophoresis, the DNA band of interest was excised from the gel. The gel slice was placed in a 1.5 ml microcentrifuge tube. Next, 10 µl membrane minding solution (4.5 M guanidine isothiocyanate) was added per 10 mg of gel slice. The mixture was subjected to vortex and has been incubated at 50-65ºC until the gel slice was completely dissolved. For binding DNA fragments of interest, the dissolved gel mixture was transferred into a SV minicolumn. Thereafter, the SV minicolumn was inserted into the collection tube and incubated at room temperature for 1 min. The SV minicolumn was centrifuged at 16000g for 1 min. Flowthrough was discarded and the SV minicolumn was washed with 700 µl membrane wash solution (10 mM potassium acetate pH 5.0, 80% ethanol and 16.7 µM EDTA pH 8.0), and centrifuged at 16000g for 1 min. Flowthrough was discarded, and the SV minicolumn was reinserted into the collection tube. The above washing process was repeated with 500 µl membrane wash solution. The collection tube was then emptied and the column assembly recentrifuged for 1 min to allow evaporation of any residual ethanol. Finally, for elution, the SV minicolumn was carefully transferred into a clean 1.5 ml microcentrifuge tube. 50 µl nuclease-free water was added to the SV minicolumn, incubated at room temperature for 1 min, and centrifuged at 16000g for 1 minute. The SV minicolumn was discarded, and DNA was stored at -20°C until use.

2.10. Extraction of *E. coli* plasmid DNA

Plasmid DNA was extracted using the qIAprep spin miniprep kit according to the manufacturer's instructions (Qiagen, UK). At first, *E. coli* culture containing the plasmid was used to inoculate a 10 ml LB medium supplemented with an appropriate antibiotic and left in a shaking incubator overnight at 37°C. The overnight culture was pelleted by centrifugation (Sorvall legend T, Thermo Scientific) at 6800g for 15 min at room temperature. Next, the pellet was resuspended in 250 µl buffer P1containing RNase A, and was transferred to a microcentrifuge tube. After that, 250 µl buffer P2 was added and mixed thoroughly by inverting the tube 4-6 times until the solution became clear. The lysis reaction was incubated no more than 5 min and subsequently, 350 µl buffer N3 was added and mixed thoroughly by inverting the tube 4-6 times. The mixture was centrifuged for 10 min at 13000g in a table-top microcentrifuge (microfuge, Sigma). The supernatant was applied to the QIAprep spin column by pipetting, centrifuged for 1 min, and the flowthrough was then discarded. The QIAprep spin column was washed by adding 500 µl buffer PB and centrifuged for 1 min, and the flowthrough was also discarded. The washing process was repeated once more with 750 µl buffer PE. The QIAprep spin column was centrifuged for 1 min to remove residual wash buffer. Subsequently, the QIAprep spin column was placed in a clean 1.5 microcentrifuge tube, and DNA was eluted by adding 50 µl buffer EB. The plasmid was stored in -20°C until use.

2.11. Restriction digests of DNA

In this study, plasmid pCEP (Guiral *et al*., 2006) was used for the genetic complementation of the mutant strains. Plasmid pCEP and inserts were digested with *NcoI* and *BamHI* (NEB, UK). The pPP2 plasmid (Halfmann *et al*., 2007) was used for

transcriptional *lacZ*-fusions. The pPP2 and inserts were digested with *SphI* and *BamHI* (NEB, UK). The reactions were done in 50 µl using ~1 µg of each plasmid/insert, 5 µl CutSmart™ buffer, 1 µl (10 U/µl) appropriate restriction enzyme and the reaction volume was topped up to 50 µl using DNase-RNase free water. The mixture was incubated at 37ºC until completely digested. Purification was then performed for digested DNA molecules fragments using the Wizard® SV Gel and PCR Clean-Up System kit (Section 2.8). Ligation reactions were accomplished in a typical ligation mixture consisting of 1:3 molar ratio of plasmid to insert, 2 µl (400 U/µl) of T4 DNA ligase (NEB) and 2 µl of 10X T4 DNA ligase reaction buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT, pH 7.5). The ligation mixture was incubated at 16°C overnight.

Moreover, pLEICS-01 was used for protein expression. This plasmid was constructed by the protein expression laboratory (PROTEX, University of Leicester) using In-Fusion® HD cloning kit (Clontech, USA) to clone the target gene into pLEICS-01. The cloning reaction mixture in 10 µl contained 50 ng/µl of insert, 100 ng/µl of pLEICS-01, 1 µl cloning enhancer, 2 µl of 5X In-Fusion enzyme premix, and was incubated in a water bath first at 37°C for 15 min, next at 50°C for 15 min, and then was put on ice.

2.12. Transformation into *E. coli*

The *E. coli* chemically competent cells One Shot® TOP10 (Invitrogen, UK) were used to achieve successful propagation, or BL21 (DE3) pLysS (Agilent Technologies, USA) for protein expression. 50 µl of competent cells were thawed on ice and placed in an already chilled 14 ml BD Falcon polypropylene round bottom tube. The ligation reaction or an aliquot of plasmid was added directly to the cells and incubated on ice for 30 min with gentle stirring every 5 min. The mixture was placed in a water bath fixed at

42°C for 45 sec and returned back on ice for 2 min instantly. Then 500 µl of LB at room temperature was added into the reaction mixtures and incubated at 37°C for 1.5 hrs in an Innova 44 shaking incubator adjusted at 225g (New Brunswick Scientific, USA). Finally, the transformation mixture was divided into two aliquots (250 µl) and plated on LA supplemented with an appropriate antibiotic. The plates were subsequently incubated at 37°C overnight in a static incubator.

2.13. Transformation into *S. pneumoniae* D39

Transformation into *S. pneumoniae* was performed according to the method of Bricker and Camilli (1999). 10 µl *S. pneumoniae* D39 culture was grown overnight in 10 ml of BHI at 37°C. The culture was diluted at 1:50 or 1:100 in 10 ml of fresh BHI broth. The bacterial suspension was incubated at 37°C, and when OD600 nm was reached approximately 0.05-0.08, bacterial culture (860 µl) was transferred into a sterile microcentrifuge tube, and then mixed with 100 µl NaOH (100 mM), 10 µl BSA (20% w/v), 10 µl $CaCl_2$ (100 mM), 2 µl competence stimulating peptide (CSP) (50 ng/µl), and 5-10 µl DNA or plasmid with a concentration of 200-400 ng/µl (Alloing *et al.*, 1996). The mixture was incubated at 37ºC for 3 hrs. An aliquot of 330 µl was sampled every hour and plated out on BAB with the appropriate antibiotic. Finally, all plates were incubated at 37°C overnight in a CO_2 incubator.

2.14. Mutation by splicing overlap extension

SOEing protocol (splicing by overlap extension) was used to carry out the strategy of allelic replacement mutagenesis (Horton, 1995; Song *et al.*, 2005). The two-step PCR procedure is based on generating a fusion between an antibiotic-resistance cassette and the upstream and downstream regions (left and right flanks) of a target gene. The first

PCR step amplifies the *aadA* or *aph* genes resistance to spectinomycin and kanamycin, respectively, and the upstream and downstream regions of the target gene. In the second step, the amplified upstream and downstream regions were joined with the amplicons representing *aadA* or *aph* using the upstream-F and downstream-R primers. The PCR product was incorporated into chromosomal DNA of *S. pneumoniae* D39 by transformation and subsequent homologous recombination with the target region.

2.14.1. Amplifying SOEing DNA fragments

The spectinomycin resistance cassette *aadA* (1158 bp) was amplified from the plasmid pDL278 (Yesilkaya, 1999) using the Spec-F/Spec-R primers, while kanamycin-resistance cassette *aph* (895 bp) was amplified from pCEP plasmid using Kan-F/Kan-R primers (Guiral *et al.*, 2006) (Table 2.11). The DNA fragments surrounding the target gene were amplified using SPDXUPstream/F and SPDXUPstream/R primers (chimeric), or SPDXDOWNstream/F and SPDXDOWNstream/R primers (chimeric), (where X indicates gene/locus number) and the PCR products were 600 bp in length. The DNA fragments were then purified by using the Wizard® SV Gel and PCR Clean-Up System (Promega, UK) as previously described in (Section 2.8). The upstream and downstream regions of the target gene in addition to antibiotic resistance genes (*aadA* and *aph*) were amplified with PrimeSTAR HS premix (Section 2.7) using the primers shown in Table 2.11.

Table 2.11. The list of primers used for amplifying SOEing fragments for genetic mutations. Italic typeface indicates *aadA* and *aph* gene homology regions.

Primers ID	Nucleotide sequence (5′-3′)
Spectinomycin/F	ATCGATTTTCGTTCGTGAAT
Spectinomycin/R	GTTATGCAAGGGTTTATTGT
Kanamysin/F	GCGCGCATGCTGGGGATCCGT
Kanamysin/R	CTAAAACAATTCATCCAGTAA
SPD0999UP/F	TGATGTTAAAAAATCAGTCGGC
SPD0999UP/R	*ATTCACGAACGAAAATCGATC*CAAGATGTTCCATCTTTAC
SPD0999DOWN/F	*ACAATAAACCCTTGCATAAC*GTCAAAAGTTAATTTTCCC
SPD0999DOWN/R	ATGAAAAAAGTCAGCGAGTC
SPD01518UP/F	CACAGGTTGTTAACGATTCTCC
SPD1518UP/R	*ATTCACGAACGAAAATCGATC*ATAATCCCATCTCATATTGTT
SPD1518DOWN/F	*ACAATAAACCCTTGCATAAC*TTTTTCAAACCTATAATTTGTA
SPD1518DOWN/R	GAAAACAATTTAAAAGAGGAGG
SPD1518UPR/kana	*ACGGATCCCCAGCATGCGCGC*CATAATCCCATCTCATATGTT
SPD1518DF/kana	*TTACTGGATGAATTGTTTTAG*TTTTTCAAACCTATAATTTGTA

2.14.2. Fusion of SOEing fragments and transformation into *S. pneumoniae*

The terminal primers (SPDXUPstream/F and SPDXDOWNstream/R) were used to fuse the amplicons of the target gene flanking regions and *aadA* or *aph* to produce a fused product. The PCR was done in 50 µl reaction mixture containing 2 µl of each of the upstream and downstream PCR products (~20 ng/µl each), 4 µl of the *aadA* or *aph* gene amplicons (~20 ng/µl each), 25 µl of 2X PrimeSTAR HS premix, 2 µl of SPDXUPstream/F and SPDXDOWNstream/R primers mixture (10 pmol/µl) and 15 µl of DNase-RNase free water as previously, described in section 2.7. The fused PCR products were analysed by agarose gel electrophoresis. The amplicons were then cut and purified using DNA purification Wizard® SV Gel and PCR Clean-Up System kit as previously described in section 2.8. The linear joined products were transformed into *S.*

pneumoniae D39, as described in section 2.13 and the transformants were plated on BAB plates supplemented either with 100 µg/ml spectinomycin or 200 µg/ml kanamycin.

2.14.3. Confirmation of successful mutation by PCR

Gene replacement in spectinomycin/kanamycin-resistant colonies was confirmed by PCR. The genomic DNA was extracted from selected spectinomycin/kanamycin-resistant colonies as previously described in section 2.5. The DNA used as a template for PCR analysis and the DNA of wild type D39 was used as a control. The SPDXUP/F and SPDXDOWN/R primers were used to amplify the entire mutated region. Furthermore, the primers Spec/F/R and Kana/F/R were used to amplify the spectinomycin cassette (*aadA*) and kanamycin cassette (*aph*), respectively, from the transformants. Additionally, SPDXUP/F and Spec/R or Kana/R were used to amplify the upstream region of each target gene and the *aadA*/*aph*, while the primers SPDXDOWN/R and Spec/F or Kana/F were used to amplify the downstream region of each target gene and the *aadA*/*aph*. The PCR was performed using HotStarTaq Plus Master Mix as described in section 2.7.

2.14.4 DNA sequencing of mutant strains

In addition to PCR, the successful mutations were further confirmed by DNA sequencing. Two new sets of primers were designed external to the terminal ends of mutated region (Table 2.12). These primers were used to confirm the successful mutations. The primer set F1 and F2 would amplify 200 nucleotides before the upstream region of the mutated target, and approximately 200 nucleotides before the start of antibiotic resistance cassette, respectively. The other set of primers R1 and R2

would amplify approximately 200 nucleotides after the end of antibiotic resistance cassette, and 200 nucleotides far from the downstream region of the mutated gene, respectively. The extracted genomic DNA of each mutant was used as the template to amplify the entire fused constructs plus 200 bp up and downstream region using their specific F1 and R2 sequencing primers. The PCR was carried out using HotStarTaq *Plus* Master Mix as described in section 2.7. The PCR products were then purified using the Wizard® SV Gel and PCR Clean-Up System (Promega, UK) section 2.8. The amplicons were sequenced at the Protein Nucleic Acid Chemistry Laboratory (PNACL), University of Leicester using the gene specific sequencing primers SPDX-seqM/F1, SPDX-seqM/F2, SPDX-seqM/R1, SPDX-seqM/R2 (where X indicates gene/locus number).

Table 2.12. Primers used for DNA sequencing.

Primers ID	Nucleotide sequence (5′-3′)
SPD0999-seqM/F1	ATCAAGTCTATTGTGGACCA
SPD0999-seqM/F2	GGCCCAGAAAGAAAGTCAAAA
SPD0999-seqM/R1	ACGAGCTCAGAAATCTGATAC
SPD0999-seqM/R2	AAAAGGGAAGTAGAAATTCAAAAG
SPD1518-seqM/F1	CAAAGTAAGCAAGACATAAA
SPD1518-seqM/F2	ATCCCCAAAACCATGTTTCA
SPD1518-seqM/R1	GGAGCTATAGATACAGTCA
SPD1518-seqM/R2	TTGGTATGAATTGGTTAACT

2.15. Genetic complementation of successful gene replacement of mutants

Genetic complementation of mutant strains was performed according to the protocol described previously to eliminate the possibility of a polar effect due to mutation (Guiral *et al.,* 2006). This protocol is based on the use of plasmid pCEP (9540 bp) which has a 2 kb DNA segment that is homologous to the pneumococcal genome, hence it incorporates the cloned genes by homologous recombination and the plasmid pCEP is

nonreplicative in *S. pneumoniae*. The 9540 bp plasmid incorporates directly into downstream of the *amiA* operon, which is believed to be a genetically silent site, and has no effect on cellular functions (Alloing *et al.,* 1996).

2.15.1. Double digest, cloning the *rgg* gene into pCEP and transformation into *E. coli*

pCEP plasmid, provided by Dr Marc Prudhomme (Toulouse, France), was extracted using the QIAprep Spin Miniprep kit as previously described in section 2.10. The target genes and their putative promoter regions were amplified using the *NcoI-BamHI* integrated primers listed in Table 2.13. The PCR was carried out using PrimeSTAR HS premix as previously described in section 2.7. The amplified PCR products were purified using the Wizard® SV Gel and PCR Clean-Up System (Promega, UK) as described in section 2.8 and then analysed by agarose gel electrophoresis. The plasmid pCEP and the inserts were double digested using *NcoI-BamHI* enzymes (Section 2.11). The digested samples were purified using the Wizard® SV Gel and PCR Clean-Up System (Promega, UK). Both digested and undigested pCEP were analysed by agarose gel electrophoresis for comparison. The inserts were ligated to pCEP (Section 2.11) and then an aliquot of the ligation reaction was transformed into *E. coli* One Shot® TOP10 chemically competent cells, plated on LA supplemented with 50 μg/ml of kanamycin and the resistant colonies were selected (Section 2.12).

Table 2.13. A list of primers used for genetic complementation. Italic typeface indicates incorporated restriction sites.

Primers ID	Nucleotide sequence (5′-3′)
SPD0999F*NcoI*C	CATG*CCATGG*ATGGAACATCTTGGAAAAGT
SPD0999R*BamHI*C	AC*GGATCC*TTAACTTTTGACAAATTTTT
SPD1518F*NcoI*C	CATG*CCATGG*TTTTTCTGTTTCCTTTCATT
SPD1518R*BamHI*C	AC*GGATCC*TTATAGGTTTGAAAAATCAT
MalF	GCTTGAAAAGGAGTATACTT
PCEPR	AGGAGACATTCCTTCCGTATC

2.15.2. Confirmation of transformation pCEP constructs and DNA sequencing

An antibiotic-resistant transformant was grown overnight in 10 ml LB supplemented with 50 μg/ml kanamycin. The QIAprep Spin Miniprep kit was used for extraction of recombinant plasmid DNA (Section 2.10). To confirm successful cloning, primers SPDXF*NcoI*C and SPDXR*BamHI*C (X refers to the gene) were used to amplify target inserts while MalF and pCEPR primers were used to amplify the up and downstream of cloning site plus target gene. Amplification was done using HotStarTaq Plus Master Mix (Section 2.7). The successful recombinant pCEP was analysed by DNA sequencing using MalF and pCEPR.

2.15.3. Transformation of pCEP constructs into pneumococcal mutants

The recombinant pCEP constructs were transformed into relevant pneumococcal mutant strains and selected on BAB plates supplemented with spectinomycin and kanamycin as described in section 2.13. The successful pCEP constructs carrying the intact copy of the target genes and their putative promoter regions were confirmed by PCR using the primers MalF and PCEPR.

2.16. *lacZ*-fusions transcriptional report

2.16.1. Double digest, cloning *rgg* into pPP2 and transformation into *E. coli*

The transcriptional *lacZ*-fusions were constructed as described previously using the integrative reporter plasmid pPP2 (Halfmann *et al*., 2007). The genetic region encompassing the putative promoter site was amplified using the primers that are modified to incorporate *SphI* and *BamHI* specific sites (Table 2.14). The PCR was done using PrimeSTAR HS premix and the products were purified using the Wizard® SV Gel and PCR Clean-Up System (Promega, UK). The PCR products were analysed by

agarose gel electrophoresis. The pPP2 plasmid and inserts were both double digested using the restriction enzymes *SphI* and *BamHI*. Purification was carried out to the digested samples using the Wizard® SV Gel and PCR Clean-Up System (Promega, UK). The insert was ligated to pPP2 and the ligation mixture was transformed into *E. coli* TOP10 chemically competent cells, and the transformants were selected on Lysogeny broth (LB) agar medium supplemented with 100 µg/ml of ampicillin.

To confirm the successful cloning, colony PCR was done for the selected transformants. FusionXF*sphI* and FusionXR*BamHI* (X refers to target gene) were used to amplify the target putative promoter regions (Table 2.14). In addition, Fusion/UF and Fusion/DR primers were designed to amplify up and downstream of the cloning site plus the insert. The recombinant pPP2 was extracted from a positive colony and further confirmed by DNA sequencing using Fusion/UF and Fusion/DR primers.

Table 2.14. A list of primers used for amplifying of transcriptional *lacZ*-fusions. Italic typeface represents incorporated restriction enzyme recognition sites.

Primers ID	Nucleotide sequence (5′-3′)	Restriction site
Fusion/UF	CTACTTGGAGCCACTATCGA	-
Fusion/DR	AGGCGATTAAGTTGGGTAAC	-
Fusion0999F*sphI*	GAC*GCATGC*AGCTTTCTAGCTTGGTCAG	*SphI*
Fusion0999R*BamHI*	ACG*GGATCC*CGTTCAATTTTATTAGCTTG	*BamHI*
Fusion1518F*sphI*	GAC*GCATG*CAAGAAGATTCCTTATAACT	*SphI*
Fusion1518R*BamHI*	ACG*GGATCC*TAAAAATTTGTCCATAATCCC	*BamHI*
Fusion *cps* F*sphI*	GAC*GCATG*CGATTATACCACATTGTGTAC	*SphI*
Fusion *cps* R*BamHI*	ACG*GGATCC*AACGTGATTTTTTAAAACGTC	*BamHI*

2.16.2. Transformation of recombinant pPP2 constructs into pneumococcal strains

The recombinant pPP2 constructs were transformed into *S. pneumoniae* D39 or mutants Δ*rgg0999*, Δ*rgg1518*, Δ*rgg0144* and Δ*rgg0939* using the protocol described in Bricker

and Camilli (1999) (section 2.13). The colonies were individually selected on blood agar plates supplemented with 15 µg/ml tetracycline for wild type or with 100 µg/ml spectinomycin plus 15 µg/ml tetracycline for the mutant strains. The successful transformation of *lacZ*-fusions into the pneumococcal genome was confirmed by PCR using HotStarTaq plus Master Mix and the Fusion/UF – Fusion/DR primers.

2.16.3. β-galactosidase activity

The β-galactosidase activity was determined in strains under study according to the method of Miller, (1972). The promoter regions transcriptionally drive the expression of the promoterless *lacZ* gene, which results in the production of β-galactosidase. The protocol shown below is used to precisely determine the level of the β-galactosidase activity in cell extracts by measuring the appearance of yellow colour over time at OD420 nm, which is attributed to hydrolysis of the chromogenic substrate 2-Nitrophenyl β-D-galactopyranoside (ONPG) by β-galactosidase to 2-Nitrophenyl with yellow colour and colourless galactose (Figure 2.1).

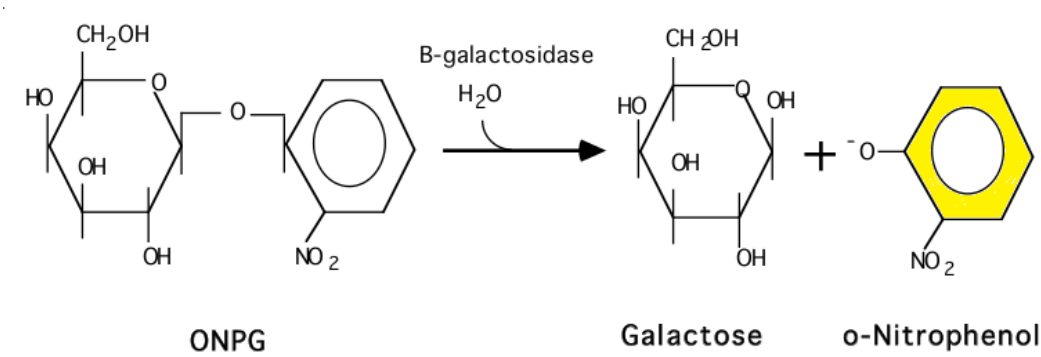


Figure 2.1. An illustration showing the hydrolysis mechanism for the chromogenic substrate 2-Nitrophenyl β-D-galactopyranoside (ONPG) to galactose and 2-Nitrophenyl by the β-galactosidase enzyme.

Pneumococcal *lacZ* reporter strains were grown microaerobically in 10 ml CDM supplemented with 55 mM of the selected sugar (galactose, mannose or glucose) as the main carbon source (Section 2.3). The cultures were grown to mid-exponential phase then 2 ml of the culture was centrifuged at 3500g (Hettich MIKRO 22R, Germany) for 15 min at 4°C. The supernatant was discarded and the pellet was then resuspended with 2 ml of chilled Z buffer Table 2.15. The OD600 nm of the resuspended cells was then measured.

Table 2.15. Composition of Z buffer per 50 ml*.

Components	Amount	Final Molarity
$Na_2HPO_4.7H_2O$	0.80 g	0.06
$NaH_2PO_4.H_2O$	0.28 g	0.04
1M KCl	0.5 ml	0.01
1M $MgSO_4$	0.05 ml	0.001
2-mercaptoethanol	0.175 ml	0.05

*Adjust the pH to 7.0 and stored at 4ºC

The cells were diluted in Z buffer in a ratio of 1:1 (500 µl of cells and 500 µl of Z buffer), and a drop of Triton X100 (4-(1,1,3,3-Tetramethylbutyl) phenyl-polyethylene glycol solution) (Sigma) was added to the mixture. The reaction mixture was vortexed and then incubated at 28°C for 5 min. After that, 200 µl of 2-Nitrophenyl β-D-galactopyranoside ONPG (4 mg/ml) was added to the reaction mixtures and the samples were incubated at 30°C. When sufficient yellow colour had developed, the reaction was inactivated by adding 500 µl of 1 M Na_2CO_3. The incubation time to develop colour was recorded and 1.5 ml of the sample was transferred to a fresh microcentrifuge tube and centrifuged for 5-10 min at 14000g (Microfuge, Sigma). The optical density of the supernatant was measured at OD420 nm, and the specific β-galactosidase activity was expressed as per unit of time per unit of volume of cell suspensions using the equation below:

$$\text{Miller Units} = 1000 \times (OD420) / (T \times V \times OD600)$$

The OD600 nm represents cell density in the washed cell suspension, T is the time of the reaction in minutes and V is the volume in ml of culture used in the assay.

2.17. Phenotypic characterisation of pneumococcal strains

2.17.1. Growth studies

Growth studies were carried out using 96 well flat-bottom cell culture plates. To do this, 198 µl of BHI, or CDM supplemented with 55 mM of galactose, glucose or mannose as main carbon source with 2 µl bacterial suspensions containing ~5 x 10^9 CFU/ml, were added into the wells in triplicate. The microtitre plate was sealed with MicroAmp® Optical Adhesive Film (Invitrogen, UK), and was incubated microaerobically inside the MultiskanTM GO Microplate Spectrophotometer (Thermo Scientific, UK), which was set to run for 20 hours at 37°C, taking a reading every hour at OD600 nm. The following equation was used to calculate maximum specific growth rates (µ) by linear regression analysis of the plots:

$$(\mu)\ h^{-1} = \ln OD_2 - \ln OD_1 / t_2 - t_1$$

ln is the natural logarithm, t= time, OD_2 is the cell density at t_2 (late exponential phase) and OD_1 is the cell density at t_1 (early exponential phase), (Neidhardt *et al.*, 1990).

2.17.2. Determination of pneumococcal capsule production

Glucuronic acid is the major component of *S. pneumoniae* type 2 capsule, hence it was assayed to determine the capsule production by following the protocol described previously (Cho *et al.,* 2009). First pneumococcal strains were inoculated into CDM supplemented either with galactose or glucose. Simultaneously, viable counts for each overnight culture were done, and 1 ml of each bacterial culture was transferred to a new microcentrifuge tube. Thereafter, 200 µl of 1% (v/v) zwittergent 3-14 detergent (Sigma-Aldrich, UK) in 100 mM citric acid pH 2.0 was mixed and incubated on a thermo-shaker at 50ºC for 20 min. The sample was centrifuged for 5 min at 14000g. Next, 300 µl of the supernatant was taken and transferred to a new microcentrifuge tube and 1200 µl absolute ethanol was then added to bring the final concentration to 80%. In order to precipitate capsular polysaccharides, the sample was incubated at 4ºC for 20 min, and then centrifuged for 5 min at 14000g. The supernatant was discarded without disturbing the pellet and left to dry. The pellet was subsequently resuspended in 200 µl deionized water, and stored at -20ºC until use (Goncalves Mdos *et al.,* 2014).

To determine the amount of capsule produced in strains under study, 125 µl of each of the purified capsule polysaccharide sample was transferred to a fresh microcentrifuge tube. Next, 750 µl of 0.025M sodiumtetraborate solution (Borax) in 93 % (v/v) sulfuric acid (Sigma-Aldrich) was added to each tube and incubated at 100ºC for 10 min on a shaker hotplate at 250g (Thermo-Shaker). The samples were cooled on ice bath to room temperature. After that, 25 µl of 0.125 % (w/v) carbazole solution (0.250 g carbazole and 200 ml absolute ethanol) was added to the reaction mixture and heated to 100ºC for 10 min on a shaker hotplate. The reaction mixture was left to cool at room temperature and the absorbance was measured by spectrophotometer at OD530 nm. The glucuronic acid concentration in each sample was calculated using a standard curve prepared with

the known glucuronic acid concentrations (0, 10, 20, 40, 60, 80 and 100 μg ml^{-1}) (Sigma-Aldrich). Finally, concentration of CPS for each strain was expressed as μg of glucuronic acid / CFU (Lai *et al.*, 2003).

2.18. *In vivo* work

2.18.1. Preparation of bacterial dose

In order to prepare the bacterial inoculum for infection of mice, a sweep of colonies was inoculated into 10 ml BHI broth in a universal tube. The cultures were then incubated statically for 16-18 hours at 37°C until OD500 nm reached 1.4-1.6. The tubes were then centrifuged at 1500g for 15 min and the supernatant was discarded using a sterile Pasteur pipette. The pellet was resuspended with 1 ml prewarmed BHI serum broth (80% v/v BHI and 20% v/v filtered foetal calf serum). 700 μl of the resuspended pellet was added to 10 ml fresh BHI serum broth to bring OD500 nm to ~0.70. The tubes were then incubated at 37°C for 5 hours until OD500 nm reached ≥ 1.6. After that, the cultures were divided into 500 μl aliquots and kept at -80°C until needed.

2.18.2 Virulence test

Female CD1 mice (Charles River, UK) at 8-10 weeks of age were used for virulence testing. To determine the virulence of pneumococcal strains, the standardised inoculum was prepared as described in section 2.17.1. Mice were lightly anaesthetised with 2.5% v/v Isoflurane (Isocare, UK) over oxygen (1.4 to 1.6 litres/min), in an anaesthetic box. To initiate infection, mice were handled vertically while 50 μl of PBS containing ~2 x 10^6 CFU pneumococci were administered progressively in both nostrils. The mice were left on their backs in a cage to recover from the effects of the anaesthetic and prevent the inoculum from flowing out the nostrils. The inoculum dose was confirmed by viable

counting (CFU) on blood agar plates post-infection. Mice were watched periodically for disease signs (hunched, piloerect, or lethargy) for 7 days. The mice that reached the extremely lethargic stage were culled humanely. The time to culling point is defined as survival time (Gaspar *et al.,* 2014). Simultaneously, the progression of bacteraemia was monitored in each mouse by taking approximately 20 µl of venous blood from dorsal tail vein after 24 and 36 h post-infection. Colonies were enumerated using serial dilution of blood in sterilised PBS, and then cultured on blood agar plates supplemented with 5% v/v defibrinated horse blood. The median survival times of mice were examined statistically using the Mann-Whitney U test.

2.18.3 Mouse colonisation model

Colonisation experiments were carried out as essentially described in section 2.17.2 but in this experiment, each mouse was held horizontally while administering the dose (2.5 x 10^5 CFU/mice in 20 µl PBS) to avoid pneumococci disseminating into the lower respiratory tract. The inoculum dose was confirmed by viable counting on blood agar plates post-infection. To determine the progress of colonisation, the following procedure was done at 0, 3 and 7 days after intranasal infection. Set groups of mice were anesthetised with 5% v/v isoflurane and were then culled. Thereafter, nasopharyngeal tissue was taken and transferred into 5 ml of sterile PBS. Tissue samples were weighed and then homogenized (ULTRA-TURRAX^R, TP 18/10, Nr.15959, UK). Viable counts in homogenized tissue were determined by plating the serial dilutions of blood onto blood agar plates as described in section 2.3.

2.19. Microarray analysis and confirmation of results using quantitative reverse transcriptase PCR (qRT-PCR)

2.19.1. Extraction of pneumococcal RNA

The pneumococcal strains were grown microaerobically in CDM supplemented with 55 mM of glucose, galactose or mannose. Pneumococcal RNA extraction was performed according to the Trizol method (Stewart *et al.,* 2002). The bacterial pellet was resuspended with 500 µl of Trizol reagent (Invitrogen, UK). Thereafter, 100 µl chloroform was added to the mixture which was vortexed for 15 seconds. The samples were transferred into lysing matrix B tubes containing 0.1 mm silica spheres (MP Biomedicals, UK). The samples were sonicated for 45 sec using a PowerLyzerTM 24 homogeniser (MO BIO, USA). Next, the samples were left at room temperature for 1-2 min and were then centrifuged at 12000g for 15 min at 4°C. The upper aqueous phase containing RNA was transferred into a fresh tube and 250 µl isopropanol was added to the samples and vortexed for 15 sec. At this stage, the samples were left at room temperature for 15 min followed by centrifugation at 12000g for 10 min at 4°C. After removing supernatant, 500 µl 75% v/v ethanol was added to the pellet and centrifuged at 12000g for 10 min at 4 °C. The supernatant was removed and the pellet was left to dry for 15 min at room temperature. Finally, extracted RNA sample was resuspended with 87.5 µl DNase-RNase free water and stored at -80°C until needed.

2.19.2. DNase treatment of RNA

To remove contaminating DNA, the extracted RNA sample was treated with DNase using Ambion TURBO DNA-freeTM kit (Invitrogen). In a microcentrifuge tube, 20 µl of total RNA (~ 1.5-2 µg) was mixed with 5 µl of 10X TURBO DNase Buffer, 2 µl of TURBO DNase (2 U/µl) and 23 µl of nuclease-free water. The resulting mixture was

incubated at 37°C for 60 min. The reaction was stopped by adding 5 µl of 25mM EDTA and incubated at 65°C for 10 min. Subsequently, the reaction mixture was incubated on ice for 5 min and centrifuged at 10000g for 1.5 min at 4°C. The supernatant containing RNA was transferred to a new tube and the concentration of each RNA sample was measured spectrophotometrically using a NanoDrop 1000 (Thermo Scientific, UK) at 230 nm. Finally, the RNA samples were stored at -80°C until use.

2.19.3. Microarrays analysis

Streptococcus pneumoniae and its isogenic mutant strains *Δrgg0999* or *Δrgg1518* were grown microaerobically in CDM supplemented with 55 mM of glucose, galactose or mannose as the sole carbon source. The impact of *Δrgg0999* or *Δrgg1518* on the transcriptome of *S. pneumoniae* was analysed relative to the wild type (Kloosterman *et al.,* 2006b). The level of mRNA expression for all strains was identified in collaboration with Prof Oscar Kuipers, University of Groningen, the Netherlands.

2.19.4. Complementary DNA Synthesis

SuperScript III reverse transcriptase (Invitrogen) was used to perform the first-strand cDNA synthesis according to the manufacturer's instructions. The DNase-treated RNA (~1µg) was mixed with 1 µl of 10 mM of dNTP mix (Promega, UK) and 1 µl of 300 ng/µl random primers (Invitrogen). The reaction volume was then topped up to 20 µl with DNase-RNase free water. Next, the mixture was heated to 65°C for 5 min. Then 4 µl of 5X first-strand buffer, 1 µl of 0.1 M dithiothreitol (DTT) and 1 µl (200 U/µl) of SuperScrip III reverse transcriptase were added to the mixture. The sample was incubated at 25°C for 5 min and incubated again at 50°C for 45 min followed by inactivation by heating to 70°C for 15 min.

2.19.5. Quantitative reverse transcriptase polymerase chain reaction (qRT-PCR)

qRT-PCR was carried out using SensiMixTM SYBR® Hi-ROX kit (Bioline, UK). Corbett RG6000 PCR system (Qiagen, UK) was used to amplify the synthesised cDNA. The reaction mixture contained 2 µl of cDNA (~15 ng/µl), 10 µl 2X SensiMixTM SYBR Hi-ROX, 2 µl of gene specific primers (Table 2.16) and 6 µl DNase-RNase free water. The tubes were placed in a thermo cycler which was set to run at 95°C for 10 min (heat activation), followed by 40 cycles of amplification consisting of denaturation at 95°C for 20 sec, annealing at 55°C for 30 sec, and then extension at 72°C for 20 sec. The levels of transcription of genes were normalised to the transcription of *gyrB.* After that, the threshold cycle C_T method was used to analyze the difference in gene expression (Livak and Schmittgen, 2001). More than twofold difference in expression was treated as significant (Yesilkaya *et al*., 2009), taking into account that some differences less than twofold are also considered to be biologically important (Chaussee *et al*., 2002).

Table 2.16. Primers used for real-time qRT-PCR.

Primers ID	Nucleotide sequence (5′-3′)
gyrB RT F	GGCTGATCCACCAGCTGAGTC
gyrB RT R	TCGTGTGGCTGCCAAGCGTG
SPD0161 RT F	CGGTTCTTACGGTTGGTCTC
SPD0161 RT R	GGTTTTCCTTTTTCGGCATA
SPD0013 RT F	TGGTACAAGAAATTACCGATGG
SPD0013 RT R	ACAGATGGCGTGAAAAACTG
SPD1985 RT F	AAGGCACTTTTGGTGACAGA
SPD1985 RT R	GCTACCCCATCTGTGACATTC
SPD0058 RT F	TGACTCTGGATGGTTTGGAA
SPD0058 RT R	AATCATCCACGATACCAGCA
SPD0126 RT F	GCAGAAGAATCTCCCGTAGC
SPD0126 RT R	TTTCTGAGCAGCTTTTGCAT
SPD1614 RT F	TTGAATATTTATCAAATGAAGAACGA
SPD1614 RT R	AGTCATATGAAACGGGATGTTG
SPD0115 RT F	GCGCAGCTTATATACCCAAA
SPD0115 RT R	TGACATCTCCAGCATTTACCA
SPD1290 RT F	TGGCAAGCTGTTTAGCCTAC
SPD1290 RT R	TCACAATGGCAAAAAGAAGC

Primers ID	Nucleotide sequence (5′-3′)
SPD1339 RT F	CAGGCATTTTCGAAGAAAGAG
SPD1339 RT R	CTTTACGGCTACCAGCCAAT
SPD0343 RT F	TCACGTGGTTACACAGTTGC
SPD0343 RT R	TGATACGACGAGGTTTTTCG
SPD0373 RT F	AGTCAGCACCAGCAGAAGTG
SPD0373 RT R	CGACGGTGGATAGATGAGAC

2.20. Purification of recombinant protein

2.20.1. Amplification of target gene for cloning and transformation

Plasmid pLEICS-01 that has hexa histidine-tagged was used for protein expression. The target gene was amplified with PrimeSTAR HS premix (Section 2.7), using the target gene primers incorporating 15 nucleotides complementary to the multiple cloning sites in pLEICES-01 (Table 2.17). The amplified target gene was analysed by agarose gel electrophoresis and then purified using the Wizard® SV Gel and PCR Clean-Up System (Promega, UK) as described in section 2.8.

Table 2.17. Primers used for cloning the target gene. Italic typeface nucleotides are complementary to cloning site in pLEICES-01.

Primers ID	Nucleotide sequence (5′-3′)
T7 Promoter-F (PNACL)	TAATACGACTCACTATAGGG
pLIECS-01-Seq-R (PNACL)	ATTAACATTAGTGGTGGTGGT
ProteinSPD1518F (This study)	*TACTTCCAATCCATG*AGATGGGATTATGGACAAAT
ProteinSPD1518R (This study)	*TATCCACCTTTACTGTCA*TAGGTTTGAAAAATCATTAA

The target gene was cloned into pLEICS-01 by Dr Yang (PROTEX, Department of Molecular and Cell Biology, University of Leicester), using In-Fusion® HD Cloning Kit (Clontech, USA). The constructed recombinant plasmid was sequenced using T7

promoter-F, pLIECS-01-Seq-R primers and proteinSPD1518F/R primers (Table 2.17) to ensure no mutations had occurred. The constructed plasmid was transformed into *E. coli* BL21 (DE3) pLysS for protein expression (Section 2.12). To confirm the expected successful transformation, the PCR was performed using the recombinant plasmid as a template, the primers from Table 2.17 and HotStarTaq plus Master Mix (Section 2.7).

2.20.2. Small-scale protein expression

E. coli BL21 (DE3) pLysS carrying the recombinant construct was streaked on LB agar supplemented with 100 μg/ml ampicillin and incubated overnight at 37°C. A single colony was cultured in 10 ml of LB supplemented with 100 μg/ml ampicillin and then incubated overnight at 37°C in a shaking incubator at 220 g. The overnight culture was diluted (1 in 10) in fresh LB broth supplemented with 100 μg/ml ampicillin, then the culture was incubated at 37°C in a shaking incubator at 220 g for 1-2 h until the OD600 nm had reached 0.5-0.6. Next, 1 ml of the culture was centrifuged, and marked as "before induction pellet" and stored at -20°C until use. The remaining culture was induced using three different concentrations (0.1, 0.5 and 1 mM) of Isopropyl β-D-1-thiogalactopyranoside (IPTG) and subsequently, the induced culture was incubated at three different temperatures (18, 30 and 37°C) for approximately 9 h until the OD600 nm had reached 1.5-1.7. At this point, 1 ml of induced culture was centrifuged, and marked as "after induction". Both pellets were resuspended in 1 ml PBS and sonicated for 10 times using a sonicator (Sanyo soniprep 150, Japan) at an amplitude of 8 microns for 15 sec, and was placed on ice for 45 sec to prevent denaturation of protein due to overheating. The cell lysate was transferred to a sterile microcentrifuge tube and centrifuged at 14000g for 15 min at 4°C. Finally, the supernatant was transferred to a

sterile microcentrifuge tube and marked as "clear cell lysate" and the pellet was marked as 'after sonication'. Both samples were stored at -20°C for later use.

2.20.3. Large-scale protein expression and purification

The *E. coli* BL21 (DE3) pLysS carrying the desired construct was grown in LB supplemented with 100 µg/ml ampicillin at 37°C in a shaking incubator at 220 g. The overnight culture was used to inoculate 800 ml LB supplemented with 100 µg/ml ampicillin and incubated at 37°C in a shaking incubator set at 220 g for approximately 2 h. When the OD600 nm reached 0.5-0.6, the culture was induced using the optimal concentration of IPTG (1 mM), and subsequently, the culture was incubated at the optimal temperature (37°C), which had been determined in small-scale expression (Section 2.20.2), until the OD600 nm reached 1.5-1.7. At this point, the induced culture was centrifuged at 10000g for 30 min at 4°C (Sorvall Evolution$_{RC}$, USA). The supernatant was discarded while the pellet was kept at -80°C. Next day, the pellet was resuspended in 40 ml of binding buffer (20 mM Tris, 150 mM NaCl, pH 7.45), then 2 tablets of complete Mini, EDTA-free protease inhibitor cocktail tablets (Roche, Germany) and 30 µl of 10X TURBO DNase Buffer were added to the pellet. The suspension was sonicated as previously described in section 2.28.3. The cells lysate was then centrifuged at 25000g for 30 min at 4°C. After that, the clear cells lysate was transferred to a fresh tube. The 10 prepacked gravity flow column His GraviTrap^{TM} (GE Healthcare, UK) was used to purify the protein. Routinely, 100 ml of the binding buffer was used for column calibration. The clear cells lysate was passed three times through the column and was then washed by 100 ml of binding buffer to remove contaminating proteins from the column. The protein sample was eluted using different concentrations

(25, 50, 100, 300, 500 mM) of imidazole elution buffer (binding buffer and of imidazole) and the desired protein was collected.

2.20.4. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE)

SDS PAGE gel was used for analysis of recombinant proteins, and prepared according to the method described by Sambrook *et al.* (1989). SDS PAGE consisted of the resolving and stacking gel as indicated in Table 2.18.

Table 2.18. Composition of resolving and stacking gel.

Components	Resolving gel	Stacking gel
Tris-HCl 1.5 M or 1M* pH8.8	3.8 ml	0.63 ml
Acrylamide bisacrylamide 30% v/v	7.5 ml	0.83 ml
Ammonium persulphate (APS) 10% w/v	0.15 ml	0.05 ml
Sodium dodecyl sulphate (SDS) 10% w/v	0.15 ml	0.05 ml
dH_2O	3.4 ml	3.4 ml
Tetramethylethylenediamine (TEMED)	0.006 ml	0.005 ml

* 1Mfor stacking gel.

The resolving gel was poured between two glass plates in a gel caster and left to polymerise. Next, the stacking gel was poured on top of the resolving gel, and a comb was directly placed in the stacking gel to create wells, and the gel was also left to polymerise. For sample loading, 20 µl of recombinant protein mixed with 8 µl of loading buffer (6.25 ml 0.5M Tris-HCl pH 6.8, 1 g SDS, 50 mg bromophenol blue, 10 ml glycerol, 10 mM DTT and 6.25 ml dH_2O), and the samples were boiled at 95 ºC for 5 min. The SDS gel glass was placed in the mini protein tetra system tank (Bio-Rad, UK), and the running buffer (Tris/Glycine/SDS buffer, Bio-Rad, Germany) was then added into the tank. The comb was taken away gently and 30 µl of each sample and protein marker Precision Plus (Bio-Rad, UK) were loaded into the wells. An electric field (200 V) was applied across the gel causing negatively charged proteins to migrate across the gel towards the positive electrode.

Next, The SDS gels were stained using staining and destaining solution (Table 2.19). To do this, the SDS gels were placed in staining solution for 30 min with gentle shaking and the staining solution was then discarded. Next, the destaining solution was added to the gel for only 15 min with gentle shaking and was then discarded. Finally, the destaining solution was re-added to SDS gels for overnight. Next day, the destaining solution was discarded, and the gels were washed with dH_2O, and then scanned.

Table 2.19. Composition of staining and destaining solution.

Components	Staining solution	Destaining solution
Coomassie brilliant blue w/v	0.4 g	-
Absolute methanol	160 ml	-
Absolute acetic acid	40 ml	40 ml
dH_2O	200 ml	260 ml
Isopropanol	-	100 ml

2.20.5. Dialysis and MALDI-TOF analysis for recombinant protein

The recombinant protein was dialysed using the binding buffer. The eluted protein solution was placed in Amicon Ultra-15 Centrifugal Filter Units (Millipore, UK) with a filter size of 10 kDa. The binding buffer (40 ml) was added and the column was centrifuged for four times at 4000g for 20 min at 4°C until only 0.5 ml of protein remained. The protein solution was divided into 20 µl aliquots, and placed in liquid nitrogen for approximately 1 min and then stored at -80°C.

To confirm protein identity, the purified recombinant protein was analysed using matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry (PNACL, University of Leicester).

2.20.6 Quantitative analysis of protein

The Bradford method was used to measure the purified recombinant protein concentration (Bradford, 1976). Serial dilutions of bovine serum albumin (BSA) were prepared (50 to 1000 µg/ml) as standards to measure protein concentration in samples. 10 µl of BSA standards were transferred into a 96 wells cell culture plate, then 10 µl of recombinant protein sample was also transferred into the same microtitre plate. Next, 190 µl of Bradford reagent (Bio-Rad, UK) was added to each well. The microtitre plate was incubated at room temperature for 10-15 min with gentle shaking. Thereafter, the OD of each well was recorded at 595 nm using a microplate reader (Bio-Rad Model 680, UK). The standard curve representing the known concentrations of BSA was used as calibrator to determine the protein concentrations in the samples.

2.21. Electrophoretic mobility shift assay (EMSA)

The presence of regulatory elements in the putative promoter regions of target genes was identified using the bacterial promoter prediction tool (BPROM) (Solovyev and Salamov, 2011) and the motif-site was localized using MEME (Bailey and Elkan, 1994). The primers were designed to amplify putative promoter sites, approximately 100-200 bp DNA fragment upstream of the gene (Table 2.20).

All DNA fragments for EMSA were amplified from chromosomal *S. pneumoniae* D39 by PCR using 5′-FAM tagged primers. The PrimeSTAR HS premix (Section 2.7) was used for performing PCR. The PCR products were purified as previously described in section 2.9 and then analysed using agarose gel electrophoresis.

Table 2.20. Primers used for amplifying the putative promoter regions of the selected genes for EMSA.

Primers	Nucleotide sequence (5′-3′)
P*cps*EMSA FAMF	ACACATCTGCTTCTAAAATA
P*cps*EMSAR	GATTAACACCTATACATTGA
gyrB EMSAFAMF	ATGACAGAAGAAATCAAAAATCTGC
gyrB EMSAR	CCTGGACGCATACGAACAG
P*cps*motif1UPR	TTTCACTATCTGCATCTTTA
P*cps*motif 1downF	TAGACATTACCGTAAAAAAG
P*cps*motif 1downR	TTCGCTTCACTTTCTGTGAA
P*cps*motif 2UPR	TAATGTCTACACCTTTTTTC
P*cps*motif 2downF	AGTGATATAATCGTAAGATG
P*cps*motif 2downR	ACTATATTAACACTTCGCTT

EMSA was done according to the protocol described previously (Chang *et al.,* 2011) and used to analyse the interaction of the transcriptional regulator (Rgg1518) with capsule locus genes (SPD_0315-SPD_0328). The solutions for EMSA including 10X TB buffer, binding buffer and Polyacrylamide gel were prepared as shown in Table 2.21, 2.22 and 2.23.

Table 2.21. Composition of 10X TB buffer* for EMSA.

Components	g/l	Final concentration
Tris base	108	89 mM
Boric acid	55	89 mM

*dissolved in one litter of dH_2O, 10X TB buffer was autoclaved, 0.5 TB running buffer was prepared by dilute 50 ml of 10X TB buffer in 950 ml dH_2O then use it to run the gel.

Table 2.22. Composition of Binding buffer for EMSA.

Components	Volume µl	Final concentration
1M Tris-HCl pH 7.5	200	20 mM
1M KCl	300	30 mM
1M DTT	10	1 mM
500 mM EDTA pH 8.0	20	1 mM
100% glycerol	1000	10%
dH_2O	8470	Adjust to 10 ml

Table 2.23. Composition of polyacrylamide gel

Components	Gel ml
10X TB buffer	0.3
Acrylamide 30%	1.6
dH_2O	4.1
APS 10%	0.075
TEMED	0.005

Electrophoresis was carried out using chilled 0.5X TB running buffer which was used for preparing the gel. The gel was pre-run without samples for 20 min at 200 V to remove all traces of ammonium persulfate. Then, the reaction mixture containing 1 µl of 10 µM of FAM-labeled *cps* probe mixed with 1 µl of serial concentrations of His-tagged Rgg1518 protein (0.2-0.6 µM) and in 20 µl DNase-RNase free water that had been incubated for 20 min at 37ºC, was loaded on polyacrylamide gel after addition of 4 µl binding buffer. DNA-protein complexes were analysed in non-denaturing condition by PAGE (8%) and run for 40-60 min at 200 V. Once the electrophoresis was done, the gel was carefully removed from the mould. All gel shifts were detected by fluorescence imaging using a Typhoon PhosphorImager (Typhoon 9400, Amersham Biosciences).

2.22. Statistical analysis

GraphPad Prism version 7 (GraphPad, California, USA) was used for statistical analysis. Data were expressed as means and ± standard error of the mean (SEM). One-way and two-way analysis of variance (ANOVA) followed by Dunnett's multiple comparison tests, were used to compare the groups for growth studies. Moreover, the Mann Whitney test was used for virulence assessment by survival time while one-way ANOVA was used to compare the groups for bacteraemia development and colonisation experiment. Significance is represented as *$p<0.05$, **$p<0.01$, ***$p<0.001$ and ****$p<0.0001$.

Chapter 3. Results

Section A: Construction of mutants

To study the role of Rgg proteins (SPD_0999 and SPD_1518) in pneumococcal survival and virulence, the genes coding for these two putative transcriptional regulators, *rgg0999* and *rgg1518* were inactivated by allelic replacement mutation. The mutant strains were characterised phenotypically by growth studies, biochemical assays and gene expression analysis. The roles of these regulators and their response to host-derived sugars were determined by constructing transcriptional reporter systems in these wild type and mutant strains. The double Rgg0999 and Rgg1518 mutant was also constructed to study if there is a synergistic effect between these two proteins.

3.1 Construction of isogenic mutants and genetically complemented strains

Gene mutation was achieved by splicing overlap extension (SOEing) (Horton, 1995). The mechanism of this method is illustrated in Figure 3.1. This strategy is more advantageous than other strategies such as plasmid-based mutation techniques. In SOEing technique there is no need for cloning and propagation of recombinant plasmid in the surrogate host. In addition, it is quick and less laborious (Horton *et al*., 1993). This SOEing method allows production of desired genetic mutations within the desired genetic location. Achieving such an objective is a difficult task with other methods such as the mariner mutagenesis technique, which introduces random insertions in the target. The SOEing method is a widely used mutagenesis technique in *S. pneumoniae* because the microbe is naturally competent, which is a physiological state that gives the pneumococcus the ability to take up free DNA from the environment, and to incorporate this acquired DNA into its genomic DNA (Seitz and Blokesch, 2014).

The SOEing method was used to delete the target gene and replace it with an antibiotic resistance gene, either with the *aadA* that encodes aminoglycoside-3-adenylyltransferase (Aad), which confers resistance to spectinomycin or, with the *aph* that encodes aminoglycoside-O-phosphotransferase (Aph) and confers resistance to kanamycin. Furthermore, both *aadA* and *aph* were used to delete and replace *rgg0999* and *rgg1518,* respectively, to generate a double knockout mutant. The SOEing mutagenesis method includes two PCR steps. In the first step of PCR, the *aadA* or *aph* gene, and the upstream and downstream regions of target genes are amplified. In the second step of PCR, the upstream and downstream regions are fused to the amplified *aadA* or *aph* gene using the upstream-F and downstream-R primers. The fused PCR product was then introduced into *S. pneumoniae* by transformation, and homologous recombination replaces the target gene with *aadA* or *aph* **(**Figure 3.1).

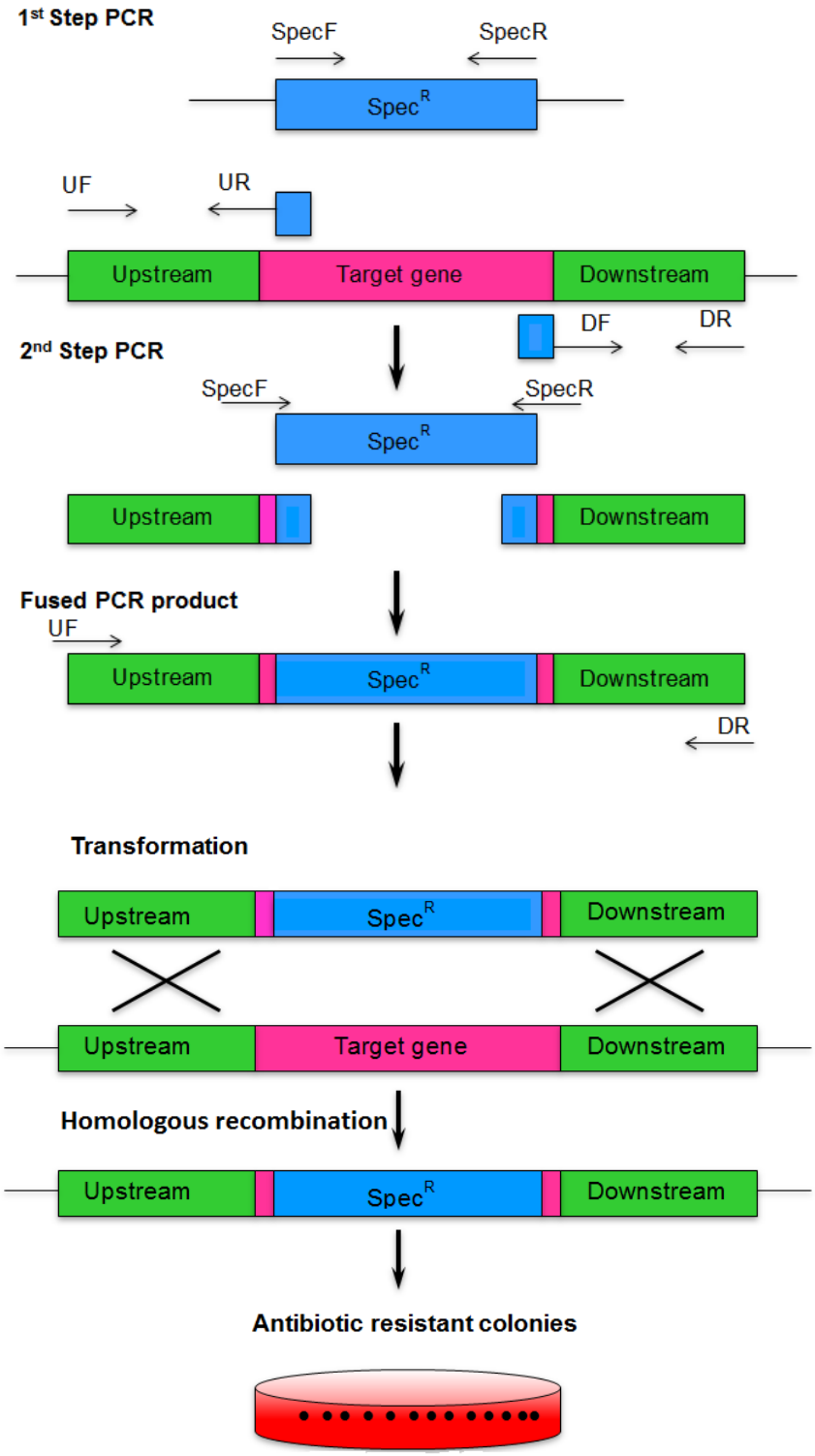


Figure 3.1. The illustration shows the steps of allelic replacement mutagenesis and gene knockout by homologous recombination. Upstream and downstream regions indicate left and right flanks respectively, (adapted from Horton, 1995; Song *et al.*, 2005).

3.1.1. Amplification of SOEing fragments

Upstream and downstream regions (600 bp) of each target genes and the *aadA* (1158 bp) or *aph* (895 bp) genes were amplified efficiently as described previously in section 2.14.1. The amplified fragments were then analysed by agarose gel electrophoresis (Figure 3.2). The results demonstrated the successful amplification of the antibiotic resistance genes (*aadA*, *aph*) and the up and downstream regions of each target gene. Lanes 1-2 and 3-4 contain the amplicons, approximately 600 bp in size, for the upstream and downstream regions of *rgg0999*, and *rgg1518*, respectively. Lanes 5 and 6 show the amplicons for *aadA* (1158 bp) from pDL278 (Yesilkaya, 1999) and *aph* (895 bp) from pCEP (Guiral *et al.*, 2006), respectively. Both *rgg0999* and *rgg1518* flanking regions shown in lanes 1-2 and 3-4 respectively, contained compatible ends with the spectinomycin cassette. While the *rgg1518* flanks shown in lanes 7-8 contained compatible ends with the kanamycin cassette and these amplicons were needed to construct kanamycin resistant *rgg1518* mutant in order to generate a double *rgg0999* and *rgg1518* mutant subsequently.

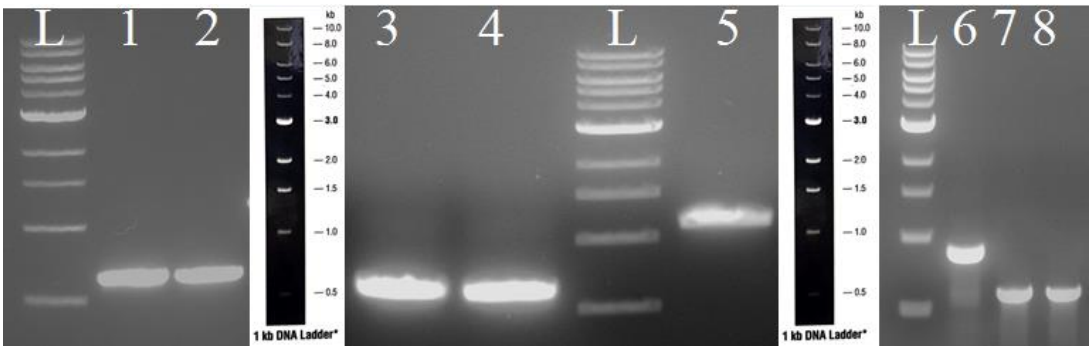


Figure 3.2. Agarose gel electrophoresis analysis shows the amplified flanking regions of each target genes. L: 500 ng of 1 kb DNA ladder (New England Biolabs, UK). Lanes 1 and 2: up- and downstream regions of *rgg0999*, lanes 3 and 4: up- and downstream regions of *rgg1518*, with compatible ends to the spectinomycin resistance gene. Lane 5: *aadA* amplified from the plasmid pDL278. Lane 6: *aph* amplified from the plasmid pCEP. Lane 7 and 8: up-downstream region and of *rgg1518*, respectively with compatible ends to the kanamycin resistance gene.

3.1.2. Fusion of SOEing fragments and transformation

The upstream and downstream regions of each gene were fused to the amplified resistance genes *aadA,* or *aph* using the gene specific SPDXUPstream/F and SPDXDOWNstream/R (X represents gene code) primers as previously described in section 2.14.2. Gel purification was performed to the fused PCR product for each target gene in order to avoid generating any nonspecific DNA products. The purified fused PCR products for target genes were subsequently analysed by agarose gel electrophoresis (Figure 3.3). The size of PCR products resulted from fusion process represented in lanes 1 and 2, approximately 2358 bp, was equal to the total size of upstream and downstream regions (approximately 600 bp each) plus 1158 bp of *aadA*, proving that the successful fusion had occurred. Lane 3 shows the fusion of upstream and downstream regions of *rgg1518* to the *aph* (895 bp), producing expected product size of approximately 2095 bp. The purified fusion products were then transformed into wild type D39 as described in section 2.14.2, and the colonies were selected on BAB plates supplemented with spectinomycin to generate mutations in *rgg0999* and *rgg1518.* The mutants were designated as *Δrgg0999* and *Δrgg1518*. To construct a double mutant, first, a mutation in *rgg1518* was constructed using *aph* as above. The resulting construct carrying mutation (lane 3) was then introduced into *Δrgg0999,* and colonies were selected on kanamycin and spectinomycin. The double mutant was designated as *Δrgg0999/1518*.

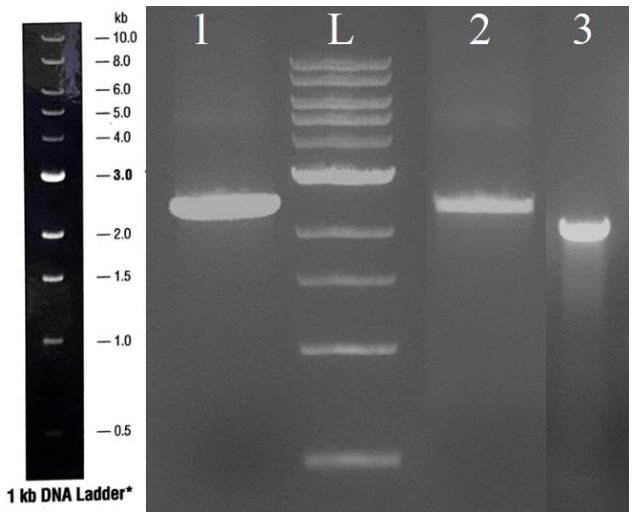


Figure 3.3. Agarose gel electrophoresis analysis shows the fused SOEing products. L: 500 ng of 1 kb DNA ladder (NEB). Lanes 1 and 2: fused SOEing fragments of approximately 2358 bp in size representing the flanking regions (up-downstream) of the genes *rgg0999* and *rgg1518* fused to the *aadA*, respectively, conferring resistance to spectinomycin. Lane 3: fused SOEing fragment of approximately 2095 bp in size representing the flanking regions of *rgg1518* fused to the *aph*, conferring resistance to kanamycin.

3.1.3. Mutation confirmation strategy

The successful construction of isogenic mutant strains was confirmed by PCR as described in section 2.14.3. The colony PCR strategy used to confirm gene replacement is illustrated in Figure 3.4 A and B. Figure 3.4A represents different primer combinations used to confirm the replacement of the target genes with the *aadA*. The gene specific SPDXUPstream/F and SPDXDOWNstream/R primers were used to amplify whole mutated regions from the extracted DNA of the putative mutants, and wild type as a control. These primers are expected to amplify a larger PCR product from the putative mutant strains compared to the wild type as the size of both deleted target genes are significantly less than the inserted *aadA* gene (1158 bp). Furthermore, the SPDXUPstream/F and Spec/R primers were used to amplify the upstream flanking region, and the *aadA*, while the SPDXDOWNstream/R and Spec/F primers amplified

the downstream flanking region and *aadA* of each target gene. This set of primers creates about 1758 bp from the successfully constructed mutants, whereas, there is no product from the D39 DNA as expected because the native pneumococcal genome lacks the *aadA* cassette. Additionally, the Spec/F and Spec/R primers were used to amplify the *aadA* (spectinomycin resistance gene) of approximately 1158 bp in size from the genomic DNA of each putative mutant strains. The double mutant (*rgg0999/1518*) was analysed using the same strategy as shown in (Figure 3.4B).

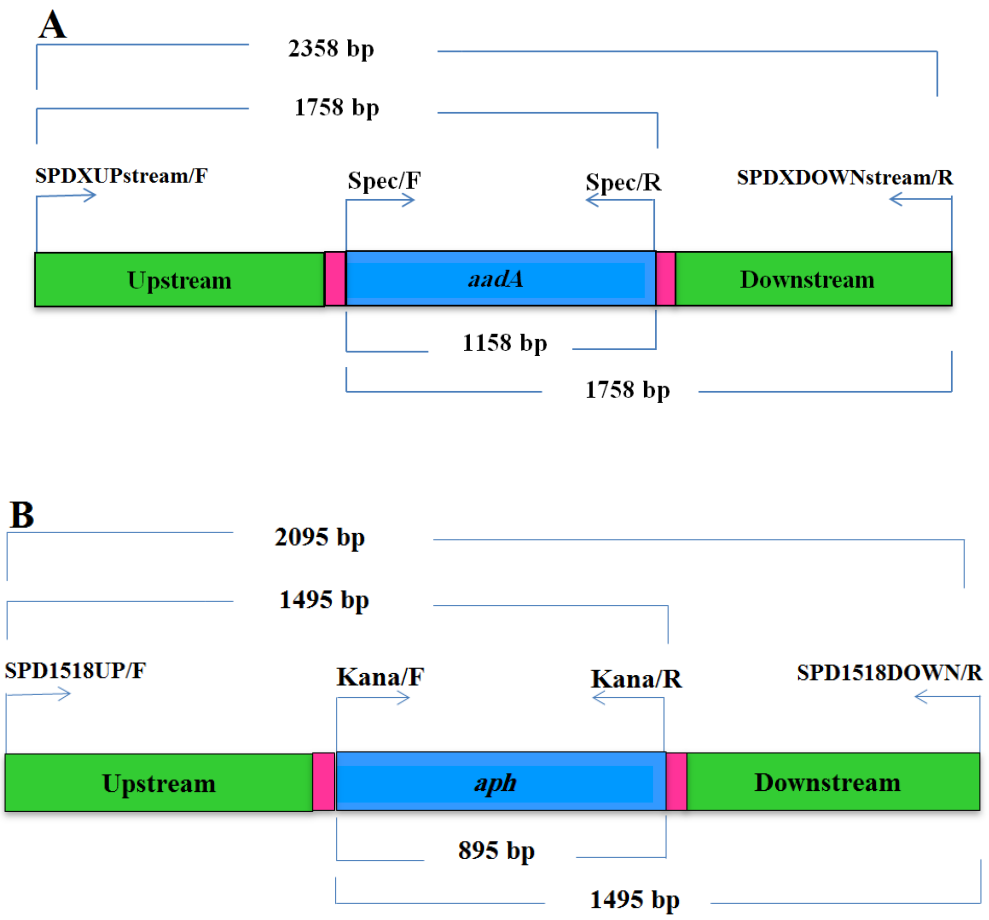


Figure 3.4. Schematic presentation shows the PCR strategy to confirm target gene replacement in mutant strains. Confirmation of successful mutation using spectinomycin resistance gene *aadA* (A) and kanamycin resistance gene *aph* (B). Upstream and downstream regions indicate the right and left flank of the target gene, respectively, and X is for target gene. The specific primers used for confirmation PCR strategy and the amplicon sizes are in line with expectation as indicated above. The sequence of each specific primer is shown in Table 2.11.

3.1.4. Confirmation of mutant strains

The putative mutant strains were grown overnight in BHI supplemented with the appropriate antibiotic (spectinomycin or kanamycin) at 37°C. The mutant genomic DNA was extracted from these cultures as described in section 2.5. The extracted genomic DNA of putative mutants was analysed by agarose gel electrophoresis (Figure 3.5).

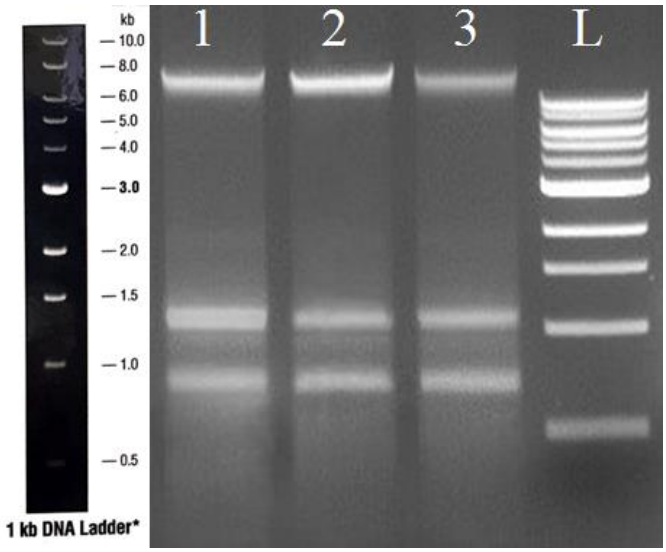


Figure 3.5. Agarose gel electrophoresis analysis showing the genomic DNA of putative mutants. Lanes 1-2: show the genomic DNA from putative mutants: *Δrgg0999* and *Δrgg1518*, respectively with the *aadA* cassette. Lane 3: shows genomic DNA extracted from *Δrgg0999/1518* with the *aadA*/*aph* cassette. L: 500 ng of 1 kb DNA ladder (NEB).

Confirmation of the successful gene replacement with *aadA* and *aph* cassettes was performed by PCR using the genomic DNA from the putative mutant strains as templates, including the wild type DNA as a control. The amplified PCR products were routinely analysed by agarose gel electrophoresis. Figure 3.6 shows successful replacement of *rgg0999* (A) and *rgg1518* (C) with the *aadA* spectinomycin resistance gene. Lane 1 in each picture shows PCR products amplified with Spec/F and Spec/R

primers from each putative mutant genomic DNA, all the amplified products had the expected size of *aadA* gene cassette (1158 bp). Lanes 2 and 3 in each picture show PCR products amplified with the SPDXUP/F and Spec/R or SPDXDOWN/R and Spec/F primers, respectively, from each putative mutant genomic DNA. All the amplified products had the expected size of 1758 bp, which is equivalent to the size of the up or downstream flanking region of each gene (600 bp) plus *aadA* (1158 bp). Lane 4 in each picture shows PCR products amplified with SPDXUP/F and SPDXDOWN/R primers from each putative mutant genomic DNA. All amplified DNA had also the expected size of 2358 bp, which is equivalent to the size of *aadA* (1158 bp) plus1200 bp of the up-downstream regions.

Picture B and D in Figure 3.6 show confirmation of successful replacement of *rgg0999* and *rgg1518* with the *aadA* spectinomycin resistance gene. Lanes 1, 2 and 3 in each picture show the absence of PCR products because the wild type genomic DNA lacks spectinomycin cassette hence no amplicons were obtained when using primers Spec/F and Spec/R, SPDXUP/F and Spec/R in addition to SPDXDOWN/R and Spec/F. Moreover, Lane 5 in each picture shows PCR products amplified with the gene specific SPDXUP/F and SPDXDOWN/R primers from the wild type DNA. All the amplified products had the expected sizes (2064 bp and 2052 bp) of the target genes (*rgg0999* (864 bp) and *rgg1518* (852 bp)) plus 1200 bp up and downstream regions.

The double mutant lacking both *rgg0999* and *rgg1518* was constructed by introducing the SOEing product carrying the mutated *rgg1518* with *aph* into *Δrgg0999* by transformation. The transformants were selected on BAB plates supplemented with 100 μg/ml spectinomycin and 200 μg/ml kanamycin. The putative double mutant genomic

DNA was extracted in order to confirm the successful gene (*rgg1518*) replacement with kanamycin cassette by PCR. Picture E and F in Figure 3.6 illustrate confirmation of successful replacement of *rgg1518* with the *aph* kanamycin resistance gene using *Δrgg0999* genomic DNA as a control. Lane 1 in picture E shows PCR products amplified with Kana/F and Kana/R primers from putative double mutant genomic DNA, the amplified products had the expected size (895 bp) of the *aph* gene cassette. Lanes 2 and 3 in picture E show PCR products amplified with the SPD1518UP/F and Kana/R or SPD1518DOWN/R and Kana/F primers, respectively, from putative double mutant genomic DNA. All the amplified products had the expected size of 1495 bp, which is equivalent to the size of the up-downstream regions of each gene (600 bp) plus the *aph* (895 bp), while lane 4 in shows PCR product amplified with gene specific SPD1518UP/F and SPD1518DOWN/R primers from putative double mutant genomic DNA. The amplified DNA had the expected size of 2095 bp, which is equivalent to the size of *aadA* (895 bp) plus1200 bp representing up and downstream regions. Lane 1 in picture F shows PCR products amplified with the gene specific SPD1518UP/F and SPD1518DOWN/R primers from the *Δrgg0999* genomic DNA. The amplified products had the expected sizes (2052 bp) of the target gene (*Rgg1518* (852 bp) plus 1200 bp representing up and downstream regions. Moreover, lanes 1, 2 and 3 in picture F show the absence of PCR product when *Δrgg0999* genomic DNA was used as a template as it lacks the kanamycin resistance cassette.

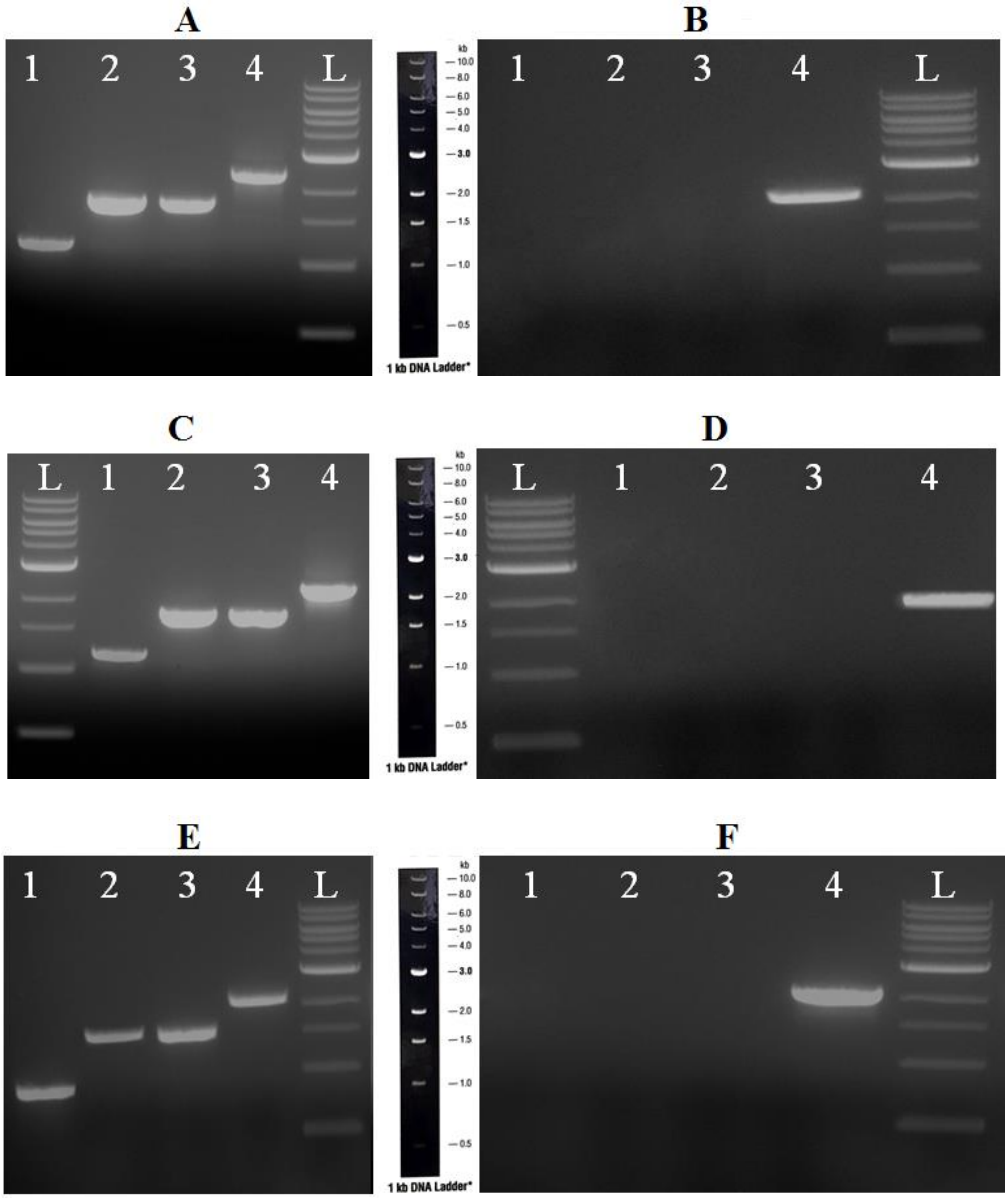


Figure 3.6. Agarose gel electrophoresis analysis confirms mutation of transcriptional regulators. (A) *Δrgg0999*, (B) control reactions for *Δrgg0999*, (C) *Δrgg1518,* (D) control reactions for *Δrgg1518*, (E) *Δrgg0999/1518* (F) control reactions for double mutant (using genomic DNA of *Δrgg0999* as a template). In (A) and (C) the amplicons were obtained using mutant strains DNA as template. Lanes 1: amplified *aadA*, lanes 2: upstream region + *aadA*, lanes 3: downstream region +*aadA*, lanes 4: upstream region + *aadA* + downstream region. In (B) and (D) using wild type DNA as template. Lane 1: WT DNA lacking *aadA*. Lanes 2: WT DNA lacking upstream region + *aadA*, lanes 3: WT DNA Lacking downstream region + *aadA*, lanes 4: upstream region + wild type gene + downstream region. In (E) the amplicons were obtained using double mutant strains DNA as a template. Lanes 1: amplicons for *aph*, lanes 2: upstream region + *aph*, lanes 3: downstream region + *aph*, lanes 4: upstream region + *aph* + downstream region. In (F) *Δrgg0999* DNA was used as a control that lacks *aph* gene. Lanes 1- 4: used similar strategy in (E). L: 500 ng of 1 kb DNA ladder (NEB).

3.1.5. DNA sequencing of mutants

DNA sequencing was done for successful gene replacements in order to ensure no undesired mutation had occurred at the junction points. The primers SPDX-seqM/F1 and SPDX-seqM/R2 (X refers to gene code) were designed to amplify 200 bp up- and downstream 'external' sequence encompassing the mutated region as shown in Figure 3.7. The second set of primers, designated as SPDX-seqM/F2 and SPDX-seqM/R1 located 228 bp up or downstream of the mutated region for Δ*rgg0999*, respectively, while for Δ*rgg1518*, the primers located 193 bp up or downstream of the mutated region, respectively.

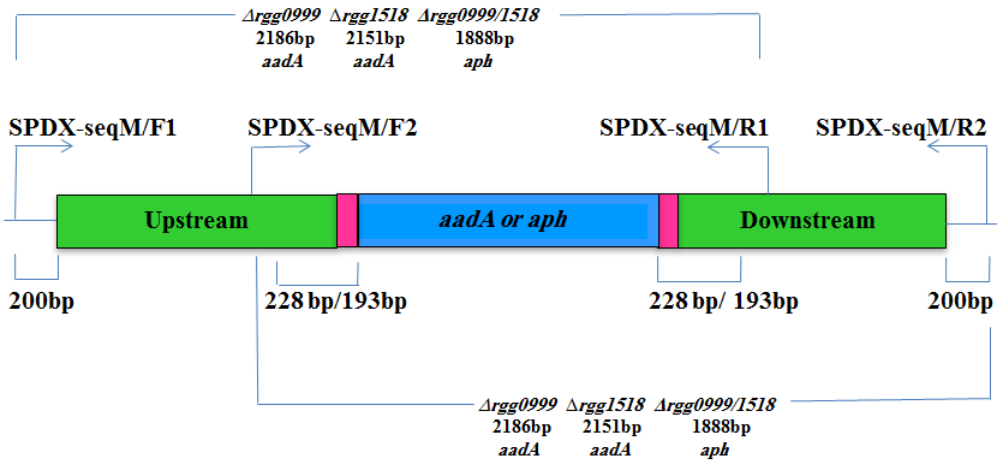


Figure 3.7. The illustration shows the sequencing strategy for mutants. X refers to gene code. The sequence of the primers is listed in Table 2.12.

As illustrated in Figure 3.8, the primers (SPDX-seqM/F1, SPDX-seqM/R1 and SPDX-seqM/F2, SPDX-seqM/R2) were used to amplify the mutated regions using the genomic DNA of each mutant strain as a template. The amplified products were then analysed using agarose gel electrophoresis. The successful amplification of the mutated regions from *Δrgg0999*, *Δrgg1518* and *Δrgg0999/1518* are shown in Figure 3.8. All the

amplified products of *Δrgg0999* represented in lane 1 and 2 had the expected amplicon size of approximately 2186 bp, which equals to the total sizes of *aadA* (1158 bp) plus 1028 bp up and downstream regions of mutated genes. The amplified products of *Δrgg1518* represented in lanes 3 and 4 had the expected size of approximately 2151 bp, which equals to the total size of *aadA* (1158 bp) plus 993 bp up and downstream regions of the mutated genes. Furthermore, all the amplified products in lane 5 and 6 obtained with the SPD1518-seqM/F1, SPD1518-seqM/R1 and SPD1518-seqM/F2, SPD1518-seqM/R2 primers, respectively using *Δrgg0999/1518* double mutant DNA as a template, which generated an estimated 1888 bp amplicon, equal in size to the *aph* (895 bp) plus 993 bp up and downstream regions of mutated genes.

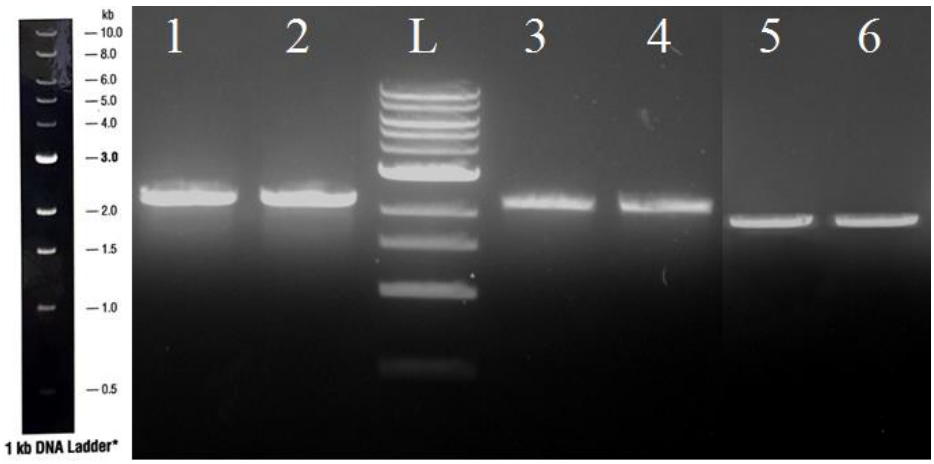


Figure 3.8. Agarose gel electrophoresis analysis showing the amplifications for the mutated regions for sequencing. Lanes 1-2 and 3-4: amplified mutated regions from *Δrgg0999*, *Δrgg1518*, using SPDX-seqM/F1, SPDX-seqM/R1 and SPDX-seqM/F2, SPDX-seqM/R2 primers, respectively. L: 500 ng of 1 kb DNA ladder (NEB). Lanes 5 and 6: amplicons representing *rgg1518* mutated region in the *Δrgg0999/1518* double mutant.

The amplified products, confirmed by agarose gel electrophoresis, were purified using Wizard® SV Gel and PCR Clean-Up System kit and then sequenced using the gene specific sequencing primers SPD0999-seqM/F1, SPD0999-seqM/R1, SPD0999-seqM/F2, SPD0999-seqM/R2, SPD1518-seqM/F1, SPD1518-seqM/R1, SPD1518-seqM/F2 and SPD1518-seqM/R2 as listed in Table 2.12. The sequencing results proved successful replacement of the target genes with the *aadA* or *aph* resistance genes, and no undesired mutations could be detected (Appendix 1).

3.2. Genetic complementation of mutants

Insertion deletion mutation can affect the expression of downstream genes, which is called a polar effect (Reyrat *et al.*, 1998). The possibility of a polar effect caused by a insertion-deletion mutation is likely to be higher if the target gene is the first gene of an operon (Guiral *et al.*, 2006). Therefore, complementation of mutant strains is needed to eliminate the potential polar effect that could have been caused by mutation. The pCEP plasmid was used for the construction of complemented strains as shown in Figure 3.9 (Guiral *et al.*, 2006). The size of pCEP is approximately 9540 bp, which is non-replicative plasmid in *S. pneumoniae*. Moreover, the multiple cloning site of pCEP is surrounded by 2 kb DNA homologous to *S. pneumoniae* D39 genome. Hence, this homology in sequence leads to an immediate successful integration of the intact copy of gene downstream of the *amiA* operon. This region is believed to be a transcriptional silent site and therefore does not affect any cellular physiological functions (Guiral *et al.*, 2006).

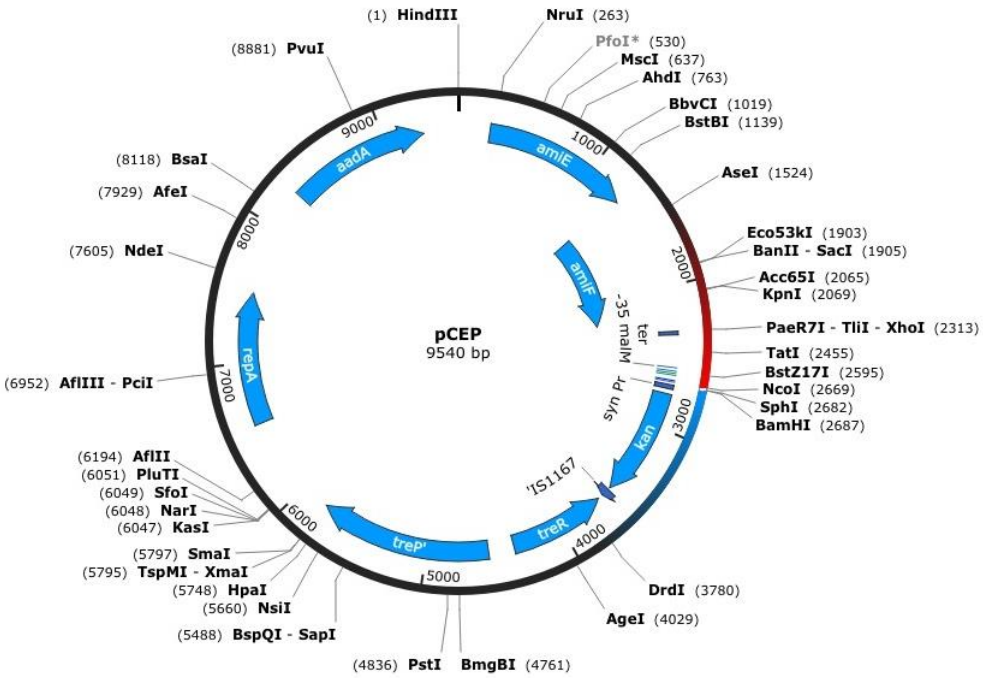


Figure 3.9. Schematic illustration shows pCEP genetic map (Guiral *et al.*, 2006). The *amiE* and *amiF* genes related to the oligopeptide permease *ami* operon. The *treR* and *treP* are segments of trehalose utilisation system. *NcoI* and *BamHI* are restriction sites. The *malR* repressor gene: maltose-inducible promoter.

3.2.1. Extraction and digestion of pCEP

The pCEP plasmid was extracted using QIAprep Spin Miniprep kit (Qiagen, UK) and was then double digested with *NcoI-BamHI* for each complementation (*Δrgg0999* and *Δrgg1518*), as described in section 2.12. The digested plasmid was subsequently purified to exclude any residual enzymes and salts from the DNA. The successful digestion of pCEP was analysed by agarose gel electrophoresis and compared to the undigested pCEP (Figure 3.10). Lane 1 shows *NcoI-BamHI* digested pCEP plasmid approximately of 9.5 kb. Lane 2 shows undigested pCEP plasmid.

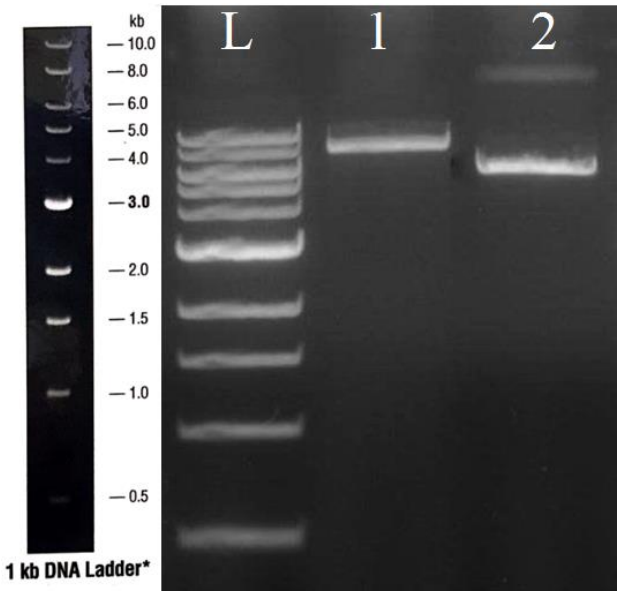


Figure 3.10. Agarose gel electrophoresis analysis demonstrates digested and undigested pCEP plasmid. L: 500 ng of 1 kb DNA ladder (NEB). Lane 1: shows *NcoI-BamHI* digested pCEP. Lane 2 shows undigested pCEP.

3.2.2. Amplification of target genes for genetic complementation

The intact copies of *rgg1518* and *rgg0999* and their putative promoter regions were amplified using the primers listed in Table 2.13. *NcoI-BamHI* specific sites were incorporated in *rgg0999* and *rgg1518* primers as described in section (2.12). The PCR products were purified using the Wizard® SV Gel and PCR Clean-Up System (Promega, UK). Agarose gel electrophoresis analysis was performed to confirm the successful amplification (Figure 3.11). *NcoI* and *BamHI* enzymes were used to double digest the purified inserts.

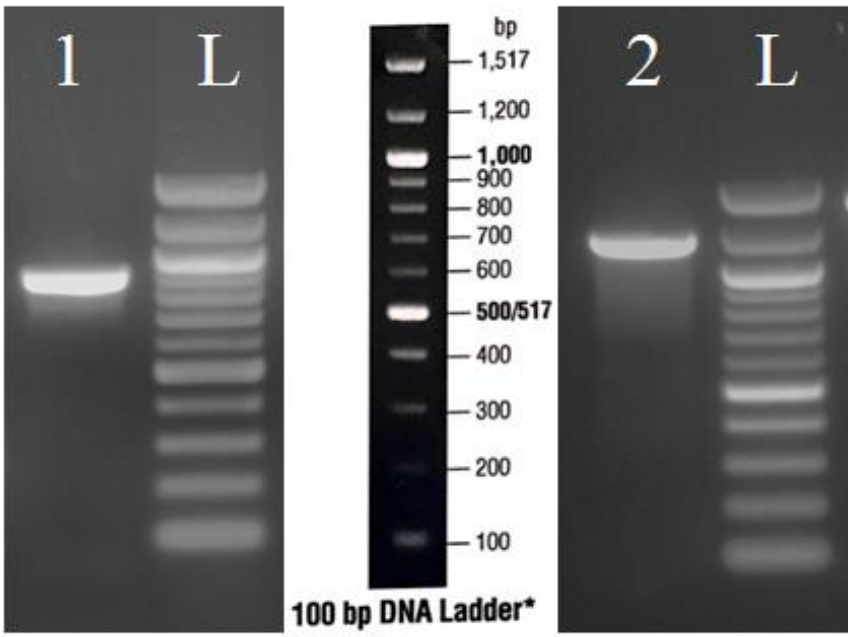


Figure 3.11. Agarose gel electrophoresis analysis showing the amplified genes for complementation of mutant strains. Lane 1: amplicons represent the putative promoter- and coding regions of *rgg0999* (864 bp). L: 500 ng of 100 bp DNA ladder (NEB). Lane 2: amplicons represent the putative promoter region and the coding region of *rgg1518* (1178 bp).

3.2.3. Construction of recombinant pCEP and transformation into *E. coli* TOP10

The inserts and vector that had compatible ends were ligated as described in section 2.11. The two ligation reactions were transferred separately into *E. coli* TOP10 chemically competent cells for propagation and to confirm cloning for successful ligation. Thereafter, the recombinant pCEP was extracted from kanamycin-resistant colonies as described in section 2.10. The successful cloning was confirmed by PCR using SPDXF*NcoI*C, SPDXR*BamHI*C primers (X refers to the gene), and MalF, PCEPR primers, whose recognition sites are localised in immediately up and downstream of the cloning site, respectively. In the absence of insert, MalF and PCEPR primers amplify about 263 bp products from the empty pCEP plasmid (Figure 3.12).

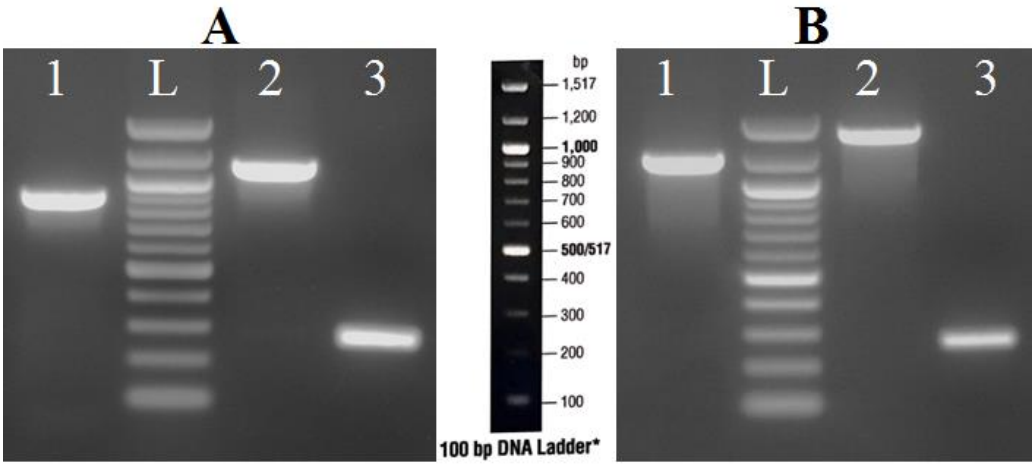


Figure 3.12. Agarose gel electrophoresis shows successful cloning of the inserts into pCEP, (A) *rgg0999* and (B) *rgg1518*. L: 500 ng of 100 bp DNA ladder (NEB). Lane 1 in A and B represents the amplicons for coding region of *rgg0999* (864 bp), and the putative promoter region and the coding region of *rgg1518* (1178 bp), respectively. Lanes 2: in each picture represent the amplified inserts plus upstream and downstream of cloning site by Mal F/ PCEPR primers, *rgg0999* (1127 bp) and *rgg1518* (1441 bp), respectively. Lanes 3: amplicons for the empty PCEP using Mal F/ PCEPR primers with a size of approximately 263bp.

The successful construction of recombinant pCEP plasmids carrying the intact copy of insert was further analysed by DNA sequencing using Mal F/ PCEPR primers. The results of sequence analysis confirmed the successful cloning of the inserts (Appendix 2).

3.2.4. Transformation of recombinant pCEP constructs into *S. pneumoniae* and PCR confirmation

The recombinant pCEP constructs were transformed into the relevant pneumococcal mutant strains, *Δrgg0999* and *Δrgg1518,* as described in section 2.15.3. The transformants resistant to spectinomycin and kanamycin were selected to confirm the successful integration of the inserts within the pneumococcal genome by PCR using

MalF and PCEPR primers, these primers have recognition sites that are localised in immediately up and downstream of the cloning site, respectively. Moreover, the MalF and PCEPR primers amplified approximately 263 bp in size for pCEP in the absence of insert. The PCR products were analysed by agarose gel electrophoresis (Figure 3.13). Lanes 1 and 2 represent the expected amplified PCR products, approximately 1127 bp and 1441 bp from complemented strains for *Δrgg0999* and *Δrgg1518*, respectively. The complemented strains were successfully confirmed by DNA sequencing and were then designated as *Δrgg0999c* and *Δrgg1518c*.

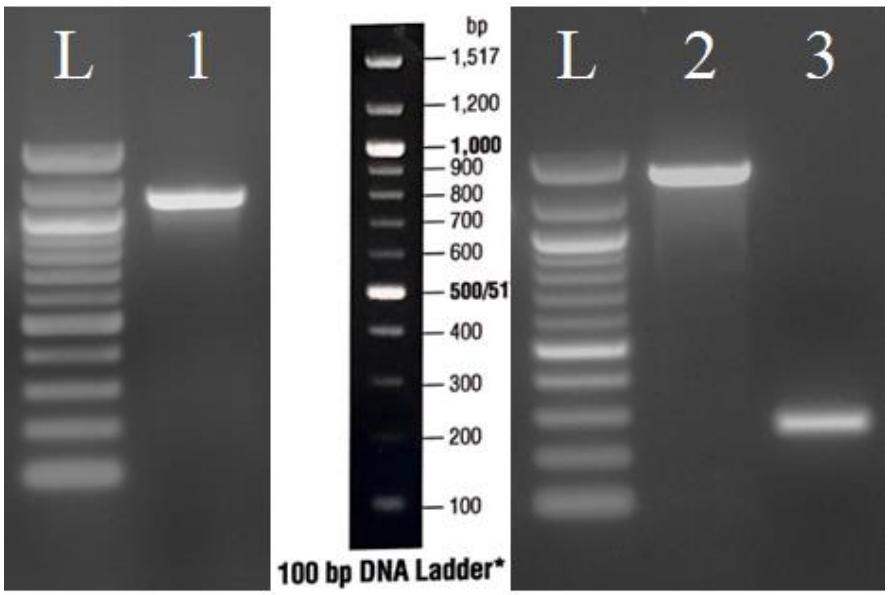


Figure 3.13. Agarose gel electrophoresis analysis showing the successful genetic complementation of mutants. L: 500 ng of 100 bp DNA ladder (NEB). Lane 1: shows amplified intact copy of *rgg0999* (1127 bp) from *Δrgg0999c*. Lane 2: shows amplified intact copy of *rgg1518* (1441 bp), including its putative promoter region from *Δrgg1518c*. Lane 3: amplified empty PCEP with a size of approximately 263 bp using MalF and PCEPR primers.

Section B: Microarrays analysis

3.3. Microarray analysis for *Δrgg0999* and *Δrgg1518*

The influence of mutation on the transcriptome of *S. pneumoniae* was examined using microarray analysis. The level of mRNA expression for all pneumococcal strains was analysed using microarray analysis in collaboration with Prof Oscar Kuipers and Irfan Manzoor, University of Groningen, the Netherlands. Microarray analysis is an ideal tool to reveal the genome wide changes due to Rgg0999 and Rgg1518 mutations. The wild type D39 and its isogenic mutants (*Δrgg0999* and *Δrgg1518*), were grown microaerobically in CDM supplemented with 55 mM of galactose or mannose as an essential source of carbon as these sugars are known to be plentiful in the respiratory tract within the structure of complex host glycans (King, 2010; Philips *et al.*, 2003). When the growth reached mid- exponential phase, the total RNA was successfully extracted, and treated with DNase as described in section 2.19.1 and 2.19.2, respectively. The total extracted RNA was analysed by agarose gel electrophoresis (Figure 3.14). The experiment was repeated with three biological replicates. Lanes 1and 2 show the total extracted RNA from wild type D39 in CDM supplemented with 55 mM galactose and mannose, respectively. Lanes 3 and 4 show the total extracted RNA from *Δrgg0999* in CDM supplemented with 55 mM galactose and mannose, respectively. Lanes 5 and 6 show the total extracted RNA from *Δrgg1518* in CDM supplemented with 55 mM galactose and mannose, respectively. The concentration of each RNA sample was measured spectrophotometrically. As these images show, the extracted RNA was intact and virtually free from the genomic DNA.

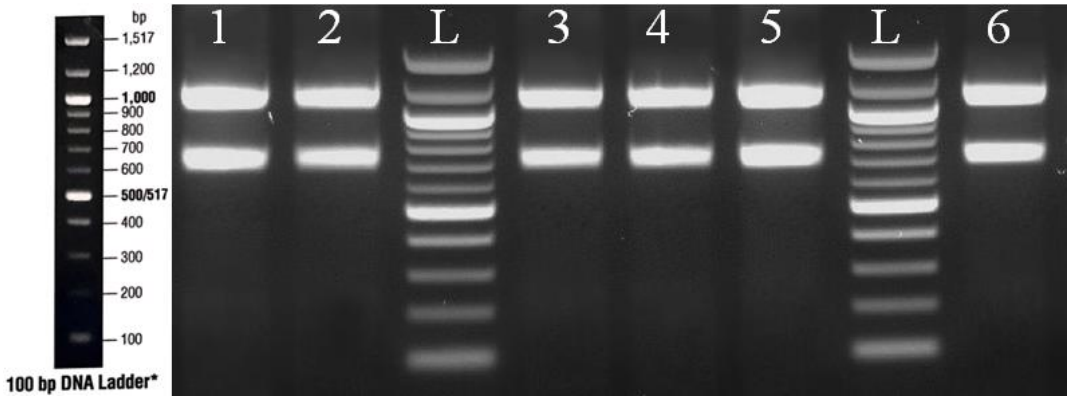


Figure 3.14. Agarose gel electrophoresis showing the total RNA extracted from the pneumococcal strains in different sugars. Lanes 1 and 2: show RNA of wild type D39 grown microaerobically in CDM supplemented with 55 mM of galactose or mannose, respectively. L: 500 ng of 100 bp DNA ladder (NEB). Lanes 3 and 4 RNA of *Δrgg0999* grown microaerobically CDM supplemented with 55 mM of galactose or mannose, respectively. Lanes 5 and 6 RNA of *Δrgg1518* grown microaerobically in CDM supplemented with 55 mM of galactose and mannose, respectively.

The results showed that the largest number of differentially expressed genes in *rgg* mutant strains was seen on galactose. Moreover, the genes upregulated in *Δrgg0999* had a substantial overlap to those upregulated in *Δrgg1518*. The Venn diagram (Figure 3.15) clearly shows the relative transcript levels in *Δrgg1518* and *Δrgg0999* in galactose were compared to *Δrgg0999* and *Δrgg1518* in mannose as generated by VENNY (http://bioinfogp.cnb.csic.es/tools/venny/index.html). On galactose, the microarray data showed that 378 genes were differentially expressed in the *Δrgg1518* compared to wild type strain: 80 genes were down-regulated and 298 were up-regulated, whereas 335 genes were differentially expressed in the *Δrgg0999* compared to wild type strain: 251 genes were down-regulated and 85 were up-regulated (Appendix 5).

The high number of down-regulated genes in *Δrgg0999* means that this regulator acts as an activator for these genes. Overall, the expression of operons that were affected by the absence of Rgg0999 included those, putatively encoding for ribosomal protein (SPD_0192-SPD_0203), hypothetical protein (SPD_0114-SPD_0124), capsule biosynthesis locus (SPD_0315-SPD_0328), tagatose 1,6-diphosphate (SPD_1050-SPD_1053), and CAAX amino terminal protease family protein (SPD_0146-SPD_0147), which have been found to have a role in biofilm formation and virulence, and to be regulated by Rgg0144 in *S. pneumoniae* (Cuevas *et al.*, 2017).

On the other hand, in Δ*rgg1518*, it was found that the expression of operons were affected by the absence of Rgg1518. The operons that putatively encode an iron (III) ABC transporter, a permease protein (SPD_0223-SPD_0224), a hypothetical protein (SPD_0946-SPD_1322), and PTS system (SPD_0424-SPD_0426) were down-regulated in the *Δrgg1518*. Interestingly, the operons (SPD_0192-SPD_0203, SPD_0114-SPD_0124, SPD_0315-SPD_0328, SPD_1050-SPD_1053 and SPD_0146-SPD_0147) were down-regulated in the *Δrgg0999* while they were up-regulated in the *Δrgg1518* relative to the wild type. Additionally, the operon (SPD_0424-SPD_0426) was down-regulated in the *Δrgg1518* while it was up-regulated in the *Δrgg0999* relative to the wild type. Thus, this shows that each Rgg has a specific role under the same environmental condition.

When the *rgg* mutants were grown on mannose, the size of putative regulon for both *rgg* (*rgg1518* and *rgg0999*) was smaller than that on galactose (Appendix5). The microarray data revealed that 51 genes were differentially expressed in the *Δrgg1518* relative to wild type strain: 44 genes were down-regulated and 7 were up-regulated. For

Rgg1518, the notable differentially expressed loci were those coding for tagatose pathway (SPD_1050-SPD_1053), a region encoding for hypothetical proteins (SPD_0114-SPD_0116), and ribosomal protein (SPD_0550-SPD_1245) which were down-regulated in the *Δrgg1518* relative to the wild type.

On the other hand, 24 genes were differentially expressed in the *Δrgg0999* relative to wild type strain, five genes were down-regulated in the *Δrgg0999* and 19 were up-regulated (Appendix 5). The operon that encodes for hypothetical proteins (SPD_0115-SPD_0116) was down-regulated in the *Δrgg0999* relative to the wild type, while the gene that encodes for capsule biosynthesis (SPD_0328) was up-regulated in the *Δrgg0999* on mannose relative to the wild type. This shows that under different environmental conditions, the same Rgg can act either as a repressor or the activator for the same target. I also found that the regulation exerted by different Rggs on the same target is influenced by the different sugars, For example, the genes that encode for hypothetical proteins (SPD_0114-SPD_0124) were up-regulated in the *Δrgg1518* on galactose while they were down-regulated in the *Δrgg0999* on mannose relative to the wild type. Interestingly, the gene that encodes for amino acid ABC transporter, permease protein (SPD_1290) was down-regulated in both Rggs (*Δrgg1518* and *Δrgg0999*) on mannose relative to the wild type.

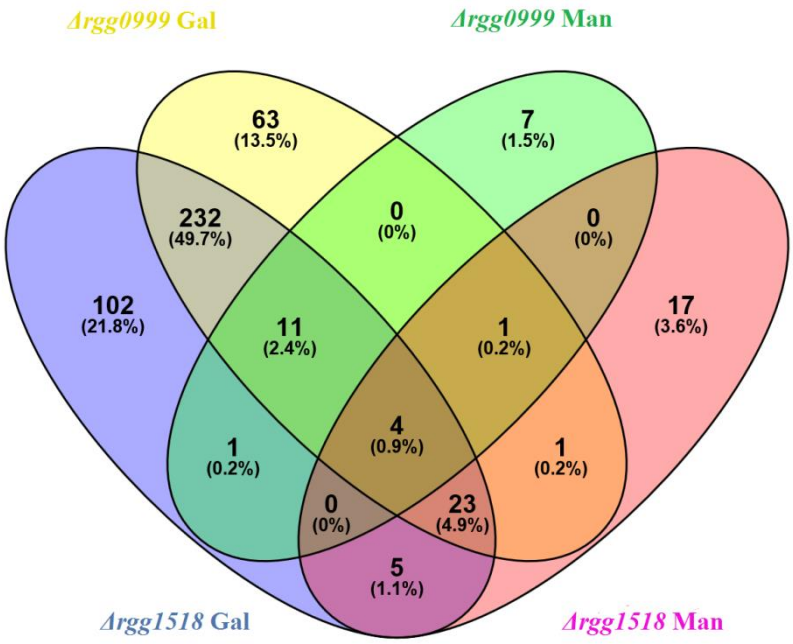


Figure 3.15. Venn diagram of the genes significantly differentially transcribed due to loss of *rgg0999* and *rgg1518*. Transcript levels in Δ*rgg0999* or Δ*rgg1518* in galactose were compared to Δ*rgg0999* or Δ*rgg1518* in mannose. For all intersections, which are not drawn to scale, the numbers of genes are indicated the total number of genes influenced in each condition tested was Δ*rgg1518* Gal: 232 genes, Δ*rgg0999* Gal: 4 genes, Δ*rgg1518* Man: 23 genes, and Δ*rgg0999* Man: 11genes. Gal: galactose, Man: mannose.

3.4. Confirmation of microarrays analysis by qRT-PCR

To confirm microarray results, the quantitative reverse transcriptase PCR (qRT-PCR) was used for selected genes. Random primers were used to synthesise the cDNA as described in section (2.19.4). DNA gyrase B gene (*gyrB*) is known as the housekeeping gene, and its expression remains constant in various environmental conditions (Goerke *et al.*, 2001). Therefore, *gyrB* was appropriate to be selected to normalise the expression of target genes. Gene specific primers from Table 2.16 were used to analyse transcriptional level of each target gene as described in section (2.19.5).

qRT-PCR data was analysed by the Comparative CT Method ($2^{-\Delta\Delta CT}$) after normalising the expression of each gene to that of *gyrB*. More than twofold difference in expression was treated as significant (Yesilkaya *et al*., 2009). Then the expression of target transcriptional regulators in the mutant strains was compared to wild type D39. The results of the qRT-PCR analysis were apparently similar to microarray data (Table 3.1 and Table 3.2).

Table 3.1. Fold difference in expression of each target gene in *Δrgg0999* or *Δrgg1518* compared to wild type D39 grown microaerobically in CDM supplemented with galactose.

Gene tag	Function	Fold difference *	
		Microarray	qRT-PCR
Δrgg0999			
SPD_0161	hypothetical protein	-2.54	-3.08 ± 0.05
SPD_0013	cell division protein FtsH	-2.35	-2.33 ± 0.01
SPD_1985	alcohol dehydrogenase, iron-containing	2.02	3.1 ± 0.1
SPD_0058	phosphoribosylamine--glycine ligase	2.48	2.39 ± 0.07
Δrgg1518			
SPD_0126	pneumococcal surface protein A	2	2.7 ± 0.1
SPD_0013	cell division protein FtsH	2.82	2.33 ± 0.01
SPD_1614	phosphate transport system regulatory protein PhoU, putative	-2.24	-2.2 ± 0.12
SPD_0058	phosphoribosylamine--glycine ligase	-2.48	-2.39 ± 0.07

* Fold changes ≥ 2 were known to be significant, ± refers to the standard deviation of six measurements. '-' indicates down regulation

Table 3.2. Fold difference in expression of each target gene in *Δrgg0999* or *Δrgg1518* compared to wild type D39 grown microaerobically in CDM supplemented with mannose.

Gene tag	Function	Fold difference *	
		Microarray	**qRT-PCR**
Δrgg0999			
SPD_0115	hypothetical protein	-2.02	-3.8 ± 0.2
SPD_1290	amino acid ABC transporter, permease protein	-2.56	-2.43 ± 0.06
SPD_1339	ATP synthase F0, B subunit	2.18	3 ± 0.2
Δrgg1518			
SPD_0343	6-phosphogluconate dehydrogenase, decarboxylating	-2	-2.9 ± 0.2
SPD_1290	amino acid ABC transporter, permease protein	-2.31	-2.43 ± 0.06
SPD_0373	hypothetical protein	1.98	2.9 ± 0.2

* Fold changes ≥ 2 were known to be significant, ± refers to the standard deviation of six measurements. '-' indicates down regulation.

Section C: *In vitro* characterisation of pneumococcal strains

3.5. Phenotypic characterisation of the mutants

There is a strong relationship between pneumococcal virulence and efficient sugar acquisition and catabolism (Yesilkaya *et al.,* 2009). The pneumococcus is usually present in the multi-sugar environment in the host tissues, and it can utilise 32 different sugars such as glucosides, galactosides and polysaccharides (Bidossi *et al.*, 2012). This allows the pathogen to utilise the available sugar in different host tissues (Bidossi *et al.*, 2012; Terra *et al.,* 2015). Glucose, galactose and mannose are abundantly available to pneumococci in the host because they are included in mucin-carbohydrate side chains (Terra *et al.*, 2010). Galactose is a very important sugar for pneumococcal biology in the respiratory tract where the microbe colonises and disseminates to cause serious diseases, because galactose forms a large percentage of mucin sugar content in the form of galactose and N-acetylgalactosamines compared to other sugars such as mannose and glucose (Terra *et al.*, 2010). In addition, galactose plays required role in the formation of the essential virulence determinant the polysaccharide capsule (Paixão *et al.*, 2015). Eighteen different sugars are involved in the formation of 90 pneumococcal capsular serotypes, and interestingly, nearly all structurally characterised capsule types contain either galactose or the sugars synthesised from the intermediates formed during galactose catabolism (Carvalho *et al.*, 2011; Paixão *et al.*, 2015).

In order to examine the effect of loss of *rgg* in pneumococcal growth, all strains were grown microaerobically in BHI or CDM supplemented with 55 mM glucose, galactose or mannose individually as the main carbon source as previously described in section 2.3.

3.5.1. Growth of pneumococcal strains in BHI

The pneumococcal strains grew well in BHI microaerobically at 37°C (Figure 3.15). The growth rate was calculated by measuring the increase in absorbance over time. There was no statistical difference either in growth yield or the rate among pneumococcal strains (p> 0.05).

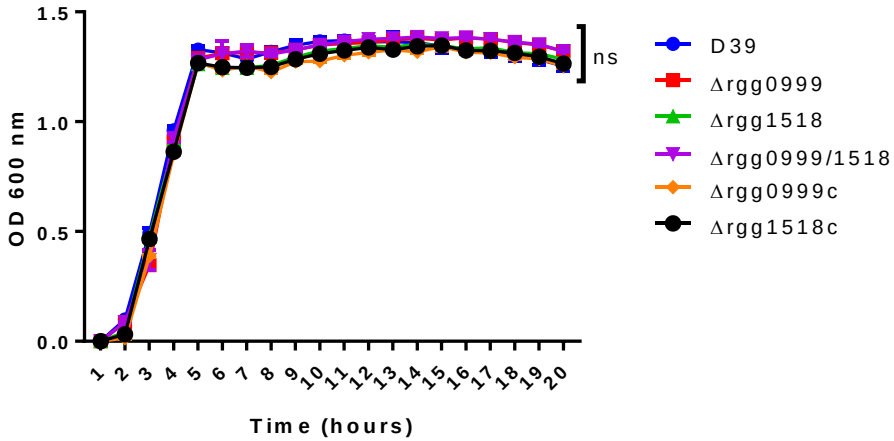


Figure 3.16. Growth profiles of pneumococcal strains in BHI under the microaerobic condition at 37ºC. There was no significant difference between the strains in the growth rate using one way ANOVA and Dunnett's multiple comparisons test, (p>0.05). The experiment was repeated using 3 replicate of 6 independent biological samples. Error bar indicates standard error of mean.

3.5.2. Growth of pneumococcal strains in CDM supplemented with 55 mM glucose, galactose or mannose.

In order to examine the role of *rgg* genes in carbohydrate utilisation, the pneumococcal strains were grown in CDM supplemented either with glucose, galactose and mannose. There was no statistical difference either in growth yield or the rate among pneumococcal strains in glucose (p>0.05). The results suggest that inactivation of *rgg*

did not decrease the glucose utilisation of the mutants under *In vitro* conditions (Table 3.4A and Figure 3.17A).

However, on galactose the growth rates of Δ*rgg0999*, Δ*rgg1518* and Δ*rgg0999/1518* (0.056±0.002, 0.037±0.002 and 0.051±0.004, n=6, respectively) were significantly lower than that of the wild type D39 (0.150±0.001), ($p<0.0001$). Moreover, the *rgg* mutants had the lowest growth yield compared to the wild type ($p<0.0001$). Complemented strains had the similar growth profile as the wild type ($p>0.05$), indicating that the observed phenotypic effect was not due to polar effects (Table 3.4B and Figure 3.17B). In CDM supplemented with mannose, the Δ*rgg0999*, Δ*rgg1518* and Δ*rgg0999/1518* displayed a lower growth rate (0.092±0.0034, 0.093±0.004 and 0.090±0.0052, n=6, respectively) compared to the wild type D39 (0.111±0.0028), ($p<0.01$), hence there is synergistic effect among Rggs. Complemented mutants had the similar phenotype as the wild type (Table 3.4C and Figure 3.17A). These results showed that *rgg0999* and *rgg1518* have important roles for growth in an environment where galactose and mannose as the sole source of carbohydrate.

Table 3.4. Growth rate (μ) h^{-1} and yield (maximal OD_{600}nm) of pneumococcal strains grown microaerobically in CDM supplemented with 55 mM (A) glucose, (B) galactose or (C) Mannose. Values are average of three replicates of six independent biological samples; '±' indicates standard error of means (SEM).

A

Strains	**Growth rate**	**Growth yield**
D39	0.181 ± 0.004	1.273 ± 0.002
Δ*rgg0999*	0.175 ± 0.011	1.248 ± 0.014
Δ*rgg1518*	0.177 ± 0.01	1.297 ± 0.0004
Δ*rgg0999/1518*	0.178 ± 0.003	1.283 ± 0.001
Δ*rgg0999c*	0.179 ± 0.002	1.259 ± 0.0006
Δ*rgg1518c*	0.171 ± 0.01	1.269 ± 0.0061

B

Strains	**Growth rate**	**Growth yield**
D39	0.150 ± 0.001	1.050 ± 0.006
Δ*rgg0999*	0.056 ± 0.002	0.654 ± 0.004
Δ*rgg1518*	0.037± 0.002	0.665 ± 0.00003
Δ*rgg0999/1518*	0.051 ± 0.004	0.551 ± 0.0006
Δ*rgg0999c*	0.157 ± 0.0007	1.040 ± 0.0006
Δ*rgg1518c*	0.15 ± 0.0002	1.04 ± 0.0004

C

Strains	**Growth rate**	**Growth yield**
D39	0.111 ± 0.0028	1.194 ± 0.0024
Δ*rgg0999*	0.092 ± 0.0034	0.977 ± 0.0037
Δ*rgg1518*	0.093 ± 0.004	0.971 ± 0.007
Δ*rgg0999/1518*	0.090 ±0.0052	0.794 ± 0.0006
Δ*rgg0999c*	0.107 ± 0.0028	1.156 ± 0.0223
Δ*rgg1518c*	0.104 ±0.0012	1.175 ± 0.0003

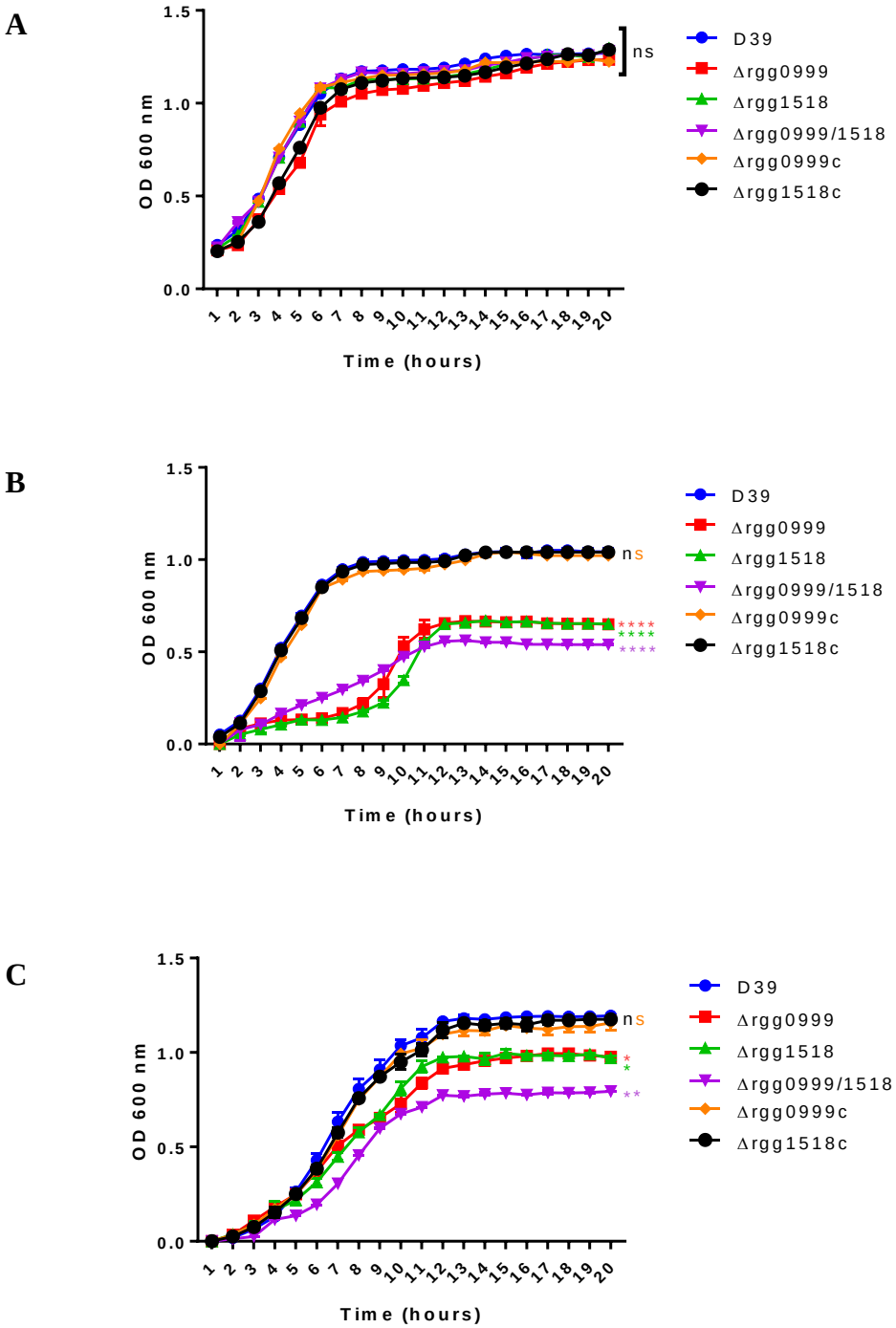


Figure 3.17. Growth curves of pneumococcal strains grown microaerobically in CDM supplemented with 55 mM (**A**) glucose, (**B**) galactose, (**C**) Mannose. Experiment was repeated using 3 replicates of 6 independent biological samples; vertical bars indicate standard error of means (SEM). Significant differences were seen by comparing the growth rates of mutant strains to the wild type D39 using one-way ANOVA and Dunnett's multiple comparisons test. (*p<0.05, **p<0.01, **** p<0.0001 relative to the wild type).

3.5.3. Capsule synthesis determination by glucuronic acid assay

Capsular polysaccharide (CPS) is the most important pneumococcal virulence determinant, which can protect pathogen from phagocytosis (Hyams *et al.,* 2010). Glucuronic acid is the major components of the capsular polysaccharide and is found in wild type D39 (Bonofiglio *et al.,* 2005). The microarray data highlighted that genes that are possibly regulated by Rgg1518 include the capsule locus genes (SPD_0315-SPD_0328) on galactose. My hypothesis was that the mutant (*Δrgg1518*) would show a reduced capsule synthesis. Hence, I determined the level of capsule synthesis by assaying for glucuronic acid in a capsule extract as this compound is found in type 2 D39 strain capsule. Therefore, glucuronic acid was extracted after the pneumococcal strains were grown in CDM supplemented either with galactose or glucose as described previously in section 2.16.2.

The production of capsular polysaccharides was the highest in wild type D39 (117.387±3.499, n=6) as well as in complemented strain Δ*rgg1518c* (112.552±4.07, n=6) while lowest in Δ*rgg1518* (63.325±2.49, n=6) comparing to the wild type ($p<0.0001$) (Table 3.5A and Figure 3.18A). In the CDM supplemented with glucose, there was no significant difference in production of capsular polysaccharides, for wild type D39, Δ*rgg1518* and Δ*rgg1518c* (71.13±6.356, 71.2±5.907, and 71.83±4.228, n=6, respectively) as shown in Table 3.5A and Figure 3.18A. The results indicated that inactivation of rgg*1518* decreases the capsule production of the mutant strain under *In vitro* conditions.

Table 3.5. Analysis of total glucuronic acid of pneumococcal strains. The strains were grown in microaerobic condition in CDM supplemented with galactose (A) or glucose (B). Glucuronic acid concentration of capsular polysaccharide was expressed as μg/10^9 CFU. Values are average of three replicates of six independent biological samples; '±' indicates standard error of mean (SEM).

Strains	(A) Glucuronic acid μg/10^9 CFU in galactose	(B) Glucuronic acid μg/10^9 CFU in glucose
D39	117.387 ± 3.499	71.13 ± 6.356
Δ*rgg1518*	63.325 ± 2.49	71.20 ± 5.907
Δ*rgg1518c*	112.552 ± 4.07	71.83 ± 4.228

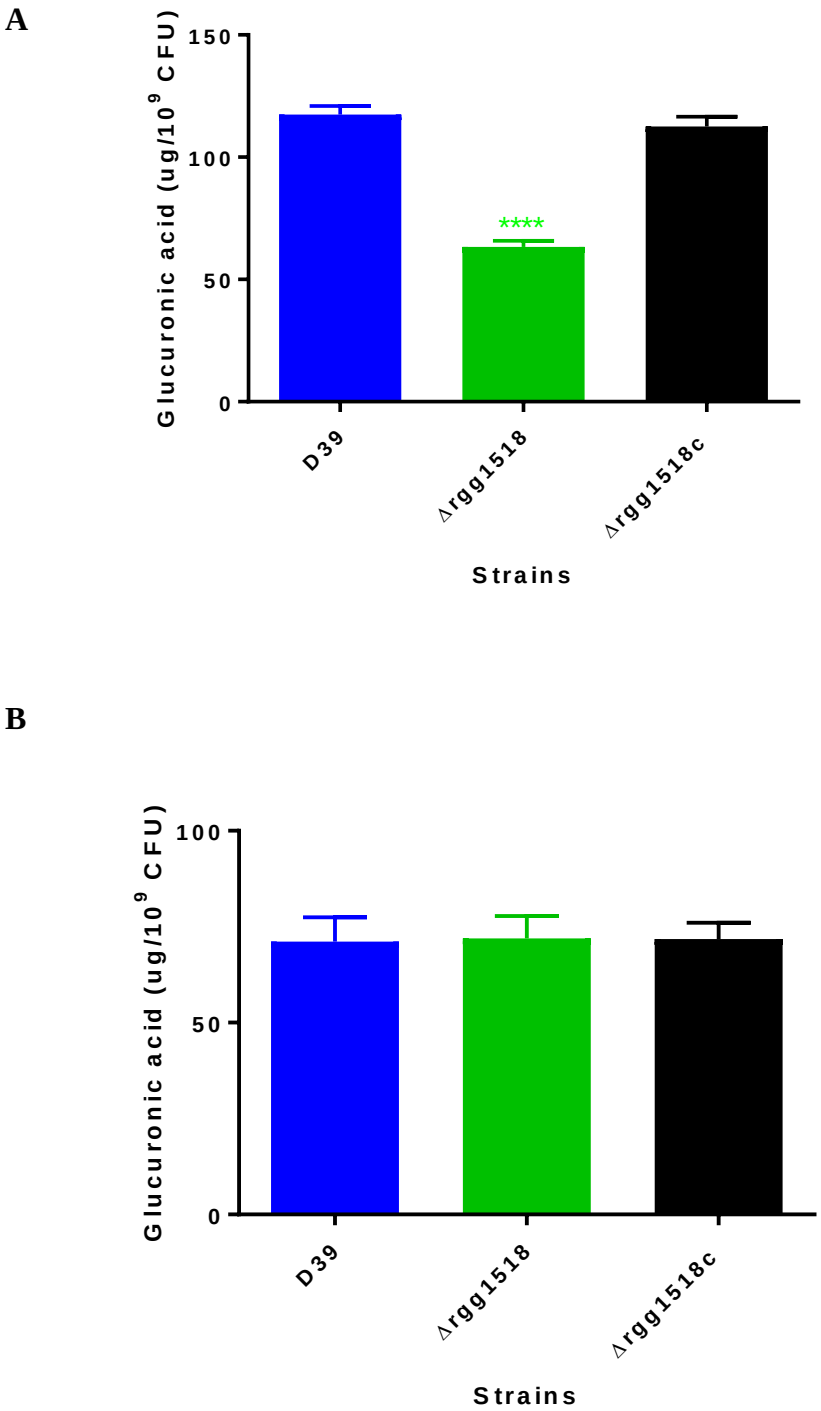


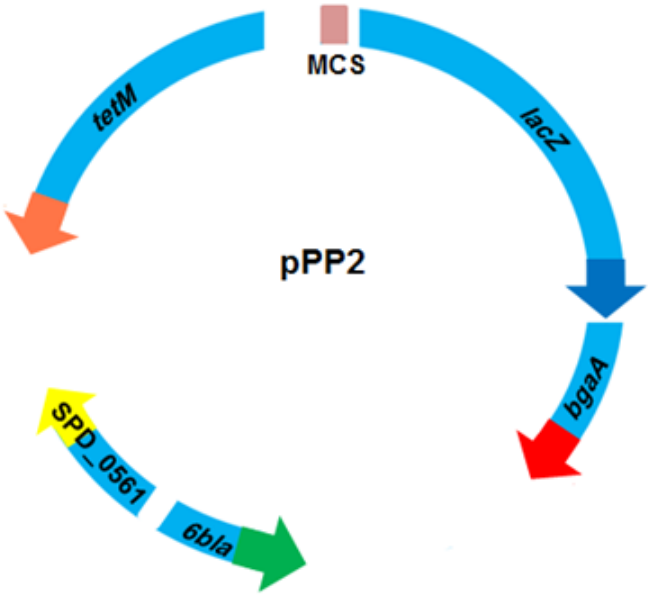
Figure 3.18. The quantification of glucuronic acid levels in pneumococcal strains grown in CDM supplemented either with galactose (A) or glucose (B) using one-way ANOVA and Dunnett's multiple comparisons test, (****$p<0.0001$, $p>0.05$). The experiment was repeated using 3 replicate of 6 independent biological samples.

3.6. Construction of *lacZ* reporter assay and β-galactosidase assay

3.6.1. Construction of *lacZ*-fusions

Transcriptional *lacZ*-fusions were used to investigate the expression level of *rggs* in different sugars. *E. coli lacZ* gene (reporter gene) coding for β-galactosidase is known to be widely used for reporter assays because β-galactosidase activity can be easily assayed (Miller, 1972). The putative promoter region of the target gene (the insert) drives the promoterless *lacZ* gene expression (Halfmann *et al.*, 2007). The constructed recombinant plasmid carrying promoter-*lacZ* fusions were transformed into *S. pneumoniae* as describe in section 2.16.2. The pPP2 plasmid has homologous regions to pneumococcal SPD_0561, and *bgaA*, which encodes for endogenous β-galactosidase. Hence, upon successful genome integration of the recombinant plasmid, pneumococcal *bgaA* is inactivated, decreasing the background galactosidase activity (Figure 3.19A), and the pneumococcus becomes tetracycline-resistant due to integration of the *tetM* (Halfmann *et al.*, 2007), (Figure 3.19C).

A



B

158

SphI KpnI EcoRI XbaI BamHI

GCATGCATCGGTACCTGCGAATTCTAGTCTAGAATCGGATCC**TAA**C**T**

SD M K H L D P V V L Q

AGT**TAA**GGAGGCAAAT**ATG**AAACATCTTGATCCCGTCGTTTTACAA

lacZ

C

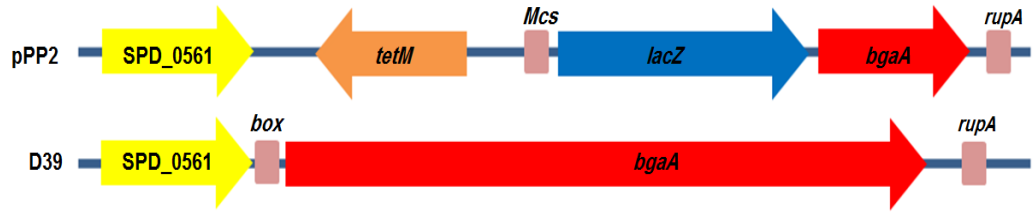


Figure 3.19. (A) Genetic map of the integrative promoter probe plasmids pPP2. (B) The nucleotide sequence of the multiple cloning sites (MCS) and the start of *lacZ* gene. (C) Reorganization of homologous recombination between recombinant pPP2 and *bgaA* region of D39. Upon integration, the endogenous *bgaA* gene is disrupted and box, one of the two repetitive elements (*box*, *rupA*) is deleted.

3.6.2. Digestion of pPP2

A double digestion of pPP2 plasmid was done using the restriction enzymes *SphI* and *BamHI* as previously described in section (2.11). Agarose gel electrophoresis analysis was used to confirm the successful pPP2 digestion and compared to the undigested pPP2 (Figure 3.20). Lane 1 shows *SphI-BamHI* digested linearised pPP2 plasmid with a size of approximately 10 kb, whereas lane 2 shows undigested pPP2.

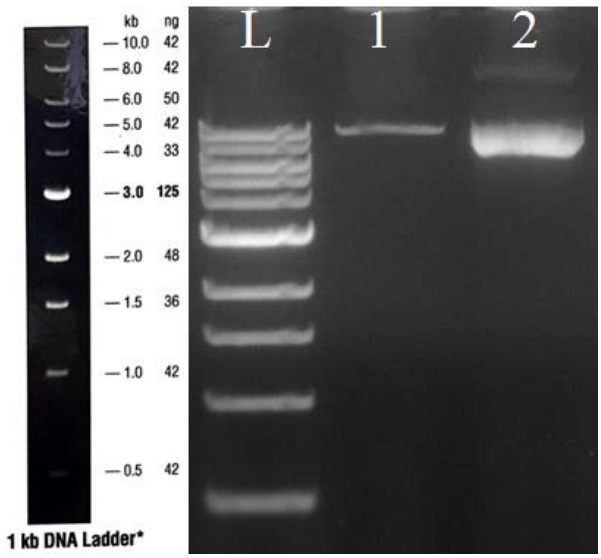


Figure 3.20. Agarose gel electrophoresis illustrates the digested and undigested pPP2 plasmid. L: 500 ng of 1 kb DNA ladder (NEB); lane 1: pPP2 digested with *SphI-BamHI*. Lane 2: undigested pPP2.

3.6.3. Amplification of putative promoter regions of *rgg* genes

Putative promoter regions of *rgg0999* and *rgg1518* in addition to 28 bp from the coding region of each target gene were amplified using primers listed in Table 2.14. These primers were modified to incorporate *SphI* and *BamHI* sites as described in section (2.18.1). The amplified PCR products were purified using the Wizard® SV Gel and PCR Clean-Up System (Promega, UK) and were subsequently analysed using agarose gel

electrophoresis. The analysis results showed that all amplicons were in consistent with the expected sizes (Figure 3.21).

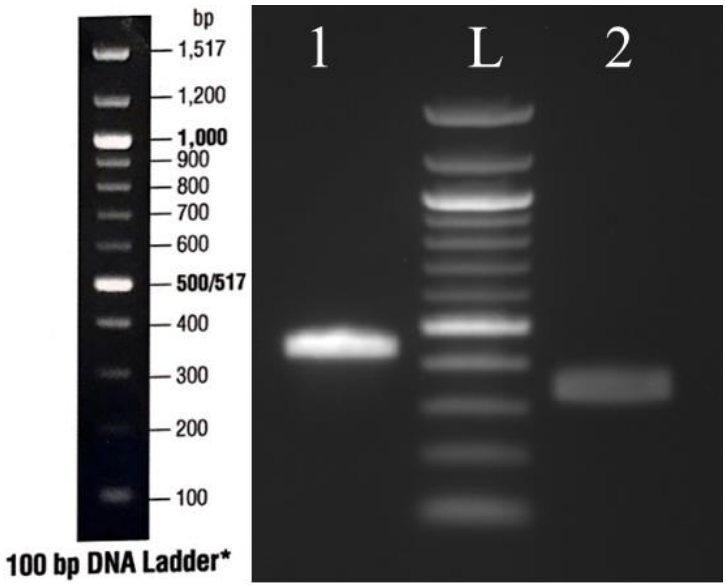


Figure 3.21. Agarose gel electrophoresis illustrates the amplified upstream regions containing the putative promoter regions of the target genes. L: 500 ng of 100 bp DNA ladder (NEB). Lanes 1-2: putative promoter regions of *rgg0999* (430 bp), and *rgg1518* (327 bp), respectively.

3.6.4. Preparation of inserts and pPP2 for ligation

The inserts were PCR amplified and double digested with *SphI* and *BamHI*. The digested inserts were ligated to the compatible ends in pPP2 as described in section 2.18.1. The ligation reaction was first transformed into *E. coli* TOP10 chemically competent cells for propagation. The successful transformation of recombinant plasmids was confirmed by PCR (Figure 3.22) using primers in Table 2.14 as previously described in section 2.18.1. For further confirmation, the pPP2 recombinant plasmids that have the target insert were sequenced using the Fusion/UF and Fusion/DR primers. The sequencing results indicated that all the recombinant plasmids have the expected inserts (Appendix 3).

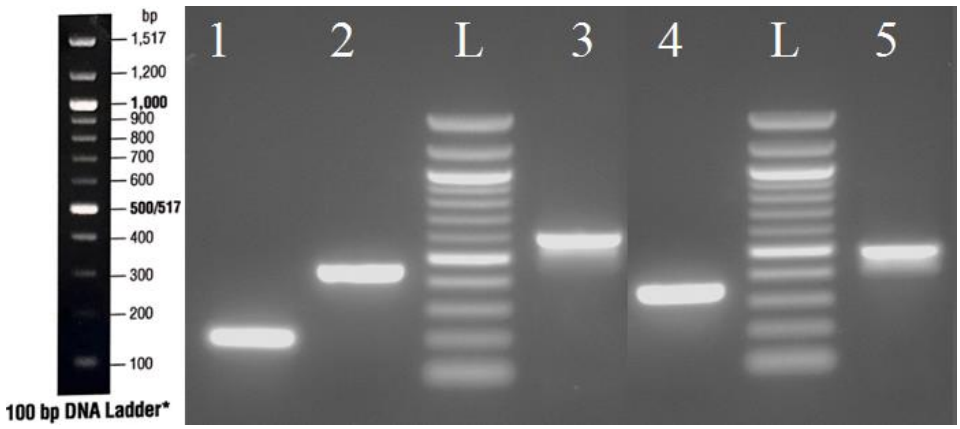


Figure 3.22. Agarose gel electrophoresis showing successful construction of recombinant pPP2. Lane 1: insert amplified from the native pPP2 using Fusion/UF and Fusion/DR primers with a size of approximately 200 bp. Lanes 2 and 4: putative promoter regions of *rgg0999* (430 bp), and *rgg1518* (327 bp), respectively. L: 500 ng of 100 bp DNA ladder (NEB). Lanes 3 and 5: amplified putative promoter regions of target genes plus upstream and downstream of cloning site by Fusion/UF and Fusion/DR primers, for *rgg0999* (630 bp), and *rgg1518* (527 bp), respectively.

3.6.5. Transformation of pPP2 recombinant plasmid into pneumococcal strains

The extracted recombinant pPP2 plasmids carrying the correct inserts were successfully sequenced and transformed into the wild type D39 using the protocol described in section 2.18.2. The transformants were selected on blood agar supplemented with tetracycline. The reporter strains were designated as P*rgg0999-lacZ*-wt and P*rgg1518-lacZ*-wt (P refers to promoter region). In addition, a native pPP2 was transformed into the genome of wild type D39 and the strain designated as pPP2-D39 in order to inactivate the native *bgaA*, which was used as an assay control. Moreover, the constructed recombinant plasmids carrying the correct inserts were transformed into the mutant strains to investigate the interaction between the transcriptional regulators of interest. The transformants were selected on blood agar supplemented with both tetracycline and spectinomycin, and the reporter strains were designated as P*rgg0999-lacZ*-Δ*rgg0999*, P*rgg0999-lacZ*-Δ*rgg1518*, P*rgg0999-lacZ*-Δ*rgg0144*, P*rgg0999-lacZ*-

Δ*rgg0939*, P*rgg1518-lacZ-*Δ*rgg0999*, P*rgg1518-lacZ-*Δ*rgg1518*, P*rgg1518-lacZ-*Δ*rgg0144* and P*rgg1518-lacZ-*Δ*rgg0939*. The successful integration of the recombinant plasmids into the pneumococcal genome was confirmed by PCR using Fusion/UF and Fusion/DR primers that were designed to amplify the target putative promoter regions plus upstream and downstream of the cloning site, respectively (Figure 3.23).

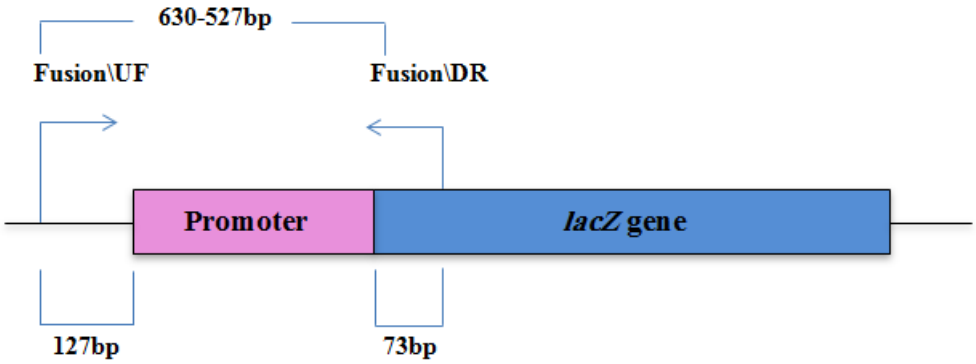


Figure 3.23. Illustration showing the strategy used to confirm the successful integration of the transcriptional lacZ fusions within the pneumococcal genome. Fusion\UF and Fusion\DR primers were used to amplify the inserts containing putative promoter regions of target genes plus upstream and downstream of the cloning site. The PCR products for *rgg0999* and *rgg1518* were 630 and 527 bp, respectively, in size while the empty pPP2 plasmid yielded PCR products of 200 bp in size using the same set of primers.

The amplified PCR products were analysed by agarose gel electrophoresis (Figure 3.24). As can be seen in picture A and B, lanes 1-5 show successful amplification of inserts representing promoter regions of each gene and upstream and downstream of the cloning site in pPP2. The PCR products in lanes 1-5 had the expected approximate sizes for *rgg0999* (630 bp) and *rgg1518* (527 bp). In addition, lane 6 in both pictures shows a promoterless fragment of the pPP2 plasmid (200 bp) using the same set of primers, Fusion\UF and Fusion\DR primers.

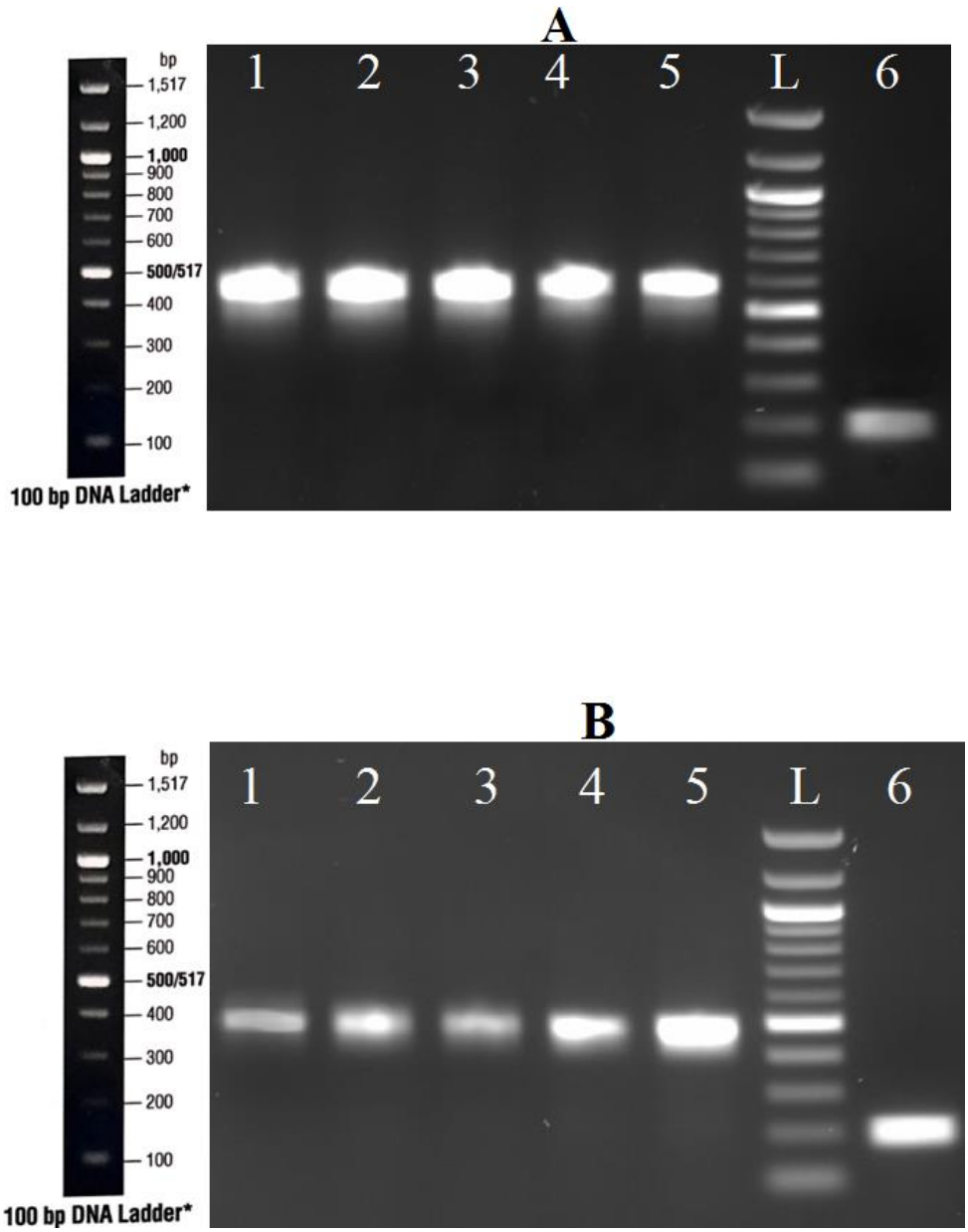


Figure 3.24. Agarose gel electrophoresis showing the successful construction of reporter strains in wild type and *rgg* mutant backgrounds. (A) P*rgg0999-lacZ*, and (B) P*rgg1518-lacZ*. Lanes 1 in A and B: putative promoter regions of target genes plus upstream and downstream of the cloning site; approximate product size for *rgg0999* was 630 bp, and for *rgg1518* it was 527 bp in wild type background. Lanes 2, 3, 4 and 5 in the picture (A) indicate the amplified putative promoter regions of *rgg0999* (630 bp) plus empty pPP2 in *rgg* mutants Δ*rgg0999*, Δ*rgg1518*, Δ*rgg0144* or Δ*rgg0939* backgrounds, respectively. Lanes 2, 3, 4 and 5 in the picture (B) demonstrate amplified P*rgg1518* (527 bp) plus empty pPP2 in *rgg* mutants in Δ*rgg0999*, Δ*rgg1518*, Δ*rgg0144* or Δ*rgg0939* backgrounds, respectively. L: 500 ng of 100 bp DNA ladder (NEB). Lanes 6A and 6B: amplified empty pPP2 using Fusion/UF and Fusion/DR primers with a size of approximately 200 bp.

3.7. Expression of transcriptional *lacZ*-fusions in the presence of different sugars

To determine the involvement of Rggs in pneumococcal biology, the expression of *rgg* genes were assayed in the presence of different host-derived sugars. The pneumococcus is usually found in the nasopharynx as a harmless commensal. The respiratory tract is covered with mucus, and mucin is the main component of mucus. Mucin is a heavily glycosylated protein and its side chain is composed of carbohydrate moieties that serve as fermentable substrates for *S. pneumoniae* (Sheehan *et al.*, 1995; Terra *et al.*, 2010). The monosaccharides glucose, galactose and mannose were used in this study to investigate whether they would induce the expression of *rgg* genes. These sugars, galactose and mannose, are present within the structure of O- and N-linked glycans that are found in the human respiratory tract, while glucose is plentiful, upto 10 mM, in the blood (Sheehan *et al.*, 1991; Rose and Voynow, 2006).

The inducibility of *rgg0999* and *rgg1518* were investigated using transcriptional reporter assays in microaerobically grown cultures in CDM supplemented with glucose, galactose or mannose (Table 3.6 and Figure 3.25). The results showed that there was no induction in β-galactosidase activity in the D39 containing pPP2 in the absence of promoter regardless of presence of galactose, mannose or glucose (2.69±0.54; 1.76±0.59 and 1.43±0.045 MU, n=3, respectively) compared to its expression in CDM alone without added sugar (1.35±0.13MU, n=3, $P>0.05$). The transcriptional reporter strain P*rgg1518*-*lacZ*-wt generated the highest β-galactosidase activity on galactose and mannose (969.49±5.88 and 503.29±12.90 MU, n=3, respectively) ($p<0.01$and $p<0.001$), whereas on glucose (176.20±3.0 MU, n=3) no significant induction of *rgg1518* could be seen relative to the no sugar control (170.53±14.9 MU, n=3) ($p>0.05$). There was no induction in β-galactosidase activity in the P*rg0999*-*lacZ*-wt on galactose, mannose or glucose (104.45±10.86; 85.59±5.56 and 80.67±7.83 MU, n=3, respectively) compared

to its expression on CDM alone (91.68±0.82 MU, n=3, P>0.05). These results show that the *rgg1518* is active when pneumococcus encounters host glycans containing galactose or mannose. On the other hand *rgg0999* is not induced by any of the tested sugars.

Table 3.6. Expression levels (in Miller units) of pneumococcal transcriptional *lacZ*-fusions to the promoter of *rgg0999* and *rgg1518* in wild type. The strains were grown microaerobically in CDM supplemented with 55 mM of glucose, galactose or mannose. The activity is expressed in nmol *p*-nitrophenol/min/ml. Values are average of three independent experiments each with three replicates. ± indicates standard error of means (SEM).

Strains	Galactose	Mannose	Glucose	CDM alone
pPP2-wt	2.69 ± 0.54	1.76 ± 0.59	1.43 ± 0.045	1.35 ± 0.13
P*rgg0999-lacZ*-wt	104.45 ± 10.86	85.59 ± 5.56	80.67 ± 7.83	91.68 ± 0.82
P*rgg1518-lacZ*-wt	969.49 ± 15.88	503.29 ± 12.90	176.20 ± 3.0	170.53 ± 1.51

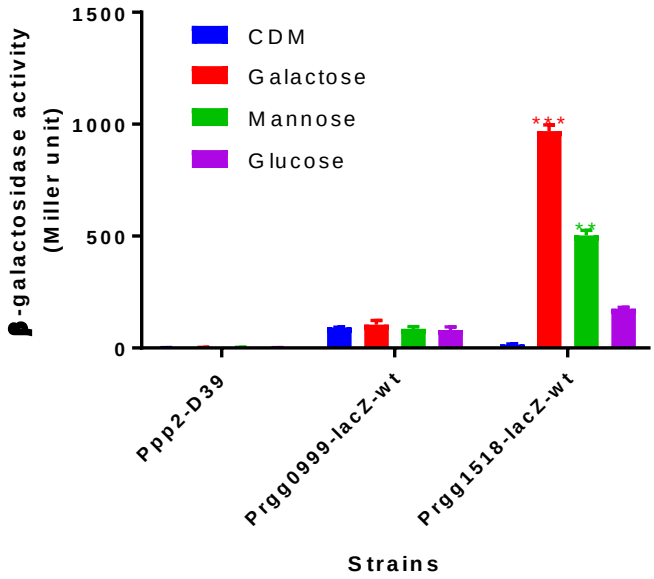


Figure 3.25. Analysis of *rgg0999* and *rgg1518* regulation using *lacZ* reporter assay. The activity of β-galactosidase under the regulation of *rgg0999* or *rgg1518* putative promoter is expressed as nmol p-nitrophenol/min/ml. Ppp2- D39 strain referred to the wild type D39 containing pPP2 in the absence of any promoter. The values are the average of three independent experiments each with three replicates vertical bars indicate standard error of means (SEM), (**p<0.01 and ***p<0.001 compared to its expression on CDM alone).

3.7.1. Role of Rgg1518 in *rggs* regulation

Several lines of evidence obtained in this study indicated that the Δ*rgg1518* was attenuated in growth galactose and mannose. Additionally, *rgg1518* promoter was induced by galactose and mannose as determined by transcriptional *lacZ* reporter fusions. Thus, these results lead us to investigate the interaction among *rgg* regulators using transcriptional *lacZ*-fusions. Therefore, P*rgg1518-lacZ* construct was introduced into Δ*rgg1518*, Δ*rgg0144*, Δ*rgg0939* and the wild type D39, generating the reporter strains P*rgg1518-lacZ*-Δ*rgg1518*, P*rgg1518-lacZ*-Δ*rgg0144*, P*rgg1518-lacZ*-Δ*rgg0939* and P*rgg1518-lacZ*-wt, respectively, and β- galactosidase activities in the reporter strains were measured when the strains were grown microaerobically in CDM supplemented with 55 mM galactose.

The β-galactosidase activity was found to be lower in P*rgg1518-lacZ*-Δ*rgg1518* (29.23±1.97 MU) compared to its expression in wild type (969.5±15.88 MU, n=3), showing that Rgg1518 is required for its own expression ($p<0.0001$). In addition, *rgg1518* expression required the presence of Rgg0144 and Rgg0939 as the expression of *rgg1518* was lower in both Δ*rgg0144* and Δ*rgg939*. P*rgg1518-lacZ*-Δ*rgg0144* and P*rgg1518-lacZ*-Δ*rgg0939* (168.95±2.53 MU, and 351.59 ± 5.47 MU, n=3, respectively) compare to the wild type (969.5±15.88 MU, n=3) ($p<0.0001$) (Figure 3.26).

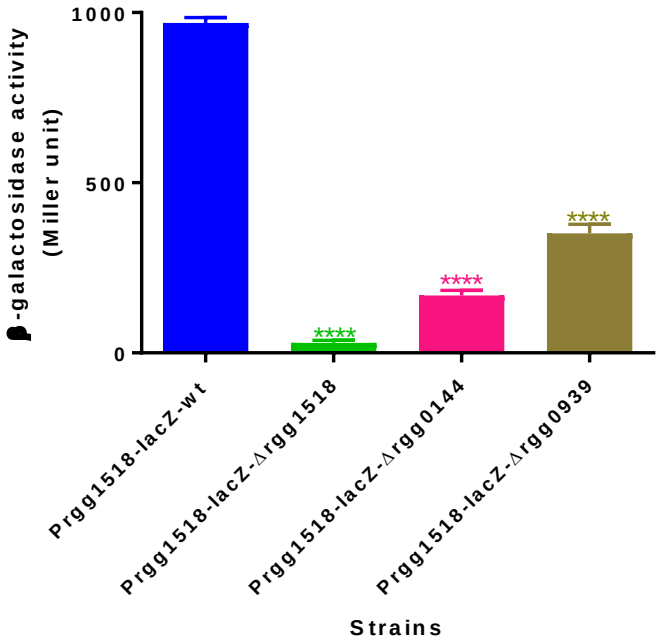


Figure 3.26. Analysis of *rgg1518* regulation using *lacZ* reporter assay. The activity of β-galactosidase, driven by *rgg1518* putative promoter in different mutant strains, is expressed as nmol p-nitrophenol/min/ml. The values are the average of three independent experiments each with three replicates vertical bars indicate standard error of means (SEM), (****p<0.0001 compared to its expression in the wild type D39).

3.7.2. Role of inactivated r*gg1518* in capsule locus genes regulation

Microarray data highlighted that Rgg1518 is involved in regulation of capsule locus genes (SPD_0315-SPD_0328). Capsule locus genes were up-regulated in the *Δrgg1518* on galactose. Galactose plays an essential role in formation of essential virulence determinant capsule (Paixão *et al.*, 2015).

The regulator (*rgg1518*) shows differential regulation in galactose and showing growth affect in this sugar. Thus, *rgg1518* is involvement in galactose metabolism. On the other hand, this regulator is unlikely to be induced by glucose, and there was no any phenotype when the mutant (Δ*rgg1518*) was grown in presence of glucose. The *rgg1518* is induced by galactose and unlikely to induce by glucose (Section 3.7) therefore, CDM supplemented with glucose was used as a control. The role of

inactivated r*gg1518* in capsule expression was investigated using transcriptional *lacZ*-fusions to assess the loss of r*gg1518* in deficiency capsule expression.

In order to generate reporter strains, putative promoter region of capsule locus genes (SPD_0315-SPD_0328) in addition to 28 bp from coding region of target gene (the insert) was amplified using primers listed in Table 2.14. These primers were modified to incorporate *SphI* and *BamHI* sites as described in section (2.18.1). The amplified PCR products were purified using the Wizard® SV Gel and PCR Clean-Up System (Promega, UK) and were subsequently analysed using agarose gel electrophoresis. The analysis results showed that amplicon was in line with the expected sizes (Figure 3.27).

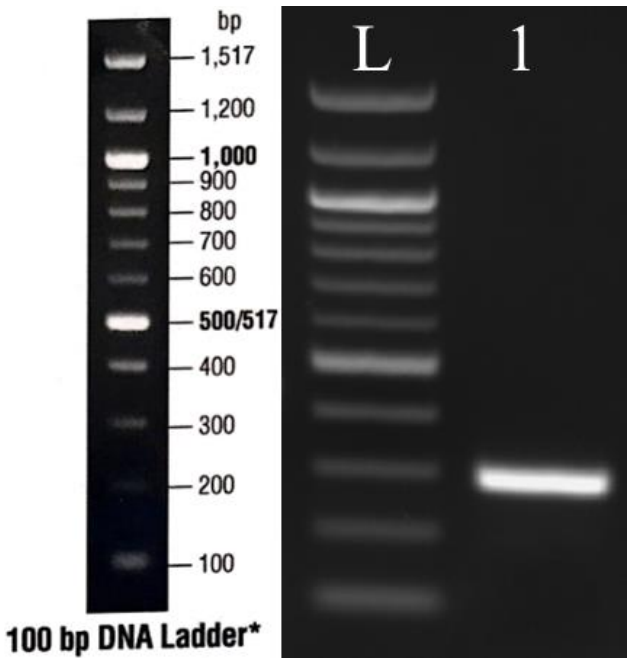


Figure 3.27. Agarose gel electrophoresis illustrates the amplified upstream regions containing the putative promoter region of the target gene. L: 500 ng of 100 bp DNA ladder (NEB). Lane 1: putative promoter region of *cps* locus (273 bp).

The insert was double digested with *SphI* and *BamHI*. The digested insert were ligated to the compatible ends in digested pPP2 as described in section 2.18.1. The ligation reaction was first transformed into *E. coli* TOP10 chemically competent cells for propagation. The successful transformation of recombinant plasmid was confirmed by PCR (Figure 3.28) using primers in Table 2.14 as previously described in section 2.18.1. For further confirmation, the pPP2 recombinant plasmid that has the target insert was sequenced using the Fusion/UF and Fusion/DR primers. The sequencing result indicated that the recombinant plasmid has the expected insert (Appendix 3).

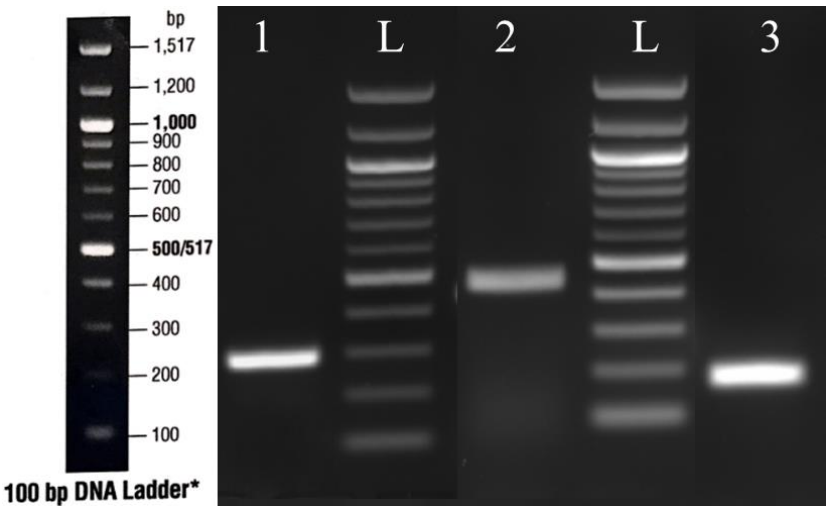


Figure 3.28. Agarose gel electrophoresis showing successful construction of recombinant pPP2. Lane 1: putative prompter region of *cps* locus (273 bp). L: 500 ng of 100 bp DNA ladder (NEB). Lane 2: amplified putative promoter region of *cps* locus plus upstream and downstream of cloning site (473 bp) by Fusion/UF and Fusion/DR primers. Lane 3: insert amplified from the native pPP2 using Fusion/UF and Fusion/DR primers with a size of approximately 200 bp.

The extracted recombinant pPP2 plasmid carrying the correct insert was successfully sequenced and transformed into the wild type D39 using the protocol described in section 2.18.2. The reporter strain was designated as P*cps*-*lacZ*-wt (P refers to promoter

region). Moreover, the constructed recombinant plasmid carrying the correct insert was also transformed into the mutant strain Δ*rgg1518*. The transformants were selected on blood agar supplemented with both tetracycline and spectinomycin, and the reporter strain was designated as P*cps-lacZ*-Δ*rgg1518*. The successful integration of the recombinant plasmids into the pneumococcal genome was confirmed by PCR using Fusion/UF and Fusion/DR primers that were designed to amplify the target putative promoter region and the up and downstream of the cloning site, respectively (Figure 3.29).

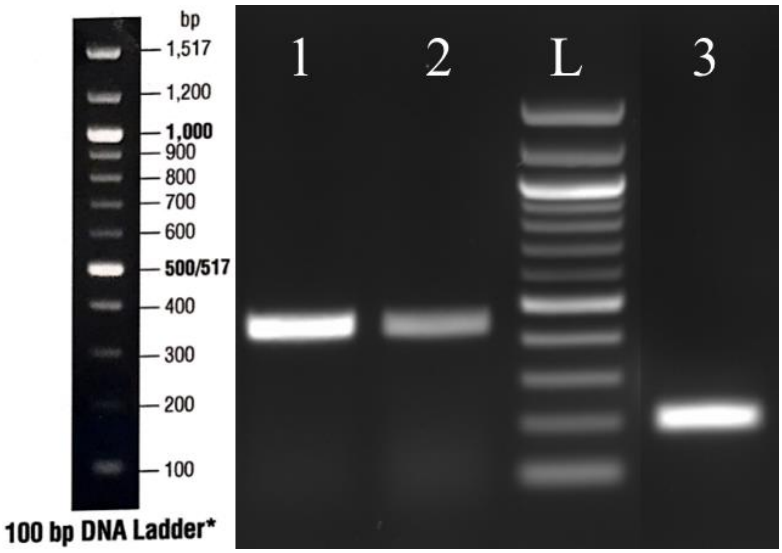


Figure 3.29. Agarose gel electrophoresis showing the successful construction of reporter strains in wild type and *rgg* mutant background. Lanes 1 and 2: putative promoter region of *cps* locus plus upstream and downstream of the cloning site; approximate product size was 473 bp, in wild type and Δ*rgg1518* background, respectively. L: 500 ng of 100 bp DNA ladder (NEB). Lane 3: amplified empty pPP2 using Fusion/UF and Fusion/DR primers with a size of approximately 200 bp.

The β- galactosidase activities in the reporter strains were measured when the strains were grown microaerobically in CDM supplemented with 55 mM either with galactose or glucose as described in section 2.16.3. The transcriptional reporter strain P*cps-lacZ*-

wt generated the highest β-galactosidase activity on galactose (71.104±2.63 MU, n=3) ($p<0.0001$), whereas on glucose (9.68 ±1.11 MU, n=3), no significant induction of *cps* locus could be seen relative to the no sugar control (7.52±1.58 MU, n=3) ($p>0.05$) (Figure 3.30A). Next I investigated Rgg1518 involvement in *cps* regulation on galactose. It was found that the β-galactosidase activity is lower in P*cps-lacZ-*Δ*rgg1518* (29.08±1.71 MU, n=3) compare to the wild type (71.104± 2.63 MU, n=3 ($p<0.0001$) (Figure 3.30B). The results indicated that in absence of *rgg1518*, the capsule locus is not induced, indicating that Rgg1518 is an activator of *cps* locus. This result was in conflict with the microarray data, which showed that Rgg1518 is a repressor of cps locus.

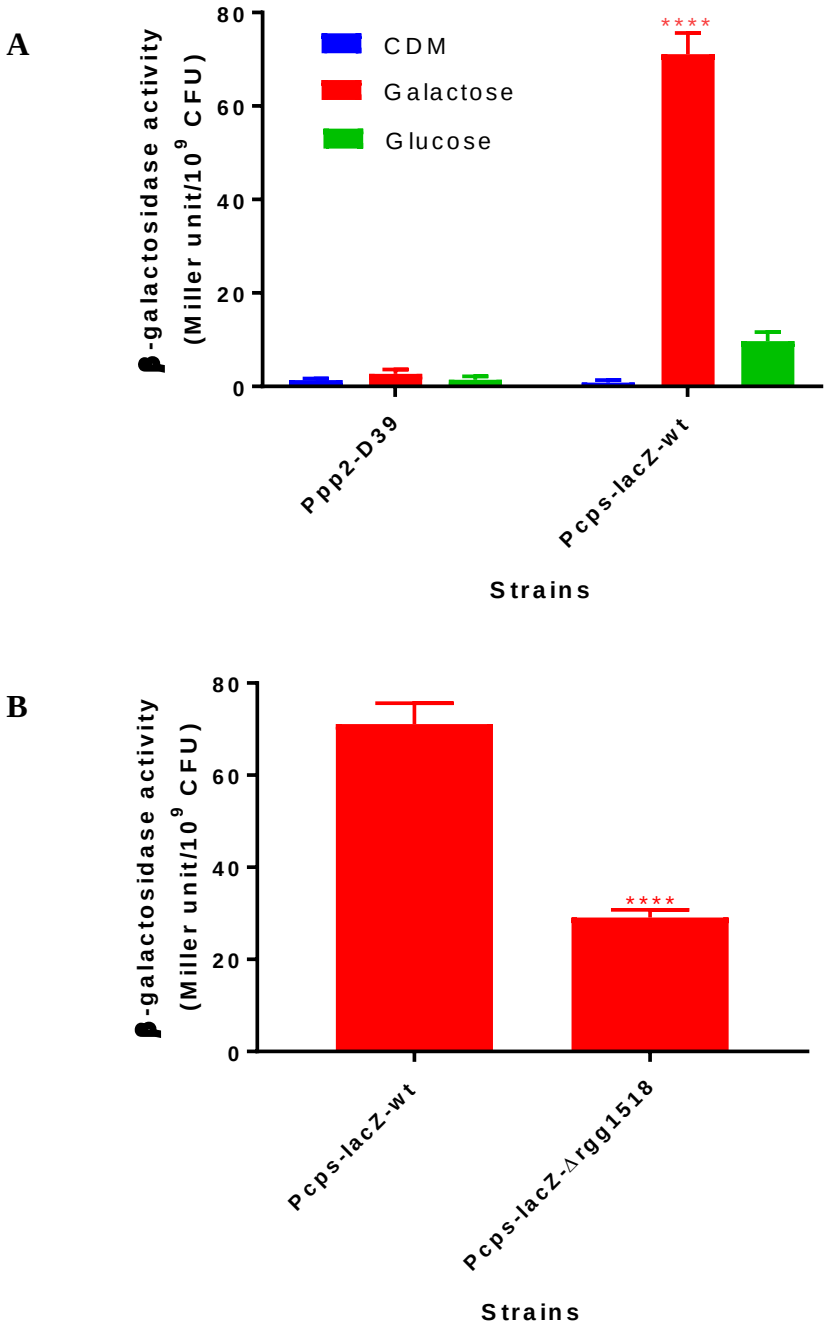


Figure 3.30. Analysis of *cps* locus using *lacZ* reporter assay. (A) Inducibility of *cps* locus in wild type D39 on galactose and glucose and (B) The expression of *cps* locus in wild type and Δ*rgg1518* background on galactose. The activity is expressed as nmol p-nitrophenol/min/ml. Ppp2-D39 strain is D39 containing native pPP2 without a *cps* promoter. The values are the average of three independent experiments each with three replicates vertical bars indicate standard error of means (SEM), (****p<0.0001 compared to its expression in the wild type D39).

Section D: DNA-protein interactions using electrophoretic mobility shift assay (EMSA)

Rgg1518 regulates a wide range of genes as determined by microarray analysis (section 3.3). It is likely that some of these genes are directly, and that others are indirectly regulated by Rgg1518. To determine the genes directly regulated by Rgg1518, I used EMSA to investigate the interaction between Rgg1518 and the selected promoters of differentially expressed genes. EMSA is a sensitive method used widely to detect protein-DNA interaction (Garner and Revzin, 1986). The method includes detecting the difference in mobility patterns of free DNA and protein-DNA complex (Hellman and Fried, 2007). There are various techniques available for studying protein-DNA interactions including Chromatin immunoprecipitation (ChIP) and nitrocellulose filter binding (Carey *et al.,* 2009). Filter binding is not suitable for experiments that requiring the use of two binding proteins at same time. Additionally, CHIP is used to detect the location of DNA binding sites on the genome for a target protein *In vivo*, it is not sensitive, and gives false positive results (Reid *et al.,* 2000; Ren *et al.,* 2000). However, EMSA is the preferred techniques because it is rapid and easy to perform.

There are various techniques used for EMSA and each technique has advantages and disadvantages associated with accuracy, and health and environment risks. The most common EMSA method used is the isotopic labeling of DNA with the ^{32}P-phosphate method, which offers high detection sensitivity, but this method suffers from the short half-life of the label and extended exposure requirements for detection, and is not preferred due to safety issues (Rodgers *et al.*, 2000). Chemiluminescence detection is another detection method that requires pre-labeling of the DNA and secondary detection methods (Jing *et al.*, 2003). The FAM method, using a fluorescent DNA probe, offers

direct detection method for observing both the DNA and protein components in the same gel-shift assay. Additionally, the required equipment of FAM method is inexpensive and it is an environment-friendly method.

In order to perform EMSA, the Rgg1518 was firstly extracted and purified, then the putative promoter regions of target genes were predicted and amplified using fluorescently labelled primers. The probe, that is the target DNA containing the putative promoter regions, and the recombinant Rgg1518 was incubated on ice before being analysed on a non-denaturing gel by electrophoresis as describe in section 3.9. I excluded Rgg0999 because it is not induced by any other tested sugars.

3.8. Production and purification of recombinant Rgg1518

3.8.1. Amplification, cloning, transformation and DNA sequencing of the target gene

PrimeSTAR HS premix and the primers listed in Table 2.17 were used to amplify *rgg1518* as described previously in section 2.7. The amplified PCR products were routinely analysed by agarose gel electrophoresis (Figure 3.31). As expected, the results obtained showed the successful amplification of the target gene. Lane 1 shows the amplicon size for *rgg1518*, approximately 879 bp. The amplicons were purified using the Wizard® SV Gel and PCR Clean-Up System (Promega, UK) (Section 2.8) to exclude enzymes, dNTPs and salts from the PCR product.

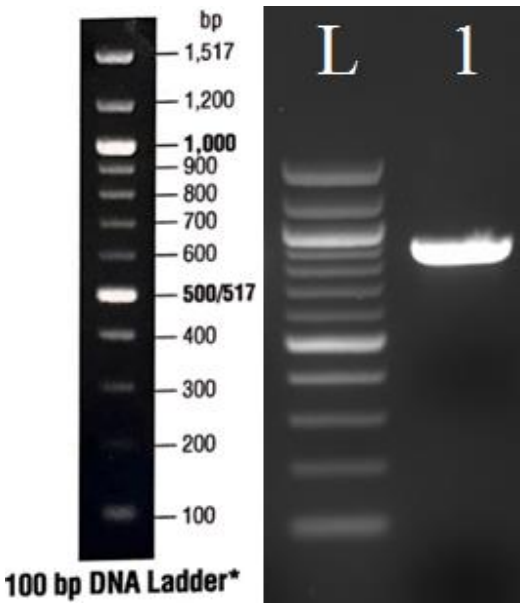


Figure 3.31. Agarose gel electrophoresis shows the amplicon for the *rgg1518*. L: 500 ng of 100 bp DNA ladder (NEB). Lane 1: amplicon representing the *rgg1518* (879 bp).

The purified amplicon for the target gene was cloned into the pLEICS-01 plasmid by PROTEX, University of Leicester. The recombinant plasmid was extracted and sequenced using T7 Promoter-F (PNACL) and pLIECS-01-Seq-R (PNACL) primers. These primers were designed to amplify the up and downstream of the cloning site, and the sequencing results showed the expected successful cloning of the gene of interest (Appendix 4). Thereafter, the recombinant plasmid was successfully transformed into *E. coli* BL21 (DE3) pLysS competent cells for protein expression, (section 2.12), and this was confirmed by PCR using ProteinSPD1518F and ProteinSPD1518R primers (Table 2.17). Eventually, the PCR product was analysed by agarose gel electrophoresis (Figure 3.32).

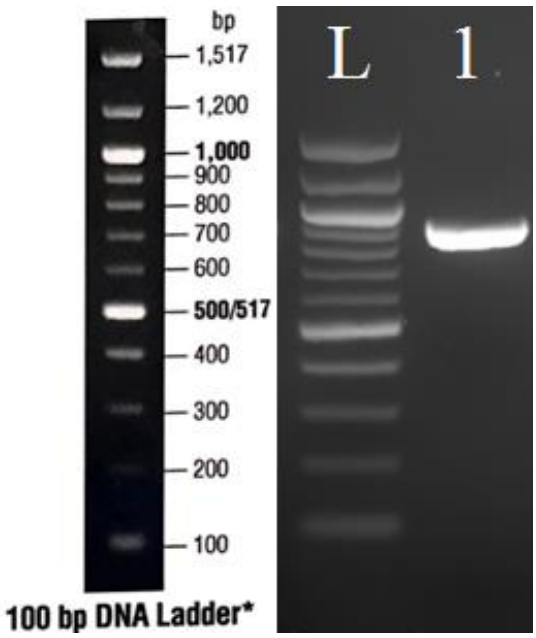


Figure 3.32. Agarose gel electrophoresis shows the amplicon for the target gene from the related recombinant plasmid. L: 500 ng of 100 bp DNA ladder (NEB). Lane 1: amplicon representing the *rgg1518* (879 bp).

3.8.2. Expression of transcriptional regulator

The expression of the transcriptional regulator (*rgg1518*) within the *E. coli* BL21 (DE3) pLysS strain containing the recombinant plasmid of interest was induced with the optimal concentration of IPTG (1 mM) when OD600 nm reached a value between 0.5-0.6. Then, the culture was incubated at 37°C for approximately 9 h until OD600 nm had reached 1.5-1.7. (Section 2.20.3). The protein sample was eluted with different concentration of imidazole buffer (25, 50, 100, 200, 300 and 500 mM) as described in section 2.20.3, and the sample was then analysed on SDS-PAGE. The results showed 300 mM imidazole elution buffer was the optimal concentration that led to achieving successful expression of the protein in purest sample state (Figure 3.35). The total molecular weight of Rgg1518 protein was calculated by summation the size of the pure target protein, Tobacco Etch Virus protease (TEV) cleavage site (957 Da) and the

histidine tag (840Da). The results showed an approximate molecular weight of 35.4 kDa for Rgg1518. The expected size for pure target protein was confirmed by SDS-PAGE analysis (Figure 3.33).

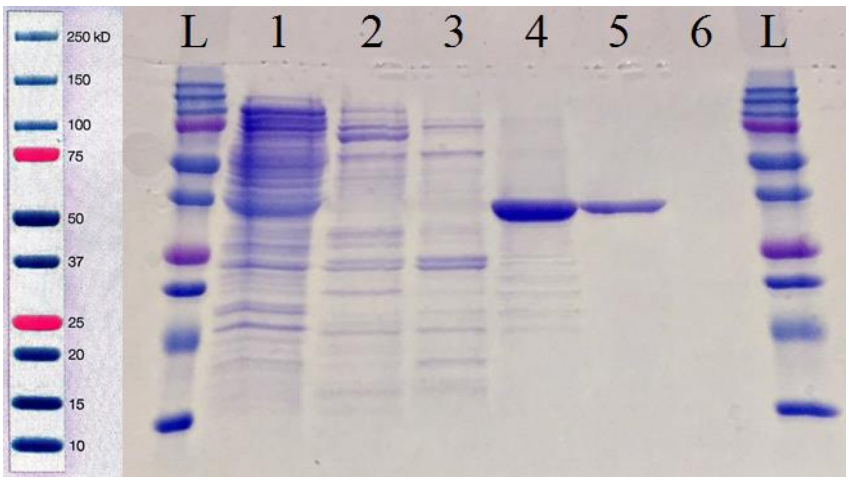


Figure 3.33. SDS-PAGE gel analysis showing the purified recombinant Rgg1518 protein (35.4 kDa). L: protein standard ladder molecular weight marker (Bio-Rad, UK). Lanes 1-6: protein fractions eluted with 25, 50, 100, 200, 300 and 500 mM imidazole elution buffer, respectively.

3.8.3. Dialysis of recombinant protein and MALDI-TOF analysis

Dialysis is important to remove residual imidazole buffer from the recombinant protein sample that could affect EMSA. The recombinant protein eluted with 300 mM imidazole elution buffer was dialysed using the binding buffer as previously described in section (2.20.5). SDS-PAGE gel analysis showed recombinant protein after dialysis, with an expected molecular weight of 35.4 kDa (Figure 3.34).

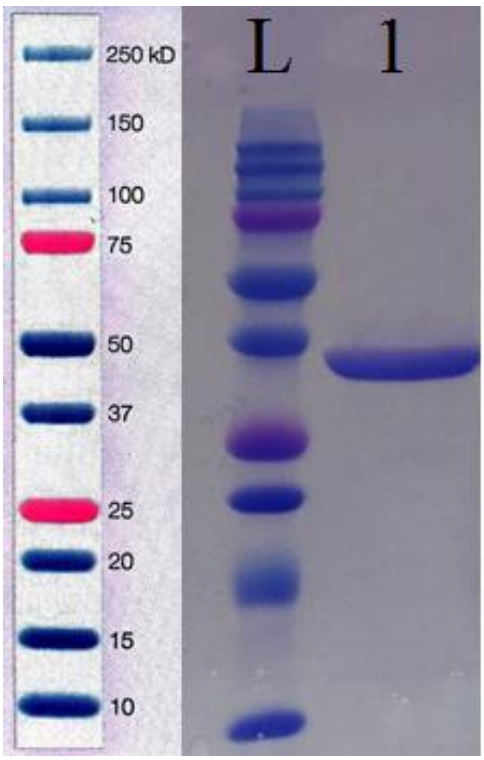


Figure 3.34. SDS-PAGE gel analysis of recombinant protein. L: protein standard ladder molecular weight marker (Bio-Rad, UK). Lane 1: Rgg1518 protein (35.4 kDa) eluted after Dialysis.

To verify the identity of the purified recombinant protein, the sample was sequenced by PNACL, University of Leicester using matrix-assisted laser desorption ionization – time of flight (MALDI-TOF). The results of Mass spectrometry for purified recombinant protein is shown in (Appendix 4). The results further verified the identity of Rgg1518.

3.9. Prediction of promoter region

The putative promoter regions of selected target genes from microarray data were identified using the bacterial promoter prediction tool (BPROM) while (MEME) was used to predict the binding motifs within analysed DNA sequence. BPROM is a bacterial promoter recognition program that predicts bacterial promoters with an accuracy of approximately 80% (Solovyev and Salamov, 2011). The MEME suite

allows the researchers to find out novel motifs in collections of unaligned nucleotide sequences, and to carry out a vast variety of other motif-based analyses. MEME detects the transcriptional factor binding motifs within the analysed DNA and protein sequence (Bailey *et al.*, 2015).

BPROM was used to detect the putative promoter regions (indicated as probe) for *cps* locus genes (SPD_0315-SPD_0328) and selected genes that appeared in the *rgg1518* regulon as determined by the microarray analysis in the presence of galactose and mannose.These genes included SPD_0378, SPD_0126, SPD_0113, SPD_1127, SPD_1634, SPD_1053, SPD_0208, SPD_1461, SPD_0404, SPD_1512, SPD_0193, SPD_0122 and SPD_1342. BPROM determined the core promoter elements (-10 and -35) within the upstream of this genes. The *cps* locus probe was used in EMSA, in addition to *gyrB* (SPD_0709 locus) which used as a control to demonstrate the specificity of binding (Figure 3.35A). In order to identify the putative motif binding site of the *cps* locus, MEME detected the putative motif binding site in the upstream of the screened genes (Figure 3.35B). The putative binding sites according to MEME program were motif 1 and motif 2 (Figure 3.35C).

A

P(SPD_0315-SPD_0328)

F

ACACATCTGCTTCTAAAATATTGTTAGAAAACGATTTGACTGTCCTGATCAATTTGTCATGTTCTTATTTCATTTTACTATATTTTTG
GTTCGCGGGAAGTCTACTAAGATACTTAAAGATGCAGATAGTGAAAAAAGGTGTAGACATTACCGTAAAAAAGTGATATAATCGT
AAGATGTTCAATGTATAGGTGTTAATC

-35 -10

R

gyrB

F

ATGACAGAAGAAATCAAAAATCTGCAGGCACAGGATTATGATGCCAGTCAAATTCAAGTTTTAGAGGGCTTAGAGGCTGTTCGTA
TGCGTCCAGG

R

B

P(SPD_0378)

-35 -10

TTTTTTCTTGCTTTCTGATCCTAAAAAAGATATACTTTGACAGTGAAATAATTTAACACTCAATAAAAATTAAAGAGCAAACTAAG
AAGCTAGACGAAGTCAGCTCAAAATACTGTTTTGAGGTTGCAGATGGAAGCTGACGCGGTTTGAAGAGATTTTCGAAGAGTATAA
ACTGCTTATAAAATAAAAAGGAGCCCTG

P(SPD_0126)

-35 -10

TCGAAAAACTTGGAGGGATAAGAAATGAAAGTAAGACCATCGGTCAAACCAATTTGCGAATACTGTAAAGTTATTCGTCGTAATG
GTCGTGTTATGGTAATTTGCCCAGCAAATCCAAAACACAAACAACGTCAAGGATAAGATAGAAAGGAGAAAAC

P(SPD_0113)

AAAAATATTGTATCATAGCAATGTAAAAAGTAAAAAAGAAAGTCGAAATAAATTTGTGAAAAAAGTTAACGTAATTTTAAAATG
AGCTTAATAAGAGGTACTATGTAGGTAGAAAATAAAGATAGGAGAAAAAGTATGAATAAAAAAATATTTTGTATTCTAGTATGTA
TTATATTTTGATTAGTCTGGCTATAATCTTTCCATGGGGCTGGCCGATATAATGTAAGGGAGTTTGAGTTGTTAAAGGAGGCTCAG
TCCTT

-35 -10

P(SPD_1127)

CTCTAAAATTTCGCAGCCATTTATTAGTAATTGCTACAGAATTCCTAGTCATTACTAGAAATGGACTAGTTTCTTTGAATAATAGA
ACTGCATAATTCTCCTATTCTAGAAGGGGAGGACCAGTATTTCTTTTATGATAGGACTAGATTGTGGTATAATAGAGAGAATAAGT
TTTTTTAGTAAGACAAAGGAGAAAATAG

-35 -10

P(SPD_1634)

-35 -10

AATCCTCCTTGATTAGGTTAGTATATCATATTTTTCTGCTTTTTACTGATAGTTTAGTAAAATTTTAGTAAAAAGGATTGACCTTGG
GAAATCCCTTGGATACAATAGAAGAAAACGATTACACGTTAAGGTGGCTTAACGGACAGTCAAAGGAGAAATCAT

P(SPD_1053)

-35 -10

AAAAAGTAAAAAATAAATAACAAAAACAAACACAAACTGTTGACTTCTTCAAACAATAGTAGTATATTATGTTTGTGAGGACTAA
TTGACTGTTTTTTAAAGACAGCCAATACACTGCTGGAATCCAGCATAGAAAAAATAAAAAGGAGTATACAAT

P(SPD_0208)

AATCAAGATACAAAGCTCGTAAAGAACAAAGCAAAATTAGGAAGTTGGAGAAGTTTGTTTACAAACAAGCCAACTTATCTATTTT
GCACAGTTCTTAGATCGTGTTCAGTTCAGCTCTTGAACTAAATAAGTATCTGAACCCCGTGAAAACTGGCCGTTCTGGCCTGACAA
TTTAACAGGAGAAAATAAAC

-35 -10

P(SPD_1461)

TTTGTATTTGCATTTATCTTAACAAAAATCACTAGGGAATGCAATTAATAACCATTTAACTATTATTTTTTAGCTATTATAAAAAAT
CTTTCCATGATAAAACGCATAATTCCAAGTTTTTTACACTTGATACTATGCGTTTTATAATTTAGAAATTTATCCTGAAAGTTATCT
TTAGAATCTATTTTTCAGTTATTATTTTAATTTTTTCAAAAATTAACTTGACTTAATTTTTTTTTTAATGTATATTAAGAGACAGGAG
GAATACAAGT

-35 -10

-10 -35

P(SPD_0404)

GATTAATGAAGGACTGAGAAATCAGTTCTTTTTTCTTTTTTAGTAAATGAAATCGGTATCTTTTTAATAAAAACAAAAATAACATT
CATAAATAAACGTTAAAATAGAAAATTCAGAAAATTTCTTCTTTTATCTTGAAAAATTTTGAAAAAATGGTATGATAGTAACAAGT
TATTTTTAAGAGGAAAGAAAGGGGAATA

P(SPD_1512) -10

-35

AGTATCATAAAAGTCAGAGAGGTTAGCTTGGAAACTAACCTCTTTTTCCTTTTCAAAATGGGGATTCTTCCTTGAAAATAATCAGT
AATTGTGCTAAAATTAAAGGAACATTCTAAAATATTCGGAATTTAAAGTAAGGAAAAAC

-35 -10

P(SPD_0193)

TCTAAAATATAAAAGAGCAGAGGCTGGTGTTTCAATCTAATTGAACACGGGCTAAACTCGGTGTGAAAAAGATAAACTTCCTAGT
GTCTGCTAGACACTGCGTCAGTTTCCTATTTTCACGTTGAGTTTGACGCCCTTTGTATCTTAGACTTGAGCATAAAAAACGCTCGTT
AAAAACTTTTTGAATAAAAAATATAGAAAAGGAACTATTTTCTC

-35 -10

P(SPD_0122)

CTAGATATGGGTATTAAAATGGAAAAAATTTTTGTTATTATTTTTTTTGTATGTCTATTTATATCATCCATAACTTTTTTAGCCTATG
ATTTTGTTAGCGAAGAAATAAAAAAAGTTGATTATTTGGATGAATGTCGTTTTTTTGATTTTAATCATAGCAATGATAATTTATCCAA
AATTAAGAAAATGATAAGGACACTGTAAAATT

-35 -10

P(SPD_1342)

GATAAATACCAGTTATAACCTAGAAATTGTGTCTTTCTAGACGATATTGAGGACAATATAATCGTAGCCCCAAAATTAGGAATCA
AGACTTATCAGGTTAAGAAAAGAAGTGATGTTGTTGATATTTAAACAATTTATTTAAACTAAAAACTCTCTATCTATTAATCCGA
GATTCCCCTCTCGGATTTTGTGTTGCTTTCATTAAATTTACAGGTAAGACTTGC

C 1

SPD_0315	GATAGTGAAA	AAAGGTG	TAGACATTAC
SPD_0378	TATAAAATAA	AAAGGAG	CCCTG
SPD_0126	GATAAGATAG	AAAGGAG	AAAAC
SPD_0113	TTGAGTTGTT	AAAGGAG	GCTCAGTCCT
SPD_1127	TTAGTAAGAC	AAAGGAG	AAAATAG
SPD_1634	ACGGACAGTC	AAAGGAG	AAATCAT
SPD_1053	GAAAAAATAA	AAAGGAG	TATACAAT
SPD_0208	TGACAATTTA	ACAGGAG	AAAATAAAC
SPD_1461	ATATTAAGAG	ACAGGAG	GAATACAAGT
SPD_0404	AAGAGGAAAG	AAAGGGG	AATA
SPD_1512	TGCTAAAATT	AAAGGAA	CATTCTAAAA
SPD_0193	AAAATATAGA	AAAGGAA	CTATTTTCTC
SPD_0122	AAGAAATAAA	AAAGTTG	ATTATTTGGA
SPD_1342	CATTAAATTT	ACAGGTA	AGACTTGCTA

C 2

SPD_0315	GTAGACATTA	CCGTAAAA	AAGTGATATA
SPD_0378	GCTTTCTGAT	CCTAAAAA	AGATATACTT
SPD_0126	TGGTAATTTG	CCCAGCAA	ATCCAAAACA
SPD_0113	AATATTGTAT	CATAGCAA	TGTAAAAAGT
SPD_1127	CCTAGTCATT	ACTAGAAA	TGGACTAGTT
SPD_1634	AAAGGATTGA	CCTTGGGA	AATCCCTTGG
SPD_1053	TGGAATCCAG	CATAGAAA	AAATAAAAAG
SPD_0208	TATCTGAACC	CCGTGAAA	ACTGGCCGTT
SPD_1461	TTGCATTTAT	CTTAACAA	AAATCACTAG
SPD_0404	TCTTCTTTTA	TCTTGAAA	AATTTTGAAA
SPD_1512	GGGGATTCTT	CCTTGAAA	ATAATCAGTA
SPD_0193	TAGACTTGAG	CATAAAAA	ACGCTCGTTA
SPD_0122	TGATTTTAAT	CATAGCAA	TGATAATTTA
SPD_1342	CCAGTTATAA	CCTAGAAA	TTGTGTCTTT

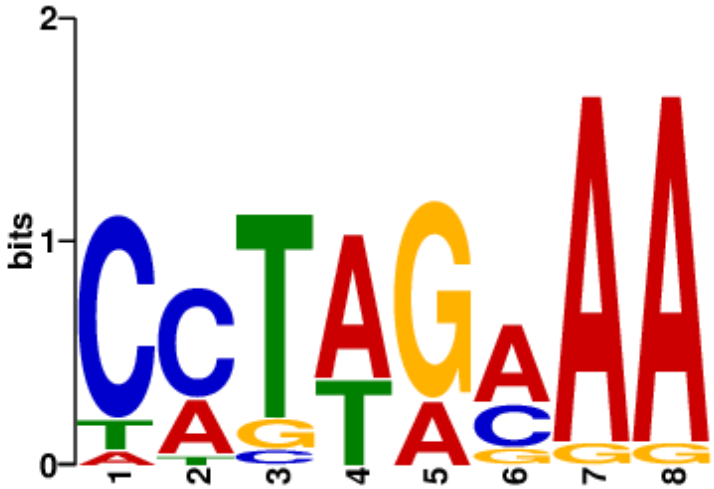


Figure 3.35. (A) The sequences of putative promoter region of *cps* locus gene which was used in EMSA, the putative binding sites according to MEME program (1) and (2) are shown in blue and red, respectively, the coding sequence of *gyrB* which was used as a control probe. The core promoter elements -10 and -35 are indicated. F: indicates forward fluorescence primer while R: refers to the reverse primer used for amplifying the probes (Table2.20). (**B**) The putative consensus sequence using MEME program to found out putative binding sites. (**C**) The putative binding sites according to MEME program (C1) motif 1 and (C2) motif 2.

3.10. Amplification of putative promoter regions

The putative promoter probe for *cps* locus and *gryB* were amplified using the primers listed in Table 2.20 as described previously in section 2.7. The *gryB* (SPD_0709 locus) encodes for gyrase and used as control probe in EMSA. The amplified PCR products were purified as described previously in section 2.9. The amplified probes were then analysed by agarose gel electrophoresis (Figure 3.36), confirming successful amplification.

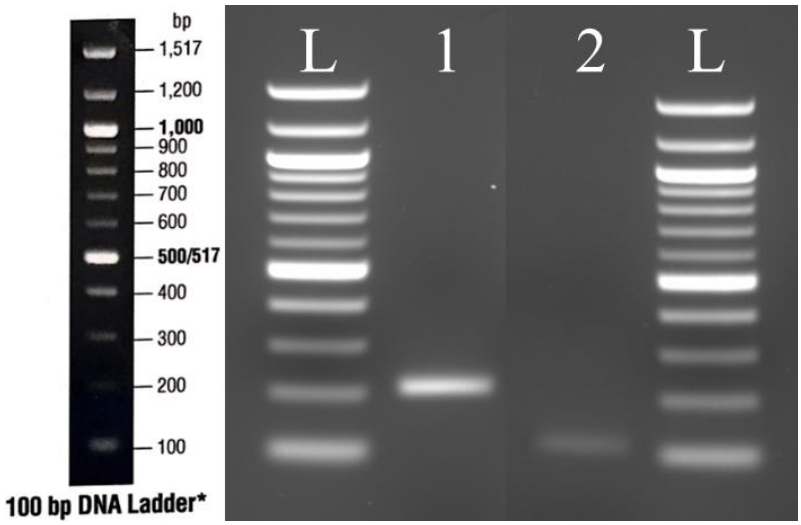


Figure 3.36. Agarose gel electrophoresis showing the amplified probes of the target genes. L: 500 ng of 100 bp DNA ladder (NEB). Lanes 1and 2: amplified products of *cps* locus gene, and *gyrB*, respectively.

3.11. Investigating direct interaction of Rgg1518 with the putative promoter of *cps* locus by electrophoretic mobility shift assay (EMSA)

3.11.1. Rgg1518 interacts directly with the putative *cps* locus promoter

This *cps* locus genes were identified to be regulated by Rgg1518 in microarray analysis and by reporter assays as previously described in section 3.3 and 3.7.2, respectively. To determine whether this regulation is direct or indirect, EMSA was performed according

to the protocol described previously (Chang *et al.*, 2011). The fluorescently labelled DNA probe for the putative *cps* locus promoter was amplified. Subsequently, 10 μM of FAM-labeled *cps* probe was mixed with increasing concentration of His-tagged recombinant Rgg1518 (0.2 - 0.6μM). The reaction mixture was incubated for 20 min at room temperature as described previously in section 2.21. The DNA-protein reactions were analysed in non-denaturing condition by PAGE. *gyrB* was used as a negative control probe as shown in Figure 3.37. The results showed that no change in mobility of *gyrB* could be detected, indicating that Rgg1518 recombinant protein could not bind to *gyrB*. Hence, Rgg1518 binding was specific for *cps* locus.

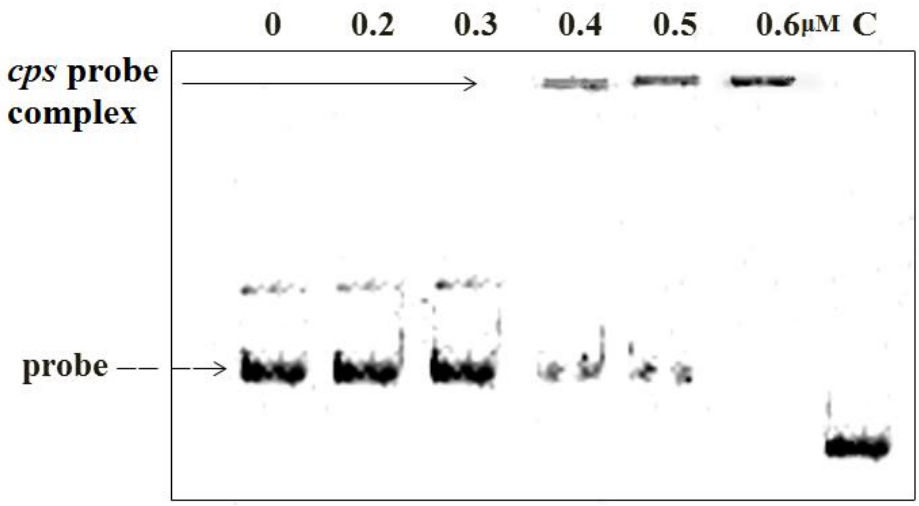


Figure 3.37. EMSA analysis showing the direct interaction of recombinant His-tagged Rgg1518 protein with the promoter of *cps* locus. 10 nM of FAM-labeled *cps* probe mixed with increasing concentration of recombinant Rgg1518 protein and incubated for 20 min at room temperature. The DNA-protein reactions were separated on non-denaturing PAGE (8%) by electrophoresis. Lane 1: DNA probe alone as a control, represented by a dotted arrow. DNA-protein interaction is seen as an upward shift of DNA probe, represented by a black arrow. Lane C: *gyrB* plus 0.6 μM Rgg1518.

3.12. Putative motif binding site Mutation

To locate the Rgg1518 binding motif in *cps* locus, I searched for a consensus binding motif in the promoter regions of differentially expressed genes identified in microarray analysis (Section 3.9). Using MEME and the putative promoter regions of genes differentially expressed putative operons on galactose/mannose, I identified two binding motifs (Figure 3.34). The putative motif binding sites 1 (AAAGGTG) and 2 (CCGTAAAA) were selected. These are near to the core promoter elements as shown in Figure 3.34A. To identify, if any of these motifs were indeed the binding site for Rgg1518, the mutation of the binding motifs 1 and 2 was carried out separately using splicing overlap extension as described in section 2.14.1. As shown in Figure 3.38A, Lanes 1A and 2A the amplified flanking regions of the putative motif binding site 1was approximately 134 bp and 105 bp , respectively. Lane 3A represents SOEing product of the putative motif binding site 1 (probe 1) approximately 239 bp. Figure 3.38B, Lanes 1B and 2B illustrate the amplified flanking regions of putative motif binding site 2, approximately 150 bp and 100 bp, respectively. Lane 3B represents SOEing product of putative motif binding site 2 (probe 2), approximately 250.

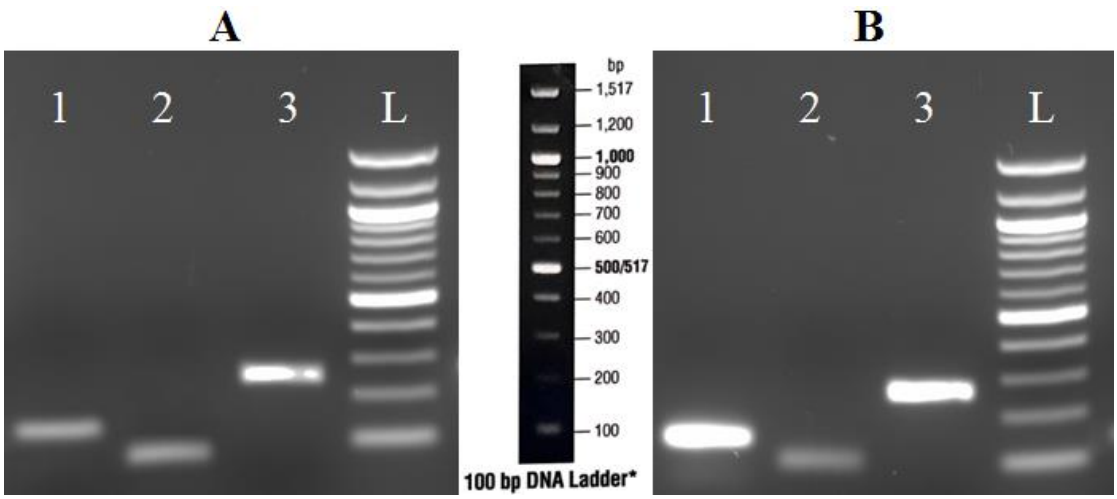


Figure 3.38. Agarose gel electrophoresis showing the amplified flanking regions and SOEing products for Rgg1518 binding analysis. (A) Putative motif binding site 1, Lanes 1A and 2A: represent amplified of up- and downstream regions, approximately 134 bp and 105 bp, respectively. Lane 3A: (probe 1) represents the fused SOEing product, approximately 239 bp. L: 500 ng of 100 bp DNA ladder (NEB). (B) Putative motif binding site 2, Lanes 1B and 2B: represent amplified of up- and downstream regions, approximately 150bp and 100 bp, respectively. Lane 3B: (probe 2) represents the fused SOEing product, approximately 250 bp.

3.13. EMSA for putative motif binding sites 1 or 2 (probe 1 or probe 2)

EMSA was used to analyse the interaction of the Rgg1518 recombinant protein either with putative motif binding sites 1 or 2 to find out the binding sequence. EMSA was carried out as described previously in section 3.11.1. The mutation of putative motif binding sites 1 or 2 was designated as probe 1 or probe 2, respectively, 10µM of FAM-labeled probe 1 or probe 2 mixed with a serial increasing concentration of His-tagged protein Rgg1518 (0.2 - 0.6µM) and incubated for 20 min at room temperature as described previously in section 2.21. The DNA-protein reactions were analysed in non-denaturing condition by PAGE. *gyrB* was used as a negative control probe in addition to *cps* probe used as a positive control as shown in Figure 3.39. The results showed that Rgg1518 can bind when putative motif binding site 2 was absent whereas it cannot bind when putative motif binding site 1 was absent. Thus, this reveals that binding motif 1 is

the binding sequence for Rgg1518 (AAAGGTG) Figure 3.38. In addition, there was no change in mobility of *gyrB* could be detected, and Rgg1518 binding also was specific with *cps* locus gene.

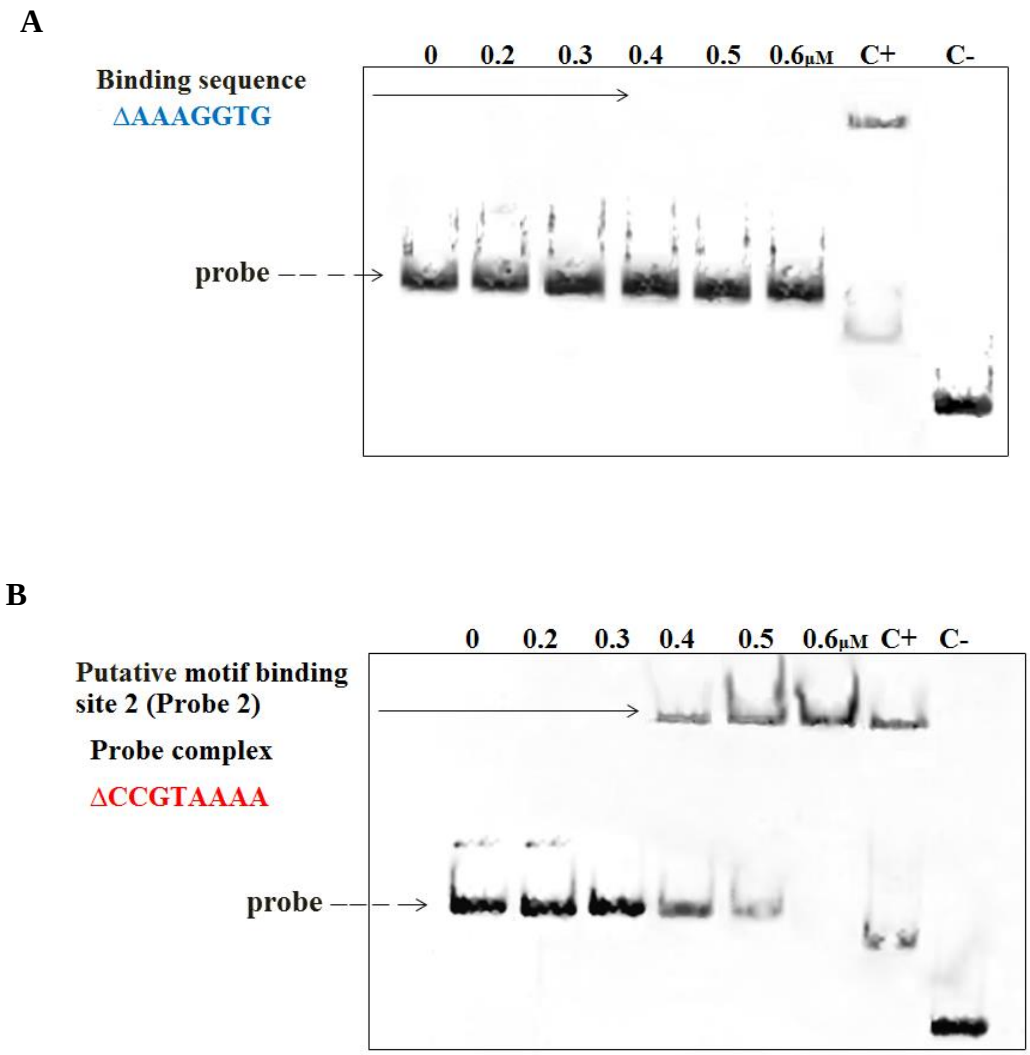


Figure 3.39. EMSA indicating the Rgg1518 interaction with different putative binding sites in *cps* locus. (A) Binding sequence of putative motif binding site 1, and (B) putative motif binding site 2. Lanes 1: contain DNA probe alone as a control, represented by a dotted arrow. DNA-protein interaction is seen as an upward shift of DNA probe, represented by a black arrow. Lane C+: intact *csp* locus probe as a positive control. Lane C-: *gyrB* plus 0.6 µM recombinant Rgg1518.

Section E: *In vivo* pneumococcal virulence studies

3.14. *In vivo* studies:

It was reported that inactivation of *rggs* attenuate *S. suis* serotype 2 (SS2) virulence significantly in a piglet infection model, and decrease the utilisation of non-glucose carbohydrates, including lactose and maltose (Zheng *et al.*, 2011). Moreover, *S. pyogenes*, Rgg (also known as RopB) is required for the expression of excreted proteins like SpeB, which has effect on host-pathogen interactions by destroying the extracellular matrix and degradation of complement factors in host (Neely *et al.*, 2003; Kappeler *et al.*, 2009). The results in this study showed the involvement of Rgg1518 in host derived sugar metabolism, capsule production and expression, which lead to it hypothesis that Rgg is involved in pneumococcal survival and virulence. Although the mutation of Rgg0999 did not show any phenotype in the assays utilised in this study, I also determined Rgg0999 involvement in colonisation and virulence. To examine Rggs role *In vivo*, I tested the Δ*rgg0999* and /or Δ*rgg1518* in a mouse model of pneumococcal pneumonia and colonisation that develops after intranasal infection. Chiavolini *et al.* 2008 demonstrated that mouse models are crucial tools to investigate the infectious disease. Mice are the most commonly used animal to investigate the pneumococcal virulence because it is easy to handle, their cost is lower, and their maintenance is not complicated. Hence, sufficient number of mice can be used to obtain statistically significant results for analysis (Mohawk and O'Brien, 2011).

3.14.1. Contribution of *Rgg* in pneumococcal virulence

In a pneumonia model, the infection through intranasal aspiration is the most common technique used as it does not need invasive surgical procedures, and it mimics the natural pneumococcal infection process in humans (Chiavolini *et al.,* 2008). The CD1 mouse strain was used because this is an outbred strain and reflects a certain extent, the heterozygosis found in human population. The contribution of Rggs to pneumococcal virulence was determined in a mouse model of pneumococcal infection as previously described in section 2.34.2. The mutant strains Δ*rgg0999*, Δ*rgg1518*, the double mutant Δ*rgg0999/1518* and the genetically complemented mutants (Δ*rgg0999c* and Δ*rgg1518c*) were tested alongside the wild type strain D39. The double mutant was used to determine if there is any synergistic affect between these two Rggs *In vivo*.

After intranasal infection, the infected mice were monitored for seven days to determine the survival time (Figure 3.40). The results showed that the median survival time of mice infected intranasally with Δ*rgg0999*, Δ*rgg1518* and Δ*rgg0999/1518* (101.1 h ± 18.58, 123.3 h ± 18.26 and 134.6 h ± 17.02 respectively, n=10) were significantly longer than the wild type infected group (47 h ± 3.12 n=10) ($p<0.05$, $p<0.01$and $p<0.001$for Δ*rgg0999*, Δ*rgg1518* and Δ*rgg0999/1518*), hence there is synergistic effect among Rggs *In vivo*. The introduction of intact copies of *rgg0999* and *rgg1518* into the mutants, *Δrgg0999* and Δ*rgg1518*, respectively, reconstituted the virulence of these strains as the median survival times of mice infected with Δ*rgg0999c* (63.8 h ± 26.1, n=5) and Δ*rgg1518c* (60.4 h ± 26.9, n=5) were not significantly different from the wild type infected cohort ($p>0.05$). The results obtained with the complemented strains rule out the possibility of polar effect of mutations, and relate to the significant attenuation

of virulence due to the loss of Rggs. These results demonstrate that the Rggs contribute pneumococcal virulence.

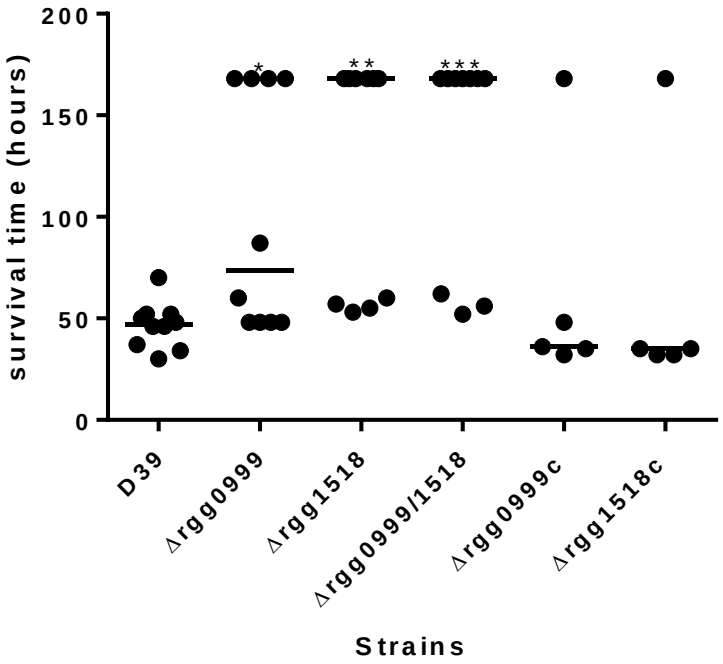


Figure 3.40. The survival time of mice infected intranasally with approximately 2 x 10^6 pneumococci/mouse. Dots represent the times each mouse became very lethargic. Horizontal lines represent the median times to the lethargic state.There are significant differences in survival times between wild type and mutant strains using Mann Whitney test, (*p<0.05, **p<0.01 and ***p<0.001).There is synergistic effect among Rggs *In vivo*

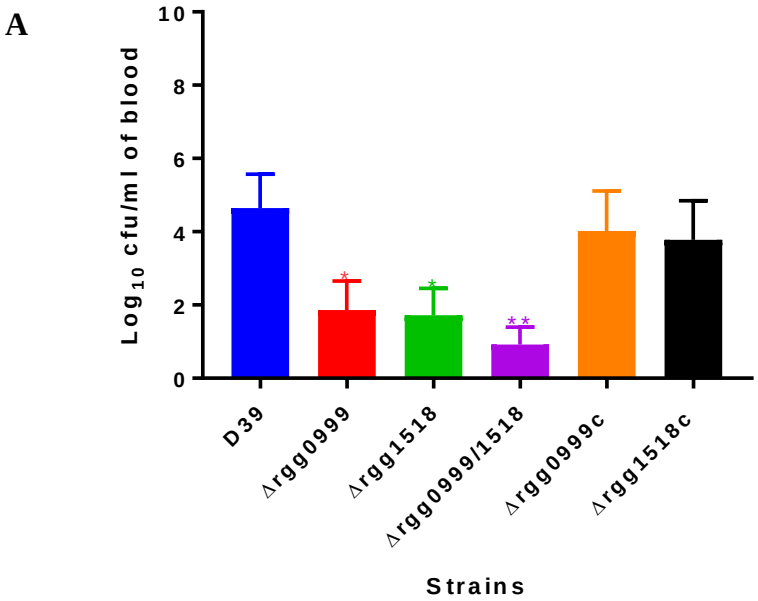
3.14.2. Level of bacteraemia in mice after infection

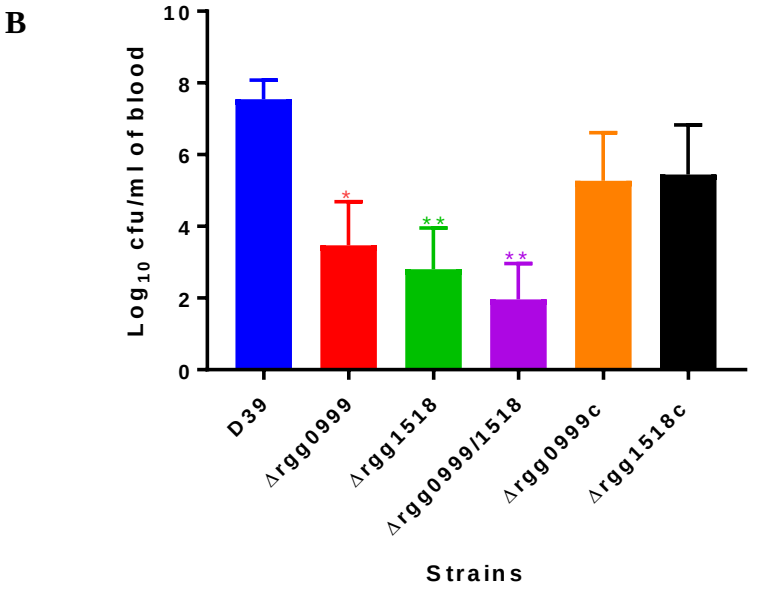
The progression of bacteraemia in mice infected with the pneumococcal strains was also determined. The effect of *Rgg* in bacteraemia was investigated by taking a sample of blood from intranasally infected mice at 24 and 48 h post-infection (Figure 3.41 A and B). The tail bleeding method was used to collect blood samples from the infected mice for culturing. The bacterial load in the Δ*rgg0999* infected group was significantly lower

at 24 and 48 h post-infection (Log_{10} 1.8±0.79 CFU/ml, Log_{10} 3.47±1.22 CFU/ml, respectively, n=10) compared to the wild type strain at 24 h (Log_{10} 4.65±0.92 CFU/ml, n=10) and 48 h post-infection (Log_{10} 7.55±0.53 CFU/ml, n=10) ($p<0.05$). The bacterial counts for the Δ*rgg1518* infected cohort was significantly lower at 24-and 48 h post-infection (Log_{10} 1.62 ±0.74 CFU/ml, Log_{10} 2.8±1.15 CFU/ml, respectively, n=10) in comparison with the wild type strain ($p<0.05$, $p<0.01$). Moreover, numbers of Δ*rgg0999/1518* in the blood of mice was significantly lower at 24-and 48 h post-infection (Log_{10} 0.92±0.47 CFU/ ml, Log_{10} 1.96±0.99 CFU/ml, respectively, n=10) compared to the wild type strain ($p<0.01$), hence there is synergistic effect among Rggs *In vivo*. Moreover the progression of bacteraemia in cohort infected with complemented strains, Δ*rgg0999c* at 24 h (Log_{10} 4.02 ± 1.09 CFU/ml, n=5) and 48 h post-infection (Log_{10} 5.27 ± 1.34 CFU/ml, n=5) and Δ*rgg1518c* at 24 h (Log_{10} 3.78 ± 1.06 CFU/ml, n=5) and 48 h post-infection (Log_{10} 5.44 ± 1.38 CFU/ml, n=5) was similar to that of wild type infected cohort ($p>0.05$). Moreover, both complemented strains Δ*rgg0999c* and Δ*rgg1518c* had significantly higher bacterial load in the blood of infected mice at 24 and 48 h of post infection compared to the mutants Δ*rgg0999* and/or Δ*rgg1518*.

The numbers of wild type D39 strain progressively increased post-infection at 48 h (Log_{10} 7.55±0.53 CFU/ml, n=10), while no mutant pneumococci could be detected in the blood of the survivors that were infected by either with Δ*rgg0999*, Δ*rgg1518* and Δ*rgg0999/1518* post infection ($p<0.0001$) (Figure 3.41C). There is no significant difference in bacterial counts for the dead mice that were infected by either with Δ*rgg0999* (Log_{10} 6.94 ±0.83 CFU/ml, n=6), Δ*rgg1518* (Log_{10} 7.0 ±0.34 CFU/ml, n=4), Δ*rgg0999/1518* (Log_{10} 6.63±0.18 CFU/ml, n=3), Δ*rgg0999c* (Log_{10} 6.65±0.34 CFU/ml,

n=4) and Δ*rgg1518c* (Log_{10} 6.8±0.36 CFU/ml, n=4) in comparison with the wild type strain (Log_{10} 7.55±0.53 CFU/ml, n=10) (p>0.05) (Figure 3.41D).





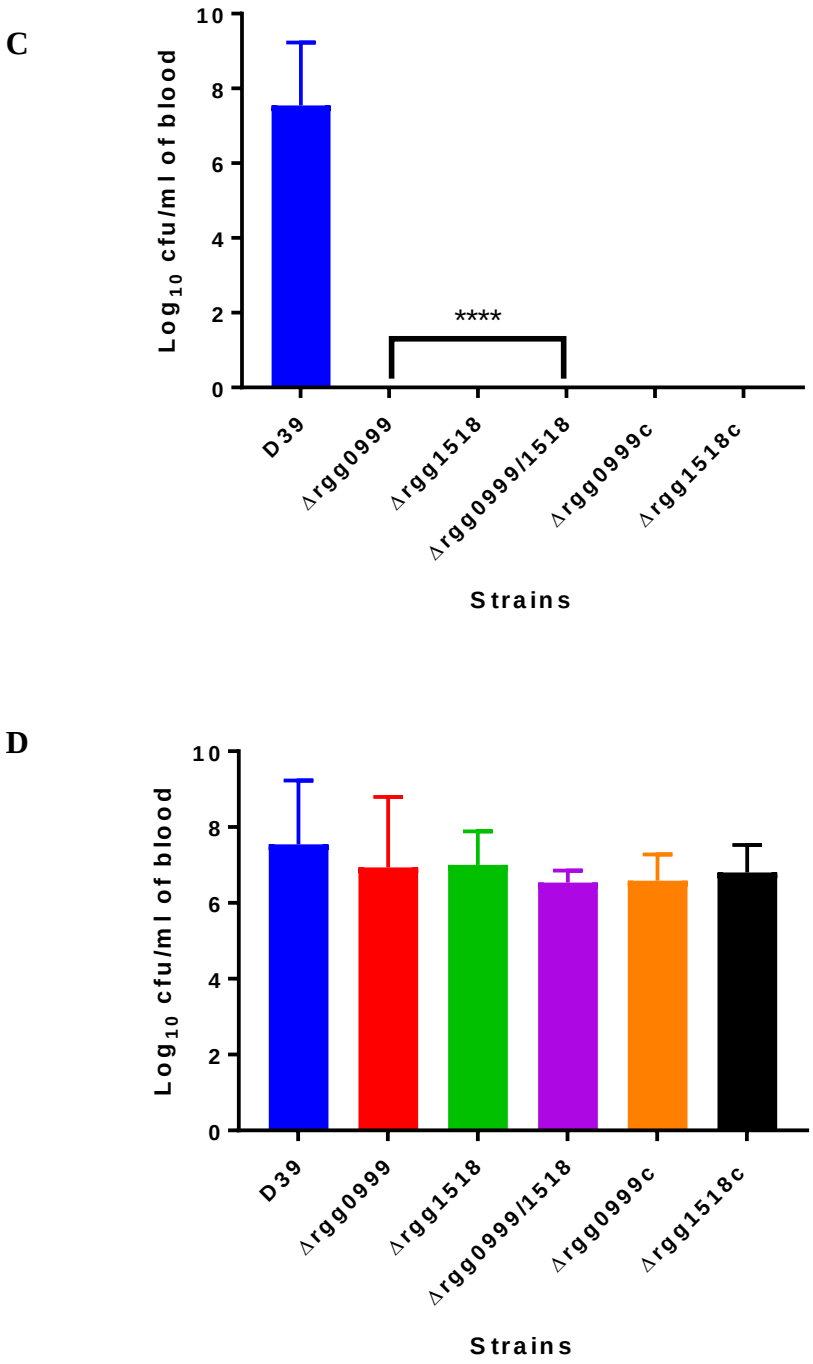
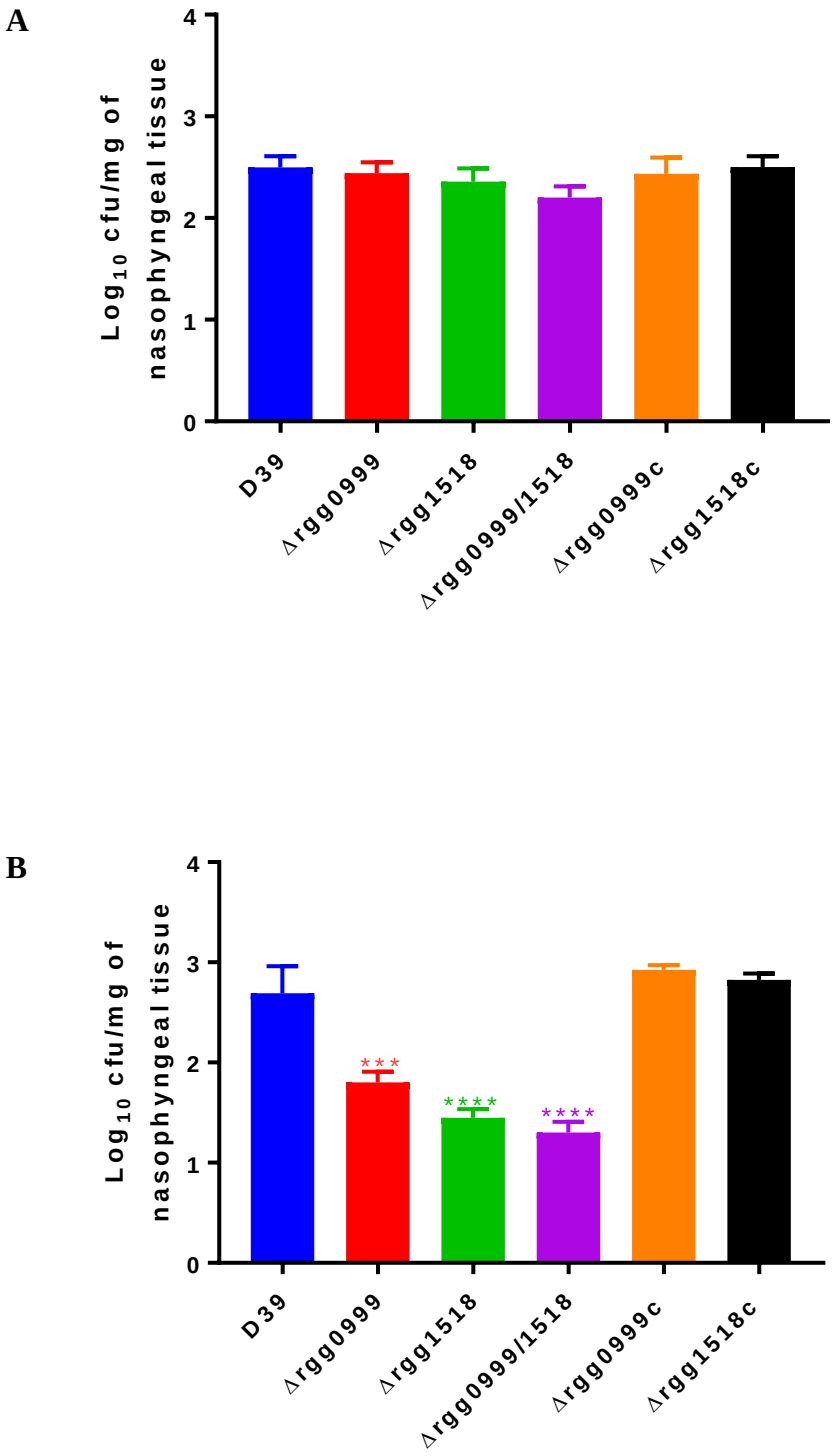


Figure 3.41. Development of bacteraemia in mice after intranasal infection with *Δrgg0999*, *Δrgg1518*, *Δrgg0999/1518* and their complemented and derivatives at 24 h (A), 48 h (B), the survivor mice (C), the dead mice (D) post-infection. Each column represents the average data for wild type D39 and mutant strains obtained from ten mice, while complemented strains were from five mice. Error bars illustrate the standard error of the mean. There are significant differences in bacterial counts between wild type and mutant strains using one way ANOVA and Tukey's multiple comparisons test, (*$p<0.05$, **$p<0.01$ and ****$p<0.0001$). While there are no significant differences in bacterial counts between wild type and the dead mice, ($p>0.05$).The complemented strains restored the characteristics back to the wild type level, ($p>0.05$).

3.14.3. Contribution of *Rggs* in pneumococcal colonisation

A pneumococcal colonisation model was used to test the ability of *rgg* mutants to survive in the nasopharynx of CD1 mice as described in section (2.17.3). Pneumococci were administered intranasally to initiate infection in mice. Serial dilution of homogenised nasopharyngeal tissue of infected mice was used to count the numbers of bacteria immediately after infection (zero time), and 3 and 7 days post-infection. The number of colony forming unit was assessed per mg of homogenised tissue. One hour after infection, the bacterial load in the nasopharyngeal tissue for all strains Δ*rgg0999*, Δ*rgg1518*, Δ*rgg0999/1518*, Δ*rgg0999c* and Δ*rgg1518c* (Log_{10} 2.44±0.104 CFU/mg, Log_{10} 2.36±0.13 CFU/mg, Log_{10} 2.20±0.108 CFU/mg, Log_{10} 2.43±0.158 CFU/mg and Log_{10} 2.5±0.108 CFU/mg, respectively, n=5) was similar to that of wild type (Log_{10} 2.49 ± 0.11 CFU/mg, n=5) ($p>0.05$) (Figure 3.42A). Furthermore, the colony counts at 3 days post-infection for Δ*rgg0999*, Δ*rgg1518* and Δ*rgg0999/1518* (Log_{10} 1.8±0.19 CFU/mg, Log_{10} 1.45±0.09 CFU/mg, and Δ*rgg0999/1518* Log_{10} 1.3±0.11 CFU/mg, respectively, n=5) were significantly lower than that of wild typeD39 strain (Log_{10} 2.69±0.27 CFU/mg, n=4), hence there is synergistic effect among Rggs *In vivo*. There was no significant difference in the bacterial counts for complemented strains, Δ*rgg0999c* (Log_{10} 2.93±0.05 CFU/mg, n=5) and Δ*rgg1518c* (Log_{10} 2.83±0.07 CFU/mg, n=5) compared to the bacterial counts of wild type D39 strain ($p<0.001$) (Figure 3.42B). On the other hand, at 7 days post-infection the colony counts for Δ*rgg0999*, Δ*rgg1518*, Δ*rgg0999/1518* Log_{10} 1.3±0.12 CFU/mg, Log_{10} 0.95±0.104 CFU/mg, and Log_{10} 0.6±0.23CFU/mg, respectively, n=5), were significantly lower than that the bacterial counts of wild type D39 strain (Log_{10} 2.975±0.17 CFU/mg, n=5) ($p<0.0001$) (Figure 3.42C). There was no significant difference in the bacterial counts for complemented strains, Log_{10} 2.93±0.05 CFU/mg, and Log_{10} 2.85±0.087 CFU/mg, respectively n=5)

compared to the bacterial counts of wild type D39 strain (p>0.05). These results clearly prove the involvement of Δ*rgg0999* and/or Δ*rgg1518* in pneumococcal colonisation.



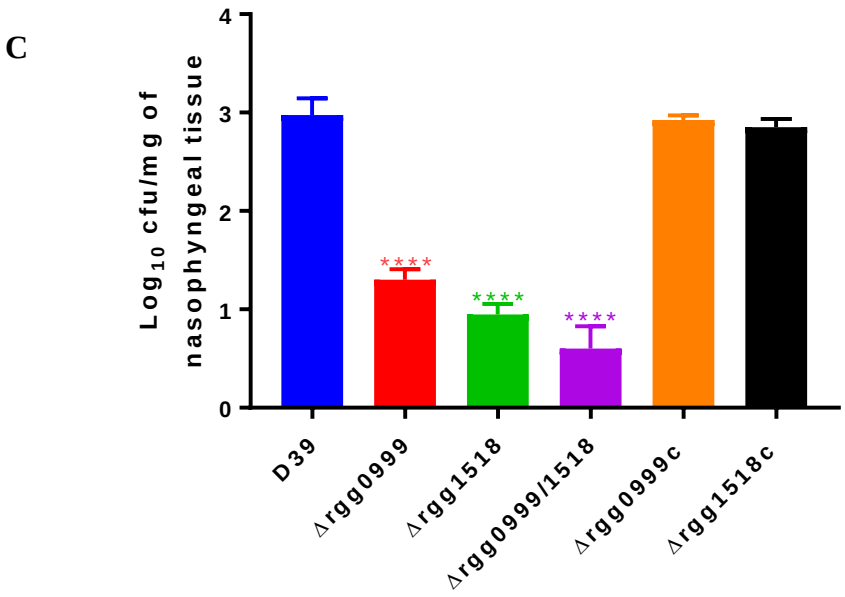


Figure 3.42. Pneumococcal strains *Δrgg0999*, *Δrgg1518* and *Δrgg0999/1518* have less ability to colonise nasopharynx. Approximately 5 x 10^5 CFU of pneumococci was used to inoculate mice. Five mice were culled at 0 time (A), day 3 (B) and day 7 (C), CFU/mg of pneumococci were defined by serial dilutions of homogenised nasopharyngeal. Each bar represents the average data collected from five mice. Error bars show the standard error of the mean. There are significant differences in bacterial count between wild type D39 and mutant strains using one-way ANOVA and Tukey's multiple comparisons test, (***p<0.001 and ****p<0.0001).

Chapter 4. Discussion

Streptococcus pneumoniae is considered to be the fourth most frequent microbial cause of fatal infections among humans, and one of the most common causes of serious diseases worldwide such as invasive pneumococcal disease and community-acquired pneumonia. Moreover, *S. pneumoniae* is considered to be a major cause of morbidity and mortality associated with pneumococcal disease throughout the world (WHO, 2003; Hanna-Wakim *et al.,* 2012). Indeed, it has been reported that the pneumococcus is responsible for the deaths of approximately 3-5 million person per year due to pneumonia, meningitis, and bacteraemia (WHO, 2014; van Deursen *et al.,* 2016). It has been also reported that 67% of the estimated 14.5 million pneumococcal disease cases among children each year are recorded in Africa and Asia, and around one million children in age under 5 died from pneumonia in 2013 (O'Brien *et al.,* 2009; WHO, 2014). Moreover, depending on the age groups, 30-60% of survivors suffer long-term sequelae including hearing loss, neurological deficits, and neuropsychological impairments (Koedel *et al.,* 2002). This microbe has a high ability to attack the host immune system and cause infection due to the emergence of multiple virulence determinants such as the capsule polysaccharides which is known as the major virulence determinant of the pneumococcus because it prevents iC3b opsonisation required for efficient phagocytosis (Nelson *et al.,* 2007). The other identified virulence factor is the thiol-activated toxins pneumolysin (PLY), which has the ability to evade the immune system of the host and promote inflammation within the infected region (Johnston, 1991). Several pneumococcal enzymes are also implicated in virulence and pathogenesis such as autolysin, neuraminidase, hyaluronidase and superoxide dismutase (SOD) (Mitchell, 2000). The autolysin enzyme, located on the cell wall of the pneumococcus, is one of several enzymes that can degrade bacterial peptidoglycan

causing cell lysis. Neuraminidase is an enzyme, also known as sialidase, which cleaves N-acetylneuraminic acid from mucin, glycolipids, glycoproteins, and oligosaccharides on host cell surface (Rosenfeld *et al.,* 1992; Kadioglu *et al.,* 2008). Rosenfeld *et al.* (1992) suggested that neuraminidase could promote colonisation by diminishing the mucus viscosity, or by letting cell surface receptors to be directly exposed to *S. pneumoniae* (Berry *et al.,* 1996). Hyaluronidase is one of the pneumococcal virulence factors that may aid *S. pneumoniae* to spread and colonise by degrading hexosaminidic linkages of hyaluronic acid, which is a component of the extracellular matrix (Kostyukova *et al.,* 1995; Berry and Paton, 2000). Berry *et al.* (2000) also suggested that there is a strong relationship between the activity of hyaluronidase and the ability to induce meningitis. All recent studies have proven that pneumococcus is still a serious and persistent threat for public health around the world. The burden of pneumococcal diseases is believed to be increasing in the future further because the pneumococcal strains develop resistance against the existing antibiotics (Hansman and Bullen, 1967; Li˜nares *et al.,* 2010). Therefore, developing effective therapeutics for this microbe is an urgent need.

In addition to traditional virulence determinants, several studies also indicated that this pathogen can adapt to diverse environmental conditions in tissues such as lung, central nervous system and ear channel, very likely using its efficient systems to sense and respond to environmental stimuli. The pneumococcus usually is present in the multi-sugar environment in the host tissues, and it can utilise 32 different sugars such as glucosides, galactosides and polysaccharides (Bidossi *et al.,* 2012). Pneumococcal host-derived glycan utilisation plays an essential role for survival and virulence. The mucin is a high-molecular-weight glycoprotein and a very rich source of galactose, N-

acetylgalactosamines, N-acetylglucosamine, mannose, fucose, sulfated sugars, and sialic acid (Terra *et al.,* 2010). During infection of respiratory tract, *S. pneumoniae* shows a tendency to rely on mucin (Bogaert *et al.,* 2004) as the main source for its nutrition requirements due to the high content of carbohydrates in mucin which makes up to 85% of total dry weight (Rose and Voynow, 2006). Galactose is very important sugar for pneumococcal biology in the respiratory tract where the microbe colonises and disseminates to cause serious diseases, because it forms a large percentage of mucin sugar content, up to 46%, in the form of galactose and N-acetylgalactosamines (Terra *et al.,* 2010). The pneumococcus also has the ability to cleave galactose from host glycans through the activity of β-galactosidase that is coded at least by two genes, *bgaA* and *bgaC* (Terra *et al.,* 2010; King *et al.,* 2010). The presence of these two genes that are responsible for β-galactosidase activity reflects the evolutionary importance of galactose for the pneumococcus.

According to the discussion above, the main hypothesis of this study was that the ability to sense and adapt to diverse environmental conditions are also important for virulence and survival. One of the most important mechanisms for adaption is transcriptional regulation. The environmental adaptation in Gram-positive bacteria is mediated either by two-component systems (TCS) (Standish *et al.,* 2007) or standalone transcription regulators (McIver, 2009). TCS consists of a membrane-bound sensor protein histidine kinase (HK) and a cytoplasmic cognate response regulator (RR); both regulate gene expression in response to changes of environment (Hoch, 2000). On the other hand, stand-alone response regulators (RRs) do not have a cognate sensor protein and they regulate gene expression by sensing the changes in intracellular conditions (Kreikemeyer *et al.,* 2003).

Rgg is one of pneumococcal transcriptional regulators. Rgg-like regulators, also known as GadR, MutR, RopB, first identified in *S. gordonii* as the regulator of glucosyltransferase gene, are a family of transcriptional factors that are widely spread in low GC Gram-positive bacteria including in *S. pyogenes* (Fleuchot *et al.,* 2011). Rgg-like protein has a preserved helix-turn-helix (HTH) motif and several invariant residues in the amino terminus that is essential for binding to the promoter regions of Rgg-regulated genes. While Rgg regulators have been studied in other streptococci, their functions in pneumococcal biology are not well studied (Sulavik *et al.,* 1992; Pulliainen *et al.,* 2008). Therefore, to increase the knowledge on pneumococcal Rggs, I studied the function of Rgg0999 and Rgg1518 in pneumococcal metabolism, colonisation, and virulence.

In this study, the *rgg0999* and *rgg1518* of *S. pneumoniae* have been studied, and both were found to be essential for pneumococcal survival and virulence in a mouse model of colonisation and bacteraemia. The study showed that Rgg1518 is involved in galactose and mannose metabolism, and it is required for activation of capsule expression. Interestingly, the presence of Rgg1518 led to decreasing the capsule production.

4.1. The importance of Rgg0999 and Rgg1518 in pneumococcal virulence

Various studies have shown that Rgg is transcriptional regulator and plays an important role in virulence and mediates non-glucose carbohydrate metabolism in various Gram-positive bacteria. Rggs were studied in detail as they have essential roles in the control of main physiological processes in streptococci other than pneumococcus (Sulavik *et al.,* 1992; Pulliainen *et al.,* 2008). Rgg of *S. pyogenes* (serotype M49) modulates the

expression of virulence factors and metabolic enzymes that may be crucial for the pathogen's adaptation to different metabolic substrates (Chang *et al.*, 2011). Additionally, *rgg* in *S. suis* serotype 2 (SS2) has critical role in virulence and decrease the utilisation of non-glucose carbohydrate, including lactose and maltose (Zheng *et al.*, 2011). The information about the role of Rggs in pneumococcal pathogenesis is scarce.

To test my hypothesis and identify the role of Rgg0999 and Rgg1518 in *S. pneumoniae*, the mutations of *rgg0999* and/or *rgg1518* were generated using splicing overlap extension (Horton, 1995). This strategy is more advantageous than other strategies such as plasmid-based mutation techniques, hence less labour intensive and quick (Horton *et al*., 1993). This SOEing method allows production of desired genetic mutations in defined locations within the whole gene. I built mutant gene construct in which upstream and downstream are fused with marker gene in this case is spectinomycin or kanamycin resistant cassette, this construct is produced by two round of RCR and then transformed in to *S. pneumoniae*. Transformation process is done by competence stimulating peptide and homologous recombination. The target gene is deleted by inserting spectinomycin or kanamycin.

In order to eliminate the potential polar effects that may be caused by mutation, genetic complementation of mutant strains was needed for phenotypic restoration of the mutant strains. Deletion-insertion mutation can passively effect on the expression of upstream or downstream genes which is called polar effect (Reyrat *et al.,* 1998). The polar effect caused by a deletion-insertion mutation is likely to be higher if the target gene is the first one of an operon (Guiral *et al.,* 2006). The pCEP plasmid (Guiral *et al.,* 2006) was used for the construction complementation of Δ*rgg0999* and *Δrgg1518*. The pCEP was

used to generate complementation strains because it is non-replicative plasmid in *S. pneumoniae*. Moreover, the multiple cloning sites of pCEP are surrounded by 2 kb DNA homologous to *S.pneumoniae* D39 genome. Hence, this homology in sequence leads to an immediate successful integration of the intact copy of gene downstream of the *amiA* operon. This is believed to be located on a transcriptional silent site and therefore do not affect any cellular physiological functions. The results showed the successful complementation as the genetically complemented mutants showed phenotype similar to the wild type D39 strain, indicating that the observed phenotypic effect was due to mutations rather than the polar effect. My study showed the regulatory role of Rggs over certain genes potentially involved in many important biological pathways, as discussed below.

In my study, the influence of *Δrgg0999* and *Δrgg1518* on the transcriptome of *S. pneumoniae* was examined using microarray analysis. The level of mRNA expression for all pneumococcal strains was analysed using microarray analysis. Microarray analysis is an ideal tool to reveal the genome wide changes due to Rgg0999 and Rgg1518 mutations. The results showed that the largest number of differentially expressed genes in *rgg* mutant strains was seen on galactose. On the other hand, when the Rgg mutants were grown on mannose, the size of putative regulon for both Rgg (Rgg1518 and Rgg0999) was smaller than that on galactose. Additionally, the results showed that each Rgg has a specific role under the same environmental condition, For example, for capsule locus genes while Rgg1518 is a repressor on galactose, Rgg0999 acts as an activator for the same locus on galactose. The microarray data also showed that under different environmental conditions, the same Rgg can act either as a repressor or the activator for the same target. For example, for SPD_0114-SPD_0124 while

Rgg1518 is a repressor on galactose, Rgg0999 acts as an activator for the same locus on mannose. This indicates apparently that regulation exerted by different Rggs on the same target gene is influenced by the different sugars. Moreover, Rgg1518 and Rgg0999 act as an activator for the same target in the same sugar, for instant, both Rggs act as an activator for the SPD_1290, encode for permease protein, on mannose. Rgg0999 and Rgg1518 seem to be having similar regulon.

4.1.1 The involvement of Rgg0999 and Rgg1518 in pneumococcal sugar metabolism

There is a strong relationship between pneumococcal virulence and efficient sugar acquisition and catabolism (Yesilkaya *et al.,* 2009). The pneumococcus is usually present in the multi-sugar environment in the host tissues, and seems it can utilise 32 different sugars such as glucosides, galactosides and polysaccharides (Bidossi *et al.,* 2012). This allows the pathogen to utilise the available sugar in different host tissues (Bidossi *et al.,* 2012; Terra *et al.,* 2015). Glucose, galactose and mannose are abundantly available to pneumococci in the host because they are included in mucin-carbohydrate side chains (Terra *et al.,* 2010). Galactose is a very important sugar for pneumococcal biology in the respiratory tract where the microbe colonises and disseminates to cause serious diseases, because galactose forms a large percentage of mucin sugar content in the form of galactose and N-acetylgalactosamines compared to other sugars such as mannose and glucose (Terra *et al.*, 2010).

In order to examine the effect of loss of *rgg* in pneumococcal growth, all strains were grown microaerobically in BHI or CDM supplemented with 55 mM glucose, galactose or mannose individually as the main carbon source. In this current study, inactivation of

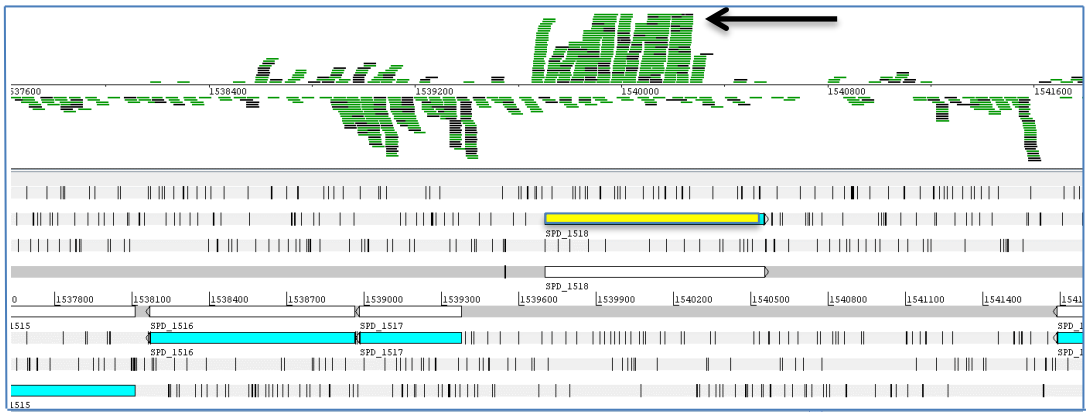
rgg0999 and/or *rgg1518* did not affect the ability of the pneumococcus to grow in BHI or CDM supplemented with glucose suggesting that *rgg0999* and *rgg1518* do not play any role in rich media, nor in glucose metabolism. However, *Δrgg0999*, *Δrgg1518* and double mutant *Δrgg0999*/1518 had significantly reduced growth rate compared to the wild type strain when grown in CDM supplemented either with galactose or mannose. The fact that Rggs have an important role in the control of galactose and mannose utilisation is emphasized by this study that revealed the expression of both *rggs* was stimulated by galactose and mannose, hence the absence of Rggs led to the significant reduction in utilisation of galactose and mannose.

Therefore, galactose and mannose have a crucial role in regulation of both *rggs*, *rgg0999* and *rgg1518*. Microarray data highlighted that genes that encode for Leloir and tagatose pathway regulated by Rgg1518 and Rgg0999, and recent study by Paixão *et al.* (2015) showed that both pathways are important for galactose utilisation. Inactivation of either Leloir or tagatose pathways resulted in an extended lag phase (18 h), or total elimination of the pneumococcal growth in CDM supplemented with galactose as sole carbon source. This result clearly underlies the importance of galactose utilisation for *In vivo* survival of the microbe. Furthermore, deletion of *rgg* led to significant differences in catabolic substrate preference during growth in complex and defined media compared to the wild type strain in *S. pyogenes* (Zheng *et al.,* 2011). In this study, lack of *rgg0999* and *rgg1518* had a significant impact on pneumococcal growth on both galactose and mannose, indicating its direct involvement in both galactose and mannose metabolism.

The inducibility of *rgg0999* and *rgg1518* was investigated using transcriptional reporter assays in microaerobically grown cultures on CDM supplemented with glucose,

galactose or mannose. The involvement of Rgg0999 and Rgg1518 in sugar metabolism was studied using *lacZ* reporter assays. The transcriptional *lacZ*-fusion results showed that the promoters of *rgg0999* and *rgg1518* were not induced in the presence of glucose. This was consistent with the growth of mutants, which grew as well as the wild type in CDM supplemented with glucose. The transcriptional *lacZ*-fusion analysis also showed that the *rgg1518* promoter activity was high on galactose and mannose, relative to its expression in wild type D39 strain, indicating that Rgg1518 was induced in the presence of these sugars. This was consistent with the growth of mutant (Δ*rgg1518*), which was attenuated in CDM supplemented either with galactose or mannose. The results also showed that the Rgg0999 activity was low on galactose and mannose, relative to its expression in wild type D39 strain, indicating that Rgg0999 unlikely to be induced by these tested sugars. Figure 4.1 shows genomic location of each individual gene (*rgg0999* and *rgg1518*) that transcript in galactose condition. Generally, gene expression analysis for sequence RNA transcripts from *S. pneumoniae* has focused on expressed, resulting in a large number of small fractions of the cDNA sequences with approximately 50 bp in size from the pneumococcal genome. Then, software Artemis programme Sanger institute pathogen genomics group was used to map these small sequences for the target genome, the transcript levels of *rgg1518* promoter are higher than that of *rgg0999*. This demonstrated that my results are consistent with software results in which that *rgg1518* was induced by galactose while *rgg0999* was not. The potential reason for these results is the putative promoter region of *rgg1518* was successfully identified in the upstream region while that of *rgg0999* has not been locating within upstream yet. Further studies will possibly lead to succeed in finding out the putative promoter region of *rgg0999*.

A



B

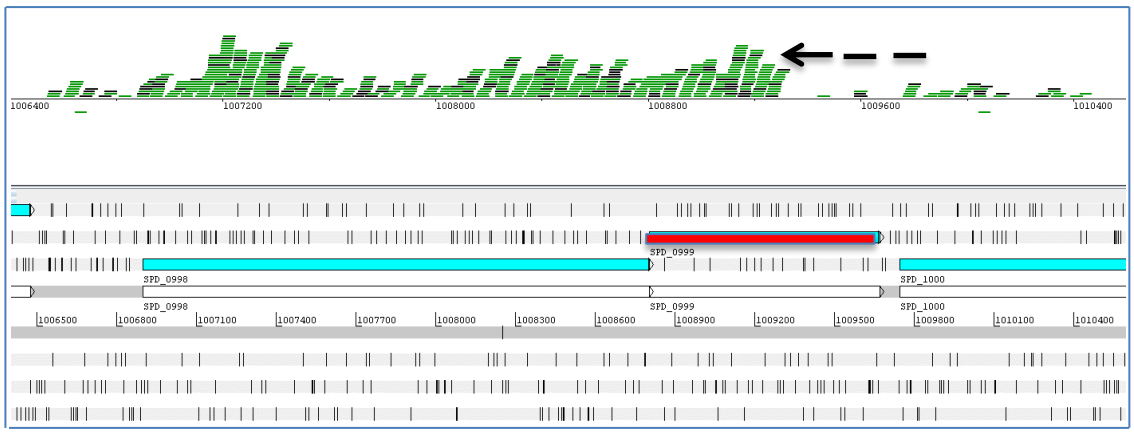


Figure 4.1. An illustration showing genomic location of each indivual gene that transcript in galactose condition using Artemis programme Sanger institute pathogen genomics group. (A) Yellow arrow indicates *rgg1518*, the transcript levels of promoter are high, represented by a black arrow. (B) Red is arrow indicates *rgg0999*, the transcript levels of promoter are low, represented by a dotted arrow. This reflects that *rgg1518* induced by galactose and *rgg0999* unlikely to be induced by this sugar.

In *S. pneumoniae* D39 genome, codes for five putative proteins (Rgg0144, Rgg0939, Rgg0999, Rgg1518, and Rgg1952) with significant homology (between 23 to 28% identity and 42 to 48% similarity at the amino acid sequence level) to what is considered as the Rgg prototype; the Rgg from *S. gordonii*. In order to investigate the interaction between the transcriptional regulators of interest, the involvement of Rgg1518 in the regulation of other *rggs* (*rgg1518, rgg0144* and *rgg0939*) was also investigated using transcriptional *lacZ*-fusions. As I mentioned above, *rgg0999* unlikely to be induced by galactose, hence it was excluded. On the other hand the *rgg1518* was induced by host-derived sugars. The results also showed that Rgg1518 is required for its own expression, and *rgg1518* expression required the presence of Rgg144 and Rgg939 in CDM supplemented with galactose. PlcR and Rgg regulators are a member of RRNPP family, Recently a PlcR/PapR homolog transcriptional regulator, known as TprA/PhrA, was studied in *S. pneumoniae*. It was found that this QS system regulates bacteriocin production loci, and involved in mucin utilisation, and galactose metabolism (Hoover *et al.*, 2015; Motib *et al.*, 2017). It was proposed that by interfering with their allosteric regulation, bacterial behavior can be modulated to discover anti-infectives. Parashar *et al.* (2015) have characterized the X-ray crystal structure of *Streptococcus dysgalactiae* Rgg protein alone and in complex with a tight-binding signaling antagonist, the cyclic undecapeptide cyclosporin A. In future, the Rgg structure can be utilised to develop antiinfectives targeting Rggs.

4.1.2. Rggs have a role capsule synthesis

S. pneumoniae can be commensal pathogen depending on its ability to colonise and be spread between humans (Weinberger *et al.*, 2008; McCuller and Tuomanen, 2001). The epithelium of the upper airways is the initial ecological niche for *S. pneumoniae*, hence

pneumococcal colonisation of the nasopharynx is considered to be the most important step to initiate the infection. For successful initiation of colonisation of the upper airways, the pneumococcus must defeat the immune system of the host, and adapt to new environmental conditions, and utilise available nutrients (Wyllie *et al.*, 2016; McCuller and Tuomanen, 2001). The pneumococcus relys on an optimal simultaneous strategy for successful colonisation and requirement for virulence factors expression, such as capsular polysaccharide.

The capsular polysaccharide is classified as important virulence factors in *S. pneumoniae*, which confer bacteria a protection against opsonophagocytosis, and plays an important role in the pneumococcal survival in different environmental conditions in the human host. (Hyams *et al.,* 2010). Glucoronic acid is the major component of capsular polysaccharide and found in wild type D39 (Bonofiglio *et al.,* 2005). Hence, I determined the level of capsule synthesis by assaying for glucuronic acid in capsule extract as this compound is found in type 2 D39 strain capsule. The microarray data highlighted that capsule locus genes (SPD_0315-SPD_0328) are possibly regulated by Rgg1518 on galactose. My hypothesis was that the inactivation of *rgg1518* may affect the capsule synthesis. This study demonstrated that Rgg1518 has a role in galactose catabolism (Section 4.1.1), and galactose catabolism is involved in capsule synthesis. Eighteen different sugars are involved in the formation of 90 pneumococcal capsular serotypes, and interestingly, nearly all structurally characterised capsule types contain either galactose or sugars synthesised from the intermediates formed during galactose catabolism (Carvalho *et al.,* 2011; Paixão *et al.,* 2015). My data indicated that inactivation of Rgg1518 decreases the capsule production of the mutant strain

(Δ*rgg1518*) under *In vitro* conditions. This was consistent with attenuation of Rgg mutants, particularly of Rgg1518, *In vivo*.

To further confirmation the role of Rgg1518 in capsule expression on galactose, I utilised transcriptional *lacZ*-fusions to assess the loss of *rgg1518* in capsule locus expression (*cps*). My data shows that P*cps*-*lacZ*-wt had high β-galactosidase activity on galactose while P*cps*-*lacZ*-Δ*rgg1518* had low β-galactosidase activity on the same sugar. This indicated that *cps* locus expression required the presence of Rgg1518. However, these results were in conflict with the microarray data, which showed that *cps* locus was up-regulated in the *Δrgg1518* on galactose. There may be a number of explanations for this. Firstly, the microarray data used mid-exponential-phase, whereas the reporter assays utilised late exponential phase cultures, consistent with the expression of quorum sensing system at this stage of bacterial growth. Secondly, in microarray analysis, the expression is measured in the mutant background relative to the wild type. While in the reporter assay the expression of *cps* locus was determined in non-native site in the mutant background and wild type. In addition, a third scenario also is possible for this observation. I speculate that there may be another protein involved, which is expected to be the activator of the *cps* locus in its native location and repressed by Rgg1518. Therefore, in microarray analysis of Rgg1518 mutant strain relative to the wild type, Rgg1518 exerted repression over this unknown protein will be eliminated and an elevated expression of *cps* locus will be observed. However, in the reporter assay, in which the *cps* expression is measured in non-native location in the genome, the unknown activator protein might not involve in regulation of P*cps*-*lacZ* in the mutant background.

This study demonstrates that *rgg1518* plays a role in capsule synthesis and regulation of capsule expression. In addition, the regulon of *rgg1518* was investigated in the presence of galactose and mannose using microarray analysis (Appendix 5). EMSA was used to detect any potential interaction between the purified Rgg1518 and the fluorescently labelled DNA probes representing the putative promoter sequence of capsule locus genes (SPD_0315-SPD_0328). The results show direct interaction between Rgg1518 and *cps* locus, and the binding sequence was AAAGGTG. Therefore, Rgg1518 directly bind with *cps* locus to form a complex which is essential for capsule expression.

4.1.3. Rggs have a role in pneumococcal survival and virulence

Related studies reported that Rgg family regulators have crucial roles in virulence. In *S. pyogenes*, Rgg (also known as RopB) is required for the expression of secreted proteins like SpeB, which has an effect on host-pathogen interactions by destroying the extracellular matrix and degradation of complement factors in host (Neely *et al.,* 2003; Kappeler *et al.,* 2009). The Rgg0939/Shp system induces the transcription of *shp* and the capsule gene locus in *S. pneumoniae*, and over expression of the Rgg system increases capsule size and reduces the biofilm formation on lung epithelial cells (Junges *et al.,* 2017). As I mentioned before, my hypothesis is that Rgg was important for pneumococcal survival and virulence. To support this hypothesis and to investigate the role of these regulatory processes in pneumococcal virulence, the pneumococcal wild type D39 and its mutants (*Δrgg0999* and/or *Δrgg1518*) were tested in a mouse model of pneumonia and colonisation. The results showed that both *rgg0999* and *rgg1518* have significant roles in pneumococcal virulence since the median survival time of mice infected intranasally with *Δrgg0999*, *Δrgg1518* or *Δrgg0999/1518* mutant were survived significantly longer than the wild type infected cohort. In addition, there was a delay in

appearance of bacteraemia in infected mice at 24 and 48 h post infection with Δ*rgg099, Δrgg1518* or Δ*rgg0999/1518* compared to the wild type strain. There was no significant difference of bacteraemia in infected animals at 24 and 48 h post infection with genetically complemented strains compared to the wild type. My *In vivo* results are supported by *In vitro* analysis of the mutants.

As mutants unable to access mucin sugars, Rggs down-regulate large number of genes involved in protein and capsule synthesis in addition to those required for iron uptake. This expression profile ensures a lower growth rate. Rggs are important because they have a role in capsule synthesis and sugars metabolism. Rgg1518 decreases the capsule production in the mutant strain, and microarray data highlighted that Rgg0999 acts as an activator for capsule locus genes. Therefore, less capsule synthesis led to less virulence due to the importance of capsule as virulence factor. My data shows that Rggs play a role in pneumococcal colonisation through their role in control of capsule expression.

With respect to colonisation, pneumococci were administered intranasally to initiate infection in mice. Serial dilution of homogenised nasopharyngeal tissue of infected mice was used to count the numbers of bacteria at the zero time, 3 and 7 days post-infection. The bacterial load in the nasopharyngeal tissue for all mutants Δ*rgg0999*, Δ*rgg1518*, Δ*rgg0999/1518* at the zero time of infection was similar to that of wild type. Furthermore, the colony counts at 3 and 7 days post-infection for mutant strain were significantly lower than that of wild type D39 strain. These results clearly prove the involvement of Rggs in pneumococcal colonisation. The reduction in colonisation and virulence in the mutants is very likely due to the inability of mutants to efficiently utilise galactose and mannose, and Rggs have regulatory roles in several genes involved

in various important biological functions. My results are in line with a previous study in Leicester shown by Motib *et al*. (2017) that pneumococcal TprA/PhrA is essential for a mouse model of virulence and colonisation. That study showed that an inability to utilise galactose either via Leloir or tagatose pathway attenuates pneumococcal colonisation and virulence (Paxio *et al.*, 2015).

In conclusion, growth studies in CDM showed that Δ*rgg0999* and Δ*rgg1518* had attenuated growth on both galactose and mannose. The transcriptional *lacZ*-fusion results showed that P*rgg0999-lacZ*-wt had low β-galactosidase activity on all host derived sugars tested, which suggests that the attenuation in growth on galactose and mannose might be due to indirect effect of Rgg0999 on the regulation of other genes involved in galactose and mannose metabolism. P*rgg1518-lacZ*-wt had high β-galactosidase activity on both galactose and mannose. The reporter assay also showed that Rgg1518 is required for its own expression, and *rgg1518* expression required the presence of Rgg0144 and Rgg0939 on galactose. My results showed that on galactose, both Rgg0999 and Rgg1518 acted as an activator of the *cps* locus. Rgg1518 interacted directly with the putative promoter region of *cps*. Based on the results obtained in this study through EMSA and transcriptional *lacZ*-fusion analysis, the following model for *rgg0999* and *rgg1518* regulation is proposed (Figure 4.2). Recently, it was demonstrated in our research group that Rgg/Shp0144 and Rgg0939/Shp0939, another peptides mediated Rggs, is important for pneumococcal colonisation and invasion through its effect on mannose and galactose utilisation and capsule synthesis (Zhi *et al.*, 2018). Rgg0144 and Rgg0939 are required for induction of Shp0144 and Shp0939, respectively. Moreveore, both Rgg0144 and Rgg0939 acted as a repressor of the *cps*

locus on mannose. In addition, both Rgg0144 and Rgg0939 interacted directly with the putative promoter region of *cps* also shown in Figure 4.2.

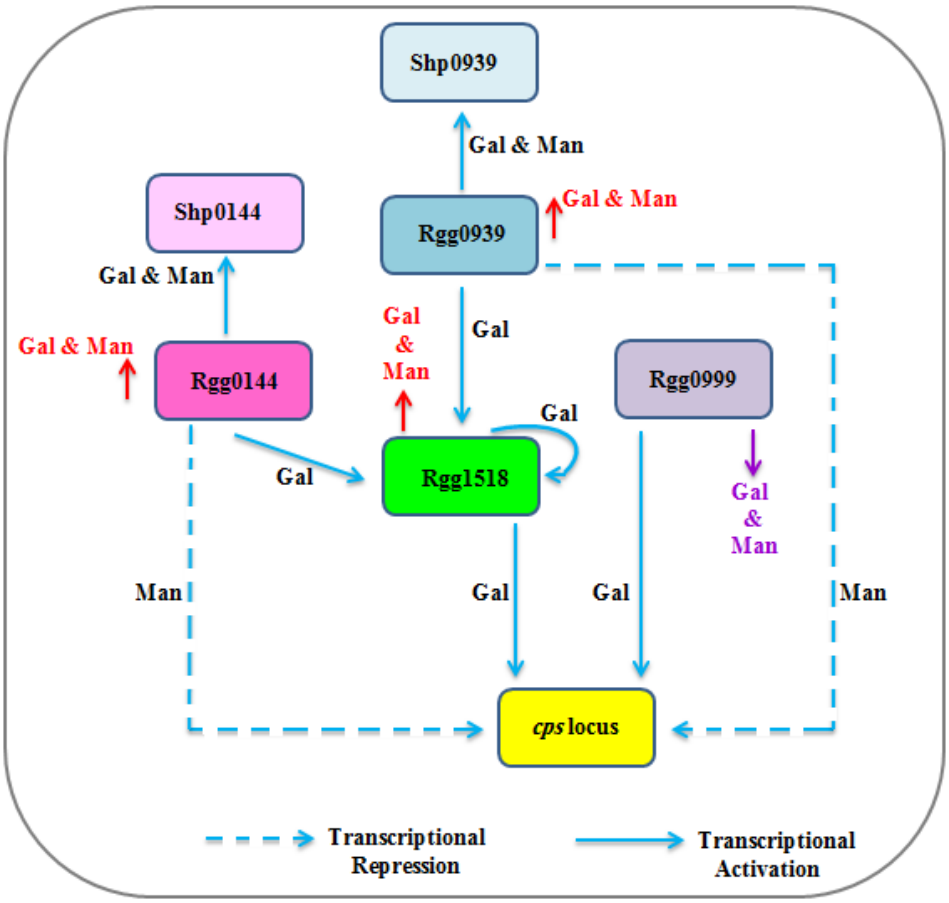


Figure 4.2. Schematic representation of proposed roles for the Rgg transcriptional regulators in *S. pneumoniae*. Rgg1518 auto-activates itself, and both Rgg0144 and Rgg0939 activate Rgg1518 on galactose. Rgg0999 and Rgg1518 activate transcription of *cps* locus on galactose. While Rgg0144 and Rgg0939 repress transcription of *cps* locus in the presence of mannose. Rgg0144 and Rgg0939 activate Shp0144 and Shp0939, respectively.These four transcriptional regulators interacted directly with the putative promoter region of *cps*. Gal: galactose, Man: mannose. Red arrow indicates increase in expression, whereas purple arrow is for decreased expression. The involvement of both Rgg0144/Shp0144 and Rgg0939/Shp0939 on galactose and mannose has been previously reported (Zhi *et al.*, 2018).

Final Remarks

In this study it was shown for the first time the role of SPD_0999 and SPD_1518 in sugar metabolism, capsule production, survival and virulence of S. pneumoniae. My study demonstrated that both *rgg0999* and *rgg1518* affect galactose and mannose utilisation, and these led to decrease pneumococcal growth abilities in the presence of galactose or mannose. This study has shown that the *rgg1518* is active when pneumococcus encounters host glycans containing galactose or mannose. Rgg1518 is required for its own expression on galactose, and *rgg1518* expression required the presence of both Rgg0144 and Rgg0939 since the expression of *rgg1518* was lower in both Δ*rgg0144* and Δ*rgg0939*. On the other hand *rgg0999* is not induced by any of the tested sugars.

It was also shown that Rgg1518 regulation of capsule locus genes is due to direct interaction of Rgg1518 with the putative promoters of capsule locus genes (SPD_0315-SPD_0328). Inactivation of *rgg1518* leads to significant reduction in the capsule production of *S. pneumoniae*. Loss of *rgg0999* and/or *rg1518* leads to significant attenuation in pneumococcal virulence and survival in nasopharynx, blood and colonisation *In vivo*.

Future work

Transcriptional regulation in pneumococcus needs more in depth study to understand how the bacterial two-component systems or stand-alone transcriptional regulators coordinate with each other to implement diverse types of regulatory processes. The results in this study showed the involvement of Rgg1518 in host derived sugar metabolism such as galactose and mannose. Galactose catabolism which has been studied in a number of Gram-positive bacteria is still not clear how *rgg1518* mediates catabolism since there are very limited reports related to pneumococcal *rggs*, and the mechanism is not fully understood. It is also important to investigate how *rgg1518* and *rgg0999* activates genes that encode for the Leloir and tagatose pathways which have a role in activate galactose catabolism. It is also worthy to investigate the gene that encode for Leloir and tagatose pathways such as SPD_1634 (encode for galactokinase) and SPD_1053 (encode for galactose-6-phosphate isomerase, LacA subunit), respectively which bind with Rgg1518 or Rgg0999 to determine if these genes are regulated directly or indirectly by rggs.

It was determined in this study that Rgg1518 activates the expression of *cps* locus in the presence of galactose, and through EMSA, the purified Rgg1518 showed direct binding with *cps* locus. Moreover, the microarray data highlighted that genes that are possibly regulated by Rgg0999 included capsule locus genes (SPD_0315-SPD_0328) on galactose. To determine whether this regulation is direct or indirect, EMSA should be used as described in section 3.11. In addition to find the Rgg0999 binding motif in *cps* locus, it should be searched for a consensus binding motif in the promoter regions of differentially expressed genes identified in microarray analysis as described in section 3.12.

It is also important to investigate the role of Rgg1518 and Rgg0999 in neuraminidase expression and biofilm formation using human airway epithelial cells as described previously (Parker *et al.*, 2009). Additionally, in depth experiments are required to study the role of Rgg1518 and Rgg0999 in regulation of pneumolysin expression using *lacZ*-fusion analysis as described in section 2.16.

The β-galactosidase activity will be tested in new constructed reporter strains, and P*rgg1518-lacZ*-wt will be used as a positive control. The new constructed reporter strains P*rgg1518-lacZ-*Δ*shp144* and P*rgg1518-lacZ-*Δ*shp939* are expected have lower β-galactosidase activity compare with P*rgg1518-lacZ*-wt. These results will demonstrate that pneumococcal *shp144 and shp939* are necessary for rgg1518. Certain Rggs, but not all, are associated with genes coding for short hydrophobic peptides (*shp*). It has been shown in other streptococci that Rgg-SHP pairs are the individual components of quorum sensing systems. For example, *Rgg2* and *Rgg3* bind to the peptide pheromones SHP2 and SHP3 in *S. pyogenes* (Parashar *et al.*, 2015). Thus, it is also we need to see whether *rgg1518* or *rgg0999* are two-component systems or stand-alone by investigating the peptide.

Appendix 1

DNA sequencing showing successful deletion of target gene and insertion into antibiotic cassette. The start of antibiotic cassette is represented by bink yellow highlighted.

(1) Δ*rgg*0999 with spectinomycin cassette

Alignment with SPD0999-SeqM-F1

```
Query  61       GTTGAGGTTGGTTACTATGACCAAACCCAAAGCAAGCTGACACCAAGTAATACGGTGCTG  120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1008063  GTTGAGGTTGGTTACTATGACCAAACCCAAAGCAAGCTGACACCAAGTAATACGGTGCTG  1008122

Query  121      GATGAACTCTGGAATGATTTCAAACTGACACCAGAAGTTGAAATCCGCAACCGTCTTGGA  180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1008123  GATGAACTCTGGAATGATTTCAAACTGACACCAGAAGTTGAAATCCGCAACCGTCTTGGA  1008182

Query  181      GCCTTCCTTTTCTCAGGAGATGATGTTAAAAAATCAGTCGGCATGCTATCTGGTGGCGAA  240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1008183  GCCTTCCTTTTCTCAGGAGATGATGTTAAAAAATCAGTCGGCATGCTATCTGGTGGCGAA  1008242

Query  241      AAAGCTCGTTTGCTTTTAGCTAAATTGTCTATGGAAAACAATAATTTTTTGATTCTGGAT  300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1008243  AAAGCTCGTTTGCTTTTAGCTAAATTGTCTATGGAAAACAATAATTTTTTGATTCTGGAT  1008302

Query  301      GAGCCGACCAACCACTTGGATATTGATAGTAAGGAAGTGCTAGAAAATGCCTTGATTGAC  360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1008303  GAGCCGACCAACCACTTGGATATTGATAGTAAGGAAGTGCTAGAAAATGCCTTGATTGAC  1008362

Query  361      TTTGATGGAACCTTGCTGTTTGTCAGTCATGATCGTTACTTTATCAATCGTGTGGCAACT  420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1008363  TTTGATGGAACCTTGCTGTTTGTCAGTCATGATCGTTACTTTATCAATCGTGTGGCAACT  1008422

Query  421      CATGTTTTGGAATTGTCTGAGAATGGTTCAACTCTCTACCTTGGAGATTACGACTACTAT  480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1008423  CATGTTTTGGAATTGTCTGAGAATGGTTCAACTCTCTACCTTGGAGATTACGACTACTAT  1008482

Query  481      GTTGAGAAGAAAGCAACAGCAGAAATGAGTCAGACTGAGGAAGCTTCAACTAGCAATCAA  540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1008483  GTTGAGAAGAAAGCAACAGCAGAAATGAGTCAGACTGAGGAAGCTTCAACTAGCAATCAA  1008542

Query  541      GCAAAGGAAGCAAGTCCAGTCAATGACTATCAGGCCCAGAAAGAAAGTCAAAAAGAAGTT  600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1008543  GCAAAGGAAGCAAGTCCAGTCAATGACTATCAGGCCCAGAAAGAAAGTCAAAAAGAAGTT  1008602
```

Alignment with SPD0999-SeqM-F2

```
Query  138   AAGATCGATTTCGTTCGTGAATACATGTTATAATAACTATAACTAATAACGTAACGTGA  197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1409  AAGATCGATTTCGTTCGTGAATACATGTTATAATAACTATAACTAATAACGTAACGTGA  1468

Query  198   CTGGCAAGAGATATTTTTAAAACAATGAATAGGTTTACACTTACTTTAGTTTTATGGAAA  257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1469  CTGGCAAGAGATATTTTTAAAACAATGAATAGGTTTACACTTACTTTAGTTTTATGGAAA  1528

Query  258   TGAAAGATCATATCATATATAATCTAGAATAAAATTAACTAAAATAATTATTATCTAGAT  317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1529  TGAAAGATCATATCATATATAATCTAGAATAAAATTAACTAAAATAATTATTATCTAGAT  1588

Query  318   AAAAAATTTAGAAGCCAATGAAATCTATAAATAAACTAAATTAAGTTTATTTAATTAACA  377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1589  AAAAAATTTAGAAGCCAATGAAATCTATAAATAAACTAAATTAAGTTTATTTAATTAACA  1648

Query  378   ACTATGGATATAAAATAGGTACTAATCAAAATAGTGAGGAGGATATATTTGAATACATAC  437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1649  ACTATGGATATAAAATAGGTACTAATCAAAATAGTGAGGAGGATATATTTGAATACATAC  1708

Query  438   GAACAAATTAATAAAGTGaaaaaaaTACTTCGGAAACATTTAAAAAATAACCTTATTGGT  497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1709  GAACAAATTAATAAAGTGAAAAAAATACTTCGGAAACATTTAAAAAATAACCTTATTGGT  1768

Query  498   ACTTACATGTTTGGATCAGGAGTTGAGAGTGGACTAAAACCAAATAGTGATCTTGACTTT  557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1769  ACTTACATGTTTGGATCAGGAGTTGAGAGTGGACTAAAACCAAATAGTGATCTTGACTTT  1828
```

(2) Δ*rgg*1518 with spectinomycin cassette

Alignment with SPD1518-SeqM-F1

```
Query  61       AAATTATACCTACTTTACGAATTAATTTGGTGACCAACTACCTGAAATGGCAGCTGATAA  120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1538961  AAATTATACCTACTTTACGAATTAATTTGGTGACCAACTACCTGAAATGGCAGCTGATAA  1539020

Query  121      ATCCCACCAATTTTGTCTTGCCATATCTGCTGCATAAACATAGCCAACCGTTGCCCAATT  180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539021  ATCCCACCAATTTTGTCTTGCCATATCTGCTGCATAAACATAGCCAACCGTTGCCCAATT  1539080

Query  181      ACCATGAGTTGAACGAGTACCACAGGTTGTTAACGATTCTCCACTACGAGCTTCTCCATT  240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539081  ACCATGAGTTGAACGAGTACCACAGGTTGTTAACGATTCTCCACTACGAGCTTCTCCATT  1539140

Query  241      AGAATTTACTATACGATGCTTAACAGTCATGGGATGTAAATTTCTACAAGGATTTAAGTT  300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539141  AGAATTTACTATACGATGCTTAACAGTCATGGGATGTAAATTTCTACAAGGATTTAAGTT  1539200

Query  301      CATGATATATTCCGCACCTTCATTTTCTTTAGTCAAATAACCTGTTCTCTTATAACTGGT  360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539201  CATGATATATTCCGCACCTTCATTTTCTTTAGTCAAATAACCTGTTCTCTTATAACTGGT  1539260

Query  361      AGTAATAACAAAATTACTAAAAGCACCTATCTGATTTGCTTCTACGAAAGTAGTTCCAAC  420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539261  AGTAATAACAAAATTACTAAAAGCACCTATCTGATTTGCTTCTACGAAAGTAGTTCCAAC  1539320

Query  421      AGCACTTATCATTAAAACAGGAATAACCattttcaataatttcttaattttcatttttc  480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539321  AGCACTTATCATTAAAACAGGAATAACCATTTTCAATAATTTCTTAATTTTCATTTTTTC  1539380

Query  481      tgtttcctttcatttttattcaAGAAGATTCCTTATAACTTCTATTTTAAATATAGACT  540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539381  TGTTTCCTTTCATTTTTTATTCAAGAAGATTCCTTATAACTTCTATTTTAAATATAGACT  1539440

Query  541      TATTCTAAACGATGGTATATCATACAAATTCCTCCTTTACTTATTATGTATAGACTCTAA  600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539441  TATTCTAAACGATGGTATATCATACAAATTCCTCCTTTACTTATTATGTATAGACTCTAA  1539500
```

Alignment with SPD1518-SeqM-F2

```
Query  172   AATATCGATTTTCGTTCGTGAATACATGTTATAATAACTATAACTAATAACGTAACGTGA  231
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1409  AAGATCGATTTTCGTTCGTGAATACATGTTATAATAACTATAACTAATAACGTAACGTGA  1468

Query  232   CTGGCAAGAGATATTTTTAAAACAATGAATAGGTTTACACTTACTTTAGTTTTATGGAAA  291
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1469  CTGGCAAGAGATATTTTTAAAACAATGAATAGGTTTACACTTACTTTAGTTTTATGGAAA  1528

Query  292   TGAAAGATCATATCATATATAATCTAGAATAAAATTAACTAAAATAATTATTATCTAGAT  351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1529  TGAAAGATCATATCATATATAATCTAGAATAAAATTAACTAAAATAATTATTATCTAGAT  1588

Query  352   AAAAAATTTAGAAGCCAATGAAATCTATAAATAAACTAAATTAAGTTTATTTAATTAACA  411
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1589  AAAAAATTTAGAAGCCAATGAAATCTATAAATAAACTAAATTAAGTTTATTTAATTAACA  1648

Query  412   ACTATGGATATAAAATAGGTACTAATCAAAATAGTGAGGAGGATATATTTGAATACATAC  471
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1649  ACTATGGATATAAAATAGGTACTAATCAAAATAGTGAGGAGGATATATTTGAATACATAC  1708

Query  472   GAACAAATTAATAAAGTGaaaaaaaTACTTCGGAAACATTTAAAAAATAACCTTATTGGT  531
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1709  GAACAAATTAATAAAGTGAAAAAAATACTTCGGAAACATTTAAAAAATAACCTTATTGGT  1768

Query  532   ACTTACATGTTTGGATCAGGAGTTGAGAGTGGACTAAAACCAAATAGTGATCTTGACTTT  591
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1769  ACTTACATGTTTGGATCAGGAGTTGAGAGTGGACTAAAACCAAATAGTGATCTTGACTTT  1828
```

(3) Δ*rgg*0999-1518 with kanamycin cassette

Alignment with SPD0999-1518-SeqM-F1

```
Query  121      ATCCCACCAATTTTGTCTTGCCATATCTGCTGCATAAACATAGCCAACCGTTGCCCAATT  180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539021  ATCCCACCAATTTTGTCTTGCCATATCTGCTGCATAAACATAGCCAACCGTTGCCCAATT  1539080

Query  181      ACCATGAGTTGAACGAGTACCACAGGTTGTTAACGATTCTCCACTACGAGCTTCTCCATT  240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539081  ACCATGAGTTGAACGAGTACCACAGGTTGTTAACGATTCTCCACTACGAGCTTCTCCATT  1539140

Query  241      AGAATTTACTATACGATGCTTAACAGTCATGGGATGTAAATTTCTACAAGGATTTAAGTT  300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539141  AGAATTTACTATACGATGCTTAACAGTCATGGGATGTAAATTTCTACAAGGATTTAAGTT  1539200

Query  301      CATGATATATTCCGCACCTTCATTTTCTTTAGTCAAATAACCTGTTCTCTTATAACTGGT  360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539201  CATGATATATTCCGCACCTTCATTTTCTTTAGTCAAATAACCTGTTCTCTTATAACTGGT  1539260

Query  361      AGTAATAACAAAATTACTAAAAGCACCTATCTGATTTGCTTCTACGAAAGTAGTTCCAAC  420
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539261  AGTAATAACAAAATTACTAAAAGCACCTATCTGATTTGCTTCTACGAAAGTAGTTCCAAC  1539320

Query  421      AGCACTTATCATTAAAACAGGAATAACCattttcaataatttcttaattttcatttttc  480
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539321  AGCACTTATCATTAAAACAGGAATAACCATTTTCAATAATTTCTTAATTTTCATTTTTTC  1539380

Query  481      tgtttcctttcatttttattcaAGAAGATTCCTTATAACTTCTATTTTAAATATAGACT  540
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539381  TGTTTCCTTTCATTTTTATTCAAGAAGATTCCTTATAACTTCTATTTTAAATATAGACT  1539440

Query  541      TATTCTAAACGATGGTATATCATACAAATTCCTCCTTTACTTATTATGTATAGACTCTAA  600
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539441  TATTCTAAACGATGGTATATCATACAAATTCCTCCTTTACTTATTATGTATAGACTCTAA  1539500

Query  601      TTTCTTTATCCCCAAAACCATGTTTCAAACCAGATAAGTTGAATCCAACTCCAAATCAAA  660
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539501  TTTCTTTATCCCCAAAACCATGTTTCAAACCAGATAAGTTGAATCCAACTCCAAATCAAA  1539560

Query  661      AATCCAGAGTTTTTCGGTAAATTCTTCAAATATTTTTTAAAGCCCATGACTTTTTCCCTC  720
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539561  AATCCAGAGTTTTTCGGTAAATTCTTCAAATATTTTTTAAAGCCCATGACTTTTTCCCTC  1539620
```

Alignment with SPD0999-1518-SeqM-F2

```
Query  182  CGCGCATGCTGGGGATCCGTTTGATTTTTAATGGATAATGTGATATAATCTTTAAATACT  241
             |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    CGCGCAAGCTGGGGATCCGTTTGATTTTTAATGGATAATGTGATATAATCTTTAAATACT  60

Query  242  GTAGAAAAGAGGAAGGAAATAATAAATGGCTAAAATGAGAATATCACCGGAATTGAAAAA  301
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTAGAAAAGAGGAAGGAAATAATAAATGGCTAAAATGAGAATATCACCGGAATTGAAAAA  120

Query  302  ACTGATC-AAAAATACCGCTGCGTGAAAGATACGGAAGGAATGTCTCCTGCTAAGGTATA  360
             ||||||| ||||||||||||||| ||||||||||||||||||||||||||||||||||||
Sbjct  121  ACTGATCGAAAAATACCGCTGCGTAAAAGATACGGAAGGAATGTCTCCTGCTAAGGTATA  180

Query  361  TAAGCTGGTGGGAGAAAATGAAAACCTATATTTAAAAATGACGGACAGCCGGTATAAAGG  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  TAAGCTGGTGGGAGAAAATGAAAACCTATATTTAAAAATGACGGACAGCCGGTATAAAGG  240

Query  421  GACCACCTATGATGTGGAACGGGAAAAGGACATGATGCTATGGCTGGAAGGAAAGCTGCC  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GACCACCTATGATGTGGAACGGGAAAAGGACATGATGCTATGGCTGGAAGGAAAGCTGCC  300

Query  481  TGTTCCAAAGGTCCTGCACTTTGAACGGCATGATGGCTGGAGCAATCTGCTCATGAGTGA  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  TGTTCCAAAGGTCCTGCACTTTGAACGGCATGATGGCTGGAGCAATCTGCTCATGAGTGA  360

Query  541  GGCCGATGGCGTCCTTTGCTCGGAAGAGTATGAAGATGAACAAAGCCCTGAAAAGATTAT  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  GGCCGATGGCGTCCTTTGCTCGGAAGAGTATGAAGATGAACAAAGCCCTGAAAAGATTAT  420

Query  601  CGAGCTGTATGCGGAGTGCATCAGGCTCTTTCACTCCATCGACATATCGGATTGTCCCTA  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  CGAGCTGTATGCGGAGTGCATCAGGCTCTTTCACTCCATCGACATATCGGATTGTCCCTA  480
```

Appendix 2

DNA sequencing of recombinant pCEP carrying the target genes and their putative promoter regions for complementation analysis.

(1) *Δrgg0999*c

Alignment with MalF

```
Query  21       ATGGAACATCTTGGAAAAGTATTTCGTGAATTTCGAACAAGTGGAAATTATTCTTTAAAG  80
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1008803  ATGGAACATCTTGGAAAAGTATTTCGTGAATTTCGAACAAGTGGAAATTATTCTTTAAAG  1008862

Query  81       GAAGCAGCAGGCGAATCCTGCTCTACCTCTCAGTTATCTCGCTTTGAGCTTGGGGAGTCT  140
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1008863  GAAGCAGCAGGCGAATCCTGCTCTACCTCTCAGTTATCTCGCTTTGAGCTTGGGGAGTCT  1008922

Query  141      GACCTGGCAGTCTCCCGTTTCTTTGAGATTTTGGATAACATTCATGTAACAATCGAAAAT  200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1008923  GACCTGGCAGTCTCCCGTTTCTTTGAGATTTTGGATAACATTCATGTAACAATCGAAAAT  1008982

Query  201      TTCATGGATAAGGCAAGGAATTTTCATAATCATGAACATGTGTCTATGATGGCACAGATT  260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1008983  TTCATGGATAAGGCAAGGAATTTTCATAATCATGAACATGTGTCTATGATGGCACAGATT  1009042

Query  261      ATCCCACTTTACTATTCAAACGATATTGCAGGTTTTCAAAAGCTTCAAAGAGAACAACTT  320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1009043  ATCCCACTTTACTATTCAAACGATATTGCAGGTTTTCAAAAGCTTCAAAGAGAACAACTT  1009102

Query  321      GAAAAGTCTAAGAGTTCGACGACTCCCCTTTATTTTGAGCTGAACTGGATTTTGCTACAA  380
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1009103  GAAAAGTCTAAGAGTTCGACGACTCCCCTTTATTTTGAGCTGAACTGGATTTTGCTACAA  1009162

Query  381      GGTCTGATTTGTCAAAGAGATGCGAGTTATGATATGAAGCAGGATGATTTGGATAAGGTA  440
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1009163  GGTCTGATTTGTCAAAGAGATGCGAGTTATGATATGAAGCAGGATGATTTGGATAAGGTA  1009222

Query  441      GCAGATTATCTCTTCAAAACAGAAGAATGGACCATGTATGAGTTGATTCTTTTCGGTAAC  500
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1009223  GCAGATTATCTCTTCAAAACAGAAGAATGGACCATGTATGAGTTGATTCTTTTCGGTAAC  1009282

Query  501      CTCTATAGTTTCTACGATGTAGACTATGTCACTCGGATTGGTAGAGAAGTTATGGAGAGG  560
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1009283  CTCTATAGTTTCTACGATGTAGACTATGTCACTCGGATTGGTAGAGAAGTTATGGAGAGG  1009342

Query  561      GAGGAATTTTACCAAGAGATTAGTCGCCATAAGAGATTAGTGTTGATTTTGGCCCTCAAT  620
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1009343  GAGGAATTTTACCAAGAGATTAGTCGCCATAAGAGATTAGTGTTGATTTTGGCCCTCAAT  1009402

Query  621      TGTTACCAGCATTGTTTAGAGCATTCTTCTTTTTATAATGCCAACTATTTTGAGGCTTAT  680
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1009403  TGTTACCAGCATTGTTTAGAGCATTCTTCTTTTTATAATGCCAACTATTTTGAGGCTTAT  1009462

Query  681      ACAGAGAAGATTATTGACAAAGGTATTAAGCTTTATGAGCGTAATGTTTTCCATTATTTA  740
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1009463  ACAGAGAAGATTATTGACAAAGGTATTAAGCTTTATGAGCGTAATGTTTTCCATTATTTA  1009522

Query  741      AAAGGTTTTGCCTTATATCAAAAAGGACAGTGTAAAGAAGGCTGTAAGCAGATGCAAGAG  800
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1009523  AAAGGTTTTGCCTTATATCAAAAAGGACAGTGTAAAGAAGGCTGTAAGCAGATGCAAGAG  1009582

Query  801      GCCATGCATATTTTTGATGTGTTAGGTCTTCCAGAGCAAGTAGCCTATTATCAGGAACAC  860
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1009583  GCCATGCATATTTTTGATGTGTTAGGTCTTCCAGAGCAAGTAGCCTATTATCAGGAACAC  1009642

Query  861      TACGA  865
                |||||
Sbjct  1009643  TACGA  1009647
```

(2) *Δrgg1518*c

Alignment with MalF

```
Query  21       TTTCATTTTTTATTCAAGAAGATTCCTTATAACTTCTATTTTAAATATAGACTTATTCTA  80
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539388  TTTCATTTTTTATTCAAGAAGATTCCTTATAACTTCTATTTTAAATATAGACTTATTCTA  1539447

Query  81       AACGATGGTATATCATACAAATTCCTCCTTTACTTATTATGTATAGACTCTAATTTCTTT  140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539448  AACGATGGTATATCATACAAATTCCTCCTTTACTTATTATGTATAGACTCTAATTTCTTT  1539507

Query  141      ATCCCCAAAACCATGTTTCAAACCAGATAAGTTGAATCCAACTCCAAATCAAAAATCCAG  200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539508  ATCCCCAAAACCATGTTTCAAACCAGATAAGTTGAATCCAACTCCAAATCAAAAATCCAG  1539567

Query  201      AGTTTTTCGGTAAATTCTTCAAATATTTTTTAAAGCCCATGACttttttcccttcttttttg  260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539568  AGTTTTTCGGTAAATTCTTCAAATATTTTTTAAAGCCCATGACTTTTTCCCTCCTTTTTG  1539627

Query  261      atataataataccacctttttctacgtttttttATAAATGTAGAAAATATTCACATTTTTC  320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539628  ATATAATAATACCACCTTTTTCTACGTTTTTTATAAATGTAGAAAATATTCACATTTTTC  1539687

Query  321      TAGGAGGAACAATATGAGATGGGATTATGGACAAATTTTTAAAGAAATTCGAAAGTCAAA  380
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539688  TAGGAGGAACAATATGAGATGGGATTATGGACAAATTTTTAAAGAAATTCGAAAGTCAAA  1539747

Query  381      AGGATTGACCCAACAAGATGTATGTGGACAAGTCATACATCGAACAACTCTAACAAATAT  440
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539748  AGGATTGACCCAACAAGATGTATGTGGACAAGTCATACATCGAACAACTCTAACAAATAT  1539807

Query  441      TGAACACGGTAAAGTTATTCCCAGTTTTGAAAACATGGTATTTCTTCTTGAACAAATTGA  500
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539808  TGAACACGGTAAAGTTATTCCCAGTTTTGAAAACATGGTATTTCTTCTTGAACAAATTGA  1539867

Query  501      TATGAGCTTGGCAGAATTCAAGTATATATGCAACGAATACCATCCTAGTAAAAGGCGAGA  560
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539868  TATGAGCTTGGCAGAATTCAAGTATATATGCAACGAATACCATCCTAGTAAAAGGCGAGA  1539927

Query  561      TATTATTGTAGAGAGCCAAAATCCGTCTACTTTTCAAGATACTAGAAAAATGGTTGAACT  620
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539928  TATTATTGTAGAGAGCCAAAATCCGTCTACTTTTCAAGATACTAGAAAAATGGTTGAACT  1539987

Query  621      CACTGAGAAATGTCAAAAATATCTTAAGACACATCACGATGTTCCTATTCAAAATATCTA  680
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539988  CACTGAGAAATGTCAAAAATATCTTAAGACACATCACGATGTTCCTATTCAAAATATCTA  1540047

Query  681      TCGTCATACAAAAATTGTCACAGAATTACGAACTAAAGGATTCAAAAACAACCACGTCTT  740
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1540048  TCGTCATACAAAAATTGTCACAGAATTACGAACTAAAGGATTCAAAAACAACCACGTCTT  1540107

Query  741      GAAAGATTTGTATGAAGAAATTTGGGACTATCTTGAACCTATGGATACATGGTACATTAG  800
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1540108  GAAAGATTTGTATGAAGAAATTTGGGACTATCTTGAACCTATGGATACATGGTACATTAG  1540167

Query  801      TGATTTGAAATTGCTTGGAACCATTCTCTTTTTCTTTCCTTCTGAAAATCTTCCCCTTCT  860
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1540168  TGATTTGAAATTGCTTGGAACCATTCTCTTTTTCTTTCCTTCTGAAAATCTTCCCCTTCT  1540227

Query  861      TATTGATAGAATTATGAAAACCATCGAGAAATATAAATACTTCCGAGAAACAAAAGCATT  920
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1540228  TATTGATAGAATTATGAAAACCATCGAGAAATATAAATACTTCCGAGAAACAAAAGCATT  1540287

Query  921      TTTATCATCTTTTTTAGCCAATCTCTCCACTGTATATTTTCAACATCATTTATTCAAAGA  980
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1540288  TTTATCATCTTTTTTAGCCAATCTCTCCACTGTATATTTTCAACATCATTTATTCAAAGA  1540347

Query  981      ATGTGAAACAATCACCCTACAACTTTTGGTATTAGCAGAAGAATTAAAAATATACGATAT  1040
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1540348  ATGTGAAACAATCACCCTACAACTTTTGGTATTAGCAGAAGAATTAAAAATATACGATAT  1540407

Query  1041     TCTTGGATTCTCTCAAGTACGCTTGGGAATACTACAACATAACTCCGATTTGATTGATAA  1100
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1540408  TCTTGGATTCTCTCAAGTACGCTTGGGAATACTACAACATAACTCCGATTTGATTGATAA  1540467

Query  1101     AGGAATAACTTTATTACGACTAACCAAAGAAGAGGCATTAGTTAAAATACTCGAAAAAGA  1160
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1540468  AGGAATAACTTTATTACGACTAACCAAAGAAGAGGCATTAGTTAAAATACTCGAAAAAGA  1540527

Query  1161     AATTAA  1166
                ||||||
Sbjct  1540528  AATTAA  1540533
```

Appendix 3

DNA sequencing of recombinant pPP2 carrying the putative promoters of target genes for transcription lacZ-fusion.

(1) Rgg (SPD_0999)

Alignment with Fusion/UF

```
Query  44       AGCTTTCTAGCTTGGTCAGTAGGTTTTTTAATTATGAAATCTAGTAATAAATAGAGTTAA  103
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1006501  AGCTTTCTAGCTTGGTCAGTAGGTTTTTTAATTATGAAATCTAGTAATAAATAGAGTTAA  1006560

Query  104      AAAGTTGAATGTTCCGAAAATATTGTATATAGTAGACTGAATCTAAAATAGTACGAAACA  163
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1006561  AAAGTTGAATGTTCCGAAAATATTGTATATAGTAGACTGAATCTAAAATAGTACGAAACA  1006620

Query  164      ATTGCTAAAACATTTATAGAAATTAATTTTACTTTCCCAATCGATTTGTTCTCATCTTAT  223
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1006621  ATTGCTAAAACATTTATAGAAATTAATTTTACTTTCCCAATCGATTTGTTCTCATCTTAT  1006680

Query  224      TTCAATCTGCTATACGTACGTAAATGATGAAAAGAATTAAATAATAGACTTGTGCGATAA  283
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1006681  TTCAATCTGCTATACGTACGTAAATGATGAAAAGAATTAAATAATAGACTTGTGCGATAA  1006740

Query  284      CTATTTGAAACAAGATTAGCTGGTTAGAAAATCCAACAAGATAGTGAAGGAATCTAATAT  343
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1006741  CTATTTGAAACAAGATTAGCTGGTTAGAAAATCCAACAAGATAGTGAAGGAATCTAATAT  1006800

Query  344      AAAATATACATAATTTAGTTTATTTTATGTTATAAAAGTTGTACTAAAAGACTTACGATT  403
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1006801  AAAATATACATAATTTAGTTTATTTTATGTTATAAAAGTTGTACTAAAAGACTTACGATT  1006860

Query  404      ATCTTGAATGCCTTAGGGAAACATGCTATACTACTTGTATGATTATTTTACAAGCTAATA  463
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1006861  ATCTTGAATGCCTTAGGGAAACATGCTATACTACTTGTATGATTATTTTACAAGCTAATA  1006920

Query  464      AAATTGAACG  473
                ||||||||||
Sbjct  1006921  AAATTGAACG  1006930
```

(2) Rgg (SPD_1518)

Alignment with Fusion/UF

```
Query  43       CAAGAAGATTCCTTATAACTTCTATTTTAAATATAGACTTATTCTAAACGATGGTATATC  102
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539402  CAAGAAGATTCCTTATAACTTCTATTTTAAATATAGACTTATTCTAAACGATGGTATATC  1539461

Query  103      ATACAAATTCCTCCTTTACTTATTATGTATAGACTCTAATTTCTTTATCCCCAAAACCAT  162
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539462  ATACAAATTCCTCCTTTACTTATTATGTATAGACTCTAATTTCTTTATCCCCAAAACCAT  1539521

Query  163      GTTTCAAACCAGATAAGTTGAATCCAACTCCAAATCAAAAATCCAGAGTTTTTCGGTAAA  222
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539522  GTTTCAAACCAGATAAGTTGAATCCAACTCCAAATCAAAAATCCAGAGTTTTTCGGTAAA  1539581

Query  223      TTCTTCAAATATTTTTTAAAGCCCATGACttttcccttcttttgatataataatacca  282
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539582  TTCTTCAAATATTTTTTAAAGCCCATGACTTTTTCCCTCCTTTTTGATATAATAATACCA  1539641

Query  283      cctttttctacgttttttATAAATGTAGAAAATATTCACATTTTTCTAGGAGGAACAATA  342
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539642  CCTTTTTCTACGTTTTTTATAAATGTAGAAAATATTCACATTTTTCTAGGAGGAACAATA  1539701

Query  343      TGAGATGGGATTATGGACAAATTTTTA  369
                ||||||||||||||||||||||||||
Sbjct  1539702  TGAGATGGGATTATGGACAAATTTTTA  1539728
```

(3) Capsule locus genes (SPD_0315-SPD_0328)

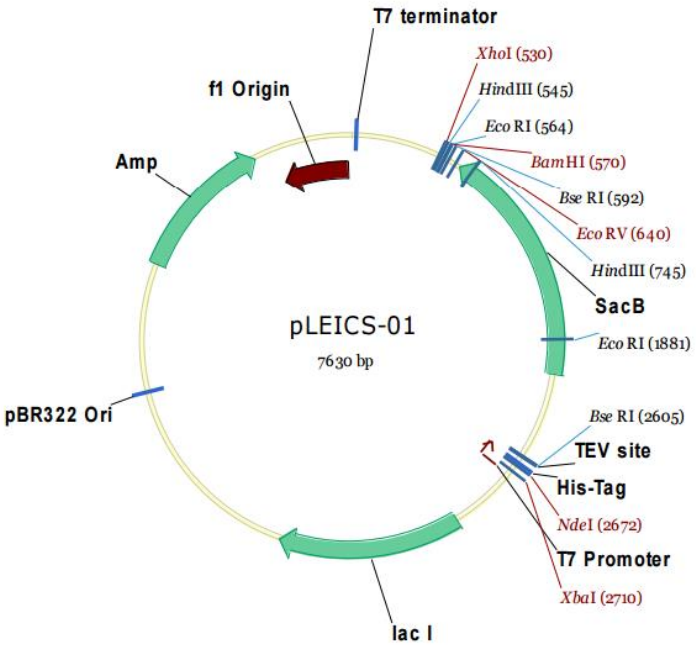
Alignment with Fusion/UF

```
Query  43       GATTATACCACATTGTGTACTATATTATATTGAAACTAGAATAGTACACATCTGCTTCTA  102
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  313499   GATTATACCACATTGTGTACTATATTATATTGAAACTAGAATAGTACACATCTGCTTCTA  313558

Query  103      AAATATTGTTAGAAAACGATTTGACTGTCCTGATCAATTTGTCATGTTCTTATTTCATTT  162
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  313559   AAATATTGTTAGAAAACGATTTGACTGTCCTGATCAATTTGTCATGTTCTTATTTCATTT  313618

Query  163      TACTATATTTTTGGTTCGCGGGAAGTCTACTAAGATACTTAAAGATGCAGATAGTGAAAA  222
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  313619   TACTATATTTTTGGTTCGCGGGAAGTCTACTAAGATACTTAAAGATGCAGATAGTGAAAA  313678

Query  223      AAGGTGTAGACATTACCGTAAAAAAGTGATATAATCGTAAGATGTTCAATGTATAGGTGT  282
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  313679   AAGGTGTAGACATTACCGTAAAAAAGTGATATAATCGTAAGATGTTCAATGTATAGGTGT  313738

Query  283      TAATCATGAGTAGACGTTTTAAAAAATCACGTT  315
                |||||||||||||||||||||||||||||||
Sbjct  313739   TAATCATGAGTAGACGTTTTAAAAAATCACGTT  313771
```

Appendix 4

(1) Genetic map of pLEICS-01 used for cloning the gene of interest and analysing the protein expression. (2) DNA sequencing of recombinant pLEICS-01carrying target gene. (3) the identity of the recombinant proteins.

(**1**)



(**2**) *Rgg* (SPD_1518)
Alignment with T7 Promoter-F

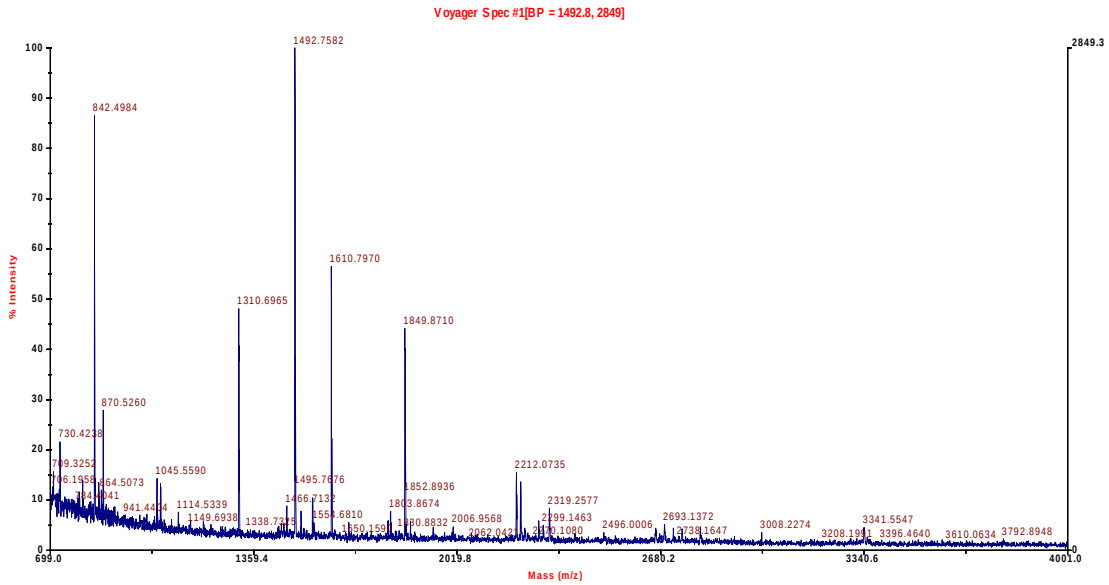
```
Query  13       ATGAGATGGGATTATGGACAAATTTTTAAAGAAATTCGAAAGTCAAAAGGATTGACCCAA  72
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539701  ATGAGATGGGATTATGGACAAATTTTTAAAGAAATTCGAAAGTCAAAAGGATTGACCCAA  1539760

Query  73       CAAGATGTATGTGGACAAGTCATACATCGAACAACTCTAACAAATATTGAACACGGTAAA  132
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539761  CAAGATGTATGTGGACAAGTCATACATCGAACAACTCTAACAAATATTGAACACGGTAAA  1539820

Query  133      GTTATTCCCAGTTTTGAAAACATGGTATTTCTTCTTGAACAAATTGATATGAGCTTGGCA  192
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539821  GTTATTCCCAGTTTTGAAAACATGGTATTTCTTCTTGAACAAATTGATATGAGCTTGGCA  1539880

Query  193      GAATTCAAGTATATATGCAACGAATACCATCCTAGTAAAAGGCGAGATATTATTGTAGAG  252
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539881  GAATTCAAGTATATATGCAACGAATACCATCCTAGTAAAAGGCGAGATATTATTGTAGAG  1539940

Query  253      AGCCAAAATCCGTCTACTTTTCAAGATACTAGAAAAATGGTTGAACTCACTGAGAAATGT  312
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1539941  AGCCAAAATCCGTCTACTTTTCAAGATACTAGAAAAATGGTTGAACTCACTGAGAAATGT  1540000

Query  313      CAAAAATATCTTAAGACACATCACGATGTTCCTATTCAAAATATCTATCGTCATACAAAA  372
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1540001  CAAAAATATCTTAAGACACATCACGATGTTCCTATTCAAAATATCTATCGTCATACAAAA  1540060

Query  373      ATTGTCACAGAATTACGAACTAAAGGATTCAAAAACAACCACGTCTTGAAAGATTTGTAT  432
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1540061  ATTGTCACAGAATTACGAACTAAAGGATTCAAAAACAACCACGTCTTGAAAGATTTGTAT  1540120

Query  433      GAAGAAATTTGGGACTATCTTGAACCTATGGATACATGGTACATTAGTGATTTGAAATTG  492
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1540121  GAAGAAATTTGGGACTATCTTGAACCTATGGATACATGGTACATTAGTGATTTGAAATTG  1540180

Query  493      CTTGGAACCATTCTCTTTTTCTTTCCTTCTGAAAATCTTCCCCTTCTTATTGATAGAATT  552
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1540181  CTTGGAACCATTCTCTTTTTCTTTCCTTCTGAAAATCTTCCCCTTCTTATTGATAGAATT  1540240

Query  553      ATGAAAACCATCGAGAAATATAAATACTTCCGAGAAACAAAAGCATTTTTATCATCTTTT  612
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1540241  ATGAAAACCATCGAGAAATATAAATACTTCCGAGAAACAAAAGCATTTTTATCATCTTTT  1540300
```

Alignment with ProteinSPD1518F primer

```
Query  36       CTTGGAACCATTCTCTTTTTCTTTCCTTCTGAAAATCTTCCCCTTCTTATTGATAGAATT  95
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1540181  CTTGGAACCATTCTCTTTTTCTTTCCTTCTGAAAATCTTCCCCTTCTTATTGATAGAATT  1540240

Query  96       ATGAAAACCATCGAGAAATATAAATACTTCCGAGAAACAAAAGCATTTTTATCATCTTTT  155
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1540241  ATGAAAACCATCGAGAAATATAAATACTTCCGAGAAACAAAAGCATTTTTATCATCTTTT  1540300

Query  156      TTAGCCAATCTCTCCACTGTATATTTTCAACATCATTTATTCAAAGAATGTGAAACAATC  215
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1540301  TTAGCCAATCTCTCCACTGTATATTTTCAACATCATTTATTCAAAGAATGTGAAACAATC  1540360

Query  216      ACCCTACAACTTTTGGTATTAGCAGAAGAATTAAAAATATACGATATTCTTGGATTCTCT  275
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1540361  ACCCTACAACTTTTGGTATTAGCAGAAGAATTAAAAATATACGATATTCTTGGATTCTCT  1540420

Query  276      CAAGTACGCTTGGGAATACTACAACATAACTCCGATTTGATTGATAAAGGAATAACTTTA  335
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1540421  CAAGTACGCTTGGGAATACTACAACATAACTCCGATTTGATTGATAAAGGAATAACTTTA  1540480

Query  336      TTACGACTAACCAAAGAAGAGGCATTAGTTAAAATACTCGAAAAAGAAATTAATGATTTT  395
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1540481  TTACGACTAACCAAAGAAGAGGCATTAGTTAAAATACTCGAAAAAGAAATTAATGATTTT  1540540

Query  396      TCAAACCTA  404
                |||||||||
Sbjct  1540541  TCAAACCTA  1540549
```

(**3**) MALDI-TOF mass spectrometry confirming the identity of the recombinant proteins.

Rgg1518



Appendix 5
Microarray analysis of *Δrgg0999* and *Δrgg1518* in comparison with wild type D39 grown in CDM supplemented with galactose and mannose.

1-Galactose

(Galactose), downrgulated in Δ*rgg0999*

Locus Tag	Name	FOLD[a]	Function
SPD_0013	ftsH	-2.35	cell division protein FtsH
SPD_0017		-2.4	
SPD_0021		-2.11	
SPD_0047		-3.15	hypothetical protein
SPD_0075		-2.48	hypothetical protein
SPD_0076		-2.11	potassium uptake protein, Trk family protein
SPD_0091		-3.26	hypothetical protein
SPD_0100		-3.25	
SPD_0113		-7.08	
SPD_0114		-7.79	hypothetical protein
SPD_0115		-9.66	hypothetical protein
SPD_0116		-3.38	hypothetical protein
SPD_0117		-2.49	hypothetical protein
SPD_0118		-2.45	hypothetical protein
SPD_0120		-2.74	hypothetical protein
SPD_0121		-3.38	hypothetical protein
SPD_0122		-2.28	hypothetical protein
SPD_0123		-4.54	hypothetical protein
SPD_0124		-2.55	hypothetical protein
SPD_0130		-2.19	metallo-beta-lactamase superfamily protein domain protein
SPD_0145		-4.45	hypothetical protein
SPD_0146		-3.7	CAAX amino terminal protease family protein
SPD_0147		-2.52	CAAX amino terminal protease family protein
SPD_0150		-2.35	ABC transporter, substrate-binding protein
SPD_0161		-2.54	hypothetical protein
SPD_0179		-2.1	lipoprotein, putative
SPD_0180		-3.22	hypothetical protein
SPD_0181		-3.37	conserved hypothetical protein TIGR00250
SPD_0182		-2.08	hypothetical protein
SPD_0192	rpsJ	-2.91	ribosomal protein S10
SPD_0193	rplC	-2.96	ribosomal protein L3
SPD_0194	rplD	-3.67	ribosomal protein L4
SPD_0195	rplW	-3.15	ribosomal protein L23
SPD_0196	rplB	-2.43	ribosomal protein L2
SPD_0197	rpsS	-3.17	ribosomal protein S19
SPD_0198	rplV	-3.24	ribosomal protein L22
SPD_0199	rpsC	-2.97	ribosomal protein S3

SPD_0200	rplP	-2.44	ribosomal protein L16
SPD_0201	rpmC	-3.08	ribosomal protein L29
SPD_0202	rpsQ	-2.64	ribosomal protein S17
SPD_0203	rplN	-2.34	ribosomal protein L14
SPD_0215	infA	-2.76	translation initiation factor IF-1
SPD_0216	rpsM	-2.74	ribosomal protein S13
SPD_0217	rpsK	-2.67	ribosomal protein S11
SPD_0218	rpoA	-2.18	DNA-directed RNA polymerase, alpha subunit
SPD_0219	rplQ	-3.85	ribosomal protein L17
SPD_0244	cdsA	-2.05	phosphatidate cytidylyltransferase
SPD_0251	rpsL	-3.61	ribosomal protein S12
SPD_0252	rpsG	-3.57	ribosomal protein S7
SPD_0262		-2.16	PTS system, mannose/fructose/sorbose family protein, IID component
SPD_0263	manM	-2.62	PTS system, mannose-specific IIC component
SPD_0264	manL	-3.33	PTS system, mannose-specific IIAB components
SPD_0273		-2.47	hypothetical protein
SPD_0315	cps2A	-3.7	integral membrane regulatory protein Cps2A
SPD_0316	cps2B	-3.49	tyrosine-protein phosphatase CpsB
SPD_0317	cps2C	-3.87	chain length determinant protein/polysaccharide export protein, MPA1 family protein
SPD_0318	cps2D	-2.64	tyrosine-protein kinase Cps2D cytosolic ATPase domain
SPD_0320	cps2T	-2.06	glycosyl transferase, group 1 family protein, putative
SPD_0326	cps2K	-3.83	UDP-glucose 6-dehydrogenase, putative
SPD_0327	cps2P	-3.96	UDP-galactopyranose mutase
SPD_0328	cps2L	-3.35	glucose-1-phosphate thymidylyltransferase
SPD_0329	rfbC	-2.58	dTDP-4-dehydrorhamnose 3,5-epimerase, putative
SPD_0330	rfbB	-2.49	dTDP-glucose 4,6-dehydratase
SPD_0334	aliA	-3.02	oligopeptide ABC transporter, oligopeptide-binding protein AliA
SPD_0336	pbp1A	-2.46	penicillin-binding protein 1A
SPD_0337	recU	-3.61	recombination protein U
SPD_0343	gnd	-2.95	6-phosphogluconate dehydrogenase, decarboxylating
SPD_0344		-2.9	DNA-binding response regulator
SPD_0365	tig	-3.14	trigger factor
SPD_0373		-2.43	hypothetical protein
SPD_0378		-2.78	enoyl-CoA hydratase/isomerase family protein
SPD_0379		-2.24	transcriptional regulator, MarR family protein
SPD_0380	fabH	-3.52	3-oxoacyl-(acyl-carrier-protein) synthase III
SPD_0381	acpP	-2.33	acyl carrier protein
SPD_0382	fabK	-3.37	trans-2-enoyl-ACP reductase II
SPD_0383	fabD	-2.94	malonyl CoA-acyl carrier protein transacylase
SPD_0393	nusB	-2.22	transcription antitermination factor NusB
SPD_0394		-2.95	hypothetical protein
SPD_0404	ilvB	-3.1	acetolactate synthase, large subunit, biosynthetic type
SPD_0406	ilvC	-2.76	ketol-acid reductoisomerase
SPD_0407		-2.96	hypothetical protein
SPD_0408		-2.37	hypothetical protein

SPD_0418		-3.34	hypothetical protein
SPD_0437		-2.06	hypothetical protein
SPD_0448	glnA	-2.68	glutamine synthetase, type I
SPD_0467	blpS	-2.12	BlpS protein
SPD_0473	blpY	-2.37	immunity protein BlpY
SPD_0474		-2.32	hypothetical protein
SPD_0533		-2.13	metallo-beta-lactamase superfamily protein
SPD_0550	rplK	-3.31	ribosomal protein L11
SPD_0551	rplA	-3.14	ribosomal protein L1
SPD_0554		-3.35	ABC transporter, ATP-binding protein
SPD_0555		-2.24	hypothetical protein
SPD_0557		-2.08	
SPD_0593		-2.22	elongation factor Tu family protein
SPD_0620	lysS	-2.59	lysyl-tRNA synthetase
SPD_0621	lctO	-2.14	lactate oxidase
SPD_0636	spxB	-2.1	pyruvate oxidase
SPD_0650	clpP	-2.2	ATP-dependent Clp protease, proteolytic subunit ClpP
SPD_0659	ftsE	-2.02	cell division ATP-binding protein FtsE
SPD_0667	sodA	-3.59	superoxide dismutase, manganese-dependent
SPD_0674	rpsP	-2.37	ribosomal protein S16
SPD_0685	gor	-2.55	glutathione-disulfide reductase
SPD_0690		-2.51	hypothetical protein
SPD_0700	pepN	-2.08	aminopeptidase N
SPD_0710	ezrA	-2.13	septation ring formation regulator EzrA
SPD_0721	folD	-2.08	methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase
SPD_0724	deoB	-2.6	phosphopentomutase
SPD_0726		-2.05	purine nucleoside phosphorylase, family protein 2
SPD_0728		-2.03	hypothetical protein
SPD_0750		-2.47	hypothetical protein
SPD_0751		-2.3	hypothetical protein
SPD_0752		-3.6	hypothetical protein
SPD_0753	pcp	-2.79	pyrrolidone-carboxylate peptidase
SPD_0757	rpsA	-2.91	ribosomal protein S1
SPD_0760	dnaX	-2.15	DNA polymerase III, gamma and tau subunits
SPD_0762	sufC	-2.22	FeS assembly ATPase SufC
SPD_0763	sufD	-3.12	FeS assembly protein SufD
SPD_0764	sufS	-5.71	cysteine desulfurases, SufS subfamily protein
SPD_0765		-4.49	SUF system FeS assembly protein, NifU family protein
SPD_0766	sufB	-3.82	FeS assembly protein SufB
SPD_0784		-2.07	type I restriction-modification system, R subunit, putative
SPD_0790	pyk	-2.25	pyruvate kinase
SPD_0797		-2.02	hypothetical protein
SPD_0835	frr	-2.69	ribosome recycling factor
SPD_0838		-2.07	PhoH family protein
SPD_0848	rpmI	-2.49	ribosomal protein L35

SPD_0900	asd	-2.59	aspartate-semialdehyde dehydrogenase
SPD_0901	dapA	-2.24	dihydrodipicolinate synthase
SPD_0913		-2.08	hypothetical protein
SPD_0959		-2.36	hypothetical protein
SPD_0968		-2.63	acetyltransferase, GNAT family protein
SPD_0969		-2.25	hypothetical protein
SPD_0970	map	-2.34	methionine aminopeptidase, type I
SPD_0977		-2.21	hypothetical protein
SPD_0989	rplU	-2.65	ribosomal protein L21
SPD_0990		-2.51	hypothetical protein
SPD_0991	rpmA	-2.89	ribosomal protein L27
SPD_0997	hup	-2.47	DNA-binding protein HU
SPD_1014		-2.15	IS630-Spn1, transposase Orf1
SPD_1037		-2.26	histidine triad protein
SPD_1042	nrdE	-3.6	ribonucleoside-diphosphate reductase, alpha subunit
SPD_1050	lacD	-2.22	tagatose 1,6-diphosphate aldolase
SPD_1051	lacC	-2.36	tagatose-6-phosphate kinase
SPD_1053	lacA	-2.84	galactose-6-phosphate isomerase, LacA subunit
SPD_1060	lepA	-2.5	GTP-binding protein LepA
SPD_1061		-2.28	serine/threonine protein phosphatase
SPD_1064		-2.04	hemolysin A, putative
SPD_1078	ldh	-2.19	L-lactate dehydrogenase
SPD_1083	vicX	-3.21	vicX protein
SPD_1084		-4.07	sensory box sensor histidine kinase
SPD_1085		-2.01	DNA-binding response regulator
SPD_1099		-2.44	amino acid ABC transporter, ATP-binding protein
SPD_1100	zwf	-2.47	glucose-6-phosphate 1-dehydrogenase
SPD_1106		-2.63	
SPD_1127	ispD	-2.45	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase
SPD_1158	gdhA	-2.93	NADP-specific glutamate dehydrogenase
SPD_1188	rplJ	-2.59	ribosomal protein L10
SPD_1193	msrAB1	-2.39	peptide methionine sulfoxide reductase msrA/msrB 1
SPD_1194	thrB	-2.37	homoserine kinase
SPD_1195	hom	-3.59	homoserine dehydrogenase
SPD_1206		-2.16	hypothetical protein
SPD_1207	tyrA	-2.79	prephenate dehydrogenase
SPD_1228		-2.05	phosphate ABC transporter, ATP-binding protein, putative
SPD_1245	rpsU	-2.01	ribosomal protein S21
SPD_1268		-2.06	hypothetical protein
SPD_1287	trxB	-2.47	thioredoxin-disulfide reductase
SPD_1290		-3.06	amino acid ABC transporter, permease protein
SPD_1298	nox	-2.25	NADH oxidase
SPD_1319		-2.06	hypothetical protein
SPD_1326	pgm	-2.04	phosphoglucomutase/phosphomannomutase family protein
SPD_1327	bta	-2.35	bacterocin transport accessory protein

SPD_1334	atpC	-2.08	ATP synthase F1, epsilon subunit
SPD_1335	atpD	-2.53	ATP synthase F1, beta subunit
SPD_1336	atpG	-2.43	ATP synthase F1, gamma subunit
SPD_1337	atpA	-2.28	ATP synthase F1, alpha subunit
SPD_1338	atpH	-3.52	ATP synthase F1, delta subunit
SPD_1339	atpF	-3.57	ATP synthase F0, B subunit
SPD_1344		-3.25	hypothetical protein
SPD_1346		-2.05	Uncharacterized BCR, putative
SPD_1360		-2.46	hypothetical protein
SPD_1368	rpsR	-2.41	ribosomal protein S18
SPD_1369	ssb	-2.31	single-strand binding protein
SPD_1370	rpsF	-3.81	ribosomal protein S6
SPD_1382		-2.46	glutathione S-transferase family protein
SPD_1385		-2.03	ABC transporter, ATP-binding protein
SPD_1404	tpiA	-2.44	triosephosphate isomerase
SPD_1439	rpsO	-2.68	ribosomal protein S15
SPD_1461	psaB	-4.9	manganese ABC transporter, ATP-binding protein
SPD_1462	psaC	-2.81	manganese ABC transporter, permease protein, putative
SPD_1463		-4.98	ABC transporter, substrate binding lipoprotein
SPD_1473		-2.86	hypothetical protein
SPD_1474	divIVA	-3.72	cell division protein DivIVA
SPD_1475	ylmH	-2.48	YlmH protein
SPD_1477	ylmF	-2.52	YlmF protein
SPD_1479	ftsZ	-2.57	cell division protein FtsZ
SPD_1480	ftsA	-2.86	cell division protein FtsA
SPD_1512	secA	-2.32	preprotein translocase, SecA subunit
SPD_1513		-3.55	
SPD_1514		-8.71	ABC transporter, ATP-binding protein
SPD_1515		-7.54	hypothetical protein
SPD_1516		-9.02	hypothetical protein
SPD_1517		-4.53	hypothetical protein
SPD_1520		-2.63	nitroreductase family protein
SPD_1524		-4.68	transcriptional regulator, GntR family protein
SPD_1525		-6.46	ABC transporter, ATP-binding protein
SPD_1526		-3.63	hypothetical protein
SPD_1581		-2.68	
SPD_1588		-3.02	hypothetical protein
SPD_1626	xth	-3.05	exodeoxyribonuclease III
SPD_1627		-3.88	
SPD_1633	galT-2	-2.49	galactose-1-phosphate uridylyltransferase
SPD_1634	galK	-2.35	galactokinase
SPD_1644		-2.06	hypothetical protein
SPD_1655	scpB	-2.04	segregation and condensation protein B
SPD_1668	amiE	-2.16	oligopeptide ABC transporter, ATP-binding protein AmiE
SPD_1669	amiD	-3.2	oligopeptide ABC transporter, permease protein AmiD

SPD_1670	amiC	-2.05	oligopeptide ABC transporter, permease protein AmiC
SPD_1671	amiA	-4.52	oligopeptide ABC transporter, oligopeptide-binding protein AmiA
SPD_1686		-2.5	
SPD_1689		-2.9	
SPD_1706		-2.09	hypothetical protein
SPD_1707		-3.42	hypothetical protein
SPD_1727		-2.41	hypothetical protein
SPD_1730		-2.17	
SPD_1739	recA	-2.12	recA protein
SPD_1759	rpoB	-2.41	DNA-directed RNA polymerase, beta subunit
SPD_1764	murA-2	-2.03	UDP-N-acetylglucosamine 1-carboxyvinyltransferase
SPD_1774	pflA	-2.09	pyruvate formate-lyase activating enzyme
SPD_1790	rpmH	-2.41	ribosomal protein L34
SPD_1795		-2.14	hypothetical protein
SPD_1898		-2.09	hypothetical protein
SPD_1939		-2.05	hypothetical protein
SPD_1942		-2.08	hypothetical protein
SPD_1949		-2.54	hypothetical protein
SPD_1955		-2.21	hypothetical protein
SPD_1963	rpmF	-4.08	ribosomal protein L32
SPD_1983		-3.09	hypothetical protein
SPD_2011	glpF	-2.23	glycerol uptake facilitator protein
SPD_2012	glpO	-3.68	alpha-glycerophosphate oxidase
SPD_2013	glpK	-3.6	glycerol kinase
SPD_2016		-2.42	TIM-barrel protein, nifR3 family protein, putative
SPD_2032		-2.23	DHH subfamily 1 protein
SPD_2044		-2.09	rod shpae-determining protein MreD, putative
SPD_2068		-3.96	serine protease
SPD_2069		-2.48	SpoJ protein
	engB	-2.36	
	glmM	-2.27	
	ppnK	-2.08	
	rpmE2	-2.04	
	tpx	-3.64	

(Galactose), upregulated in Δ*rgg0999*

Locus tag	Name	FOLD	Function
SPD_0058	purD	2.48	phosphoribosylamine--glycine ligase
SPD_0059	purE	2.18	phosphoribosylaminoimidazole carboxylase, catalytic subunit
SPD_0066		2.24	PTS system, IIB component
SPD_0067		2.02	PTS system, IIC component
SPD_0068		2.12	PTS system, IID component
SPD_0069		2.2	PTS system, IIA component
SPD_0070	agaS	2.03	sugar isomerase domain protein AgaS

SPD_0138		2.03	glycosyl transferase, group 1 family protein
SPD_0139		2.1	glycosyl transferase, group 2 family protein
SPD_0141		2.33	hypothetical protein
SPD_0163		2.09	hypothetical protein
SPD_0167	ribB	2.02	3,4-dihydroxy-2-butanone 4-phosphate synthase/GTP cyclohydrolase II
SPD_0168	ribE	2.01	riboflavin synthase, alpha subunit
SPD_0223		2.24	iron(III) ABC transporter, permease protein
SPD_0228		2.2	transcriptional regulator, AraC family protein
SPD_0276		2.6	hypothetical protein
SPD_0278		2.07	hypothetical protein
SPD_0287		2	hyaluronate lyase precursor
SPD_0293		2.25	PTS system, IIA component
SPD_0302		2.03	hypothetical protein
SPD_0364		2.15	ABC-type polar amino acid transport system, ATPase component
SPD_0424		2.09	PTS system, cellobiose-specific IIC component
SPD_0425		2.27	hypothetical protein
SPD_0426	lacF-1	2.17	PTS system, lactose-specific IIA component
SPD_0458	hrcA	2.82	heat-inducible transcription repressor HrcA
SPD_0459	grpE	4.69	heat shock protein GrpE
SPD_0460	dnaK	2.08	chaperone protein DnaK
SPD_0496		2.36	cell filamentation protein Fic-related protein
SPD_0520		2.03	transposase, putative, truncation
SPD_0527		2.37	oxidoreductase, putative
SPD_0592	rsuA-2	2.07	ribosomal small subunit pseudouridine synthase A
SPD_0603		2.06	peptidase, M50 family protein
SPD_0613		2.11	hypothetical protein
SPD_0616		2.16	amino acid ABC transporter, ATP-binding protein
SPD_0628	ssrA	2.54	transcriptional activator TenA, TENA/THI-4 family protein
SPD_0640		2.42	
SPD_0744		2.07	
SPD_0773		2.49	PTS system, fructose specific IIABC components
SPD_0831		2.15	hypothetical protein
SPD_0843	celA	2.14	competence protein CelA
SPD_0891		2.14	
SPD_0893		2.1	hypothetical protein
SPD_0929		2	hypothetical protein
SPD_0938		2.3	hypothetical protein
SPD_0942		2.13	hypothetical protein
SPD_0945		2.2	AMP-binding enzyme, putative
SPD_0946		2.18	hypothetical protein
SPD_0951		2.14	hypothetical protein
SPD_1035		2.2	hypothetical protein
SPD_1072		2.06	hypothetical protein
SPD_1185		2.11	hypothetical protein
SPD_1254		2.01	hypothetical protein

SPD_1255		2.03	ABC transporter, ATP-binding protein
SPD_1257		2.09	hypothetical protein
SPD_1261		2.05	hypothetical protein
SPD_1266		2.18	hypothetical protein
SPD_1271		2.07	hypothetical protein
SPD_1322		2.09	hypothetical protein
SPD_1325		2.28	
SPD_1356		2.08	
SPD_1409		2.18	sugar ABC transporter, ATP-binding protein
SPD_1424	trpC	2.05	tRNA pseudouridine synthase A
SPD_1470		2.08	hypothetical protein
SPD_1587		2.06	transcriptional activator, putative
SPD_1594		2	transcriptional regulator
SPD_1599	tenA	2.02	indole-3-glycerol phosphate synthase
SPD_1611		2.05	hypothetical protein
SPD_1612	galE-2	2.1	UDP-glucose 4-epimerase
SPD_1613	galT-1	2.03	galactose-1-phosphate uridylyltransferase
SPD_1677	rafE	2.18	sugar ABC transporter, sugar-binding protein
SPD_1711		2.13	single-strand binding protein family protein
SPD_1733		2.15	hypothetical protein
SPD_1750	truA	2.02	multimeric flavodoxin WrbA (general function prediction only)
SPD_1750	wrbA	2.03	multimeric flavodoxin WrbA (general function prediction only)
SPD_1855		2.06	hypothetical protein
SPD_1901		2.17	transposase, putative
SPD_1917		2.2	hypothetical protein
SPD_1974		2.01	hypothetical protein
SPD_1980		2	
SPD_1985		2.02	alcohol dehydrogenase, iron-containing
SPD_1987		2.12	fucolectin-related protein
SPD_1989		2.18	PTS system, IID component
SPD_1996		2.23	fucose operon repressor, putative
SPD_2009		2.02	hypothetical protein

(Galactose), downregulated in Δ*rgg1518*

Locus tag	Name	Fold	Gene Name
SPD_0036		-2.19	hypothetical protein
SPD_0058	purD	-2.48	phosphoribosylamine--glycine ligase
SPD_0059	purE	-2.22	phosphoribosylaminoimidazole carboxylase, catalytic subunit
SPD_0105		-2.21	hypothetical protein
SPD_0141		-2.19	hypothetical protein
SPD_0143		-2.17	UDP-glucose 6-dehydrogenase, putative
SPD_0163		-2	hypothetical protein
SPD_0167	ribB	-2.15	3,4-dihydroxy-2-butanone 4-phosphate synthase/GTP cyclohydrolase II

SPD_0223		-2.12	iron(III) ABC transporter, permease protein
SPD_0224		-2	iron(III) ABC transporter, permease protein
SPD_0228		-2.25	transcriptional regulator, AraC family protein
SPD_0276		-2.6	hypothetical protein
SPD_0293		-2.09	PTS system, IIA component
SPD_0364		-2.09	ABC-type polar amino acid transport system, ATPase component
SPD_0372		-2.09	sodium:alanine symporter family protein
SPD_0424		-2.04	PTS system, cellobiose-specific IIC component
SPD_0425		-2.41	hypothetical protein
SPD_0426	lacF-1	-2.81	PTS system, lactose-specific IIA component
SPD_0449		-2.03	hypothetical protein
SPD_0470	blpC	-2.01	peptide pheromone BlpC
SPD_0496		-2.33	cell filamentation protein Fic-related protein
SPD_0509		-2.07	hypothetical protein
SPD_0520		-2.21	transposase, putative, truncation
SPD_0527		-2.36	oxidoreductase, putative
SPD_0531		-2	
SPD_0573	msrAB2	-2	peptide methionine sulfoxide reductase msrA/msrB 2
SPD_0592	rsuA-2	-2.33	ribosomal small subunit pseudouridine synthase A
SPD_0603		-2.14	peptidase, M50 family protein
SPD_0613		-2.28	hypothetical protein
SPD_0616		-2.36	amino acid ABC transporter, ATP-binding protein
SPD_0626		-2.17	ABC transporter, ATP-binding protein
SPD_0640		-2.33	
SPD_0647		-2.05	transcriptional regulator, TetR family protein
SPD_0744		-2.26	
SPD_0772		-2.32	1-phosphofructokinase, putative
SPD_0843	celA	-2.09	competence protein CelA
SPD_0891		-2.15	
SPD_0893		-2.14	hypothetical protein
SPD_0903	xylH	-2.03	4-oxalocrotonate tautomerase
SPD_0929		-2	hypothetical protein
SPD_0938		-2.34	hypothetical protein
SPD_0942		-2.27	hypothetical protein
SPD_0945		-2.13	AMP-binding enzyme, putative
SPD_0946		-2.06	hypothetical protein
SPD_0951		-2.07	hypothetical protein
SPD_1035		-2.1	hypothetical protein
SPD_1182		-2.07	hypothetical protein
SPD_1254		-2.38	hypothetical protein
SPD_1257		-2.32	hypothetical protein
SPD_1261		-2.18	hypothetical protein
SPD_1266		-2.13	hypothetical protein
SPD_1271		-2.13	hypothetical protein
SPD_1322		-2.04	hypothetical protein

SPD_1325		-2.57	
SPD_1354		-2.13	hypothetical protein
SPD_1424	truA	-2.29	tRNA pseudouridine synthase A
SPD_1470		-2.01	hypothetical protein
SPD_1471		-2.1	hypothetical protein
SPD_1503		-2.25	hypothetical protein
SPD_1518		-2.07	transcriptional activator, Rgg/GadR/MutR family protein
SPD_1587		-2.13	transcriptional activator, putative
SPD_1599	trpC	-2.05	indole-3-glycerol phosphate synthase
SPD_1611		-2.17	hypothetical protein
SPD_1614		-2.24	phosphate transport system regulatory protein PhoU, putative
SPD_1677	rafE	-2.08	sugar ABC transporter, sugar-binding protein
SPD_1711		-2.23	single-strand binding protein family protein
SPD_1731		-2.01	hypothetical protein
SPD_1733		-2.22	hypothetical protein
SPD_1750	wrbA	-2.13	multimeric flavodoxin WrbA (general function prediction only)
SPD_1789		-2.01	cell wall surface anchor family protein
SPD_1796		-2.11	L-asparaginase, putative
SPD_1798		-2	DNA-binding response regulator
SPD_1855		-2.17	hypothetical protein
SPD_1901		-2	transposase, putative
SPD_1917		-2.12	hypothetical protein
SPD_1974		-2.06	hypothetical protein
SPD_1980		-2.34	
SPD_1985		-2.14	alcohol dehydrogenase, iron-containing
SPD_1987		-2.24	fucolectin-related protein
SPD_1989		-2.09	PTS system, IID component

(Galactose), uprgulated in Δ*rgg1518*

Locus Tag	Name	FOLD	Function
SPD_0013	ftsH	2.82	cell division protein FtsH
SPD_0017		4.48	
SPD_0020		2.89	
SPD_0021		2.28	
SPD_0047		3.46	hypothetical protein
SPD_0064		2.62	transcriptional regulator, GntR family protein
SPD_0091		2.74	hypothetical protein
SPD_0100		2.38	
SPD_0113		5.88	
SPD_0114		7.19	hypothetical protein
SPD_0115		8.53	hypothetical protein
SPD_0116		2.09	hypothetical protein
SPD_0117		2.33	hypothetical protein
SPD_0118		2.46	hypothetical protein

SPD_0120		2.47	hypothetical protein
SPD_0121		2.82	hypothetical protein
SPD_0123		3.6	hypothetical protein
SPD_0124		2.41	hypothetical protein
SPD_0126	pspA	2	pneumococcal surface protein A
SPD_0145		3.39	hypothetical protein
SPD_0146		3.66	CAAX amino terminal protease family protein
SPD_0147		2.74	CAAX amino terminal protease family protein
SPD_0148		2.27	transporter, major facilitator family protein
SPD_0150		2.57	ABC transporter, substrate-binding protein
SPD_0179		2.35	lipoprotein, putative
SPD_0180		4.3	hypothetical protein
SPD_0181		6.06	conserved hypothetical protein TIGR00250
SPD_0182		2.18	hypothetical protein
SPD_0192	rpsJ	4.51	ribosomal protein S10
SPD_0193	rplC	5.62	ribosomal protein L3
SPD_0194	rplD	7.46	ribosomal protein L4
SPD_0195	rplW	4.73	ribosomal protein L23
SPD_0196	rplB	3.67	ribosomal protein L2
SPD_0197	rpsS	6.72	ribosomal protein S19
SPD_0198	rplV	8.56	ribosomal protein L22
SPD_0199	rpsC	7.81	ribosomal protein S3
SPD_0200	rplP	3.69	ribosomal protein L16
SPD_0201	rpmC	7.17	ribosomal protein L29
SPD_0202	rpsQ	8.11	ribosomal protein S17
SPD_0203	rplN	6.94	ribosomal protein L14
SPD_0204	rplX	5.41	ribosomal protein L24
SPD_0205	rplE	6.66	ribosomal protein L5
SPD_0206	rpsN	4.41	ribosomal protein S14
SPD_0207	rpsH	5.83	ribosomal protein S8
SPD_0208	rplF	4.77	ribosomal protein L6
SPD_0209	rplR	3.12	ribosomal protein L18
SPD_0210	rpsE	3.66	ribosomal protein S5
SPD_0211	rpmD	4.34	ribosomal protein L30
SPD_0212	rplO	4.82	ribosomal protein L15
SPD_0213	secY	2.66	preprotein translocase, SecY subunit
SPD_0215	infA	4.94	translation initiation factor IF-1
SPD_0216	rpsM	3.89	ribosomal protein S13
SPD_0217	rpsK	4.22	ribosomal protein S11
SPD_0218	rpoA	3.43	DNA-directed RNA polymerase, alpha subunit
SPD_0219	rplQ	8.71	ribosomal protein L17
SPD_0244	cdsA	2.08	phosphatidate cytidylyltransferase
SPD_0251	rpsL	5.28	ribosomal protein S12
SPD_0252	rpsG	4.74	ribosomal protein S7
SPD_0253	fusA	3	translation elongation factor G

SPD_0256		2.18	conserved hypothetical protein TIGR00053
SPD_0257		2.9	hypothetical protein
SPD_0263	manM	2.59	PTS system, mannose-specific IIC component
SPD_0264	manL	2.52	PTS system, mannose-specific IIAB components
SPD_0274	rplM	3.17	ribosomal protein L13
SPD_0275	rpsI	3.25	ribosomal protein S9
SPD_0314		2.01	
SPD_0315	cps2A	5.86	integral membrane regulatory protein Cps2A
SPD_0316	cps2B	4.55	tyrosine-protein phosphatase CpsB
SPD_0317	cps2C	6.36	chain length determinant protein/polysaccharide export protein, MPA1 family protein
SPD_0318	cps2D	4.11	tyrosine-protein kinase Cps2D cytosolic ATPase domain
SPD_0319	cps2E	2.5	undecaprenylphosphate glucosephosphotransferase Cps2E
SPD_0320	cps2T	2.56	glycosyl transferase, group 1 family protein, putative
SPD_0326	cps2K	4.05	UDP-glucose 6-dehydrogenase, putative
SPD_0327	cps2P	5.9	UDP-galactopyranose mutase
SPD_0328	cps2L	5.44	glucose-1-phosphate thymidylyltransferase
SPD_0329	rfbC	3.9	dTDP-4-dehydrorhamnose 3,5-epimerase, putative
SPD_0330	rfbB	2.96	dTDP-glucose 4,6-dehydratase
SPD_0331	rfbD	3.45	dTDP-4-dehydrorhamnose reductase
SPD_0332		2.28	
SPD_0333		2.4	hypothetical protein
SPD_0334	aliA	4.26	oligopeptide ABC transporter, oligopeptide-binding protein AliA
SPD_0336	pbp1A	3.08	penicillin-binding protein 1A
SPD_0337	recU	4.58	recombination protein U
SPD_0342		2.02	hypothetical protein
SPD_0343	gnd	2.51	6-phosphogluconate dehydrogenase, decarboxylating
SPD_0344		2.85	DNA-binding response regulator
SPD_0365	tig	4.18	trigger factor
SPD_0378		3.32	enoyl-CoA hydratase/isomerase family protein
SPD_0379		4.45	transcriptional regulator, MarR family protein
SPD_0380	fabH	7.19	3-oxoacyl-(acyl-carrier-protein) synthase III
SPD_0381	acpP	2.66	acyl carrier protein
SPD_0382	fabK	4.54	trans-2-enoyl-ACP reductase II
SPD_0383	fabD	4.67	malonyl CoA-acyl carrier protein transacylase
SPD_0385	fabF	2.96	3-oxoacyl-[acyl-carrier-protein] synthase II
SPD_0387	fabZ	3.61	beta-hydroxyacyl-(acyl-carrier-protein) dehydratase FabZ
SPD_0388	accC	2.16	acetyl-CoA carboxylase, biotin carboxylase
SPD_0389	accD	2.88	acetyl-CoA carboxylase, carboxyl transferase, beta subunit
SPD_0390	accA	2.72	acetyl-CoA carboxylase, carboxyl transferase, alpha subunit
SPD_0394		2.56	hypothetical protein
SPD_0404	ilvB	2.89	acetolactate synthase, large subunit, biosynthetic type
SPD_0406	ilvC	2.83	ketol-acid reductoisomerase
SPD_0407		3.09	hypothetical protein
SPD_0408		2.08	hypothetical protein
SPD_0418		3.19	hypothetical protein

SPD_0420	pflB	2.73	formate acetyltransferase
SPD_0448	glnA	2.97	glutamine synthetase, type I
SPD_0460	dnaK	2.09	chaperone protein DnaK
SPD_0467	blpS	2.43	BlpS protein
SPD_0550	rplK	5.61	ribosomal protein L11
SPD_0551	rplA	4.58	ribosomal protein L1
SPD_0554		2.09	ABC transporter, ATP-binding protein
SPD_0556		2.1	hypothetical protein
SPD_0557		2.22	
SPD_0560		2.11	PTS system, IIB component, putative
SPD_0561		2.68	PTS system, IIC component, putative
SPD_0562	bgaA	2.13	beta-galactosidase precursor, putative
SPD_0576		2	conserved hypothetical protein TIGR01440
SPD_0577	zmpB	2.42	zinc metalloprotease ZmpB
SPD_0593		2.34	elongation factor Tu family protein
SPD_0594		2.1	hypothetical protein
SPD_0620	lysS	2.28	lysyl-tRNA synthetase
SPD_0636	spxB	2.02	pyruvate oxidase
SPD_0650	clpP	2.4	ATP-dependent Clp protease, proteolytic subunit ClpP
SPD_0652	livJ	2.06	branched-chain amino acid ABC transporter, amino acid-binding protein
SPD_0667	sodA	3.05	superoxide dismutase, manganese-dependent
SPD_0674	rpsP	2.89	ribosomal protein S16
SPD_0690		2.61	hypothetical protein
SPD_0705		2.16	DnaQ family exonuclease/DinG family helicase, putative
SPD_0710	ezrA	2.48	septation ring formation regulator EzrA
SPD_0716		2.01	IS630-Spn1, transposase Orf1
SPD_0721	folD	2.19	methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase
SPD_0724	deoB	2.32	phosphopentomutase
SPD_0726		2.72	purine nucleoside phosphorylase, family protein 2
SPD_0728		2.56	hypothetical protein
SPD_0750		2.61	hypothetical protein
SPD_0751		2.64	hypothetical protein
SPD_0752		4.33	hypothetical protein
SPD_0753	pcp	2.59	pyrrolidone-carboxylate peptidase
SPD_0756		2.34	
SPD_0757	rpsA	3.4	ribosomal protein S1
SPD_0760	dnaX	2.14	DNA polymerase III, gamma and tau subunits
SPD_0762	sufC	2.4	FeS assembly ATPase SufC
SPD_0763	sufD	3.61	FeS assembly protein SufD
SPD_0764	sufS	5.72	cysteine desulfurases, SufS subfamily protein
SPD_0765		4.49	SUF system FeS assembly protein, NifU family protein
SPD_0766	sufB	3.41	FeS assembly protein SufB
SPD_0784		2.48	type I restriction-modification system, R subunit, putative
SPD_0785		2.27	hypothetical protein
SPD_0790	pyk	2.45	pyruvate kinase

SPD_0835	frr	3.8	ribosome recycling factor
SPD_0838		2.52	PhoH family protein
SPD_0847	infC	2.24	translation initiation factor IF-3
SPD_0848	rpmI	4.05	ribosomal protein L35
SPD_0900	asd	2.3	aspartate-semialdehyde dehydrogenase
SPD_0905		2.4	acetyltransferase, GNAT family protein
SPD_0913		2.66	hypothetical protein
SPD_0959		2.81	hypothetical protein
SPD_0968		2.66	acetyltransferase, GNAT family protein
SPD_0969		2.65	hypothetical protein
SPD_0970	map	2.21	methionine aminopeptidase, type I
SPD_0989	rplU	3.66	ribosomal protein L21
SPD_0990		3.62	hypothetical protein
SPD_0991	rpmA	3.83	ribosomal protein L27
SPD_0996		2.04	DegV family protein
SPD_1014		2.27	IS630-Spn1, transposase Orf1
SPD_1037		2.62	histidine triad protein
SPD_1039	ptsI	2	phosphoenolpyruvate-protein phosphotransferase
SPD_1042	nrdE	3.71	ribonucleoside-diphosphate reductase, alpha subunit
SPD_1049	lacT	2.13	transcription antiterminator LacT
SPD_1050	lacD	4.14	tagatose 1,6-diphosphate aldolase
SPD_1051	lacC	3.57	tagatose-6-phosphate kinase
SPD_1052	lacB	3.46	galactose-6-phosphate isomerase, LacB subunit
SPD_1053	lacA	4.73	galactose-6-phosphate isomerase, LacA subunit
SPD_1060	lepA	2.32	GTP-binding protein LepA
SPD_1061		2.7	serine/threonine protein phosphatase
SPD_1064		2.31	hemolysin A, putative
SPD_1078	ldh	2.2	L-lactate dehydrogenase
SPD_1083	vicX	3.29	vicX protein
SPD_1084		4.52	sensory box sensor histidine kinase
SPD_1085		2.11	DNA-binding response regulator
SPD_1099		2.46	amino acid ABC transporter, ATP-binding protein
SPD_1100	zwf	2.33	glucose-6-phosphate 1-dehydrogenase
SPD_1106		2.5	
SPD_1125	pck	2.78	choline kinase
SPD_1127	ispD	3.22	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase
SPD_1158	gdhA	2.03	NADP-specific glutamate dehydrogenase
SPD_1175		2.07	hypothetical protein
SPD_1188	rplJ	3.23	ribosomal protein L10
SPD_1193	msrAB1	2.37	peptide methionine sulfoxide reductase msrA/msrB 1
SPD_1194	thrB	2.15	homoserine kinase
SPD_1195	hom	3.21	homoserine dehydrogenase
SPD_1206		2.31	hypothetical protein
SPD_1207	tyrA	3.1	prephenate dehydrogenase
SPD_1228		2.33	phosphate ABC transporter, ATP-binding protein, putative

SPD_1245	rpsU	2.31	ribosomal protein S21
SPD_1246	nagB	2.89	glucosamine-6-phosphate isomerase
SPD_1268		3.5	hypothetical protein
SPD_1269		2	hypothetical protein
SPD_1290		2.56	amino acid ABC transporter, permease protein
SPD_1298	nox	2.24	NADH oxidase
SPD_1309	pgdA	2.26	peptidoglycan GlcNAc deacetylase
SPD_1310		2.25	hypothetical protein
SPD_1318	tuf	2.01	translation elongation factor Tu
SPD_1319		2.37	hypothetical protein
SPD_1326	pgm	2.27	phosphoglucomutase/phosphomannomutase family protein
SPD_1338	atpH	4.53	ATP synthase F1, delta subunit
SPD_1339	atpF	4.65	ATP synthase F0, B subunit
SPD_1340	atpB	2.01	ATP synthase F0, A subunit
SPD_1344		2.88	hypothetical protein
SPD_1346		2.63	Uncharacterized BCR, putative
SPD_1349	murC	2.04	UDP-N-acetylmuramate--alanine ligase
SPD_1365		2.08	hypothetical protein
SPD_1368	rpsR	3.31	ribosomal protein S18
SPD_1369	ssb	2.9	single-strand binding protein
SPD_1370	rpsF	5.23	ribosomal protein S6
SPD_1400		2.26	hypothetical protein
SPD_1403	lytC	2.07	1,4-beta-N-acetylmuramidase, putative
SPD_1404	tpiA	3.33	triosephosphate isomerase
SPD_1410		3.18	
SPD_1439	rpsO	2.35	ribosomal protein S15
SPD_1461	psaB	3.72	manganese ABC transporter, ATP-binding protein
SPD_1462	psaC	2.56	manganese ABC transporter, permease protein, putative
SPD_1463		5.74	ABC transporter, substrate binding lipoprotein
SPD_1473		2.85	hypothetical protein
SPD_1474	divIVA	5.48	cell division protein DivIVA
SPD_1475	ylmH	2.85	YlmH protein
SPD_1477	ylmF	2.63	YlmF protein
SPD_1479	ftsZ	3.25	cell division protein FtsZ
SPD_1480	ftsA	2.63	cell division protein FtsA
SPD_1512	secA	2.64	preprotein translocase, SecA subunit
SPD_1513		4.09	
SPD_1514		9.48	ABC transporter, ATP-binding protein
SPD_1515		8.42	hypothetical protein
SPD_1516		9.86	hypothetical protein
SPD_1517		4.67	hypothetical protein
SPD_1520		2.54	nitroreductase family protein
SPD_1522		2.11	hypothetical protein
SPD_1524		5.06	transcriptional regulator, GntR family protein
SPD_1525		6.46	ABC transporter, ATP-binding protein

SPD_1526		2.87	hypothetical protein
SPD_1542		2.23	serine/threonine protein kinase
SPD_1554		2.3	iojap-related protein
SPD_1566		3.13	hypothetical protein
SPD_1581		2.88	
SPD_1588		2.68	hypothetical protein
SPD_1626	xth	2.36	exodeoxyribonuclease III
SPD_1627		2.81	
SPD_1633	galT-2	2.03	galactose-1-phosphate uridylyltransferase
SPD_1642	proWX	2.08	choline transporter (glycine betaine transport system permease protein)
SPD_1644		2.13	hypothetical protein
SPD_1655	scpB	2.47	segregation and condensation protein B
SPD_1667	amiF	2.43	oligopeptide ABC transporter, ATP-binding protein AmiF
SPD_1668	amiE	3.23	oligopeptide ABC transporter, ATP-binding protein AmiE
SPD_1669	amiD	5.09	oligopeptide ABC transporter, permease protein AmiD
SPD_1670	amiC	2.38	oligopeptide ABC transporter, permease protein AmiC
SPD_1671	amiA	7.63	oligopeptide ABC transporter, oligopeptide-binding protein AmiA
SPD_1683		2.07	
SPD_1686		5.53	
SPD_1687		2.02	
SPD_1688		3.75	
SPD_1689		4.07	
SPD_1691		3.14	
SPD_1692		2.24	
SPD_1694		3.28	
SPD_1696		2.07	
SPD_1698		3.23	
SPD_1707		3.63	hypothetical protein
SPD_1726	ply	2.57	pneumolysin
SPD_1727		4.81	hypothetical protein
SPD_1728		2.13	hypothetical protein
SPD_1730		2.1	
SPD_1737	lytA	2.01	autolysin/N-acetylmuramoyl-L-alanine amidase
SPD_1739	recA	2.44	recA protein
SPD_1795		2.4	hypothetical protein
SPD_1870		2.06	
SPD_1879		2.46	
SPD_1898		2.69	hypothetical protein
SPD_1939		2.07	hypothetical protein
SPD_1942		2.36	hypothetical protein
SPD_1949		2.6	hypothetical protein
SPD_1955		2.08	hypothetical protein
SPD_1963	rpmF	4.87	ribosomal protein L32
SPD_1983		3.41	hypothetical protein
SPD_2011	glpF	2.2	glycerol uptake facilitator protein

SPD_2012	glpO	4.02	alpha-glycerophosphate oxidase
SPD_2013	glpK	4.35	glycerol kinase
SPD_2016		2.1	TIM-barrel protein, nifR3 family protein, putative
SPD_2020		2.27	DNA-binding response regulator
SPD_2032		2.73	DHH subfamily 1 protein
SPD_2041	tsf	2.99	translation elongation factor Ts
SPD_2042	rpsB	4.98	ribosomal protein S2
SPD_2066		2.6	
SPD_2068		5.84	serine protease
SPD_2069		2.55	SpoJ protein
	engB	2.58	
	glmM	2.4	
	ppnK	2.03	
	tpx	2.17	

2-Mannose

(Mannose), downrgulated in Δ*rgg0999*

Name	Name	FOLD	Function
SPD_1439	rpsO	-2.18	ribosomal protein S15
SPD_0115		-2.02	hypothetical protein
SPD_0116		-3.11	hypothetical protein
SPD_0537		-2.57	matrixin family protein
SPD_1290		-2.56	amino acid ABC transporter, permease protein

(Mannose), upregulated in Δ*rgg0999*

Locus	Name	FOLD	Function
SPD_1339	atpF	2.18	ATP synthase F0, B subunit
SPD_1338	atpH	2.26	ATP synthase F1, delta subunit
SPD_0328	cps2L	2.69	glucose-1-phosphate thymidylyltransferase
SPD_2037	cysK	2.26	cysteine synthase A
SPD_1504	nanA	2.11	sialidase A precursor
SPD_0201	rpmC	2.12	ribosomal protein L29
SPD_0202	rpsQ	2.08	ribosomal protein S17
SPD_0197	rpsS	2.03	ribosomal protein S19
SPD_0093		2.28	hypothetical protein
SPD_0094		2.11	hypothetical protein
SPD_0150		2.49	ABC transporter, substrate-binding protein
SPD_0257		2.51	hypothetical protein
SPD_0373		2	hypothetical protein
SPD_0728		2.19	hypothetical protein
SPD_1084		2.08	sensory box sensor histidine kinase
SPD_1491		2.6	hypothetical protein
SPD_1515		2.03	hypothetical protein
SPD_1516		2.21	hypothetical protein
SPD_2059		2.2	

(Mannose), downrgulated in Δ*rgg1518*

Locus	Name	FOLD	Function
SPD_0316	cps2B	-2.09	tyrosine-protein phosphatase CpsB
SPD_1149	crcB1	-1.98	CrcB protein
SPD_0395	efp	-2.07	translation elongation factor P
	engB	-2.18	
SPD_0343	gnd	-2	6-phosphogluconate dehydrogenase, decarboxylating
	groEL	-2.2	
SPD_0458	hrcA	-2.01	heat-inducible transcription repressor HrcA
SPD_1053	lacA	-2.72	galactose-6-phosphate isomerase, LacA subunit
SPD_1052	lacB	-4.21	galactose-6-phosphate isomerase, LacB subunit
SPD_1050	lacD	-3.45	tagatose 1,6-diphosphate aldolase
SPD_1246	nagB	-2.19	glucosamine-6-phosphate isomerase
SPD_0542	pepV	-2.06	dipeptidase PepV
SPD_0420	pflB	-3.23	formate acetyltransferase
SPD_1461	psaB	-2.27	manganese ABC transporter, ATP-binding protein
SPD_0550	rplK	-2.2	ribosomal protein L11
SPD_0989	rplU	-3.16	ribosomal protein L21
SPD_1245	rpsU	-1.98	ribosomal protein S21
SPD_0021		-1.92	
SPD_0104		-2.62	LysM domain protein
SPD_0114		-2.11	hypothetical protein
SPD_0116		-2.64	hypothetical protein
SPD_0146		-2.02	CAAX amino terminal protease family protein
SPD_0182		-1.9	hypothetical protein
SPD_0418		-2.02	hypothetical protein
SPD_0557		-1.95	
SPD_0765		-2.3	SUF system FeS assembly protein, NifU family protein
SPD_0913		-2.34	hypothetical protein
SPD_0990		-2.24	hypothetical protein
SPD_1290		-2.31	amino acid ABC transporter, permease protein
SPD_1291		-2.69	ArsC family protein
SPD_1293		-2.01	acetyltransferase, GNAT family protein
SPD_1402		-2.22	non-heme iron-containing ferritin
SPD_1473		-1.91	hypothetical protein
SPD_1513		-3.58	
SPD_1514		-6.35	ABC transporter, ATP-binding protein
SPD_1515		-2.93	hypothetical protein
SPD_1516		-5.31	hypothetical protein
SPD_1517		-11.04	hypothetical protein
SPD_1566		-1.97	hypothetical protein
SPD_1834		-2.3	alcohol dehydrogenase, iron-containing
SPD_1962		-2.3	hypothetical protein
SPD_0764	sufS	-2.14	cysteine desulfurases, SufS subfamily protein
SPD_1567	trx	-2.86	thioredoxin

SPD_2041	tsf	-1.97	translation elongation factor Ts

(Mannose), upregulated in Δ*rgg1518*

Locus	Name	FOLD	Function
SPD_0373		1.98	hypothetical protein
SPD_0489		1.92	hypothetical protein
SPD_0519		2.06	hypothetical protein
SPD_0606		1.95	hypothetical protein
SPD_1170		-1.5	oligopeptide ABC transporter,oligopeptide-binding protein
SPD_1783		1.92	hypothetical protein
SPD_2054	recF	1.9	recF protein

a. FOLD ≥ 2 or ≤ -2.0 (Δrgg*0999* or Δrgg*1518* compared with D39 wild type). All P-values are <0.001.

References:

Abraham-Van, P. B. (2004) Review of pneumococcal conjugate vaccine in adults: implications on clinical development. *Vaccine*, 22, 1362-1371.

Afzal, M., Shafeeq, S. and Kuipers, O.P. (2014) LacR is a repressor of lacABCD and LacT is an activator of lacTFEG, constituting the lac gene cluster in *Streptococcus pneumoniae*. *Applied and Environmental Microbiology*, 80, 5349-5358.

Agaisse, H., Gominet, M., Okstad, O.A., Kolsto, A.B. and Lereclus, D. (1999) PlcR is a pleiotropic regulator of extracelular virulence factor gene expression in *Bacillus thuringiensis*. *Molecular Microbiology*, 32, 1043-1053.

Alloing, G., Granadel, C., Morrison, D. and Claverys, J. (1996) Competence pheromone, oligopeptide permease, and induction of competence in *Streptococcus pneumoniae*. *Molecular Microbiology*, 21, 471-478.

Al-Mazrou, Y., Twum-Danso, K., Al Zamil, F. and Kambal, A. (2005) *Streptococcus pneumoniae* serotypes/serogroups causing invasive disease in Riyadh, Saudi Arabia: extent of coverage by pneumococcal vaccines. *Annals of Saudi Medicine*, 25, 94-99.

Araj, G.F., Bey, H.A., Itani, L.Y. and Kanj, S.S. (1999) Drug-resistant *Streptococcus pneumoniae* in the Lebanon: implications for presumptive therapy. *International Journal of Antimicrobial Agents*, 12, 349-354.

Arbique, J.C., Poyart, C., Trieu-Cuot, P., Quesne, G., da Glo´ria, M., Carvalho, S., Steigerwalt, A.G., Morey, R.E., Jackson, D., Davidson, R.J. and Facklam., R.R. (2004) Accuracy of phenotypic and genotypic testing for identification of *Streptococcus pneumoniae* and description of *Streptococcus pseudopneumoniae* sp. *Journal of Clinical Microbiology*, 42, 4686-4696.

Austrian, R. (1999) The pneumococcus at the millennium: not down, not out. *The Journal of Infectious Diseases*, 179, Suppl, S338-341.

Austrian, R. (1981) Some observations on the pneumococcus and on the current status of pneumococcal disease and its prevention. *The Journal of Infectious Diseases*, 3, S1-S17.

Bailey, T.L., Johnson, J., Grant, C.E. and Noble, W.S. (2015) The MEME Suite. *Nucleic Acids Research*, 43, W39-49.

Bailey, T.L. and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings International Conference on Intelligent Systems for Molecular Biology*, 2, 28-36.

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

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

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

Bergmann, S., Rohde, M., Chhatwal, G. and Hammerschmidt, S. (2001) α-Enolase of *Streptococcus pneumoniae* is a plasmin(ogen)-binding protein displayed on the bacterial cell surface. *Molecular Microbiology*, 40, 1273-1287.

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

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

Bessen, D.E., Manoharan, A., Luo, F., Wertz, J.E. and Robinson, D.A. (2005) Evolution of transcription regulatory genes is linked to niche specialization in the bacterial pathogen *Streptococcus pyogenes*. *Journal of Bacteriology*, 187, 4163-4172.

Bidossi, A., Mulas, L., Decorosi, F., Colomba, L., Ricci, S., Pozzi, G., Deutscher, J., Viti, C. and Oggioni, M. (2012) A functional genomics approach to establish the complement of carbohydrate transporters in *Streptococcus pneumoniae*. *PloS One*, 7, e33320.

Bisno, A. L. (1991) Medical progress: group A streptococcal infections and acute rheumatic fever. *The new England Journal of Medicine*, 325, 783-793.

Bogaert, D., Groot, R. De and Hermans, P.W.M. (2004) *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. *The Lancet. Infectious Diseases*, 4, 144-154.

Bonofiglio, L., Garcia, E. and Mollerach, M. (2005) Biochemical characterization of the pneumococcal glucose 1-phosphate uridylyltransferase (GalU) essential for capsule biosynthesis. *Current Microbiology*, 51, 217-221.

Bradford, M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, 72, 248-254.

Bricker, A. and Camilli, A. (1999) Transformation of a type 4 encapsulated strain of *Streptococcus pneumoniae*. *FEMS Microbiology Letters*, 172, 131-135.

Briles, D.E., Ades, E., Paton, J.C., Sampson, J.S., Carlone, G.M., Huebner, R.C., Virolainen, A., Swiatlo, E. and Hollingshead, S.K. (2000) Intranasal immunization of mice with a mixture of the pneumococcal proteins PsaA and PspA is highly protective against nasopharyngeal carriage of *Streptococcus pneumoniae*. *Infection and Immunity*, 68, 796-800.

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

Calix, J.J., and Nahm, M.H. (2010) A new pneumococcal serotype, 11E, has a variably inactivated *wcjE* gene. *The Journal of Infectious Diseases*, 202, 29-38.

Calix J.J, Porambo R.J, Brady A.M., Larson, T.R., Yother, J., Abeygunwardana, C. and Nahm, M.H. (2012) Biochemical, genetic and sero-logical characterization of two capsule subtypes among *Streptococcus pneumoniae* serotype 20 strains: discovery of a new pneumococcal serotype. *Journal of Biological Chemistry*, 287, 27885-27894.

Camara, M., Boulnois, G., Andrew, P. and Mittchell, T. (1994) A neuraminidase from *Streptococcus pneumoniae* has the features of a surface protein. *Infection and Immunity*, 62, 3688-3695.

Carbon, C., Vanrensburg, G, D., Hagberg, L., Fogarty, C., Tellier, G., Rangaraju, M. and Nusrat, R. (2006) Clinical and bacteriologic efficacy of telithromycin in patients with bacteremic community-acquired pneumonia. *Respiratory Medicine*, 100, 577-585.

Carey, M.F., Peterson, C.L. and Smale, S.T. (2009) Chromatin immunoprecipitation (chip). Cold Spring Harbor Protocols, 2009, pdb. prot5279.

Carvalho, S., Kloosterman, T., Kuipers, O. and Neves, A. (2011) CcpA ensures optimal metabolic fitness of *Streptococcus pneumoniae*. *PloS One*, 6, e26707.

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

Chang, J.C., LaSarre, B., Jimenez, J.C., Aggarwal, C. and Federle, M.J. (2011) Two group A streptococcal peptide pheromones act through opposing Rgg regulators to control biofilm development. *PLoS Pathogen*, 7, e1002190.

Chiavolini, D., Pozzi, G. and Ricci, S. (2008) Animal models of *Streptococcus pneumoniae* disease. *Clinical Microbiology Reviews*, 21, 666-685.

Claverys, J. P. and Havarstein, L. S. (2002) Extracellular-peptide control of competence for genetic transformation in *Streptococcus pneumoniae*. *Frontiers*. *Bioscience*, 7, 1798-1814.

Core, L. and Perego, M. (2003) TPR-mediated interaction of RapC with ComA inhibits response.

Regulator-DNA binding for competence development in *Bacillus subtilis*. *Molecular Microbiology*, 49, 1509-1522.

Centers for Disease Control and Prevention. (2010a) Active Bacterial Core Surveillance report, Emerging Infections Program Network, *Streptococcus pneumoniae*. Available at: http://www.cdc.gov/abcs/reports-findings/survreports/spneu10.html. Accessed 1 December 2011.

Centers for Disease Control and Prevention. (2015) Epidemiology and Prevention of Vaccine-Preventable Diseases. Hamborsky J, Kroger A, Wolfe S, eds. 13th ed., Washington DC: Public Health Foundation, 2015.

Centers for Disease Control and Prevention. (2010b) Licensure of a 13-Valent Pneumococcal Conjugate Vaccine (PCV13) and Recommendations for Use Among Children-Advisory Committee on Immunization Practices (ACIP). *Morbidity and Mortality Weekly Report*, 59, 258-261.

Centers for Disease Control and Prevention, C. F. D. C. A. P. (2005b) *Streptococcus pneumoniae disease in CDC, Atlanta, USA*.

Chao, Y., Marks, L.R., Pettigrew, M.M. and Hakansson, A.P. (2014) *Streptococcus pneumoniae* biofilm formation and dispersion during colonization and disease. *Frontiers in Cellular and Infection Microbiology*, 4, 194.

Charpentier, E. and Tuomanen, E. (2000) Mechanisms of antibiotic resistance and tolerance in *Streptococcus pneumoniae*. *Microbes and Infection*, 2**,** 1855-1864.

Chauvaux, S. (1996) CcpA and HPr (ser-P): mediators of catabolite repression in *Bacillus subtilis*. Res. *Microbiology*, 147, 518-522.

Chaussee, M.S., Sylva, G.L., Sturdevant, D.E., Smoot, L.M., Graham, M.R., Watson, R.O. and Musser, J.M. (2002) Rgg influences the expression of multiple regulatory loci to coregulate virulence factor expression in *Streptococcus pyogenes*. *Infection and Immunity*, 70, 762-770.

Chhatwal, G.S., (2002) Anchorless adhesins and invasins of gram-positive bacteria: a new class of virulence factors. *Trends in Microbiology*, 10, 205-208.

Chimalapati, S., Cohen, J.M., Camberlein, E., MacDonald, N., Durmort, C., Vernet, T., Hermans, P.W.M., Mitchell, T. and Brown, J.S. (2012) Effects of deletion of the *Streptococcus pneumoniae* lipoprotein diacylglyceryl transferase gene lgt on ABC transporter function and on growth *In vivo*. *PLoS One*, **7**, e41393.

Cho, S., Law, J. and N.G, C. (2009) Effect of Growth at Sub-lethal Concentrations of Kanamycin on the Cell Membrane Integrity and Amount of Capsular Glucuronic Acid in Wild-type *Escherichia coli* and Strain with a cpsB Mutation. *Journal of Experimental Microbiology and Immunology*, 13**,** 29-35.

Clore G.M. and Venditti V. (2013) Structure, dynamics and biophysics of the cytoplasmic protein-protein complexes of the bacterial phosphoenolpyruvate: sugar phosphotransferase system.*Trends in Biochemical Sciences*, 38, 515-530.

Cook, L.C. and Federle, M.J. (2014) Peptide pheromone signaling in *Streptococcus* and *Enterococcus*. *FEMS Microbiology Reviews*, 38, 473-492.

Cook L.C., LaSarre, B. and Federle E, M. J. (2013) Interspecies communication among commensal and pathogenic streptococci. *MBio,* 4**,** e00382-13.

Cuevas, R., Eutsey, R., Kadam, A., West-Roberts, J.A., Woolford, C.A., Mitchell, A.P., Mason, K.M. and Hiller, N.L. (2017) A novel streptococcal cell-cell communication peptide promotes pneumococcal virulence and biofilm formation. *Molecular Microbiology*, 105, 554-571.

Cvitkovitch, D.G., Li, Y.H. and Ellen, R.P. (2003) Quorum sensing and biofilm formation in *Streptococcal infections*. *The Journal of Clinical Investigation*, 112, 1626-1632.

Dahesh, S., Hensler, M.E., Van Sorge, N.M., Gertz, R.E. Jr., Schrag, S., Nizet. V. and Beall, B.W. (2008) Point mutation in the group B streptococcal pbp2x gene conferring decreased susceptibility to beta-lactam antibiotics. *Antimicrobial Agents and Chemotherapy*, 52, 2915-2918.

Davidson, A.L., Dassa, E., Orelle, C. and Chen, J. (2008) Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiology and Molecular Biology Reviews*, 72, 317-364, table of contents.

Declerck, N., Bouillaut, L., Chaix, D., Rugani, N., Slamti, L., Hoh, F., Lereclus, D. and Arold, S.T. (2007) Structure of PlcR: Insights into virulence regulation and evolution of quorum sensing in Gram-positive bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 18490-18495.

de Saizieu, A., Gardes, C., Flint, N., Wagner, C., Kamber, M., Mitchell, T.J., *et al.* (2000) Microarray-based identification of a novel *Streptococcus pneumoniae* regulon controlled by an autoinduced peptide. *Journal of Bacteriology*, 182, 4696-4703.

Dowson, C., Coffey, T. and Spratt, B. (1994) Origin and molecular epidemiology of penicillin‐binding‐proteinmediated resistance to β-lactam antibiotics. *Trends in Microbiology*, 2, 361-366 (1994).

Eze, U.A., Mustapha, A. and Nworie, A. (2013) *Streptococcus pneumoniae* virulence factor and their role in pathogenesis. *Continental Journal of Medical Research*, 7, 27-45.

Falkenhorst, G., Remschmidt, C., Harder, T., Hummers-Pradier, E., Wichmann, O. and Bogdan, C. (2017) Effectiveness of the 23-valent pneumococcal polysaccharide vaccine (PPV23) against pneumococcal disease in the elderly: systematic review and meta-analysis. *PLoS One*, 12, e0169368.

Fedson, D.S. (2014) Preventing non bacteremic pneumococcal pneumonia in older adults: historical background and considerations for choosing between PCV13 and PPV23. *Human Vaccines & Immunotherapeutics*, 10, 1322-1330.

Felmingham, D. (2004) Comparative antimicrobial susceptibility of respiratory tract pathogens. *Chemotherapy*, 50 Suppl 1, 3-10.

Fischetti, V.A. (2016) M protein and other surface proteins on streptococci. *Streptococcus pyogenes* Basic Biology to Clinical Manifestations (Ferretti, J.J., Stevens, D.L. and Fischetti, V.A., eds). *University of Oklahoma Health Sciences Center, Oklahoma City*, 27-53.

Fleuchot, B., Gitton, C., Guillot, A., Vidic, J., Nicolas, P., Besset, C., Fontaine, L., Hols, P., Leblond-Bourget, N., Monnet, V. and Gardan, R. (2011) Rgg proteins associated with internalized small hydrophobic peptides: anew quorum-sensing mechanism in streptococci. *Molecular Microbiology*, 80, 1102-1119.

Garcia, P., Gonzalez, M.P., Garcia, E., Lopez, R. and Garcia, J. (1999) LytB, a novel pneumococcal murein hydrolase essential for cell separation. *Molecular Microbiology*, 31, 1275-1281.

Garner, M. and Revzin, A. (1986) The use of gel electrophoresis to detect and study nucleic acid-protein interactions. *Trends in Biochemical Sciences*, 11, 395-396.

Gaspar, P., Al-Bayati, F.A.Y., Andrew, P.W., Neves, A.R. and Yesilkaya, H. (2014) Lactate dehydrogenase is the key enzyme for pneumococcal pyruvate metabolism and pneumococcal survival in blood. *Infection and Immunity*, 82, 5099-109.

Giammarinaro, P. and Paton, J.C. (2002) Role of RegM, a homologue of the catabolite repressor protein CcpA, in the virulence of *Streptococcus pneumoniae*. *Infection and Immunity*,70, 5454-5461.

Gohar, M., Faegri, K., Perchat, S., Ravnum, S., Økstad ,O.A., Gominet, M., Kolstø, A-B. and Lereclus, D. (2008) The PlcR regulon of *Bacillus cereus*. *PLoS ONE*, 3,1-9.

Gominet, M., Slamti, L., Gilois, N., Rose, M. and Lereclus, D. (2001) Oligopeptide permease is required for expression of the *Bacillus thuringiensis* PlcR regulon and for virulence. *Molecular Microbiology*, 40, 963-975.

Goncalves Mdos, S., Delattre, C., Balestrino, D., Charbonnel, N., ElboutachfaitiL, R., Wadouachi, A., Badel, S., Bernardi, T., MichaudI, P. and Forestier, C. (2014) Anti-biofilm activity: a function of *Klebsiella pneumoniae* capsular polysaccharide. *PLoS One*, 9, e99995.

Gorke, B. and Stulke, J. (2008) Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. Nature Rev. *Microbiology*, 6, 613-624.

Goerke, C., Bayer, M.G. and Wolz, C. (2001) Quantification of bacterial transcripts during infection using competitive reverse transcription-PCR (RT-PCR) and LightCycler RT-PCR. *Clinical and Diagnostic Laboratory Immunology*, 8, 279-282.

Grenha, R., Slamti, L., Nicaise, M., Refes, Y., Lereclus, D., and Nessler, S. (2013) Structural basis for the activation mechanism of the PlcR virulence regulator by the quorum-sensing signal peptide PapR. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 1047-1052.

Grossiord, B., Vaughan, E.E., Luesink, E. and Vos, W.M. de. (1998) Genetics of galactose utilisation via the Leloir pathway in lactic acid bacteria. *Le Lait*, 78, 77-84.

Guenzi, E., Gasc, A.M., Sicard, M.A. and Hakenbeck, R. (1994) A twocomponent signal transducing system is involved in competence and penicillin susceptibility in laboratory mutants of *Streptococcus pneumoniae*. *Molecular Microbiology*, 12, 505-515.

Guiral, S., Hénard, V., Laaberki, M.H., Granadel, C., Prudhomme, M., Martin, B. and Claverys, J.P. (2006) Construction and evaluation of a chromosomal expression platform (CEP) for ectopic, maltose-driven gene expression in *Streptococcus pneumoniae*. *Microbiology*, 152, 343-349.

Halfmann, A., Hakenbeck, R. and Brückner, R. (2007) A new integrative reporter plasmid for *Streptococcus pneumoniae*. *FEMS Microbiology Letters*, 268, 217-224.

Halfmann, A., Schnorpfeil, A., Müller, M., Marx, P., Günzler, U., Hakenbeck, R. and Brückner, R. (2011) A ctivity of the Two-Component Regulatory System CiaRH in *Streptococcus pneumoniae* R6. *Journal of Molecular Microbiology and Biotechnology*, 20, 96-104.

Hampton, L.M, Farley, M.M, Schaffner, W., Thomas, A., Reingold, A., Harrison, L.H., Lynfield, R., Bennett, N.M., Petit, S., Gershman, K., Baumbach, J., Beall, B., Jorgensen, J., Glennen, A., Zell, E.R. and Moore, M. (2012) Prevention of antibiotic-nonsusceptible *Streptococcus pneumoniae* with conjugate vaccines. *The Journal of Infectious Diseases*, 205, 401-411.

Hansman, D. and Bullen, M. (1967) A resistant pneumococcus. *The Lancet*, 2, 264-265.

Hansman, D., Glasgow, H. N., Sturt, J., Devitt, L. and Douglas, R. M. (1971) Pneumococci insensitive to penicillin. *Nature*, 230, 407-408.

Hanna-Wakim, R., Chehab, H., Mahfouz, I., Nassar, F., Baroud, M., Shehab, M., Pimentel, G., Wasfy, M., House, B., Araj, G., Matar, G., Dbaibo, G. and Lebanese Inter-Hospital Pneumococcal Surveillance Program. (2012) Epidemiologic characteristics, serotypes, and antimicrobial susceptibilities of invasive *Streptococcus pneumoniae* isolates in a nationwide surveillance study in Lebanon. *Vaccine*, 30, G11-G17.

Havarstein, L.S., Coomaraswamy, G. and Morrison, D.A. (1995a) An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae*. *Proceedings of the National Academy of Sciences of the United States of America*, 92, 11140-11144.

Havarstein, L.S., Diep, D.B. and Nes, I.F. (1995b) A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. *Molecular Microbiology*, 16, 229-240.

Hellman, L. and Fried, M. (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nature Protocols*, 2, 1849-1861.

Henrichsen, J. (1995) Six newly recognized types of *Streptococcus pneumoniae*. *Journal of Clinical Microbiology*, 33, 2759-2762.

Hendriksen, W.T., Bootsma, H.J., Diepen van, A., Estevao, S., Kuipers, O.P., Groot de, R. and Hermans, P.W.M. (2009) Strain-specific impact of PsaR of *Streptococcus pneumoniae* on global gene expression and virulence. *Microbiology*, 155, 1569-1579.

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

Hirst, R. A., Yesilkaya, H., Clitheroe, E., Rutman, A., Dufty, N., Mitchell, T. J., O'callaghan, C. and Andrew, P.W. (2002) Sensitivities of human monocytes and epithelial cells to pneumolysin are different. *Infection and Immunity*, 70**,** 1017-1022.

Ho, C.H., and Lin, T.Y. (2005) Pneumococcal vaccines. *Chang Gung Medical Journal*, 28, 765-772.

Hoch, J.A. (2000) Two-component and phosphorelay signal transduction. *Current Opinion in Microbiology*, 3, 165-170.

Holmes, A., McNab, R., Millsap, K.W., Rohde, M., Hammerschmidt, S., Mawdsley, J.L. and Jenkinson, H.F. (2001) The pavA gene of *Streptococcus pneumoniae* encodes a fibronectin-binding protein that is essential for virulence. *Molecular Microbiology*, 41, 1395-1408.

Honeyman, A.L., Cote, C.K. and Curtiss, R. (2002) Construction of transcriptional and translational lacZ gene reporter plasmids for use in *Streptococcus mutans*. *Journal of Microbiological Methods*, 49, 163-171.

Hondorp, E. R. and McIver, K.S. (2007) The Mga virulence regulon: infection where the grass is greener. *Molecular Microbiology*, 66l, 1056-1065.

Hondorp, E.R, Hou, S.C., Hause, L.L., Gera, K., Lee, C.E. and McIver, K.S. (2013) PTS phosphorylation of Mga modulates regulon expression and virulence in the group A *Streptococcus*. *Molecular Microbiology*, 88, 1176-1193.

Hoover, S.E., Perez, A.J., Tsui, H.-C.T., Sinha, D., Smiley, D. L., DiMarchi R.D., Winkler, ME. and Lazazzera, B.A. (2015) A new quorum-sensing system (TprA/PhrA) for *Streptococcus pneumoniae* D39 that regulates a lantibiotic biosynthesis gene cluster. *Molecular Microbiology*, 97, 229-243.

Horton, R. (1995) PCR-mediated recombination and mutagenesis. SOEing together tailor-made genes. *Molecular Biotechnology*, 3, 93-99.

Horton, R., Ho, S., Pullen, J., Hunt, H., Cai, Z. and Pease, L. (1993) Gene splicing by overlap extension. *Methods in Enzymology*, 217, 270-279.

Hoskins, J., Alborn Jr, W.E., Arnold, J., Blaszczak, L.C., Burgett , S., DeHoff, B.S., Estrem, S.T., *et al.* (2001) Genome of the bacterium *Streptococcus pneumoniae* strain R6. *Journal of Bacteriology*, 183, 5709-5717.

Hostetter, M.K. (1986) Serotypic variations among virulent pneumococci in deposition and degradation of covalently bound C3b: implications for phagocytosis and antibody production. *The Journal of Infectious Diseases*, 153, 682-693.

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

Huang, S.S., Johnson, K.M., Ray, G.T., Wroe, P., Lieu, T.A., Moore, M.R., Zell, E.R., Linder, J.A., Grijalva, C.G., Metlay, J.P. and Finkelstein, J.A. (2011) Healthcare utilization and cost of pneumococcal disease in the United States. *Vaccine*, 29, 3398-3412.

Ibrahim, Y.M., Kerr, A.R., Mc Cluskey, J. and Mitchell, T.J. (2004) Control of virulence by the two-component system CiaR/H is mediated via HtrA, amajor virulence factor of *Streptococcus pneumoniae*. *Journal of Bacteriology*, 186, 5258-5266.

Ibrahim, M., Guillot, A., Wessner, F., Algaron, F., Besset, C., Courtin, P., Gardan, R. and Monnet, V. (2007a) Control of the Transcription of a Short Gene Encoding a Cyclic Peptide in *Streptococcus thermophilus*: a New Quorum-Sensing System? . *Journal of Bacteriology*, 189, 8844-8854.

Ibrahim, M., Nicolas, P., Bessieres, P., Bolotin, A., Monnet, V. and Gardan, R. (2007b) A genome-wide survey of short coding sequences in streptococci. *Microbiology*, 153, 3631-3644.

Irwin, C.R., Schor, S.L. and Ferguson, M.W. (1994) Effects of cytokines on gingival fibroblasts *In vitro* are modulated by the extracellular matrix. *Journal of Periodontal Research*, 29, 309-317.

Iyer, R. and Camilli, A. (2007) Sucrose metabolism contributes to *In vivo* fitness of *Streptococcus pneumoniae*. *Molecular Microbiology*, 66, 1-13.

Iyer, R., Baliga, N.S. and Camilli, A. (2005) Catabolite control protein A (CcpA) contributes to virulence and regulation of sugar metabolism in *Streptococcus pneumoniae*. *Journal of* Bacteriology, 187, 8340-8349.

Jansen, A.G.S.C., Rodenburg, G.D., de Greeff, S.C, Hak, E., Veenhoven, R.H., Spanjaard, L., Schouls, L.M., Sanders, E.A. and van der Ende, A. (2009) Invasive pneumococcal disease in the Netherlands: syn-dromes, outcome and potential vaccine benefits. *Vaccine*, 27, 2394-2401.

Jansson, P.E., Lindberg, B., Anderson, M., Lindquist, U. and Henrichsen, J. (1988) Structural studies of the capsular polysaccharide from *Streptococcus pneumoniae* type 2, a reinvestigation. *Carbohydrate Research*, 182, 111-117.

Jedrzejas, M.J. (2001) Pneumococcal virulence factors: structure and function. *Microbiology and Molecular Biology Reviews*, 65, 187-207.

Jeong, J.K., Kwon, O., Lee, Y.M., Oh, D.B., Lee, J.M., Kim, S., Kim, E.H., Le, T.N., Rhee, D.K. and Kang, H.A. (2009) Characterization of the *Streptococcus pneumoniae* BgaC protein as a novel surface beta-galactosidase with specific hydrolysis activity for the Galbeta1-3GlcNAc moiety of oligosaccharides. *Journal of Bacteriology*, 191, 3011-3023.

Jiang, M., Shao, W., Perego, M., and Hoch, J.A. (2000) Multiple histidine kinases regulate entry into stationary phase and sporulation in *Bacillus subtilis*. *Molecular Microbiology*, 38, 535-542.

Jing, D., Agnew, J., Patton, W. F., Hendrickson, J. and Beechem, J. M. (2003) A sensitive two‐color electrophoretic mobility shift assay for detecting both nucleic acids and protein in gels. *Proteomics*, 3, 1172-1180.

Johnston, R.B.Jr. (1991) Pathogenesis of pneumococcal pneumonia. *Reviews of Infectious Diseases*, 13, S509-S517.

Junges, R., Salvadori, G., Shekhar, S., Amdal, H. A., Periselneris, J. N., Chen, T., Brown, J.S. and Petersen, F.C. (2017) A quorum-sensing system that regulates *Streptococcus pneumoniae* biofilm formation and surface polysaccharide production. *mSphere*, 2, e00324-17.

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

Kahya, H.F., Andrew, P.W. and Yesilkaya, H. (2017) Deacetylation of sialic acid by esterases potentiates pneumococcal neuraminidase activity for mucin utilization, colonization and virulence. *PLoS Pathogens*, 13, e1006263. doi:10.1371/ journal.ppat.1006263.

Kappeler, K. V., Anbalagan, S., Dmitriev, A. V., Mcdowell, E. J., Neely, M. N. and Chaussee, M. S. (2009) A naturally occurring Rgg variant in serotype M3 *Streptococcus pyogenes* does not activate speB expression due to altered specificity of DNA binding. *Infection and Immunity*, 77, 5411-5417.

Kawamura, Y., Hou, XG., Sultana, F., Miura, H. and Ezaki, T. (1995) Determination of 16S rRNA sequences of *Streptococcus mitis* and *Streptococcus gordonii* and phylogenetic relationships among members of the genus *Streptococcus*. *International Journal of Systematic and Evolutionary Microbiology*, 45, 406-408.

Kelly, P.S., Jiaqi, J., Roy, R.L., Sherman, M.W. and Yuval, K. (2017) Ritornello: high fidelity control-free chromatin immunoprecipitation peak callin. *Nucleic Acids Research*, gkx799.

Kerr, A.R., Paterson, G.K., McCluskey, J., Iannelli, F., Oggioni, M.R., Pozzi, G. and Mitchell1, T.J. (2006) The contribution of PspC to pneumococcal virulence varies between strains and is accomplished by both complement evasion and complement independent mechanisms. *Infection and Immunity*, 74, 5319-5324.

King, S.J. (2010) Pneumococcal modification of host sugars: a major contributor to colonization of the human airway? *Molecular Oral Microbiology*, 25, 15-24.

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

Kirkham, L.A., Kerr,A.R., Douce,R.G., Paterson, G.K., Dilts, D.A., Liu, D. and Mitchell1, T.J. (2006) Construction and immunological characterization of a novel nontoxic protective pneumolysin mutant for use in future pneumococcal vaccines. *Infection and Immunity*, 74, 586-593.

Kloosterman, T.G., Bijlsma, J.J.E., Kok, J. and Kuipers, O.P. (2006a) To have neighbour's fare: extending the molecular toolbox for *Streptococcus pneumoniae*. *Microbiology*, 152, 351-359.

Kloosterman, T.G., Hendriksen, W.T., Bijlsma, J.J.E., Bootsma, H.J., Hijum, S.A.F.T. van, Kok, J., Hermans, P.W.M. and Kuipers, O.P. (2006b) Regulation of glutamine and glutamate metabolism by GlnR and GlnA in *Streptococcus pneumoniae*. *The Journal of Biological Chemistry*, 281, 25097-25109.

Koedel, U., Scheld, W.M. and Pfister, H.W. (2002) Pathogenesis and pathophysiology of pneumococcal meningitis. *The Lancet. Infectious Diseases*, 2, 721-736.

Koetje, E.J., Hajdo-Milasinovic, A., Kiewiet, R., Bron, S. and Tjalsma, H. (2003) A plasmid-borne Rap–Phr system of Bacillus subtilis can mediate cell-density controlled production of extracellular proteases. *Microbiology*, 149, 19-28.

Konings, W.N., Poolman, B. and Veen, H.W. van. (1994) Solute transport and energy transduction in bacteria. *Antonie van Leeuwenhoek*, 65, 369-380.

Kostyukova, N.N., Volkova, M.O., Ivanova,V.V. and Kvetnaya, A.S. (1995) Astudy of pathogenic factors of *Streptococcus pneumoniae* strains causing meningitis. *FEMS Immunology and Medical Microbiology*, 10, 133-137.

Kozlowicz, B.K., Dworkin, M. and Dunny, G.M. (2006) Pheromone-inducible conjugation in *Enterococcus faecalis*: a model for the evolution of biological complexity? *International Journal of Medical Microbiology*, 296, 141-147.

Kreikemeyer, B., McIver, K.S. and Podbielski, A. (2003) Virulence factor regulation and regulatory networks in *Streptococcus pyogenes* and their impact on pathogen-host interactions. Trends in Microbiology, 11, 224-232.

Lai, Y.-C., Peng, H.-L. and Chang, H.-Y. (2003) RmpA2, an activator of capsule biosynthesis in Klebsiella pneumoniae CG43, regulates K2 cps gene expression at the transcriptional level. *Journal of Bacteriology*, 185**,** 788-800.

Lange, R., Wagner, C., de Saizieu, A., Flint, N., Molnos, J., Stieger, M., Caspers, P., Kamber, M., Keck,W., and Amrein, K.E. (1999) Domain organization and molecular characterization of 13 two-component systems identified by genome sequencing of *Streptococcus pneumoniae*. *Gene*, 237, 223-234.

Langvas-Nielsen, A. (1944) Fermentation power of pneumococci. *Acta Pathologica Microbiologica Scandinavica*, 21, 370-373.

Lasarre, B., Aggarwal, C. and Federle, M.J. (2013) Antagonistic Rgg regulators mediate quorum sensing via competitive DNA binding in *pyogenes*. *mBio*, 3, e00333-12.

Laux, A., Sexauer, A., Sivaselvarajah, D., Kaysen, A. and Brückner, R. (2015) Control of competence by related non-coding csRNAs in *Streptococcus pneumoniae* R6. *Frontiers in Genetics*, 6, 246. https://doi.org/10.3389/fgene.2015.00246.

Lawrence, S. L., Feil, S. C., Morton, C. J., Farrand, A. J., Mulhern, T. D., Gorman, M. A., Wade, K. R., Tweten, R. K. and Parker, M. W. (2015) Crystal structure of *Streptococcus pneumoniae* pneumolysin provides key insights into early steps of pore formation. *Scientific Reports*, 5, 14352.

Lereclus, D., Agaisse, H., Gominet, M., Salamitou, S. and Sanchis, V. (1996) Identification of a *Bacillus thuringiensis* gene that positively regulates transcription of the phosphatidylinositol-specific phospholipase c gene at the onset of the stationary phase. *Journal of Bacteriology*, 178, 2749-2756.

Li˜nares, J., Ardanuy, C., Pallares, R. and Fenoll. A. (2010) Changes in antimicrobial resistance, serotypes and genotypes in *Streptococcus pneumoniae* over a 30-year period. *Clinical Microbiology and Infection*, 16, 402-410.

Livak, K. and Schmittgen, T. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C (T)) Method. *Methods*, 25, 402-408.

Long, Y., Hou, Y., Feng, X., Yang, D., Ding, C. and Zhen, Y. (2012) Research progress of *Streptococcus pneumoniae* capsular polysaccharide-protein conjugate vaccines. *African Journal of Microbiology Research*, 6, 1364-1369.

Loo, C.Y., Corliss, D.A. and Ganeshkumar, N. (2000) *Streptococcus gordonii* biofilm formation: identification of genes that code for biofilm phenotypes. *Journal of Bacteriology*, 182, 1374-1382.

Lysenko, E., Ratner, A., Nelson, A. and Weiser, J. (2005) The role of innate immune responses in the outcome of interspecies competition for colonization of mucosal surfaces. *PLoS Pathogens*, 1, 1-e1.

MacLeod, C.M. and Krauss, M.R. (1950) Relation of virulence of pneumococcal strains for mice to the quantity of capsular polysaccharide formed *In vitro*. *The Journal of Experimental Medicine*. 92, 1-9.

Manco, S., Hernon, F., Yesilkaya, H., Paton, J.C., Andrew, P.W. and Kadioglu1, A. (2006) Pneumococcal neuraminidases A and B both have essential roles during infection of the respiratory tract and sepsis. *Infection and Immunity*, 74, 4014-4020.

Mashburn-Warren, L., Morrison, D.A. and Federle, M.J. (2012) The cryptic competence pathway in *Streptococcus pyogenes* is controlled by a peptide pheromone. *Journal of Bacteriology*, 194, 4589-4600.

Marion, C., Burnaugh, A.M., Woodiga, S.A. and King, S.J. (2011) Sialic acid transport contributes to pneumococcal colonization. *Infection and Immunity*, 79, 1262-1269.

Marion, C., Stewart, J.M., Tazi, M.F., Burnaugh, A.M., Linke, C.M., Woodiga, S.A. and King, S.J. (2012) *Streptococcus pneumoniae* can utilize multiple sources of hyaluronic acid for growth. *Infection and Immunity*, 80, 1390-1398.

Marks, L.R., Parameswaran, G.I., and Hakansson, A.P (2012a) Pneumococcal interactions with epithelial cells are crucial for optimal biofilm formation and colonization *In vitro* and *In vivo*. *Infection and Immunity*, 80, 2744-2760.

Marongiu, A., Hasan, O., Ali, A., Bakhsh, S., George, B., Irfan, N., Minelli, C., Canova, C., Schofield, S., De Matteis, S. and Cullinan, P. (2016) Are welders more at risk of respiratory infections? Findings from a cross-sectional survey and analysis of medical records in shipyard workers: the WELSHIP project. *Thorax*, 71, 581-582.

Martinón-Torres, F., Czajka, H., Center, K.J., Wysocki, J., Majda-Stanislawska, E., Ome~naca, F., Bernaola I. E., Blazquez, G. D., Concheiro-Guis_an, A., Gimenez-Sanchez, F., Szenborn, L., Giardina, P.C., Patterson, S., Gruber, W.C., Scott, D.A. and Gurtman, A. (2015) 13-Valent Pneumococcal ConjugateVaccine (PCV13) in Preterm Versus Term Infants. *Pediatrics*, 135, e876-886.

McCuller, J.A. and Tuomanen, E.I. (2001) Molecular pathogenesis of pneumococcal pneumonia. *Frontiers of Bioscience*, 6, D877-889.

McIver, K.S. (2009) Stand-alone response regulators controlling global virulence networks in *Streptococcus pyogenes*. *Contributions to Microbiology*, 16, 103-119.

Meyers, J.A., Sanchez, D., Elwell, L.P. and Falkow, S. (1976) Simple agarose gel electrophoretic method for the identification and characterization of plasmid deoxyribonucleic acid. *Journal of Bacteriology*, 127, 1529-1537.

Miles, A., Misra, S. and Irwin, J. (1938) The estimation of the bactericidal power of the blood. *The Journal of Hygiene*, 38, 732-749.
Miller, J.H. (1972) Experiments in molecular genetics. *Cold Spring Harbor Laboratory*, pp. 352-355.

Miller, M.B. and Bassler, B.L. (2001) Quorum sensing in bacteria. *Annual Review of Microbiology*, 55, 165-199.

Mitchell, A.M., and Mitchell, T.J. (2010) *Streptococcus pneumoniae*. virulence factors and variation. *Clinical Microbiology and Infection*, 16, 411-418.

Mitchell, T.J. (2000) Virulence factors and the pathogenesis of disease caused by *Streptococcus pneumoniae*. *Research in Microbiology*, 151, 413-419.

Mohawk, K. L. and O'Brien, A. D. (2011) Mouse models of Escherichia coli O157: H7 infection and shiga toxin injection. *BioMed Research International*, 2011.

Morona, J.K., Paton, J. C., Miller, D.C. and Morona, R. (2000) Tyrosine phosphorylation of CpsD negatively regulates capsular polysaccharide biosynthesis in *Streptococcus pneumoniae*. *Molecular Microbiology*, 35, 1431-1442.

Motib, A., Guerreiro, A., Al-Bayati, F., Manzoor, I., Shafeeq, S., Kadam, A., Kuipers, O., Hiller, L., Cowen, T., Piletsky, S., Andrew, P.W. and Yesilkaya, H. (2017) Modulation of quorum sensing in a Gram-Positive pathogen by linear molecularly imprinted polymers with anti-infective properties. *Angewandte Chemie International Edition*, 56, 16555 -16558.

Morton, N. and Swartz, M.D. (2004) Bacterial Meningitis-A View of the Past 90 Years. *The New England Journal of Medicine*, 351**,** 1826-1828.

Musher, D.M. and Thorner, A.R. (2014) Community-acquired pneumonia. *The New England Journal of Medicine*, 371, 1619-1628.

Neely, M.N., Lyon, W.R., Runft, D.L. and Caparon, M. (2003) Role of RopB in growth phase expression of the SpeB cysteine protease of *Streptococcus pyogenes*. *Journal of Bacteriology*, 185, 5166-5174.

Neidhardt, F., Ingraham, J. and Schaechter, S. (1990) Physiology of the Bacterial Cell: A Molecular Approach. *Sinauer Associates Inc*, Sunderland.

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

Neves, A.R., Pool, W.A., Solopova, A., Kok, J., Santos, H. and Kuipers, O.P. (2010) Towards enhanced galactose utilization by *Lactococcus lactis*. *Applied and Environmental Microbiology*, 76, 7048-7060.

Normark, H. and Normark, S. (2002) Antibiotic tolerance in pneumococci. *Clinical Microbiology and Infection*, 8, 613-622.

O'Brien, K.L., Wolfson, L.J., Watt, J.P., Henkle, E., Deloria-Knoll, M., McCall, N., Lee, E., Mulholland, K., Levine, O.S. and Cherian, T. (2009) Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years:global estimates. *Lancet*, 374, 893-902.

Paixão, L., Oliveira, J., Veríssimo, A., Vinga, S., Lourenço, E.C., Ventura, M.R., Kjos, M., Veening, J.-W., Fernandes, V.E., Andrew, P.W., Yesilkaya, H. and Neves, A.R. (2015) Host glycan sugar-specific pathways in *Streptococcus pneumoniae*: galactose as a key sugar in colonisation and infection. *PloS One*, 10, e0121042.

Parashar, V., Aggarwal, C., Federle, M.J. and Neiditch, M.B. (2015) Rgg protein structure-function and inhibition by cyclic peptide compounds. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 5177-5182.

Park, I.H., Park, S., Hollingshead, S.K., and Nahm, M.H. (2007) Genetic basis for the new pneumococcal serotype, 6C. *Infection and* Immunity, 75, 4482-4489.

Paterson, G.K., Blue, C.E. and Mitchell, T.J. (2006) Role of two-component systems in the virulence of *Streptococcus pneumoniae*. *Journal of Medical Microbiology*, 55, 355-363.

Perez-Pascual, D., Gaudu, P., Fleuchot, B., Besset, C., Rosinski-C hupin, I., Guillot, A., Monneta,V. and Gardana, R. (2015) RovS and Its Associated Signaling Peptide Form a Cell-To-Cell Communication System Required for *Streptococcus agalactiae* Pathogenesis. *mBio*, 6, e02306-14.

Pestova, E.V., Håvarstein, L.S. and Morrison, D.A. (1996) Regulation of competence for genetic transformation in *Streptococcus pneumoniae* by an auto-induced peptide pheromone and a two-component regulatory system. *Molecular Microbiology*, 21, 853-862.

Perego, M., Higgins, C.F., Pearce, S. R., Gallagher, M.P. and Hoch, J. A. (1991) The oligopeptide transport system of *Bacillus subtilis* plays a role in the initiation of sporulation. *Molecular Microbiology*, 5, 173-185.

Perego, M. (1997) A peptide export–import control circuit modulating bacterial development regulates protein phosphatases of the phosphorelay. *Proceedings of the National Academy of Sciences of the United States of America*, 94, 8612- 8617.

Perry, J.A., Cvitkovitch, D.G. and Levesque, C.M. (2009) Cell death in *Streptococcus mutans* biofilms: a link between CSP and extracellular DNA. *FEMS Microbiology Letters*, 299, 261-266.

Pessi, G., Blumer, C. and Haas, D. (2001) lacZ fusions report gene expression, don't they? *Microbiology*, 147**,** 1993-1995.

Pettigrew, M.M., Fennie, K.P., York, M.P., Daniels, J. and Ghaffar, F. (2006) Variation in the presence of neuraminidase genes among *Streptococcus pneumoniae* isolates with identical sequence types. *Infection and Immunity*, 74, 3360-3365.

Philips, B.J., Meguer, J.X., Redman, J. and Baker, E.H. (2003) Factors determining the appearance of glucose in upper and lower respiratory tract secretions. *Intensive Care Medicine*, 29, 2204-2210.

Postma, P.W., Lengeler, J.W. and Jacobson, G.R. (1993) Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiological Reviews*, 57, 543-594.

Pottathil, M., and Lazazzera, B.A. (2003) The extracellular Phr peptide-Rap phosphatase signaling circuit of *Bacillus subtilis*. *Frontiers in Bioscience*, 8, d32-d45.

Price, K. E. and Camilli, A. (2009) Pneumolysin localizes to the cell wall of *Streptococcus pneumoniae*. *Journal of Bacteriology*, 191, 2163-2168.

Price, C.E., Zeyniyev, A., Kuipers, O.P. and Kok, J. (2012) From meadows to milk to mucosa - adaptation of *Streptococcus* and *Lactococcus* species to their nutritional environments. *FEMS Microbiology Reviews*, 36, 949-971.

Pulliainen, A. T., Hytonen, J., Haataj, S. and Finne, J. (2008) Deficiency of the Rgg regulator promotes H_2O_2 resistance, AhpCF-mediated H_2O_2 decomposition, and virulence in Streptococcus pyogenes. *Journal of Bacteriology*, 190, 3225-3235.

Qi, F., Chen, P. and Caufield, P.W. (1999) Functional analyses of the promoters in the lantibiotic mutacin II biosynthetic locus in *Streptococcus mutans*. *Applied and Environmental Microbiology*, 65, 652-658.

Regev-Yochay, G., Trzcinski, K., C. M. Thompson, Lipsitch, M. and Malley, R. (2007) SpxB is a suicide gene of *Streptococcus pneumoniae* and confers a selective advantage in an *In vivo* competitive colonization model. *Journal of Bacteriology*, 189, 6532-6539.

Reid, J.L., Iyer, V.R., Brown, P.O. and Struhl, K. (2000) Coordinate regulation of yeast ribosomal protein genes is associated with targeted recruitment of Esa1 histone acetylase. *Molecular Cell*, 6, 1297-1307.

Reinert, R.R., Reinert, S., Vander Linden, M., CIL, M.Y., Allahham, A. and Appelbaum, P. (2005) Antimicrobial susceptibility of *Streptococcus pneumoniae* in eight European countries from 2001 to 2003. *Antimicrobial Agents and Chemotherapy*, 49, 2903-2913.

Ren, B. and Szalai, A.J. (2003) Thomas O, Hollingshead SK, Briles DE. Both family 1 and family 2 PspA proteins can inhibit complement deposition and confer virulence to a capsular serotype 3 strain of *Streptococcus pneumoniae*. *Infection and Immunity*, 71, 75-85.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T.L., Wilson, C.J., Bell, S.P. and Young, R.A. (2000) Genome-wide location and function of DNA binding proteins. *Science*, 290, 2306-2309.

Reyrat, J.M., Pelicic, V., Gicquel, B. and Rappuoli, R. (1998) Counterselectable markers: untapped tools for bacterial genetics and pathogenesis. *Infection and Immunity*, 66, 4011-4017.

Richter, S.S. and Musher, D.M. (2017) The ongoing genetic adaptation of *Streptococcus pneumoniae*. *Journal of Clinical Microbiology*, 55, 681-685.

Rocha-Estrada, J., Aceves-Diez, A.E., Guarneros, G., and de la Torre, M. (2010) The RNPP family of quorum-sensing proteins in Gram-positive bacteria. *Applied Microbiology and Biotechnology*, 87, 913-923.

Rodgers, J. T., Patel, P., Hennes, J. L., Bolognia, S. L. and Mascotti, D. P. (2000) Use of biotin-labeled nucleic acids for protein purification and agarose-based chemiluminescent electromobility shift assays. *Analytical Biochemistry*, 277, 254-259.

Rosenfeld, R.M., Doyle, W.J., Swarts, J.D., Seroky, J. and Pinero, B.P. (1992) Third-generation cephalosporins in the treatment of acute pneumococcal otitis media. An animal study. *Archives of Otolaryngology–Head & Neck Surgery*, 118, 49-52.

Rose, M.C. and Voynow, J.A. (2006) Respiratory tract mucin genes and mucin glycoproteins in health and disease. *Physiological Reviews*, 86, 245-278.

Rubins, J. B., and E.N. Janoff. (1998) Pneumolysin: a multifunctional pneumococcal virulence factor. *Journal of Laboratory and Clinical Medicine*, 131, 21-27.

Saito, H. and Miura, K. (1963) Preparation of transforming deoxyribonucleic acid by phenol treatment. *Biochimica et Biophysica Acta*, 72, 619-629.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning, A Laboratory manual. Cold Spring Harbor Laboratory Press, *Cold Spring Harbor*, New York.

Samen, U.M., Eikmanns, B.J. and Reinscheid, D.J. (2006) Thetranscrip-tional regulator RovS controls the attachment of *Streptococcus agalactiae* to human epithelial cells and the expression of virulence genes. *Infection and Immunity*, 74, 5625-5635.

Sandgren, A., Albiger, B., Orihuela, C.J., Tuomanen, E., Nor-mark, S., and Henriques-Normark, B. (2005) Virulence in mice of pneumococcal clonal types with known invasive disease potential in humans. *The Journal of Infectious Diseases*, 192, 791-800.

Sanders, J.W., Leenhouts, K., Burghoorn, J., Brands, J.R., Venema, G. and Kok, J. (1998) A chloride-inducible acid resistance mechanism in Lactococcus lactis and its regulation. *Molecular Microbiology*, 27, 299-310.

Scalise, M., Pochini, L., Giangregorio, N., Tonazzi, A. and Indiveri, C. (2013) Proteoliposomes as tool for assaying membrane transporter functions and interactions with xenobiotics. *Pharmaceutics*, 5, 472-497.

Schneider, E. (2001) ABC transporters catalyzing carbohydrate uptake. *Research in Microbiology*, 152, 303-310.

Seitz, P. and Blokesch, M. (2014) DNA transport across the outer and inner membranes of naturally transformable *Vibrio cholerae* is spatially but not temporally coupled. *mBio*, 5, e01409-14.

Shaper, M., Hollingshead, S.K., Benjamin, W.H. and Briles, D.E. (2004) PspA protects Streptococcus pneumoniae from killing by apolactoferrin, and antibody to PspA enhances killing of pneumococci by apolactoferrin. *Infection and Immunity*, 72, 5031-5040.

Sheehan, J.K., Richardson, P.S., Fung, D.C., Howard, M. and Thornton, D.J. (1995) Analysis of respiratory mucus glycoproteins in asthma: a detailed study from a patient who died in status asthmaticus. *American Journal of Respiratory Cell and Molecular Biology*, 13, 748-756.

Sheehan, J.K., Thornton, D.J., Somerville, M. and Carlstedt, I. (1991) Mucin structure. The structure and heterogeneity of respiratory mucus glycoproteins. *The American Review of Respiratory Disease*, 144, S4-9.

Sleeman, K., Knox, K., George, R., Miller, E., Waight, P., Griffiths, D., Efstration, A., Broughton, K., Mayon-White, R. T., Moxon, E. R., Crook, D. W., Public Health Laboratory, S. and Oxford Pneumococcal Surveillance, G. (2001) Invasive pneumococcal disease in England and Wales: vaccination implications. *The Journal of Infectious Diseases*, 183**,** 239-246.

Solovyev, V. and Salamov, A. (2011) Automatic Annotation of Microbial Genomes and Metagenomic Sequences. In *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies* (ed. by Li, R.). *Nova Science Publishers*, pp. 61-78.

Somerville, G.A. and Proctor, R.A. (2009) At the crossroads of bacterial metabolism and virulence factor synthesis in Staphylococci. *Microbiology and Molecular Biology Reviews*, 73, 233-248.

Song, J.-H., Ko, K., Lee, J.-Y., Baek, J., Oh, W., Yoon, H., Jeong, J.-Y. and Chun, J. (2005) Identification of essential genes in *Streptococcus pneumoniae* by allelic replacement mutagenesis. *Molecules and Cells*, 19, 365-374.

Spellerberg, B., Cudell, D.R., Sandors, J., Peare, B.J., Idanpaan-Heikkila, I., Roseno, C. and Masure, H.R. (1996) Pyruvate oxidase, as a determinant of virulence in *Streptococcus pneumoniae*. *Molecular microbiology*, 19, 803-813.

Standish, A.J., Stroeher, U.H. and Paton, J.C. (2007) Pneumococcal Two-Component Signal Transduction System RR/HK06 Regulates CbpA and PspA by Two Distinct Mechanisms. *Journal of Bacteriology*, 189, 5591-5600.

Steinmoen, H., Knutsen, E. and Havarstein, L.S. (2002) Induction of natural competence in *Streptococcus pneumoniae* triggers lysis and DNA release from a subfraction of the cell population. *Proceedings of the National Academy of Sciences of the United States of America*, 99**,** 7681-7686.

Stewart, G., Wernisch, L., Stabler, R., Mangan, J., Hinds, J., Laing, K., Young, D. and Butcher, P. (2002) Dissection of the heat-shock response in *Mycobacterium tuberculosis* using mutants and microarrays. *Microbiology*, 148, 3129-3138.

Stock, A.M., Robinson,V.L. and Goudreau, P.N. (2000) Twocomponent signal transduction. *Annual Review of Biochemistry*, 69, 183-215.

Stoodley, P.,Sauer, K., Davies, D.G., and Costerton, J.W. (2002) Biofilmsas complex differentiated communities. *Annual Review of Microbiology*, 56, 187-209.

Sulavik, M., Tardif, G. and Clewell, D. (1992) Identification of a gene, rgg, which regulates expression of glucosyltransferase and influences the Spp phenotype of *Streptococcus gordonii* Challis. *Journal of Bacteriology*, 174, 3577-3586.

Tan, T. Q. (2012) Pediatric Invasive Pneumococcal Disease in the United States in the Era of Pneumococcal Conjugate Vaccines. *Clinical Microbiology Reviews*, 25, 409-419.

Talbot, U., Paton, A.W. and Paton, J.C. (1996) Uptake of *Streptococcus pneumoniae* by respiratory epithelial cells. *Infection and Immunity*, 64, 3772-3777.

Terra, V.S., Homer, K.A., Rao, S.G., Andrew, P.W. and Yesilkaya, H. (2010) Characterization of novel beta-galactosidase activity that contributes to glycoprotein degradation and virulence in *Streptococcus pneumoniae*. *Infection and Immunity*, 78, 348-357.

Terra, V.S., Zhi, X., Kahya, H.F., Andrew, P.W. and Yesilkaya, H. (2015) Pneumococcal 6-Phospho-beta-Glucosidase (BglA3) Is Involved in Virulence and Nutrient Metabolism. *Infection and immunity*, 84, 286-292.

Tettelin, H., Nelson, K.E., Paulsen, I.T., *et al*. (2001) Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science*, 293, 498-506.

Throup, J.P., Koretke, K.K., Bryant, A.P., Ingraham, K.A., Chalker, A.F., Marra, Y. Ge, A., Wallis, N.G., Brown, J.R., Holmes, D.J., Rosenberg, M. and Burnham, K.R. (2000) A genomic analysis of two-component signal transduction in *Streptococcus pneumoniae*. *Molecular Microbiology*, 35, 566-576.

Thurlow, L.R., Thomas, V.C., Narayanan, S., Olson, S., Fleming, S.D. and Hancock, L.E. (2010) Gelatinase contributes to the pathogenesis of endocarditis caused by *Enterococcus faecalis*. *Infection and Immunity*, 78, 4936-4943.

Toma, S., Del Blue, M., Pirola, A. and Grandi, G. (1986) *nprRl* and *nprR2* regulatory regions for neutral protease expression in *Bacillus subtilis*. *Journal of Bacteriology*, 167, 740-743.

Tong, H.H., Blue, L.E., James, M.A. and Demaria, T.F. (2000) Evaluation of the virulence of a *Streptococcus pneumoniae* neuraminidase-deficient mutant in nasopharyngeal colonization and development of otitis media in the chinchilla model. *Infection and immunity*, 68, 921-924.

Trappetti, C., McAllister, L.J., Chen, A., Wang, H., Paton, A.W., Oggioni, M.R., McDevitta, C.A. and Patona, J.C. (2017) Autoinducer 2 signaling via the phosphotransferase FruA drives galactose utilization by *Streptococcus pneumoniae*, resulting in Hypervirulence. *mBio*, 8, e02269-16.

Uehara, H., Yoneda, Y., Yamane, K. and Maruo, B. (1974) Regulation of neutral protease productivity in *Bacillus subtilis*: transformation of high protease productivity. *Journal of Bacteriology*, 119, 82-91.

van Deursen, A.M., van den Bergh, M.R., Sanders, E.A. and Carriage Pilot Study G. (2016) Carriage of *Streptococcus pneumoniae* in asymptomatic, community-dwelling elderly in the Netherlands. *Vaccine*, 34, 4-6.

von Heijne G. (1986) A new method for predicting signal sequence cleavage sites. *Nucleic Acids Research*, 14, 4683-4690.

Wani, J., Gilbert, J., Plaut, A. and Weiser, J. (1996) Identification, cloning and sequencing of the immunoglobulin A1 protease gene of *Streptococcus pneumoniae*. *Infection and Immunity*, 64, 3967-3974.

Weinberger, D.M., Dagan, R., Givon-Lavi, N., Regev-Yochay, G., Malley, R. and Lipsitch, M. (2008) Epidemiologic evidence for serotype-specifi c acquired immunity to pneumococcal carriage. *The Journal of Infectious Diseases*, 197, 1511-1518.

Wyllie, A.L., Van Henten, S.C., Duurland, C.L., Weinber, D.M., Sanders, E.A., Lipsitch, M. and Trzcinski, K. (2016) Significance of serotype on interactions between *Streptococcus pneumoniae* and the human host. *Molecular surveillance of pneumococcal carriage in all ages*.

Weycker, D., Strutton, D., Edelsberg, J., Sato, R. and Jackson, L.A. (2010) Clinical and economic burden of pneumococcal disease in older US adults. *Vaccine*, 28, 4955-4960.

WHO. International travel and health: pneumococcal disease. (2012) World Health Organization. Available from: http://www.who.int/ith/diseases/ pneumococcal/en/ [accessed 03.05.12].

WHO. Pneumonia. (2014) Vol. 2015. http://www.who.int/mediacentre/factsheets/fs331/en/ [accessed 09.07.15].

Wiggins, R., Hicks, S.J., Soothill, P.W., Millar, M.R. and CorfieldD, A.P. (2001) Mucinases and sialidases: their role in the pathogenesis of sexually transmitted infections in the female genital tract. *Sexually transmitted infections*, 77, 402-408.

Wilkening, R.V., Capodagli, G.C., Khataokar, A., Tylor, K.M., Neiditch, M.B. and Federle, M.J. (2017) Activating mutations in quorum-sensing regulator Rgg2 and its conformational flexibility in the absence of an intermolecular disulfide bond. *Journal of Biological Chemistry*, 292, 20544-20557.

Woodbury, C.P. and Hippel, P.H.Von. (1983) On the determination of deoxyribonucleic acid-protein interaction parameters using the nitrocellulose filterbindingassay. *Biochemistry*, 22, 4730-4737.

Wright, A.V. and Axelsson, L. (2012) Lactic Acid Bacteria: An Introduction. In *Lactic Acid Bacteria: Microbiological and Functional Aspects* (ed. by Salminen, S., Wright, A.V., Lahtinen, S. & Ouwehand, A.). *Taylor & Francis Group*, Boca Raton, pp. 2-14.

Yesilkaya, H., Spissu, F., Carvalho, S.M., Terra, V.S., Homer, K.A., Benisty, R., Porat, N., Neves, A.R. and Andrew, P.W. (2009) Pyruvate formate lyase is required for pneumococcal fermentative metabolism and virulence. *Infection and Immunity*, **77**, 5418-5427.

Yesilkaya, H. (1999) Studies on the role of superoxide dismutase (SOD) in the virulence of *Streptococcus pneumoniae* and the effects of interferon gamma on sensitivity of phagocytes to the toxin pneumolysin. PhD Thesis. *Department of Infection, Immunity and Inflammation,. University of Leicester*, UK.

Zhang, Q., Bagrade, L., Bernatoniene, J., Clarke, E., Paton, J.C., Mitchell, T.J., Nunez, D.A. and Finn, A. (2007) Low CD4 T cell immunity to pneumolysin is associated with nasopharyngeal carriage of pneumococci in children. *The Journal of Infectious Diseases*, 195, 1194-1202.

Zheng, F., Ji, H., Cao, M., Wang, C., Feng, Y., Li, M., Pan, X., Wang, J., Qin, Y., Hu, F. and Tang, J. (2011) Contribution of the Rgg transcription regulator to metabolism and virulence of *Streptococcus suis* serotype 2. *Infection and Immunity*, 79, 1319-1328.

Zhi, X., Abdullah, I.T., Gazioglu, O., Manzoor, I., Shafeeq, S., Kuipers, O.P., Hiller, N.L., Andrew, P.W. and Yesilkaya, H. (2018) Rgg-Shp regulators are important for pneumococcal colonization and invasion through their effect on mannose utilization and capsule synthesis. *Scientific Reports*, DO-10.1038/s41598-018-24910-1. UR-https://doi.org/10.1038/s41598-018-24910-1.

Conferences

1. I attended to the induction lectures in October /2014 and these are so useful for my study. The equipment in our department are good and helpful that help me to do my experiments well. We have a computer room in our lab that enables us to read new books and papers online. All department facilities were available and necessary to do my research.

2. I attended Society for General Microbiology Conference at Birmingham ICC.


Certificate of attendance

This certificate confirms that BUSHRA STILLA.

attended the SOCIETY FOR GENERAL MICROBIOLOGY CONFERENCE.

at BIRMINGHAM ICC.

Meeting Organiser SGM.

Date 30/3/2015 – 2/4/2015

SOCIETY FOR GENERAL MICROBIOLOGY

3. I attended Postgraduate Conference at University of Leicester, Maurice Shock Medical Sciences Building.

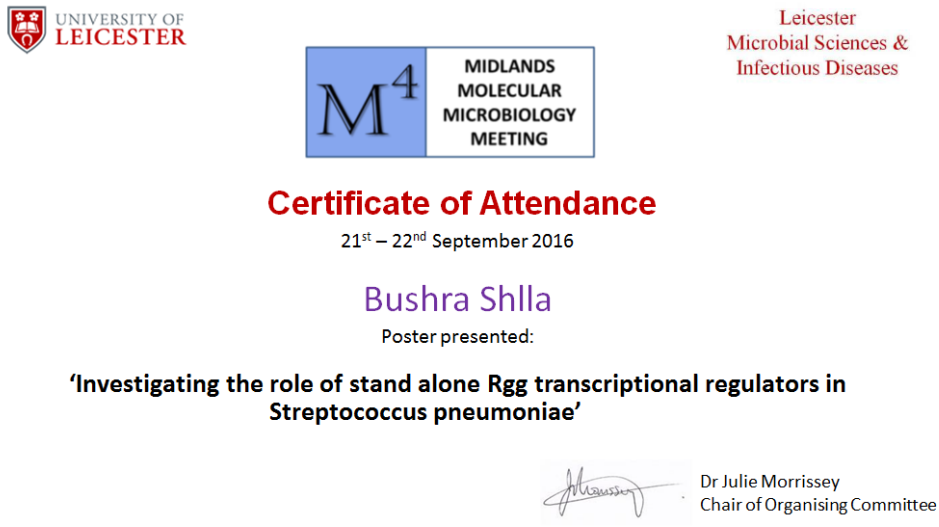
13th -14th April 2015.

4. I attended Respiratory Research Day, Held in conjunction with the Nottingham Centre for Respiratory Research and the Leicester Respiratory Biomedical Research Unit, City Hospital Campus, University of Nottingham, 6th May 2015.

5. I did oral presentation in Postgraduate Conference at University of Leicester, Maurice Shock Medical Sciences Building. 18 – 20 April 2016.

6. I attended Respiratory Research Day, Held in conjunction with the Nottingham Centre for Respiratory Research and the Leicester Respiratory Biomedical Research Unit, Stamford Court, Leicestershire, 4th May 2016.

7. I presented poster in the M4 Midland Molecular Microbiology Meeting, Leicester Microbial Sciences & Infection Diseases. 21st-22nd September 2016.


UNIVERSITY OF LEICESTER

Leicester Microbial Sciences & Infectious Diseases

M4 MIDLANDS MOLECULAR MICROBIOLOGY MEETING

Certificate of Attendance

21st – 22nd September 2016

Bushra Shlla

Poster presented:

'Investigating the role of stand alone Rgg transcriptional regulators in Streptococcus pneumoniae'

Dr Julie Morrissey
Chair of Organising Committee

8. I did oral presentation in Postgraduate Conference at University of Leicester, Maurice Shock Medical Sciences Building. 27th – 29th March 2017.

9. I presented poster in the Annual Conference 2017. 3-6 April, EICC, Edinburgh, UK.

Training

Training	Date	Outcome
Academic English for Medicine, Biological Sciences and Pychology Postgraduate Research Student	14th October 2014	Introduction about academic writing.
Academic English for Medicine, Biological Sciences and Pychology Postgraduate Research Students	21st October 2014	I learned that there are different way to arrange reference
Conducting your Literature Search	27th October 2014	I learned how to use library of university
Academic English for Medicine, Biological Sciences and Pychology Postgraduate Research Students	28th October 2014	I learned, present (design slides and the rule of present such as body language and eyes contact)
Academic English for Medicine, Biological Sciences and Pychology Postgraduate Research Students	02 November 2014	I write essay about 500 word in my subject
The Literature Review	03 November 2014	I learned that a literature review is an evaluative comparison of different research.
Academic English for Medicine, Biological Sciences and Pychology Postgraduate Research Students	11th November 2014	I learned, write citation in the text
Academic English for Medicine, Biological Sciences and Pychology Postgraduate Research Students	18th November 2014	I learned , write reference in the list of reference
Academic English for Medicine, Biological Sciences and Pychology Postgraduate Research Students	20th December 2014	I write essay about 1000 word in my subject for this reason **I got certificate Merit.**
RefWorks	11th February 2015	It introduction to refwork program and registration (make ID for Login)
Designing a Poster	18th February 2015	I learned , design poster in a good way and how can I choose its colours
Advanced Microscopy and Bio-Imaging Part 1	03 March 2015	It introduction to different types of fluorescence microscope.
Advanced Microscopy and Bio-Imaging Part 2	10th March 2015	It includes the more advanced fluorescence microscopy.
Advanced RefWorks	19th March 2015	I learned, insert references into text
Home Office licence Module 1-4	20th – 22th May 2015	I learned a lot of information. **Briefly** Handle the animals. Roles of license. Injection into Subcutaneous and Intravenous. Basic surgical techniques.
Protein Expression and Purification	24th March 2015	Purification protein and send protein to sequence

Training	Date	Outcome
Conducting your literature search	29th April 2016	It includes Apply your literature search plan. Review your literature search plan as appropriate. Keyword for search Phrase, index term and limits
Editing and Finishing the PhD Thesis - NOT FOR FIRST YEAR STUDENTS	29th April 2016	I learned Balance the number of words. Use passive voice and past or present tense. Proof reading. Put each chapter in separate file and keep at least 3 copies of everything.
Making research information come to you	03 May 2016	Saved search history. Decide how, where and when to check the feeds. Require away to receive the feed-Rss reader.
Who is citing who?	05 May 2016	Citation databases Scopus, Web of Science and Google Scholar
Preparing for the viva (CSE and CMBSP)	11th May2016	Examiners (External, internal) Re-read thesis (what strengths and weaknesses)
Listeria monocytogenes: a unique model in biology	17th March 2017	*Listeria monocytogenes* is the species of pathogenic bacteria, Gram-positive bacterium. Studies have highlighted the power of comparative genomics in virulence studies.