
# The Role of the Signaling Peptide VP1 in the Virulence of *Streptococcus pneumoniae*

Rolando A. Cuevas

A thesis submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Department of Biological Sciences
Carnegie Mellon University
Pittsburgh, PA
USA

Advisor: N. Luisa Hiller, PhD

## Abstract

The bacteria *Streptococcus pneumoniae*, also known as the pneumococcus, is a common and important human pathogen responsible for millions of deaths worldwide. This pathogen can colonize and reside in the back of the nasopharynx asymptomatically as part of the nasopharyngeal resident microflora. From the nasopharynx, the bacteria can disseminate to the middle ear and sinus to cause mild diseases, or spread into the lungs to cause severe diseases like pneumonia. Therefore, *Streptococcus pneumoniae* represents an important public health concern.

In the mucosa, the pneumococcus forms multicellular surface-associated structures called biofilms. These structures play a crucial role in the survival of bacteria during infection. Biofilms are recalcitrant to antibiotic treatment, facilitate the adhesion and cohesion of the cells to the host cells and each other, and hide bacterial cells from immune surveillance. Colonization is a population-level behavior, and its success depends on cell to cell communication and host-environment sensing. Colonization is also a prerequisite for dissemination and disease, thus the molecules that control colonization are high-value candidates for therapeutic interventions. In the pneumococcus, there are two well-characterized classes of secreted peptides that mediate signaling. In the first group, a peptide binds to surface-exposed histidine kinase receptors on the triggering a cascade, while in the second group, the signaling peptide enters the cells and binds directly to a transcription factor. Additionally, signaling peptides can also be categorized based on sequence features. Some signaling peptides characterized in pneumococci are the competence stimulating peptide (CSP), the bacteriocin inducing peptide (BIP), and the short hydrophobic peptides (SHP) family. The messages carried by these signaling molecules is not fully understood and constitutes a significant gap in the knowledge of pneumococcal biology.

We identified the product of an uncharacterized peptide-encoding gene as a putative signaling molecule using bioinformatics tools. We also found that this gene is also present across multiple pneumococcal strains. We termed the gene product Virulence Peptide 1 (VP1). We determined that VP1 is a signaling peptide, highly

induced in the presence of host cells and *in vivo*, promotes biofilm development, and serves as a potent virulence determinant.

To gather insight into the regulation of *vp1*, we analyzed the genomic organization of the *vp1* neighborhood. We demonstrated that *vp1* expression is under the control of a previously uncharacterized transcription factor, the regulator Rgg144. We found that pneumococci also encode a novel peptide belonging to the family of Rgg-cognate activator peptides SHP, which lay in between *rgg144* and *vp1* operons. We found that the pair Rgg144/SHP144 positively regulates the expression of the *vp1* operon.

Subsequently, we built on the molecular mechanism of the VP1 function to reveal novel bacterial and host molecules that enhance adherence and colonization. Transcriptional studies revealed that VP1 triggers the expression of operons involved in the transport and metabolism of hyaluronic acid (HA), a glycosaminoglycan present in the host extracellular matrix. HA is a crucial component of the extracellular matrix that is present on the apical surface of epithelial cells in the human airways. Previous studies have reported that molecules involved in the acquisition of HA contribute to colonization. Our findings suggest that host HA serves as an anchor for pneumococcal cells and that genes involved in transport and metabolism of HA promote adherence. Finally, our results are strongly supported by studies in a murine model of colonization.

In conclusion, we describe a molecular pathway utilized by pneumococci to adhere to the host cell and promote colonization. We provide substantial evidence that supports our hypothesis that VP1 is a secreted signaling molecule. Our data indicate that VP1 plays a pivotal role in pneumococcal biofilm development and behavior by orchestrating gene expression. Our studies are implemented using a virulent strain from the clinically relevant PMEN1 lineage. We conclude that VP1 is a novel streptococcal regulatory peptide that controls biofilm development and pneumococcal pathogenesis. Based on our evidence, we propose that VP1 serves as a marker for colonization and a target for drug design.

# Table of Contents

# List of Figures

## List of Tables



## List of Supplementary Figures

## List of Supplementary Tables

# Abbreviations

| | |
|---|---|
| A549 | Lung Cancer Epithelial Cells |
| ABC | ATP-binding cassette |
| BgaA | β-galactosidase A |
| BIP | bacteriocin inducing peptide |
| BLAST | Basic local alignment search tool |
| bp | Base pair |
| Cbp | Choline-binding proteins |
| CcpA | Carbon catabolite protein A |
| CDC | Center for Disease Control |
| CDM | Chemically defined medium |
| CFU | Colony forming unit |
| CMEE | Chinchilla middle ear epithelial cells |
| *Cps* | capsular polysaccharide synthesis locus |
| CSP | Competence stimulating peptide |
| DNA | Deoxyribonucleic acid |
| ECM | Extracellular matrix |
| Eng | Endo-α-N- acetylgalactosaminidas |
| EPS | Extracellular polymeric substance matrix |
| GAG | Glycosaminoglycans |
| Gal | Galactose |
| GAPDH | glyceraldehyde 3-phosphate dehydrogenase |
| GlcNAc | N-acetyl glucosamine |
| Glu | Glucose |
| HA | Hyaluronic acid |
| HK | Histidine kinase receptor |
| HysA | Hyaluronidase |
| IgG | Immunoglobulin G |
| IPD | Invasive pneumococcal disease |
| KEGG | Kyoto encyclopedia of genes and genomes |
| LytA | Autolysin A |
| Man | Mannose |
| MAST | Motif Alignment and Search Tool |
| MEME | Multiple Expectation Maximization for Motif Elicitation |
| MLEE | Multilocus enzyme electrophoresis |
| MLST | Multilocus sequence typing |
| NanA | neuraminidase A |
| NCBI | National center for biotechnology information |
| NET | Neutrophil extracellular trap |
| OD | Optical density |
| Opp | Oligo peptide permease |
| ORF | Open reading frame |
| PavA | Pneumococcal adhesion and virulence A |
| PavB | Pneumococcal adhesion and virulence B |
| PCR | Polymerase chain reaction |
| PCV | Pneumococcal conjugate vaccine |
| PCV10 | PHiDCV |
| PCV13 | Pneumococcal conjugate vaccine 13-valent |
| PCV7 | Pneumococcal conjugate vaccine 7-valent |
| PFGE | Pulse field gel electrophoresis |
| PHiDCV | pneumococcal NTHi protein D conjugate vaccine 10-valent |
| PiaA | pneumococcal iron acquisition A |
| PiuA | Pneumococcal iron uptake A |
| Playa | Pneumolysin |
| PMEN | Pneumococcal Molecular Epidemiology Network |
| PpmA | Putative proteinase maturation protein A |
| PptAB | SHP transporter (GAS) |
| PsaA | Pneumococcal surface antigen A |
| PspA | Pneumococcal surface protein A |
| PspC | Pneumococcal surface protein B |
| PTS | Phosphotransferase system |
| PTS-HAL | Phosphotransferase system - HA locus |
| QS | Quorum sensing |
| Rgg | regulator genes of glucosyl- transferase |
| RNA | Ribonucleic acid |
| RR | Response regulator |
| RRNPP | Response regulator systems |
| S.p. | *Streptococcus pneumoniae* |
| SHP | Short hydrophobic peptide |
| SlrA | Streptococcal lipoprotein rotamase A |
| StrH | β-Nacetylglucosaminidase |
| sVP1 | Synthetic VP1 peptide |
| TCS | Two-Component regulatory Systems |
| TF | Transcription factors |
| UNIPROT | The Universal Protein Resource |
| VFDB | Virulence factors data base |
| VP1 | Virulent peptide 1 |
| WGS | Whole genome sequencing |
| WHO | World Health Organization |
| ZmpB | Zinc metalloprotease B |

# Acknowledgments

Since my very first day in the lab, I have always received a tremendous amount of support and assistance from my advisor, Dr. N. Luisa Hiller. Throughout the years, she has granted me with unconditional support, guidance, and advises. But even more important, she gave me the freedom to explore new ideas, challenge old concepts, and new avenues for my dissertation project. Her constant encouragement has made me a better scientist, a critic of myself, and a critic of others. The preparation in this lab, while challenging as any lab here at CMU, was also intimate and tailored to prepare me not only as a researcher but also as a mentor and leader for future generations. Thanks, Luisa! Thanks for making science fun and enjoyable! Thanks for spice things up! And thanks for being the glue that keeps the lab together and moving!

I am also grateful to my thesis committee members, Dr. Aaron Mitchell and Dr. Frederick Lanni, for their invaluable support and insightful suggestions. Dr. Mitchell, thank you for always encouraging me to make executive decisions about my research, for the plentiful of pieces of advice, and support me when I needed to move my research forward. Dr. Lanni has always reminded me that the beauty, and the devil, are in the details. Thank you for always show interest and joy for science in our conversations. I would also like to thank my external committee member, Dr. John Williams, for his willingness to help me reach my graduation goals.

I am also grateful for the members of the Hiller Lab, past and present. Thank you, guys, for being alongside me and make this place my second home. Thank you, Dr. Anagha Kadam, for patiently teaching me caring for our experimentation animals and for guiding me into the NanoString craft. Thank you, Surya, for showing me how always to be persistent with your goals: you are an example of a well-rounded student and person! Thank you, Rory, your technical assistance ship was invaluable. I would also like to thank our former volunteer researcher, Elnaz, for our fun time in the lab and for helping sculpting my mentorship skills. I appreciate your support, and I am delighted that you love becoming a graduate student! Especially, I want to thank you, Sarah, for being such a stellar friend and a great lab mate, especially on those difficult days of winter. I truly appreciate all the support you gave me when I needed it the most.

Thank you, Dr. Haibing Teng. Now that you retired (and well deserved you have it), I will miss those endless conversations about microscopy, how to get the best informational possible image, or simply just complaining because the ‘scope broke once again… and again… and again. Also, I would like to thank everyone in the main office, particularly Ena Miceli, for keeping me registered and away from TA duties.

Most importantly, I would like to thank my parents for showing me the road to success since I was little and for always encouraging me to continue with my education. Thank you for allowing me to follow my dreams, even thousands of miles away from home. Gracias por todo!

This acknowledgment will be incomplete without paying tribute to my precious wife, Karina, and my beloved son Lucas. In your smiles, I encounter the courage that I needed to walk all these years of graduate school.

Last but not least, I wish to express my sincere thanks to all my credit card companies to believe that, one day in the distant future, I will honor all my debts.

# PART I - INTRODUCCTION

# CHAPTER 1 - *Streptococcus pneumoniae*, epidemiology and virulence

## 1.1 *Streptococcus pneumoniae* as an invasive organism

The bacteria *Streptococcus pneumoniae* is an important human pathogen, responsible for millions of deaths worldwide. It is estimated that this bacterium, also known as the pneumococcus, is responsible for more deaths worldwide than any other single pathogen (1). Further, it is one of the leading causes of death in the United States (2). A recent study about the global impact of lower respiratory infections, determined that the pneumococcal impact on global morbidity was higher than all other pathogens combined, including respiratory syncytial virus, *Haemophilus influenzae type* B, *influenza virus*, among others. In total, it was responsible for 1.18 million deaths across the globe (3,4). This pathogen can colonize and reside in the nasopharyngeal tract asymptomatically. It can also spread to other tissues, including the lungs, sinus, ears, the brain, and bloodstream (5,6). It causes mild to severe diseases: they range from pneumonia and otitis media in children to life-threatening conditions, including community-acquired pneumonia, meningitis, and sepsis (7–9). The pneumococcus affects children and their caretakers, elderly, and immunocompromised patients across the globe (10).

Pneumococcal vaccines, including Pneumovax and Prevnar, have been widely employed since 2000. Yet, they have only a limited effect at decreasing the incidence of carriage and middle ear infections, as well as infection of the elderly population (11). Furthermore, vaccination has led to a significant increase in vaccine-escape serotypes by both serotype-switching and the emergence vaccine-escape strains (12–16). However, despite the extensive vaccination programs that have helped to reduce the rate of invasive pneumococcal disease significantly, the pneumococcus continues to cause a variety of common clinical syndromes, and the pneumococcal disease remains a serious threat to human health.

## 1.2 The bacterium Streptococcus pneumoniae

*Streptococcus pneumoniae*, a Gram-positive bacterium, was initially denominated *Diplococcus pneumoniae,* from 1920 onwards. It was not until 1974 when Howard and Gooder addressed years of discussion about its particular structural morphology, which is very similar to *Streptococcus faecalis* (17), and renamed *Diplococcus pneumoniae* to its current designation, *Streptococcus pneumoniae* (18). The pneumococcus is an elongated coccus-shaped bacterium. It has a slightly pointed outer curvature, and they are usually observed as an arrangement of pairs of cocci (hence its original designation as diplococcus), ranging between 0.5 and 1.25 micrometers in size. They typically form short-length unbranched chains. Individual pneumococcal cells are non-motile and they do not form spores. This bacterium distinguishes itself from other streptococci bacteria due to its ability to breakdown red blood cells (alpha-hemolysis) (19). Like other streptococci, they can ferment sugars to lactic acid and hydrolase inulin (19,20).

The genome of pneumococci is a closed, circular DNA structure. It contains between 2.0 and 2.1 million base pairs; this number varies across strains. Recent estimates indicate that the number of genes per isolate ranges from 1900 to 2400 genes (21). At the genomic level, there is extensive diversity among pneumococcal lineages. The core genome encodes an estimated 500-1100 genes (clusters of orthologues) (22–25). The pneumococcal pangenome encodes an estimated 5000-7000 genes (22–25). These genomic variations contribute to differences in colonization and virulence potential (26). In stark contrast to its genomic variability, plasmids have not been detected in resistant pneumococcal strains; all genetics changes that confer fitness advantages are therefore chromosomal. The pneumococcus distinguished itself by the ability to undergo genetic transformation; that is, the bacterium is naturally competent and able to acquire recombine naked DNA from the environment into its genome (27). The combination of high genomic diversity with natural competence contributes to the plasticity of the pneumococcal genome.

The cell wall and the capsule constitute the major exposed surface components in the pneumococcus. The cell wall is a complex array of peptidoglycan which encloses the bacterium, and is composed of a triple-layer of peptidoglycan and teichoic acid.

Structurally, its primary function is to protect the bacterial cells from bursting and maintain the cell shape. The pneumococcal cell wall anchors to the bacterial membrane via lipoteichoic acid. The cell wall also anchors several surface-exposed proteins that function in transport processes and bacteria-host interactions. Several antibiotics target critical components of the cell wall structure; they may inhibit its synthesis and subsequently cause bacterial lysis. The outermost layer that surrounds the bacteria is the capsule which it completely envelops the pneumococcal cells and shields the bacteria from the host immune system, largely determining the bacterial virulence (28). The capsule is composed exclusively of polysaccharide chains, with a thickness ranging from 50 to 200 nm (29). With only two exceptions (serotypes 3 and 37), the capsule is linked to the cell wall via a glycosidic bond to β-D-N-acetylglucosamine (30) and is a major player in the bacterial immune evasion. The polysaccharide capsule is the major virulence factor in the pneumococcus. It is the basis for the serotyping of pneumococci, except for non-typeable pneumococci, which lack the genes required for the capsule synthesis.

Over the past decades, many pneumococcal virulence determinants have been identified, yet the driving factors that differentiate commensal and pathogenic states remain unclear. This question is the primary motivation for this dissertation. How is pneumococcal virulence regulated? In this regard, intense attention has been paid to the pneumococcal capsule, and the majority of the implemented pneumococcal vaccines targeted this structure. Several major virulent functions have been ascribed to the bacterial capsule, these include bacterial protection against opsonophagocytic clearance by neutrophils and macrophages, protection against neutrophil-extracellular traps, and entrapments in the host mucus. In particular, the extensive biochemical diversity of the capsule structure leads to extensive antigenic variation. However, capsular diversity alone does not fully explain virulence. Genomic studies attest to a large amount of variation between pneumococcal strains that contribute toward a considerable difference in the protein content and how they are regulated. Moreover, the interest in additional virulence factors, elements that contribute to the pathogen survival in vivo, has increased with the appreciation that pneumococcal proteins pose

promising alternatives for the development of pneumococcal serotype-independent vaccines.

Multiple virulence determinants have been associated with contributing to pneumococcal infection, and its number has dramatically increased over the years. For example, extensive research has been invested in understanding the biology of the pore-forming toxin pneumolysin and its role in the disruption of the immune system (31). Or the biology and role in virulence of the autolysin LytA (32). Other important virulent factors are the adhesin molecules pneumococcal virulence protein A and B (PavA and PavB) (33,34) However, bacterial adhesion is not exclusively mediated by classical adhesin molecules; they can are also be mediated by moonlighting metabolic enzymes such as enolase (35,36) or by surface-exposed enzymes such as β-galactosidase BgaA (37). These are just a few selected examples to illustrate the complexity of the pneumococcal biology in terms of the understating of its virulence determinants. They will be further discussed in section 1.5.

In conclusion, extensive literature serves to illustrate the complex interplay between pneumococcal components and the host. Unraveling these interactions may explain not only the striking differences between strains but also help understand the featured underlying pneumococcal pathogenesis.

## 1.3 Epidemiology, vaccines and drug resistance

Lower respiratory infections caused by the pneumococcus continue to be a substantial concern despite the advent of pneumococcal vaccines and antibiotics (3,4). The pneumococcus has been and continues to be one of the main bacterial pathogens in both children and elderly in the US (38). The abusive and irresponsible utilization of antibiotics for decades has facilitated the rise of antibiotic-resistance strains (39). Furthermore, decreasing the rates of carriage is challenging due the absence of symptoms during commensal colonization.

The diverse and sometimes severe, clinic manifestation of the pneumococcal disease, as local outbreak or pandemics, has generates increase incentives for the

pneumococcal research (40,41). In this section, I introduce epidemiological concepts and data regarding the pneumococcus impact on human disease. I discuss the consequences of the abuse of antibiotics on pneumococcal diversity and medical advances to combat pneumococcal infections. Finally, I describe diseases associated with the pneumococcus.

### 1.3.1 Epidemiology of the pneumococcus

As a pathogen, *Streptococcus pneumoniae* is a leading cause of invasive infections such as meningitis, sinusitis, and pneumonia in adults and the causative agent of middle ear infections in children. It also is a major cause of bloodstream infections (8,9), and one of the most important community-acquired pathogens, affecting especially the elderly people (42). The burden of pneumococcal disease is estimated to be over one million fatalities per year, and the majority of deaths occur in children younger than five years (43–45). The Center for Disease Control (CDC) classifies *S. pneumoniae* as a severe epidemiological threat due to its resistance to drugs in the penicillin and erythromycin groups (46,47) as well as less commonly used antibiotics like macrolides, lincosamides, tetracycline, and co-trimoxazole (48). This classification indicates that the pneumococcus requires active public health monitoring and prevention programs, as well as prompt and sustained action in cases of epidemic outbreaks (39).

Furthermore, the World Health Organization (WHO), have released an updated recommendation list to guide the prioritization of research, discovery, and development of new antibiotics for several bacterial threats, including *S. pneumoniae* (40,41). It highlights this bacterium as an important community-acquired pathogen, affecting especially the elderly people (42).

A recent global study on lower respiratory infections determined that of the 2.56 million deaths around the globe, the pneumococcus contributed to the morbidity more than any other respiratory pathogens combined: it was responsible for an estimated 1.18 million deaths (3,4). For children younger than two years old, the susceptibility to infections is not only due to an immature immune system, but also to gene mutations

that result in defects in either the innate or adaptive immune system. Furthermore, people with immune deficiencies, such as sickle cell disease, are at high risk to acquire an invasive pneumococcal disease (49,50).

In summary, while the introduction of PCV vaccines had helped to alleviate the burden of invasive pneumococcal diseases, the high carriage observed in children indicates that the vaccine effectiveness is limited and highlights the gaps in our understanding of the serotype's contribution to pneumococcal disease, especially in undeveloped countries.

### 1.3.2 Antibiotic resistance of pneumococci

In 1879, *Streptococcus pneumoniae* was simultaneously isolated by the U.S. Army physician George M. Sternberg and French chemist Louis Pasteur (51). Since then, the pneumococcus had remained susceptible to penicillin and most other antibiotics. It was not until mid-1940 that penicillin-resistant strains were isolated in the laboratory. However, the accelerated global consumption and irresponsible use of antibiotics in humans and animal feedstocks catalyzed the rise, and severity, of antibiotic-resistant strains. The first clinically isolated penicillin-resistant strain was reported in New Guinea in 1967, and it shows moderated sensitivity to penicillin (52). Subsequently, highly resistant pneumococcal isolates were reported from South Africa in 1977 (53), followed by the massive emergence of penicillin-resistance strains reported around the globe between 1980s and 1990s (13,54,55). The emergence of penicillin-resistant clones was mainly due to the lack of prescription control and irresponsible use. The rise of these clones led to the use of alternatives to penicillin. Among the other options were macrolides, non-beta lactam antibiotics, and fluoroquinolones. In this context, it is also of concern that pneumococci have acquired tolerance to vancomycin, an antibiotic used as an ultimate line of defense against multi-resistant Gram-positive bacteria (56).

After the 1980s outbreak, several epidemiological surveillance protocols have facilitated the tracking of those penicillin-resistant clones and the identification of globally disseminated strains. In 1997, the Pneumococcal Molecular Epidemiology

Network (PMEN) was established to organize and catalog those different strains and survey its international propagation. Among the several strains and lineages, the PMEN1 lineage is clinically important because they are frequently present in carriage sampling, they are prevalent worldwide, and they are multidrug-resistant (57). The first clone of the PMEN1 lineage was isolated in 1984 from a 64-year-old woman with breast cancer during an episode of pneumonia in a hospital in Barcelona (58). This clone was identified as the penicillin-resistant strain Spain 23F-1, and since then, the clone has propagated globally. Subsequently, the PMEN1 lineage has acquired resistance to other antibiotics, including chloramphenicol, tetracycline, and most recently fluoroquinolones and macrolides (59).

The emergence of these multi-drug resistant clones and their global dissemination has further complicated both the disease treatment and bacterial eradication, making the pneumococci a major public health concern.

### 1.3.3 Pneumococcal carriage

It is accepted that the colonization of nasopharynx is a prerequisite to disease. The pneumococcus can remain asymptomatically in the nasopharynx for months. In children, the pneumococcus can be found frequently as part of the nasopharyngeal microbiota, while in adults, its presence is low. Over the years, carriage studies have been undertaken in various countries to monitor the relationship between infection and disease and the effect of vaccine prevention without a universal agreement (60–63). In 2014, Adegbola *et al*. determined that the prevalence of *S. pneumoniae* carriage in children younger than five years ranged from 20% to 93.4% in low-income countries, after a bibliographic investigation of African and Asian populations that included medical reports from 1990 to 2012 (60). That same year, a study from Usuf *et al.* determined an overall pneumococcal prevalence of 63.2% in children younger than five years old in Africa (61). A more recent study estimates that the carriage among children younger than five years old is 75% approximately, irrespective of the prophylactic status of the population (74% pre-vaccine era versus 76% post-vaccine era) (62).

Other factors may precede and influence pneumococcal carriage. It has been shown that seasons, smoking (including passive smoking in children), co-infection with other pathogens, such as viruses, or co-colonization with different pneumococcal serotypes all contribute to increase pneumococcal carriage (64–67)**.** Carriage becomes a severe concern in children of pre-school age and attending daycare centers, where the incidence of infection rapidly spread across infants (68), while drug-resistant strains may represent up to 50% of the detected strains. Notwithstanding the actual carriage rates, it is accepted that the pneumococcus can progress to invasive pneumococcal diseases (IPD) when it disseminates from the nasopharynx to tissues (69,70). And yet, the molecular mechanisms that promote dissemination over carriage remain largely unknown.

### 1.3.4 Diseases associated to the pneumococcus

As mentioned in the previous section, the pneumococcus can cause severe IPD, including meningitis, sepsis, and pneumococcal endocarditis. While the etiology of such diseases can be attributed to various types of pathogens, or to a combination of them, this section will cover pneumococcal-related aspects of IPD.

***Pneumococcal Pneumonia:*** Pneumonia is a worldwide, substantial public health problem and a leading cause of illness and death in people of all ages, representing an enormous socio-economic burden for healthcare systems. According to recent WHO data, pneumonia is the single most significant infectious cause of death in children worldwide, accounting for 15% of all deaths of children under five years old (71). Pneumonia is an infection of the alveolar space, characterized by excessive infiltration of immune cells, pathogens, and secretions in the alveoli (pus and fluid), which combined, results in gas exchange impairment. While viruses, bacteria, and fungi can also cause pneumonia, a recent study , which compared the role of *S. pneumoniae*, *H. Influenzae* type B, RSV and influenza virus on the global incidence and causes of lower respiratory infections, determined that the pneumococcus is responsible for over a half of all reported death events, with an estimated 1.18 million deaths cause solely by

the pneumococcus (3,4). Pneumonia caused by bacteria can be treated with antibiotics. However, the rise of antibiotic-resistant strains has aggravated the situation (40,41).

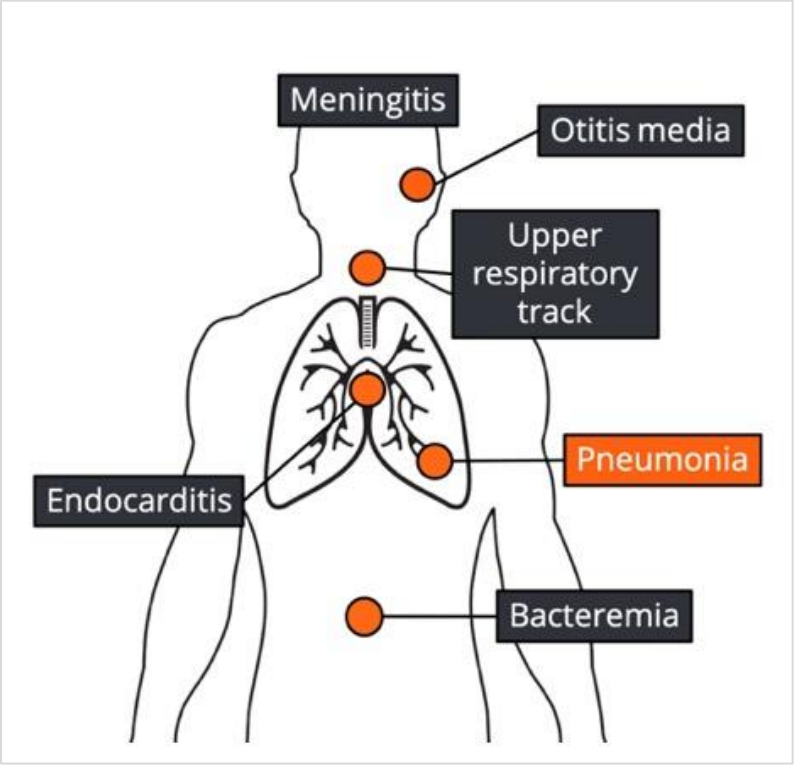



**Figure 1.1 Diseases caused by the pneumococcus**
The pneumococcus colonized the upper respiratory tract, residing in the nasopharynx and the middle ear. From the nasopharynx, it can spread to the lung to cause pneumonia. Should the pneumococcus overcome the lung immune defense, the bacteria can progress from pneumonia to invasive pneumococcal diseases (IPD). Systematic dissemination cause sepsis, endocarditis, and meningitis.

The pneumococcus utilizes a diverse arsenal of adhesin molecules to reach the lungs and attach to the alveoli surface. Many pneumococcal surface molecules serve as adhesins, mediating attachment of the bacteria to host cells (72–74). Plasmin- and fibronectin-binding proteins A and B (PfbA and PfbB) bind to the extracellular matrix and are required for adherence to lung and laryngeal epithelial cells (75,76). The pneumococcus expresses several surface-exposed glycosidases. They have been shown to contribute to bacterial adhesion not only via modification of the host surface that reveals receptors but also via direct binding (77). Once in the lung, cilia sweep the

pneumococcus and any debris or upwards and out of the bronchial tree and trachea to be ultimately expelled by cough. The bacteria are also cleared by antimicrobial peptides and the resident alveolar macrophages. Should these systems fail to clear the bacterium, or when the pneumococcal load is sufficiently high to overwhelm the clearance capacity of alveolar macrophages, the adaptative immune response is triggered. The combined action of T cells, neutrophils, and macrophages allows the removal of the bacteria and the establishment of immune memory for future infections.

The pneumococcus is associated with several other endobronchial infections, including suppurative lung disease, bronchiectasis, and protracted bacterial bronchitis (78). It has been extensively reported that the pneumococci is presence in diseases ranging from endobronchitis (79) to pediatric arthritis (80,81). It has become a recurrent problem after surgical intervention and transplantation, where pneumonia and pneumococcal bacteremia are common post-surgical manifestation of disease (82,83).

In the elderly population, community-acquired pneumonia is often the result of a compromised general state, in which the immune system has been deeply depressed by preexisting infections, such as influenza (84). This pre-existing depressed state enhances the susceptibility to pneumococcal disease, which leads the toll for infectious death in the US (42). Also, seasonal and preexisting viral infections constitute a significant contribution to pneumococcal disease (85). A high proportion of community-acquired pneumonia is caused by viral co-infections (84), infections in which the pneumococcus remains the most frequent pathogen during these influenza-associated outbreaks (86). It is now accepted that the pneumococcus takes advantage of the influenza-induced immune state of patients to spread into the lung to cause pneumonia (87). Unfortunately, there is not enough active surveillance programs of influenza-pneumonia co-infections during influenza season.

***Otitis Media*:** *S. pneumoniae* can disseminate from the back of nasopharynx into the middle ear space to cause otitis media through the Eustachian tube. *S. pneumoniae* and *H. influenzae* are the predominant pathogens associated with otitis media, whereas *S. pneumoniae* is the predominant agent during acute otitis media (88). Otitis media is primarily a pediatric infectious disease, and it is the most common reason for doctor

visitation. The diagnosis is usually based on symptoms such as otalgia, fever, erythema of the tympanic membrane, and ear secretions. Based on these symptoms, pediatricians can quickly diagnose the infection and initiated a treatment, which depends on the age of the child, laterality (one or both ears), and the severity of the disease. Antibiotics like amoxicillin are usually prescribed (89). Infants and young children are more susceptible to otitis media due to immature immune system as well as an anatomically immature Eustachian tube. In children, the Eustachian tube is shorter, wider, and horizontal relative to the nose, impeding a proper drainage of bacteria and debris. As the child grows, the maturation of the Eustachian tube (as well as the immune system) decreases the risk of otitis media (90,91). Of the spectrum of ear infections, chronic otitis media is particularly alarming due to the well-documented consequences of hearing loss. Furthermore, chronic otitis media may lead to verbal and learning disabilities if not properly treated. In children, chronic otitis media often occurs after acute otitis media, albeit it may also arise without preceding symptomatic infection, leading to persistent otitis media with ear secretions (92). A common method to treat acute and chronic otitis media is the minimally-invasive insertion of ear tubes in the tympanic membrane which help to alleviate the otitis symptoms. However, it has been shown that the pneumococcus persists in ear tubes. In addition, cochlear implants, the method of choice to rehabilitating severely to profoundly deaf children (93), also are affected by persistent bacteria (94).

It has been suggested that both chronic and acute otitis media are biofilm-related diseases; it has been also suggested that biofilms facilitate the disease onset since these structures offer many advantages to the pathogen, including increased antimicrobial resistance and host immune evasion. Biofilm presence has been confirmed in otitis media animal models and on the tympanic membrane of children (95).

***Pneumococcal Endocarditis****:* Infective endocarditis is a chronic infection of the heart valves or endocardium (96–98). In the hearth, the presence of pneumococcus can cause a congestive heart failure and myocardial infarction (99–101). The incidence of pneumococcus-caused heart disease is unknown, while the pathogenicity of such complications is poorly understood (102–104). Life-threatening acute cardiac

complications are more common in pneumococcal infection compared to other bacterial infections (105), and high morbidity and mortality rates are associated with pneumococcal endocarditis (106). Several reports have shown the presence of penicillin-resistant pneumococcal clones in the heart (107,108). Pneumococcal infections post-surgery have also been reported (102–104,109), even after the implantation of prosthetic heart valves (110).

There is an ongoing debate about the exact pathogenesis of pneumococcal infective endocarditis. It has been shown that pneumococcal serotypes 1, 8, and 12 are the most commonly found strains (106), although it has been suggested that any clonal type may cause infective endocarditis (111). Further, molecular studies have shown that pneumococci can translocate across the myocardial tissue to cause microlesions and disruption of the cardiac function (5). It has been suggested that bacterial adhesion, a feature of the biofilm mode of growth (112), may be required for disease onset (111). The America Heart Association recommends vaccination for heart compromised patients. Without prompt treatment, infective endocarditis is usually lethal (113–115). Thus, infective endocarditis remains a grave health problem and pneumococcal infections are an associated risk factor.

***Pneumococcal Conjunctivitis***: In the eye, the pneumococcus can infect the conjunctival epithelial layers, and from there, can penetrate to deeper layers of the eye, such as the vitreous body and the cornea, causing endophthalmitis (116) and keratitis (117), respectively. One distinctive difference is that the strains colonizing the eye carry an extensively different gene composition, which features a total absence of gene encoding for the capsule, making the isolates non-typeable (118–120). Non-typeable pneumococci have been reported in outbreaks of conjunctivitis among university students and military recruits as early as the 90s (121,122). Symptoms of bacterial conjunctivitis include redness of the eye, purulent discharge lasting all day, edema, and pain (123).

### 1.3.5 Vaccines against *Streptococcus pneumoniae*

Capsule production is indispensable for pneumococcal virulence. To date, over ninety distinct capsular types have been described, and strains with different capsule serotypes greatly differ in their capacity to cause disease (124–126). Pneumococcal vaccine formulations are based on various capsular polysaccharide derived from the most prevalent pneumococcal strains. Several vaccines have been introduced over the years during the last century (127). The last of the classic generation of pneumococcal vaccine was Pneumovax (PPSV23), introduced in 1983. This vaccine contains polysaccharide isolated from 23 different strains. Despite the broad spectrum of Pneumovax, vaccines based solely in polysaccharides are poorly immunogenic in high-risk groups such as immunocompromised persons, children, and elderly people (128).

A new generation of pneumococcal vaccine was introduced in 2000. The first of them was called *Prevnar* (also known as PCV7, 7vCRM; *Prevenar/Prevnar*™) (Pfizer Inc., New York, USA). It differs from the previous vaccines in that pneumococcal polysaccharides are conjugated to the diphtheria protein that function as a carrier and helps to boost the immune response and memory. *Prevnar* is a heptavalent conjugate vaccine, which targeted the seven most common serotypes (4, 6B, 9V, 14, 18C, 19F, and 23F). Prevnar was then updated to a tridecavalent formulation (*Prevnar* 13™) in 2010, which includes polysaccharide isolated from the original seven strains while introduce protection against six additional virulent strains (the serotypes 1, 3, 5, 6A, 19A and 7F). Although Prevnar covers less strain variants compared to Pneumovax, Prevnar is regarded as more effective vaccine since it generated a strong response due to the polysaccharide conjugation to a protein. Furthermore, it has been shown that Prevnar 13 is also able to reduce the nasal carriage on vaccinated individuals (129). A third conjugated vaccine called *Synflorix* (also known as PCV10 or PHiD-CV) (GlaxoSmithKline, Rixensart, Belgium), was authorized by European Medicines Agency in 2009 for use in the European Union. This decavalent formulation (1, 4, 5, 6B, 7F, 9V, 14, 18C, 19F and 23F) differs from its PCV counterpart in that the pneumococcal polysaccharides are conjugated to the protein D from non-typeable *Haemophilus*

*influenzae,* to tetanus toxin or to the diphtheria toxin. Thus, *Synflorix* also protects against *H. influenzae.*

The success of PPSV23, PCV7, PCV13, and PCV10 was accomplished with the introduction of a complementary and comprehensive prevention program that targeted high-risk populations, and included the recommendation of use multiple-dose of the vaccine (130). Currently, the PCV13 is administrated to all infants beginning at two months of age, receiving a four-dose series of vaccines total. Catch-up vaccination is recommended for children younger than age five years who did not receive the vaccine on schedule. Also, one dose of PCV13 vaccine should be administered to adults age 65 years or older who have not been previously vaccinated. The CDC also recommends one dose of PCV13 to persons ages 19 through 64 years who have not previously received PCV13 and who are at the highest risk of severe pneumococcal disease. It has been demonstrated that both PCV and PPSV are safe vaccines, extending the recommendation to even pregnant women. Considerations are recommended solely to people who may experience (rare) cases of a severe allergic reaction to one dose of the vaccine (47,131,132).

The pneumococcal vaccine has significantly reduced the rates of invasive pneumococcal disease in all age groups. However, the complexity of the polysaccharide and capsular switch between strains pose a growing problem for the complete eradication of pneumococcus (10). Economic factors also impede the eradication of this pathogen: the complexity of manufacturing a multi formulation conjugate vaccine and its global administration (133). Different groups are currently investigating new approaches for next-generation, protein-based vaccines against pneumococcus. The goal is to generate a vaccine that confers universal immunity against the different pneumococcal serotypes. It has been shown that naturally acquired immunity against the pneumococcus is dependent on exposure to pneumococcal protein virulent epitopes (134).

The more attractive pneumococcal candidate antigen is the pore-forming toxin pneumolysin due to it is a key virulent factor and potent initiator of host immune response (62,135). An elevated antibodies titer against pneumolysin is detected in otitis-

prone children (136). However, native pneumolysin cannot be used in vaccines due to its intrinsic cytolytic activity which requires chemical modifications (137). Several other pneumococcal virulent factors are or have been investigated as potential vaccines, such as pneumococcal surface protein A (PspA), autolysin, and Pneumococcal surface antigen A (PsaA) (138). The majority of the candidates are either present at the bacterial membrane, like the two iron uptake ABC transporters PiaA and PiuA (139,140) and the peptide ABC transporter AliA (141), or associated to the cell wall like PspC. It has been shown that immunization with purified PspC shows protection against sepsis via antibodies cross-reactive to PspA (142,143). Immunization with either the native or recombinant the surface-exposed protein neuraminidase (NanA) protects against the nasopharyngeal colonization and subsequent pneumococcal otitis media (144,145). Secreted molecules like the pneumococcal proteins IgA1 protease (IgA1p), enzyme that naturally acts against host IgA, produced and elevated antibody response after systemic infection. This observation lead to the development of an experimental trivalent vaccine (SlrA-IgA1p-PpmA) with proven protection against fatal pneumococcal pneumonia in mice (146–148). Several other formulations have been developed; one of them uses serotype-4 polysaccharide linked to the pneumococcal proteins NanA, PiuA, and SP0148. This formulation has shown effectively closer to available polysaccharide-based vaccines (149). More recently, another formulation, serotypes-independent, multiple-antigen formulation (GroEL, PspA, Hsp70, pneumolysin) have been reported to protect against pneumonia and sepsis (135)

In conclusion, the introduction of conjugated vaccines into childhood immunization programs has contributed to reducing the incidence of pneumococcal-associated diseases. While this strategy has helped to reduce infections caused by other bacterial pathogens such as *H. influenzae* type b and group C meningococcal, *S. pneumoniae* remains a major public health problem and a leading cause of morbidity and mortality worldwide. The large number of pneumococcal serotypes, the incredible genomic diversity, and the increasing antibiotic resistance along with the different strategies that the pneumococcal utilized during the invasion and spread pose a great challenge in developing strategies to control and tackle pneumococcal-associated

diseases. While the development of next-generation vaccines is active, little has been translated to prophylactic deployment.

### 1.3.6 Emergence of a pneumococcal treat

Numerous surveillance studies have documented the global spread of penicillin-resistant pneumococcus clones (13,150–155). Over the years, ninety-four unique outbreaks have occurred between 1916 and 2017 globally (156). Numerous studies have shown that despite the considerable diversity among strains, only a handful of highly successful clones have emerged and achieved global spread (13,154,157). Several high-resolution molecular techniques have been utilized to typing and catalog strains based on their similarities. Fingerprinting techniques such as multilocus enzyme electrophoresis (MLEE), multilocus sequence typing (MLST), and pulse-field gel electrophoresis (PFGE), have been broadly used to understand the pneumococcus epidemiological dynamic.

MLEE was originally developed to detect polymorphism in humans. It relays in the electrophoretic mobilities of enzymes isolated from patients' samples. This technique was then used to differentiate between pneumococcal strains (158,159). Due to the low resolution to discriminate between specific allelic variants, it was later replaced by MLST, which detects variations in DNA sequences amplified from multiple genomic *loci* via PCR. PFGE relies on the mobility of genomic DNA previously digested with SmaI and resolved by electrophoresis.

In response to the utilization of several fingerprinting techniques and to generate and organized strain catalog, the Pneumococcal Molecular Epidemiology Network (PMEN) was established in 1997. This entity is responsible for the global surveillance of antibiotic-susceptible and antibiotic-resistant clones. They release guidelines for the recognition, classification, and standardization of terminology of different clones of *Streptococcus pneumoniae* (160). The PMEN has utilized these molecular tools (PFGE MLST, penicillin-binding protein presence) to group clones. While the initial classification release included only 16 antibiotic-resistant strains, the classification now

includes antibiotic-sensitive strains as well; there are 43 distinct clones described currently (161).

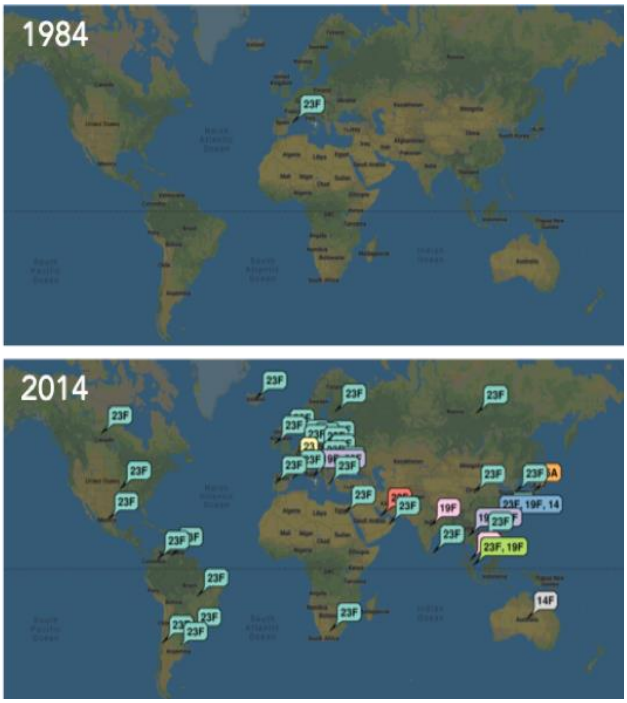



**Figure 1.2 Global dissemination of the lineage PMEN1**
The map depicts the place of initial isolation in 1984 and the current known worldwide distribution. Colors represent the serotype. Modified from PMEN Consortium website. https:// www.pneumogen.net/

One of the very well characterized lineages is represented by clones of the Spain23F or PMEN1. It was responsible for the Spanish epidemic of the 1980s (57) and has since spread to North and South America, Europe, Asia, Africa, and Australia (57). Over the past 30 years, the PMEN1 lineage has distinguished itself among all the other known, disease-associated, strains, by evolving a multi-drug-resistant profile and more recently, it has started to show vaccine-escape traits (157).

The PMEN1 lineage corresponds to the MLST 81, and it is also known as Spain23F, Spain23F-1, and SPN23F. This lineage represents a significant percentage of carriage clones, and it has been shown to cause a broad spectrum of invasive disease, including meningitis and sepsis in children and adults (162). The first PMEN1 member, a penicillin-resistant clone, was isolated in 1984 from a 64 years-old woman with breast cancer during an episode of pneumonia in a hospital in Barcelona, Spain (58). The PMEN1 lineage has subsequently acquired resistance to chloramphenicol and tetracycline, while more recent reports have shown that some PMEN1 clones also display resistance to fluoroquinolones and macrolides (10,157). They are predominantly

of serotype 23F, but there are also types 19A, 19F, 14, and 3 as a result of capsular switching. It has been shown that PMEN1 isolates have acquired a variable mechanism of resistance to antibiotics and to serotype vaccines through the years due to its high frequency of DNA donation to other pneumococcal lineages (155). The emergence of this type of multi-drug resistant clones has further complicated the treatment of pneumococcal-associated diseases and bacterial eradication.

At the genomic level, there is an extensive diversity in the genomes of *Streptococcus pneumoniae* (163,164). However, the clones in the PMEN1 lineage are closely grouped into phylogenetic lineages (155,165). The clones in this lineage not only represent a significant percentage of carriage clones; they also cause a wide spectrum of invasive disease including meningitis and sepsis in children and adults (162).

In conclusion, a few lineages of the pneumococcus have emerged as major threats, and the significant clinical incidence and international spread of the PMEN1 lineage place it among major public health concerns.

### 1.3.7 Other relevant clinical isolated – The TIGR4 clone

The *Streptococcus pneumoniae* TIGR4 (The Institute for Genomic Research) is commonly used as a model; it is a bench-friendly strain and was the first pneumococcal genome sequenced (166). The TIGR4 isolate is a highly virulent, type 4 encapsulated, Gram-positive bacterium. It was originally isolated from a 30-year-old male patient affected with systemic pneumococcal infection in Kongsvinger, Norway 1991 (167). This isolate was initially denominated as JNR.7/87 (168), then as strain sKNR.7/8 (169,170); and later as strain N4 (171). TIGR4 received its final name in a seminal report from the Fraser group in 2001 (166). Subsequently, other identifiers have been ascribed to this strain, including AE005672, ATCC BAA-334, and NC_003028. Nowadays, it is a commonly used laboratory strain, even though it was a clinically relevant strain.

## 1.4 The *Streptococcus pneumoniae* variable life style

The pneumococcus is an example of biological simplicity. With a small array of cleverly adapted toolset, it is capable of invading potentially every human organ. The understanding of the pneumococcal invasion strategy not only had served useful to tackle pneumococcal invasion, but also to understand invasion in virtually all respiratory pathogens. *Streptococcus pneumoniae, Haemophilus influenzae*, and *Neisseria meningitidis,* to name a few, share similar strategies to combat oral flora and to conquest humans from nasopharynx to the lung.

To champion colonization and survival, the pneumococcus utilizes a varied armament of pheromones and adhesion molecules to coordinate the population. The bacteria also adapt a biofilms mode of growth that provides enhanced resistance to antibiotics and help them to escape the immune surveillance. Biofilms play a major role in pneumococcal infection by forming mono- or polymicrobial settlements in adenoids and the middle ear during disease (92,172–175).

### 1.4.1 Transition from the back of the nose to the lower respiratory tract

This section will discuss the pneumococci ability to efficiently colonize the back of human nasopharynx and subsequent dissemination through the host's mucosa tissues to spread to different organs. Evasion of the immune surveillance also plays an essential role in the bacterial survival in the host.

Colonization is a complex interplay between the pathogen and the host. Colonization depends on the binding of the pneumococcus to the host epithelium and evasion of host defense. Surface exposed proteins, called adhesins, play a crucial role in pneumococcal pathogenesis by facilitating direct interaction between the bacteria and the host (72–74). They mediate binding to several components of the mucosa, including mucins, extracellular matrix (ECM), and serum components. Adhesin expression correlates with the stage of the infection as well as the tissue that the bacterium is colonizing. Section 1.5 will expand on the multiple adhesins that the pneumococcus utilizes to achieve colonization.

The pneumococcus has to combat the harsh mucus conditions. During the invasion, the pneumococcus exhibits a series of changing population-level phenotypes. The capsule helps in preventing the bacterium from being recognized and engulfed by the immune system as well as being snared and then removed by the mucus-mediated clearance (70,176). It is in the nasopharynx where the pneumococcus displays two of its distinctive macroscopic features; it undergoes phase-variation and forms biofilms on the mucosal surface. Phase-variation is a reversible phenotypic variation in colony opacity and encapsulation, where the capsule gets thinner to prevents from being trapped in the mucosa (177). Various polymer-like mucin glycoproteins are present in the mucosa. Sialic acid (N-acetylneuraminic acid) is attached to these molecules conferring them a net negative charged. These negative glycoproteins bind to positively charged particles, like bacteria, entrapping them. Then, through ciliary beating, those particles are removed (178). Thus, by reducing the capsule, the bacteria avoid get trapped in the mucosa. Also, during transit into the mucus layer, the pneumococcus releases pneumolysin, a toxin that inhibits ciliary beating. Hence, facilitating the pneumococcal arrival to the epithelial layer (179).

In the nasopharynx the pneumococcus is not alone. It is part of the commensal polymicrobial communities. To outcompete other species also present in the nasopharynx, pneumococci cells utilize an armament of secreted molecules. It is here where the pneumococcus also engages predatory warfare against other microorganisms by secreting bacteriocidal peptides (bacteriocins) (180–182).

Once the pneumococcus penetrates the mucus barrier and reaches the epithelial surface, the shielding capsule gets thinner to expose adhesins that allow it to interact and attach the epithelial extracellular matrix (ECM). The bacterium downregulates the genes required for capsule synthesis and sheds the remaining shielding polysaccharides (183,184). The interaction between the adhesins and their target set in motion signaling cascades that result in attachment and biofilm formation, determining the course of infection. Meanwhile, the bacteria also release an artillery of glycosidases, including neuraminidases (NanA, NanB, and NanC), β-glucosidases (BgaA), and β-N-glucosaminidases (SrtH). These enzymes not only deglycosylate components out of the ECM, they reduce the viscosity of the mucus and prevent entrapment (185,186).

Additionally, the bacteria can also utilize the released carbohydrates as carbon source (187,188).

Should pneumococcus overcome all the barriers that the mucus imposes, the bacteria can now firmly attach to epithelial cells, replicate, and then form biofilms. While in biofilms communities, the pneumococcus lives in a state of balance with the mucosal immune system. Biofilms facilitate pneumococcal colonization in multiple ways: they hide the bacteria from immune surveillance, increase resistance to antimicrobials, and provide a platform for DNA acquisition (189–193). Once it reaches maturation, and in response to a changing host environment (temperature, virus infections, depressed immune system), the biofilm releases pneumococcal cells (194). Pneumococcus then descends from the nasopharynx through the pharynx towards the trachea, helped by the host aspiration. The pneumococcus can now disperse into the large airways, bronchial bifurcations, and alveoli (195).

In the alveoli, the just-arrived bacteria encounter the lung macrophages. Their function is to protect the lung mucosa from pathogens by phagocytosing them. These cells are also professional antigen-presenting cells and release cytokines that initiate the alarm signaling. This cells also recruit neutrophils and lymphocytes, and coordinate the resolution of inflammation. The alveoli mucosa also contains multiple antimicrobial agents, including β-defensins and copious amount of secreted immunoglobulin A (IgA), which interfere with bacterial adhesion. In response to IgA, the pneumococcus releases the protease IgA1p, which cleaves IgA, inhibiting the IgA-mediated opsonization (148,196). Neutrophils are also recruited, and they are essential to kill the pneumococcus. Neutrophils are specialized white cells which function is to release extracellular DNA traps (NETs) to snare pathogens and kill them via the release of antimicrobial compounds. Interestingly, pneumococci counterattack by degrading NETs via the secretion of endonuclease A (EndA) (197). The alveolar epithelia secrete lysozymes and phospholipases, which, in synergism with β-defensins, participate in lysing bacteria. Lactoferrin, which functions directly on the pneumococcal cell as an antimicrobial agent (198–200), also impedes that bacteria may utilize iron by removing it from the environment, strategy known as nutritional immunity (201,202).

The pneumococcus responds to host threats by expressing an arsenal of surface-exposed proteins (72–74). They facilitate not only adhesion, but also protect the bacterium by sequestering complement regulatory proteins, and protect against oxidative stress (203–207). A particular event happens once the pneumococcus reaches the epithelial surface. The bacterium can then translocate via transcytosis to the deep epithelial layers and subsequently, to the vasculature. The bacterium exploits two mechanism to accomplish translocation. The bacterium's adhesin choline-binding protein A (CbpA, also known as PspC), binds to the surface exposed polymeric immunoglobulin receptor (pIgR) on the epithelial cells. This interaction triggers a signaling cascade that leads to the internalization and translocation of the bacterium across the epithelial cells (208–210). The second mechanism involves the exploitation of the platelet-activating factor receptor (PAFR) pathway, which binds to phosphorylcholine moieties (PCho) presents in the pneumococcal membrane teichoic and lipoteichoic acids. The mimicry of the natural PAFR ligand by the pneumococcus facilitate binding to the receptor and subsequent internalization and transcytosis through the endothelial cells (211). While these translocation events are rare, they emerge as one explanation to how the pneumococci reach the host's bloodstream (212).

The amount of damage to the lung and the alveolar epithelium is the primary determinant of the progression from pneumonia to IPD. In the alveoli, the pneumococcus rapidly multiplies and spread to establish a pneumonia state. The bacteria release the toxin pneumolysin that binds host cells and then oligomerizes, forming large transmembrane pores (213). The pneumococcus also produces an extensive amount of hydrogen peroxide via a pyruvate oxidase encoded by *spxB*, which helps to disrupt the alveolar epithelium further (214). In conjunction, pneumococcus induces a strong inflammatory response.

The progression from pneumonia to IPD can be accelerated when the bacterium encounters pulmonary epithelium previously erode by a preexisting viral infection (85). Bacterial translocation into the blood can also be enhanced by damage to the pulmonary epithelium directly. The integrity of the epithelial barrier can be disrupted by the intense inflammation; it has been shown that inflammation downregulates the host

tight junctions. Therefore, both inflammation and damage infringed directly by the pneumococcus, can provide access to deep connective tissue and vasculature.

Once the pneumococcus leakages into the bloodstream, the bacteria encounter the coagulation complement system. To evade opsonization and subsequent phagocytosis, the pneumococcus initiates a second phase variation. Thereby becoming more encapsulated and reducing the expression of adhesins. The charges in the capsule strongly repel complement deposition, thus limiting phagocytosis of the bacteria (205,215,216). The pneumococcal surface proteins PspA, CbpA, and pneumolysin, also participate in the inhibition of the complement deposition by binding directly key complement activators, sequestering them, and in doing so, inhibiting their interaction with other complement components that triggers complement deposition (215). Due to the rich environment in the bloodstream, pneumococci copiously replicate, causing bacteremia.

### 1.4.2 Pneumococcal biofilms

Every bacterium has to collaborate in multicellular communities to survive. They have to thrive not only with the dynamic host environment and the immune clearance but also with the challenge of living with diverse microbes residing in the same ecological niche. Bacterial biofilms facilitate the survival of pathogens in their environment. It is estimated that up to 80% of the earth's microbial biomass lives in the form of biofilm (217). Biofilms play a crucial role in bacterial virulence during infection. Biofilms are sessile, adherent structures that serve as communities where pneumococcal cells interact not only with each other but also with other species, collaboratively, or competitively (112). The pneumococcus forms biofilms on the mucosal of the nasopharynx and the middle ear, during asymptomatic colonization (174,218,219). It has been shown that in the middle ear, the pneumococcus usually forms polymicrobial biofilm communities with other bacteria, including *H. influenzae* and *M. catarrhalis* (220). Otitis media, in all its manifestations, is explained by the fundamental nature of the biofilms. They challenge the treatment with antibiotics since biofilms become increasingly recalcitrant to antibiotic treatment, facilitate the adhesion

and cohesion of the cells to the host cells and each other, and hide bacterial cells from immune surveillance (221). Thus, posing biofilms as an important problem for children worldwide.

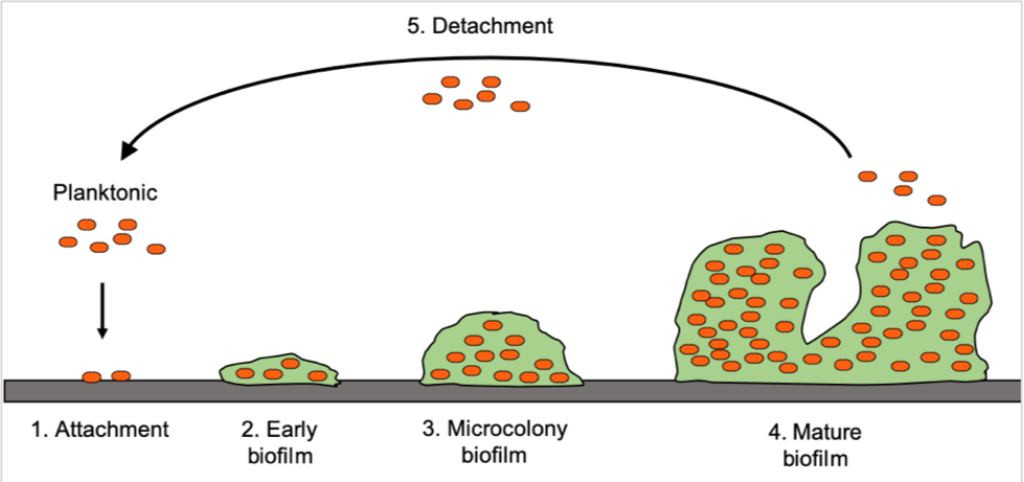



**Figure 1.3 Schematic of a biofilm formation and dispersion.**
**1.** Free-floating bacteria reversible attach to a surface forming an early biofilm. **2.** The bacteria irreversible attach to the surface and start producing the extracellular polymeric substance matrix (EPS). **3.** Biofilm grows and forms multilayer microcolonies. **4.** Mature biofilms. Many of the microcolonies fuse augmenting the total biomass. **5.** Some cells detach from the biofilm commencing the dispersion.

Bacteria living in biofilms are phenotypically distinct from their planktonic counterparts. Several *in vitro* models have been developed to study pneumococcal biofilms using systems that make bacteria grow in inorganic surfaces like cellulose, polystyrene, or glass (222,223). The next step in biofilm research was the inclusion of a continuous flow chamber that allows the constant influx of nutrients, imposes mechanical stress, and recover of metabolites released by the biofilms (224,225). This model is commonly used to study different aspects of pneumococcus biology, such as the effect of antibiotics, bespoke clone mutants, spatial and temporal gene expression, and secretion of extracellular polymeric substance matrix (EPS). While striking transcriptional differences were shown during either biofilm or planktonic mode of growth, pneumococcal bacteria exhibited an attenuated invasive capacity (226). It was not until the group of Marks *et al.* demonstrated that bacteria dispersed from a biofilm

display increased proneness for tissue dissemination and pathogenesis when compared to cells from the biofilm layer or planktonic cultures (227,228). The realization that biofilms' behavior depends largely on experimental conditions has led to the development of innovative experimental systems that resemble *in vivo* invasion, like biofilm growing on epithelial cells (229–231).

Bacterial cells reside adhered to a substratum shielded by the EPS matrix (232). The EPS physically restrict the diffusion of compounds into the biofilm, allowing the bacterial cells to remain and persist to antibiotics, antimicrobial compounds, as well as keeping the host’s immune system at bay (221). The EPS matrix is composed of lipids, proteins, and nucleic acids, but the exact composition differs between bacterial species (233). In *S. pneumoniae* it has been shown the extracellular DNA (eDNA) is an important component of the matrix (223,234). Competence drives the secretion of eDNA by activating both bacteriocin (235) and fratricide mechanisms (236,237), and by the production of the autolytic lysozyme C, LytC (223).

Genetics and proteomic changes have been reported in biofilms compared to planktonic (226). It was then shown that pneumococcal biofilms express exclusive genes involved in cell wall and isoprenoid biosynthesis, as well as purine and pyrimidine nucleotide biosynthesis (238). Furthermore, it is now recognized the genetic and molecular heterogenicity in the bacterial population residing in biofilms makes the study and research of biofilms challenging. Compared to free-floating bacteria, bacteria in biofilms are protected from the host immune system. It has been shown that, in general, the pneumococcus reduces the general protein production in biofilm, which suggest that, by adopting a diminished metabolic state, it facilitates the immune response escape (226). Orihuela and colleagues also showed that serum obtained from patients with pneumococcal pneumonia reacted stronger to planktonic cell lysates while showing a poor response to biofilm lysates. Further, they showed that by immunizing mice with killed biofilm, their serum responded stronger to biofilm. From this observation, the study concludes that the immune response against pneumococcal biofilm is diminished by reducing or hissing any potential epitope that the immune system may utilize to mount a proper response (226). Also, pneumococcal cells express twenty-one ABC transporters that function as exporters of antimicrobial peptides, as well as signaling

peptides or drugs, from the cytoplasm, hampering, even more, the immune response (187). Finally, the biofilm structure physically limits the diffusion of antimicrobial compounds into the biofilm: the charged polymers of the biofilm matrix might hamper diffusion or the tunneling system in the biofilm could route antimicrobial agents from deeply embedded cells (239)

In conclusion, pneumococcal biofilms are a highly dynamic and heterogeneous community. Bacteria use this structure to strategically evade and survive the immune vigilance posing these structures as an important mechanism for bacterial persistence *in vivo* and of high clinical importance.

## 1.5 Surface molecules and virulence factors

Virulence factors are molecules produced by microorganisms to achieve survival. It was originally considered that the capsule was the primary virulence factor of pneumococci. Over the years, it has become clear that virulence factors and the susceptibility of the host determines the course of bacterial invasion. As a result, the number of characterized virulence factors has increased dramatically. Many pneumococcal surface molecules have been shown to be virulence factors that mediate attachment of the bacteria to host cells (**Figure 1.6**) (72–74). Here, I will discuss relevant virulence factors, whose expression determines biofilm capacity and attachment, to ultimately achieve the survival of the bacteria over the host (the capsule is discussed in section 1.2).

The pneumococcal surface is decorated with a variety of proteinaceous molecules with diverse functions. Surface proteins are classified into five groups based on the molecule or structure to which they attach or bind. These are 1) transmembrane proteins, 2) membrane-anchored-proteins (lipoproteins), 3) cell-wall-anchored proteins (LPXTG-anchored proteins), 4) choline-binding proteins, and 5) non-classical surface proteins (21). Due to its unique, multilayered peptidoglycan structure, the majority of the known surface-exposed proteins are located at the cell-wall. The most versatile of this group are the choline-binding proteins family. They remain anchored to the pneumococcal cell-wall via interaction with the choline residue present on both cell-wall-

associated teichoic acids and the membrane-associated lipoteichoic acids (240). The pneumococcus encodes 10 to 15 choline-binding depending on the strain (72), including, PspA, PspC, and LytA. Some of the most relevant surface protein will be described below.

**Pneumococcal surface protein A (PspA).** This choline-binding protein is detected on the surface of virtually all pneumococcal strains (241). PspA inhibits the complement activation by interfering with the binding of complement component C3 on the surface of the bacteria. Hence, inhibiting the complement deposition and protecting the bacteria by blocking complement-mediated opsonization (215,242). It has also been shown that PspA also binds to apolactoferrin (the iron free-form of lactoferrin), sequestering from solution, and minimizing the of growth inhibition by iron depletion (243). Thus, a PspA mutant shows an attenuated virulence by reducing the bacterial capacity of scape, both complement deposition and iron-depletion mediated nutritional immunity (201,202). Similarly, the widely distributed cell wall associated serine protease A (PrtA) is capable of degrading apolactoferrin (244), and it is required for bacterial virulence in a model of animal sepsis (245).

**Pneumococcal surface protein C (PspC).** PspC is a multifunctional cell-surface choline-binding protein. Due to its polymorphism and function, it is also annotated as CbpA due to its strong binding to choline. PspC also binds the polymeric immunoglobulin receptor (pIgR) (212,246–248), thus it is also called secretory pneumococcal surface protein A (SpsA) (208–210). Its unique ability of binding pIgR is complemented with its capacity of interacting with the complement regulatory protein factor H, thus bridging bacterial attachment to epithelial cells (209). Furthermore, factor H-coated pneumococci are able to evade the host complement and immune attack (249). The bacterium exploits these mechanisms to accomplish translocation. The bacterium binds pIgR on the epithelial cells via PspC, triggering a signaling cascade that leads to the internalization and translocation of the bacterium across the epithelial cells (209,250). By utilizing the Factor-H-mediated update, pneumococci also accomplish internalization and presumably, translocation (208). In fact, a *pspC* deficient mutant has a diminished nasopharyngeal colonization capacity (251) as well as virulence in an animal model of sepsis (252).

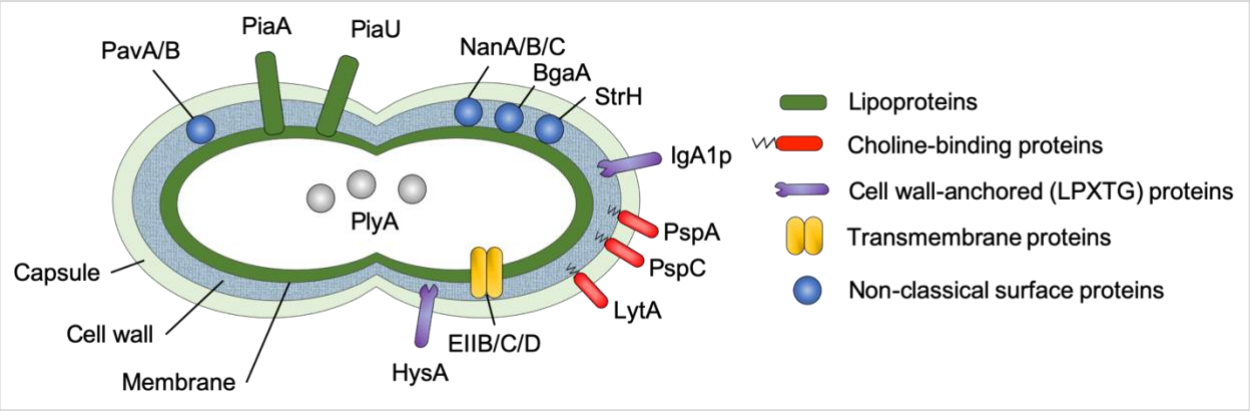



**Figure 1.4 Schematic of surface protein types and pneumococcal virulence factors**

The capsule, the cell wall, and bacterial membrane are depicted. Choline-binding proteins are Pneumococcal surface protein (PspA and PspC) and autolysin (LytA); lipoproteins are Pneumococcal iron acquisition A (PiaA) and pneumococcal iron uptake A (PiuA); cell wall-anchored proteins are Immunoglobulin A1 protease (IgA1p), hyaluronate lyase (HysA), and Pneumococcal adhesion and virulence A and B (PavA and PavB); non-classical surface proteins are b-galactosidase (BgaA), neuraminidases A, B and C (NanA, NanB and NanC), and β-N-acetylglucosaminidase (StrH); transmembrane proteins are PTS component (EIIB/C/D); cytosolic Pneumolysin (PlyA) is also depicted. Image adapted from Kadioglu *et al*., 2008).

**The pneumococcal autolysin LytA (LytA).** This choline-binding enzyme is a member of a widespread group of cell wall-degrading amidases that play roles in a variety of physiological functions associated with cell wall growth, cell wall turnover, and cellular separation during cell division (253). In the pneumococcus, LytA is the major autolysin. It cleaves the amide bond between N-acetylmuramic acid and the first amino acid of the stem peptide, L-alanine, of pneumococcal peptidoglycan. Autolysin plays a direct role in virulence by mediating the release of peptidoglycan, teichoic acid, and pneumolysin after bacterial lysis (32,254,255). LytA also enhances bile solubility: this particular feature has been historically used to diagnose pneumococcal infection (256). The pneumococcus encodes two other autolysins: LytB and LytC. LytB is involved in cell separation during growth as a mutant deficient in *lytB* displays long pneumococcal chains (257). Additionally, LytB localizes to the poles of the bacteria during division

(258). LytC, the less studied of the three enzymes, has been proposed as an autolysin required for growth in the nasopharynx. This idea came after the realization that the activity of this enzyme is maximum at 30°C, a temperature observed in the nasopharynx (259).

**Pneumococcal surface antigen A (PsaA).** PsaA is a cell-membrane metal-binding lipoprotein, component of a transport system involved in the manganese uptake. This surface molecule was originally thought to be an adhesin due to the fact that a *psaA* mutant displayed a reduced attachment capacity to epithelial lung cells (260) and reduced virulence capacity (261). It is now accepted that the deficiency was due to a pleiotropic effect due to manganese deficiency and its effect on the expression of other genes (262). Furthermore, it is believe that PsaA may have a role in pneumococcal resistance to oxidative stress resulting presumably, from the antimicrobial action of the host immune response (261,263).

**Pneumococcal iron acquisition A (PiuA)** and **Pneumococcal iron uptake A (PiaA).** These molecules are lipoprotein metal-binding components encoded by two independent operons (*piu* and *pia*), whose functions are the uptake of iron (139). These proteins were discovered by Brown *et al*., who originally termed them as Pit1A and Pit2A (Pneumococcal iron transport). They subsequently showed that immunization with these purified proteins highly protects against infection. This protection was shown to be mediated by antibodies resulting in an augmented complement-mediated opsonophagocytosis of the bacteria (139,140), leading to a reduction in respiratory infections in a mice model (264). Due to its high immunogenic capacity, these two lipoproteins were incorporated in an experimental pneumococcal vaccine, whose formulation has been shown to be as effective as commercially available vaccines (149).

**Immunoglobulin A1 protease (IgA1p).** This LPXTG-anchored surface associated-enzyme contributes to limits the effectiveness of the host immune response. This enzyme can reduce IgA1-mediated complement-mediated opsonization by cleaving IgA (148,196). It is worth noticing that in the serum, opsonization of the bacteria can still be achieved by the deposition of IgG, which is not targeted by IgA1p

and is highly induced after infections or immunization with pneumococcal vaccines (265). IgA1p enhances bacterial attachment to host cells and is required for successful colonization and infection (148,196). Due to its high immunogenic capacity, IgA1p was incorporated in an experimental vaccine along with the pneumococcal proteins PpmA (Putative proteinase maturation protein A) and SlrA (Streptococcal lipoprotein rotamase A) (146–148). The last two have been shown to contribute to pneumococcal colonization of the nasopharynx and airways in mice models (266). Pneumococci genome encodes three additional zinc metalloproteases termed as ZmpB, ZmpC, and ZmpD. Similar to AgA1p, they are all anchored to the bacterial cell wall. However, ZmpB does not play a role in complement-mediated killing (267), whereas it was reported that interacts with collagen IV instead (268). ZmpC is a metalloprotease involved in the removal of the host mucins from the epithelium, thereby enhancing bacterial attachment (269). ZmpD is variably present among pneumococci and its function and substrate specificity are unknown (270,271)

**Pneumococcal adherence and virulence factor A** and **B (PavA** and **PavB).** These proteins are representative of a subset of adhesins that bind to extracellular matrix components. PavA and PavB localized on the pneumococcal outer cell surface and has been suggested to serve as a pneumococcal adhesin for fibronectin and facilitate adherence to epithelial and endothelial cells (272). Also, it has been shown that PavA protects against cytoskeleton-mediated phagocytosis by dendritic cells (273), and it has a role in the maintenance of bacterial colonization and sepsis (274). Similarly, PavB plays a role in colonization and infection of the respiratory tract (33).

**Neuraminidases**. The Pneumococcus encodes several surface-exposed glycosidases (non-classical proteins), which have been shown to contribute to adhesion not only via modification of the host surface that reveal receptors but also via direct binding (77). The glycosidase **NanA** is required for colonization of the nasopharynx (275), is conspicuously expressed for the majority of pneumococcal strains (276), and it can remove sialic acid from soluble proteins, such as lactoferrin and the immunoglobin IgA2 (277). NanA also promotes biofilm formation (278). It has been shown that another pneumococcal neuraminidase, **NanB** functions at a much lower pH than NanA (279). Further, **NanC**, the less commo of the known neuraminidases, is more frequently

detected in isolates from cerebrospinal fluid than in carriage isolates (276). These observations have led to the hypothesis that the neuraminidases are required at different instances during bacterial infection: NanA seems required during nasopharynx colonization, NanB seems required for dissemination into the lower airways, while NanC seems required for invasive dissemination. However, it has been also shown that pneumococcal glycosidases including NanA, NanB, BgaA (β-galactosidase), StrH (β-N-acetylhexosaminidase), LacA (galactose 6-phosphate isomerase), among others, work all in concert to release sugars found in complex host glycans (280).

It has been suggested that by changing the glycosylation patterns of the host cell, the bacteria may expose host surface receptors for potential interactions (281). However, this idea is currently challenged by the evidence that the glycosidase **BgaA**, required for the release of terminal galactose residues from glycoconjugates, promote attachment to host cells (37). Further, BgaA seems to function in a manner independent of its enzymatic activity, whereas attachment is mediated via the catalytic module of the enzyme (282). In this line, attachment can also be mediated by other non-adhesin molecules, such as enolase (35,36). Thereby, the pneumococcal can employ non-adhesin molecules and glycosidases for attachment to the host.

**Hyaluronidase A (HysA).** Hyaluronidases belong to the LPXTG-anchored cell wall-bound glycosidases. HysA, also known as SpnHL, Hyl, and HylA, breaks hyaluronic acid disaccharides present in hyaluronan molecules of the mammalian connective tissue and the extracellular matrix (283). In the pneumococcus, the processing and internalization of hyaluronic acid are carried out by a unique set of enzymes: a surface associated hyaluronidase, a transport system, and a set of intracellular enzymes (**Figure 1.7**) (187,284). Several genes participating in the metabolism of hyaluronic acid have been linked to bacterial virulence. Specifically, *hysA*, the transporter-encoding genes (EIIB, EIIC, EIID) and an intracellular hydrolase (*ugl)*, are all required for: hyaluronic acid utilization, biofilm formation, thus contributing to virulence (285,286). By removing components off the ECM, HysA may potentially unmask cell-surface receptors, which are then accessible for pneumococcal adhesins, facilitating the pneumococcal penetration into host cells. A similar mechanism was proposed for the neuraminidases (281). However, a *hysA* deficient strain showed little to

no changes in virulence compared to its parental wild-type strain in animal infection model (254,287). This observation suggests that a functional redundancy may compensate the lack of HysA. In fact, a more recent study linked the genes encoded in the transporter operon to HysA-independent attachment to components of the ECM of epithelial cells of (231).

**Pneumolysin A (PlyA, ply).** PlyA is the major pneumococcal toxin. It is produced as a soluble monomer that binds to cholesterol present on the membrane of the target cells and oligomerizes forming a large ring-shaped transmembrane pore (213). It is widely distributed among all clinical pneumococci isolates, and the sequence is well-conserved. Pneumolysin modulates a plethora of cytolytic activities. Physiological, PlyA has several functions: it inhibits the ciliary beating on the respiratory epithelium, facilitating bacterial dissemination by reducing the bronchial ability to clear the mucus (179); it is able to inhibit the oxygen-dependent respiratory burst in monocytes and peroxide production (288); induces the cytokines production leading to activation of the cell-mediated immunity (289); and participates in the inhibition of the complement deposition (290), among others. Due to its profound effect on the host physiology, pneumolysin is also a potent initiator of host immune response (62,135), in fact, elevated antibody titers against pneumolysin are detected in otitis-prone children (136). Pneumolysin can also break the integrity of bronchial epithelium and alveoli by disrupting their tight junctions causing alveolar edema and hemorrhage, thus, facilitating pneumococcal dissemination from the lungs into the bloodstream (291–293). In a mouse model of acute pneumonia, pneumolysin was shown to be essential for pneumococci survival: deletion of *plyA* will lead to attenuated virulence of pneumococcus (255,294,295).

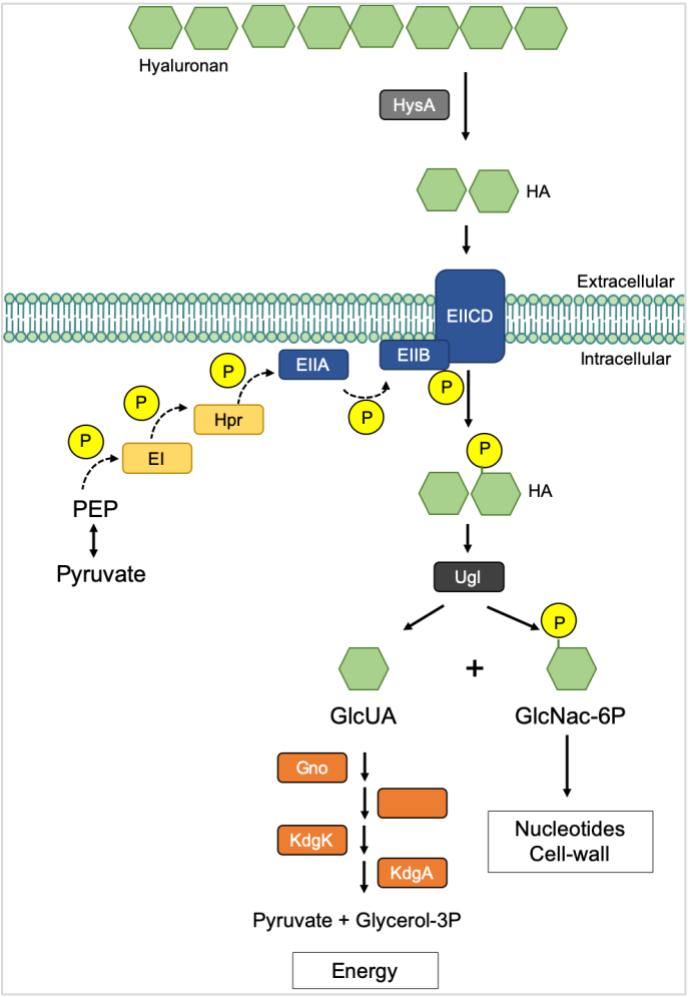



**Figure 1.5 Hyaluronic acid is imported by a PTS mechanism in *S. pneumoniae.***

The internalization of hyaluronic acid (HA, green) across the membrane is depicted for a phosphoenolpyruvate-dependent phosphotransferase system (PTS). Hyaluronan is depolymerized to the disaccharide HA acid by the exposed HysA. HA is incorporated into the cytoplasm and is phosphorylated across the cytoplasmic membrane by a PTS. The PTS is composed of two general soluble proteins enzyme I (EI) and the histidine protein (HPr) (yellow). Enzymes IIA (EIIA) and IIB (EIIB) are the cytoplasmic phosphocarriers, and the enzymes IIC (EIIC) and IID (EIID) create the membrane channel (blue). Internalized HA is degraded to constituent monosaccharides (N-acetyl-D-glucosamine and D-glucuronic acid) by Ugl. Resultant glucuronic acid is metabolized to pyruvate and glyceraldehyde-3-phosphate by subsequent reactions of isomerase, reductase, kinase and aldolase. N-acetyl-D-glucosamine is recycled for other functions. Abbreviations: PEP, phosphoenolpyruvate. GlcNAc, N-acetyl-D-glucosamine. GlcUA, D-glucuronic acid. HysA, Hyaluronidase. HA, hyaluronic acid. P enclosed in a yellow circle, phosphate group.

In conclusion, the majority of surface associated molecules in the pneumococcus also function as virulence determinants. The examples discussed in this section make us appreciate the pneumococcal biological simplicity. With a small toolset at its hands, this bacterium is capable of attaching and invading every human cell and organ. Due to its diversity, and sometimes redundancy, virulent factors may have more or less impact on pneumococcal survival. While pneumolysin and the pneumococcal capsule are the most important virulence factor, the interplay between all the virulence (and non-virulence) factors cannot be understated.

## 1.6 Bacterial cell-cell communication

To adapt and survive to the dynamics of an ever-changing environment, bacteria utilize a diverse armament of pheromones for self-communication, adaptation, and colonization, and to outcompete other microorganisms (296). These mechanisms involve the production and release of small molecules that set-in motion a variety of microbial responses. Such coordination is achieved via the secretion of auto-inducing molecules, collectively called quorum sensing (QS) molecules (297–304). QS enables coordination of activities across the community to simultaneously and synchronously regulate gene expression and, therefore, behavior on a community-wide scale. This population-level behavior is coordinated by small molecules that accumulate in the extracellular milieu and subsequently activate signal transduction pathways in the bacteria (174). A large variety of bacterial processes is controlled by QS, including the development and dispersal of biofilms, the triggering of pathogenic attack, and the assault on neighbors to acquire new genetic material (297–304).

Central to population-level behaviors is the ability to sense and respond to the environment clues, which play a crucial role in this response. By sensing the extracellular environment, the bacteria can modify its gene expression and adjust their metabolism to survive environmental changes. Gram-positive bacteria coordinate its response via signaling cascades that involve the secretion of extracellular peptides. QS signaling systems are critical for DNA uptake, colonization, immune evasion, and virulence (180–182). They also orchestrate community behaviors to cause disease (305). All of the QS peptides described to date among *Firmicutes* are synthesized by the ribosome and may differ in their processing, secretion, and signaling abilities (306). Several types of QS pathways have been described, and they may operate within a species to control a wide variety of processes. The RRNPP systems (expanded below) have been extensively investigated in several species, including *B. subtilis*, *E. faecalis*, and the *B. cereus* group, where they play in general a regulatory role (307).

In general, the genomic organization of QS systems follows a common pattern: a) the transcriptional regulators and the signaling pro-peptide are both encoded in proximity in the bacterial chromosome, and b), the mature signaling peptide is derived

from a pro-peptide which is enzymatically cleaved. There are two well-characterized classes of secreted signaling peptides in pneumococcus: the RRNPP-regulated signaling system and the Gly-Gly family of signaling peptides. In the RRNPP-regulated system, a signaling peptide enters the cells and binds directly to cytosolic transcription factors modulating its activity (307). While Gly-Gly family of signaling peptides bind outside the cell to a surface-exposed kinase, triggering a intracellular cascade (308,309). The next sections will highlight the differences and similarities between these two family of peptides.

### 1.6.1 An overview on RRNPP-regulated communication

The RRNPP systems have been extensively investigated in several species, including *B. subtilis*, *E. faecalis*, and the *B. cereus* group, where they play, in general, a regulatory role (307). Although it has been shown that this family of transcription factors controls diverse processes in different bacterial species, they share a common structural organization. Based on these structural similarities, in 2007 it was proposed that these proteins belong to a superfamily collectively denominated as RRNPP, which stands for Rgg (Regulator Gene of Glucosyltransferase in *S. gordonii*), Rap (*Bacillus* response regulator aspartate), NprR (*Bacillus* neutral protease regulator), PlcR (*Bacillus* phosphatidylinositol-specific phospholipase C gene regulator), and PrgX (Enterococcus pheromone-responsive transcription factor) (298). Their genomic organization follows a common pattern: the transcriptional regulators and the signaling pro-peptide are both encoded in proximity in the bacterial chromosome. The mature signaling peptide is derived from a pro-peptide, which is enzymatically cleaved during secretion (307). This process is carried out by cytosolic or membrane-anchored proteases or by ABC transporters, which also may function as proteases (310). Pheromones that bind to RRNPP regulators are subsequently transported into the cytoplasm by the oligopeptide permease Opp or Ami/Ali in *B. subtilis* and *S. pneumoniae,* respectively (311,312). In the cytoplasm, the pheromones interact directly with an RRNPP family member to modulate gene expression (313,314). It is noteworthy that this family comprises all

Gram-positive QS systems in which the signaling peptide binds directly to the transcriptional regulator inside the recipient cell (315).

There are multiple RRNPP systems encoded in the pneumococcal pangenome (316–318), and cognate peptides have been identified for two of the regulators: the quorum sensing peptide PhrA that binds the PlcR-like negative regulator TprA and controls the expression of an antimicrobial peptides gene cluster (312), and the paralogue system PhrA2/TprA2 for which a cross-talk with the former system has been reported (319) and it has also been shown to promote survival of the pneumococcus in the lung (319,320). Next section will cover the RRNPP system called Rgg/SHP.

### 1.6.2 The Rgg/SHP system in the pneumococcus

Rgg proteins are a distinct subgroup of transcriptions factors widespread in the Firmicute species, including the Streptococcaceae, Lactobacillales, Listeriaceae, and Enterococcaceae. In all those species, Rgg transcription factors plays a regulatory function (307,321,322). Furthermore, phylogenetic analyses suggested that they share common ancestor (314,323,324). Structurally, this family is characterized by an N-terminal helix-turn-helix motif (HTH), which binds directly to DNA. The C-terminal of the transcription factors bears a tetratricopeptide series of alpha-helices repeats, which works as a regulatory domain and form the pocket (298,307).

Rgg transcription factors are activated upon binding to a cognate peptide called SHP (short hydrophobic peptide). In this architecture, the mature SHP peptide is derived from a pro-peptide, which is encoded in the Rgg vicinity. After translation, it is exported in a pro-peptide form. For example, in *Streptococcus agalactiae* export is carried out by an ABC transporter denominated PptAB (325,326). In the pneumococcus, the exporter remains to be determinate. The secreted SHP is enzymatically cleaved during exit of the cell. An extracellular protease can further process it. Next, the mature peptide is internalized and binds to its cognate Rgg, leading to altered expression of genes regulated by the Rgg in the receiver cell (307,327). SHP peptides fall into three categories, based on the nature of the charged residue and organization relative to the Rgg. Pneumococcal genomes only encode for SHPs type I and III (307). These SHPs

are predicted to be processed and secreted by ABC transporters and internalized by members of the conserved oligopeptidase permease family (326). For example, in *S. mutants* the aminopeptidase PepO (SPR1491 in the pneumococcus) stops the signal by degrading the SHP peptide (326). The molecular components of trafficking provide clues regarding the likelihood that signaling peptides are functional across strain and species barriers (328). By compartmentalizing the pre-processed and mature form, the system avoids self-cytosolic signaling, and the bacteria respond only to the extracellular mature peptide once it gets internalized.

Rgg regulators (historically called MutR-like or MutA-like regulators) control a variety of bacterial processes, including regulation of the virulence and biofilm capacity in *S. pyogenes,* stress response in *L. lactis* (329)*,* and bacteriocin production in *S. mutans* (330), biofilm and capsule production (331), virulence (332), and colonization (318), among others. It has been suggested that they may also work as global regulators in some organisms (333). The signaling and specificity of the Rgg/SHP system depend largely on the secreted pheromone (334). They can be either activators or repressors of transcription. For instance, in *S. pyogenes* there are two well-characterized Rgg/SHP systems: Rgg2/SHP2 and Rgg3/SHP3. While related, both Rgg transcription factors control transcription with antagonistic activities; Rgg3 binds to the promoter of both *shp2* and *shp3* and represses their transcription independent of the pheromone presence. When SHP3 is internalized, Rgg3 releases from the promoters, and transcription of *shp2* and *shp3* is unblocked. Then, the pair Rgg2/SHP2 can now initiates the transcription of both *shp2* and *shp3* as well as other downstream genes. The abundance of SHP2 and SHP3 leads to a positive feedback loop, resulting in an amplified Rgg/SHP circuit (335,336).

In the pneumococcus, Rgg/SHP systems have been implicated in oxidative stress and metabolism (316). The best characterized Rgg/SHP systems in the pneumococcus are Rgg144 and Rgg939. These transcription factors respond to changes in nutritional status of the host by inducing virulent factors and regulating the capsule genes and play a role in colonization and virulence (318,331,332). The activity of Rgg144 (nomenclature based on gene number in the strain D39, SPD_0144) is regulated by a cognate peptide SHP144 (the number denotes Rgg/SHP

correspondence), which is secreted from cells and internalized into producing and neighboring cells where it binds to Rgg144 (318,332). Rgg144 is a negative regulator of the capsule synthesis. The transcription factor binds a predicted promoter upstream of the capsular-like genes, inhibiting the expression of capsular-like genes. This inhibition may lead to a reduction of the capsule size, as seen in type 2-capsule strains (318), and presumably, increased bacterial attachment: a thicker capsule may obstruct attachment by restricting exposure of bacterial receptors (337). Rggs may work as an interconnected network: a cross-talk (**Figure 1.4**) was suggested between Rgg144, Rgg939, and a less characterized Rgg1518 (318). Both QS system, Rgg144/SHP144 and SHP939/Rgg939, reduce the expression of a gene cluster encoding capsule-like genes, while it has been shows that pneumococci reduces the capsule during colonization and biofilm formation (183,184). In line with these observations, murine models showed that both of these QS systems, SHP144/Rgg144 and SHP939/Rgg939, play a role in colonization of the nasopharynx and contribute to the pathogenesis of the pneumococcus.

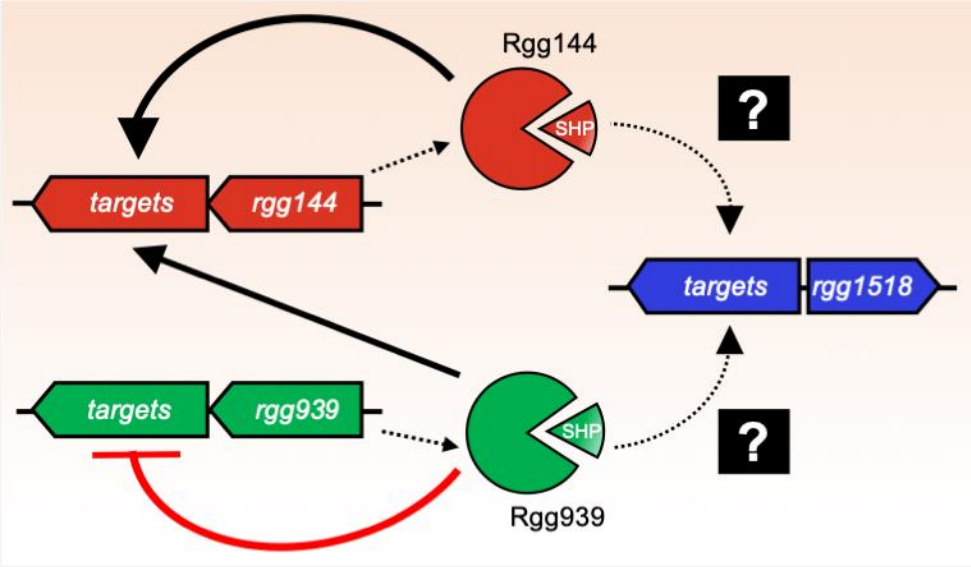



**Figure 1.6 Interconnectivity between the transcription factors Rgg144 and Rgg939**
The Rgg transcription factors and their respective operon are color-coded. Black lines indicate transcriptional upregulation. Dashed lines indicate potential interconnection between the Rgg systems. Schematic based on the works published by Cuevas in 2017 and Zhi in 2018 **(318,332)**.

### 1.6.3 Gly-Gly signaling peptides

Pneumococcus secretes small peptides that influence colonization, virulence, and adaptation via effects on competence, intra-species competition, and biofilm development. Havarstein *et al.* pioneered the field of secreted signaling peptides with the discovery of the competence stimulatory peptide (CSP) (338), which stimulates the natural pneumococcal competence, that is, the ability to acquire extracellular DNA and incorporate it into their genome (27). The group established that a leader sequence in CSP (LSXXELXXIXGG) is cleaved off from peptide, in a process concomitant with the externalization of CSP (339). They further established that the cleavage occurs after the double glycine sequence (Gly-Gly).

Two-component regulatory system (TCS) predominantly coordinate and translate the effect and expression of these peptides. In this architecture, the peptide binds to the surface-exposed of the histidine kinase receptor (HK, component one). The binding stimulates autophosphorylation of the cytosolic HK tail. Then, the HK transfer the phosphate group to a transcription factor, activating it. Finally, the transcription factor (also known as response regulator or RR), moves to the bacterial chromosome to regulate expression (340). The pneumococcal genome encodes for thirteen TCSs (annotated from TCS01 to TCS13) (341,342), and they are usually found grouped in operons with either *hk-rr* or *rr-hk* configuration (21,343). Diverse functions have been associated with TCSs that includes antibiotic resistance, competence, immune evasion, virulence, among others (344).

Pneumococcal antimicrobial peptides, collectively termed bacteriocins or bacteriocidal peptides, feature a Gly-Gly sequence. With the exception of few bacteriocins, which contains a sec-dependent leader sequence (345,346), the vast majority of bacteriocins class I (lantibiotics) and class II (non-lantibiotics) bacteriocins are transcribed as immature peptides and contains a Gly-Gly sequence (347,348).

The Gly-Gly feature has been identified in various peptides that function as signaling molecules. They have been shown to conduct multiple bacterial behaviors, including competition, virulence, and colonization (307,327). The best characterized

Gly-Gly peptides in the pneumococcus are CSP, the pheromone bacteriocin inducing peptide (BIP), and more recently, the virulent peptide 1 (VP1).

The pneumococcus is naturally able to bind, internalize, and integrate into its chromosome extracellular DNA by recombination (349). This capacity, known as competence, is stimulated by the Gly-Gly pheromone CSP (encoded by *comC*). The leader peptide of the immature CSP is cleaved after the Gly-Gly sequence by the transporter ComAB during secretion (338). The now mature and active peptide binds to the exposed HK receptor ComD, triggering its autophosphorylation. Then, HK phosphorylates the transcription factor ComE, leading to the induction of early competence genes, including the operon *comAB* and the TCS *comCDE*. Note that CSP also stimulates its expression. Additionally, ComE also stimulates the expression of the alternative sigma factor ComX, which drives the expression of late competence genes (338,350), hence, modulating several pneumococcal phenotypes including transformation, metabolism and biofilm capacity (351,352).

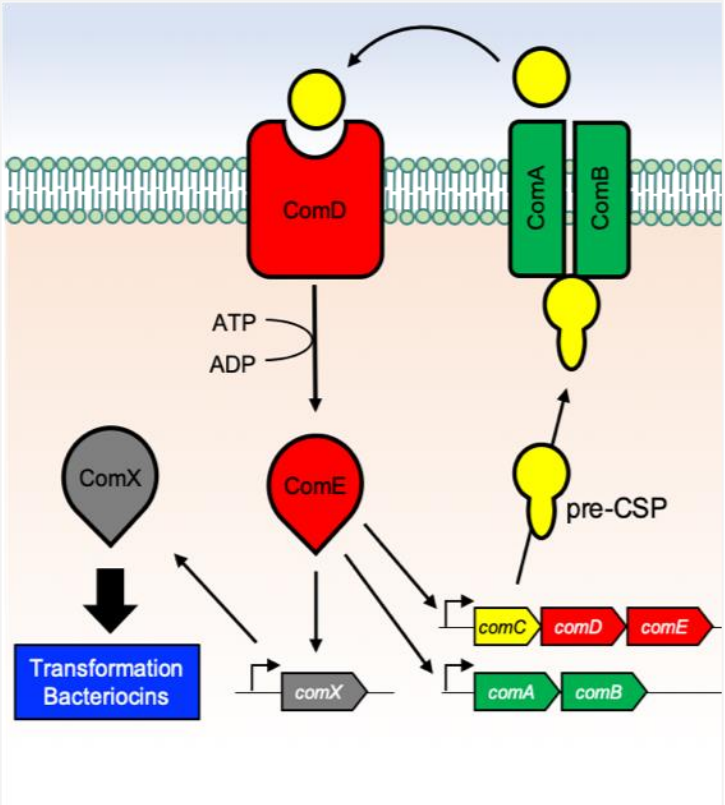



**Figure 1.7 The Gly-Gly peptide CSP control of competence**
Immature CSP is processed and exported by the ABC-transporter ComAB. The mature CSP then binds to the histidine kinase receptor ComD, which in turn, phosphorylates and activates the transcription factor ComE. Phosphorylated ComE stimulates the transcription of *comAB*, *comCDE* amplifying the signal. ComE also estimates the transcription of *comX* (and other early competence genes). ComX stimulates the transcription of genes required for DNA update (competence). The expression of bacteriocidal genes (bacteriocins and fratricide molecules) is also stimulated. Bacteriocidal peptides are then released to induce lysis and DNA release in non-competent bacterial cells.

Among the inhibitory bacteriocins in the pneumococcus, the pneumocin mediates intraspecies competition. The pheromone BIP, encoded by *blpC*, controls the expression of an array of genes including bacteriocin-encoding genes, immunity-to-bacteriocin genes, as well as genes involved in thee peptide processing and signaling (353,354). The immature BIP possesses a Gly-Gly sequence, and the immature peptide is processed and secreted via the transporter BlpAB. The active BIP binds the HK receptor BlpH, which autophosphorylates and transfers the phosphorylation to its RR pair, the transcription factor BlpR. This interaction leads to an upregulation of the entire *blp* locus via a positive feedback loop, resulting in expression of diverse bacteriocins and related bacteriocins immunity proteins (353,354).

More recently, we described the Gly-Gly peptide VP1, a new member of a family of Gly-Gly peptides with a highly conserved N-terminal sequence, which is exported into the extracellular milieu, presumably also via ABC transporter(s) (332). The *vp1* gene is widely distributed across pneumococcal strains, as well as encoded in related streptococcal species. The gene encoding VP1 is a virulence determinant and is highly upregulated in the presence of host cells where it promotes bacterial attachment and biofilm development (230,231,332).

## 1.7 Aims of the thesis

The overall aims of this dissertation are to evaluate the role of the secreted peptide VP1 in pneumococcal virulence, elucidate how *vp1* expression is regulated, how *vp1* mediates gene expression, and to determine if selected targets mediate phenotypes ascribed to *vp1*. To determine the role of *vp1* in pneumococcal virulence, the expression profile of animals challenged with a wild-type strain and *vp1* mutant strains was analyzed, and survivability was determined. To elucidate the impact of *vp1* in pneumococcal expression, several isogenic *vp1* mutants were constructed. Strains were phenotypically characterized, and whole-genome expression analysis was performed. In addition, alleles distribution across several strains and closely-related species was studied. Preliminary observations indicated that *vp1* affects phenotypes associated with pneumococcal colonization, specifically attachment to epithelial cells and biofilm development. Thus, experimental conditions mimicking host environments were developed. To determine how *vp1* is regulated, the genomic landscape surrounding vp1 was analyzed. Genome expression analysis on biofilms identified that a locus involved in the processing and acquisition of hyaluronic acid is one of the main VP1 targets. The locus was termed PTS-HAL. To determine whether PTS-HAL mediates the phenotypes ascribed to *vp1*, double-mutants of *vp1* and PTS-HAL were constructed, as well as a strain where the PTS-HAL is overexpressed in a *vp1* deletion mutant. These strains were challenged in their capacity to adhere to epithelial cells in a condition that the predicted host receptors are reduced. Finally, to determine the impact of *vp1* and PTS-HAL in nasopharyngeal colonization, *vp1* and PTS-HAL mutants and wild-type strains were characterized in a mouse model of pneumococcal infection.

## 1.8 References


1. UNICEF. Ending Preventable Child Deaths from Pneumonia and Diarrhoea by 2025 The integrated Global Action Plan for Pneumonia and Diarrhoea ( GAPPD ). UNICEF, World Health Organization; 2013.
2. Kochanek K, Murphy S, Xu J, Tejada-Vera B. Deaths: Final data for 2014. National vital statistics reportsDeaths: Final data for 2014. National vital statistics reports, National Center for Health Statistics. 2016;65(4).
3. Troeger C, Blacker B, Khalil I. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Infect Dis. 2018;18(11):1191–210.
4. Murdoch DR, Howie SRC. The global burden of lower respiratory infections: making progress, but we need to do better. Lancet Infect Dis. 2018;18(11):1162–3.
5. Brown AO, Mann B, Gao G, Hankins JS, Humann J, Giardina J, et al. Streptococcus pneumoniae Translocates into the Myocardium and Forms Unique Microlesions That Disrupt Cardiac Function. PLOS Pathogens. 2014 Sep 18;10(9):e1004383.
6. Loughran AJ, Orihuela CJ, Tuomanen EI. Streptococcus pneumoniae: Invasion and Inflammation. Microbiol Spectr. 2019;7(2).
7. Cartwright K. Pneumococcal disease in western Europe: burden of disease, antibiotic resistance and management. Eur J Pediatr. 2002 Apr;161(4):188–95.
8. Coates H, Thornton R, Langlands J, Filion P, Keil AD, Vijayasekaran S, et al. The role of chronic infection in children with otitis media with effusion: Evidence for intracellular persistence of bacteria. Otolaryngology - Head and Neck Surgery. 2008;138(6):778–81.
9. Thornton RB, Rigby PJ, Wiertsema SP, Filion P, Langlands J, Coates HL, et al. Multi-species bacterial biofilm and intracellular infection in otitis media. BMC Pediatrics. 2011;11(1):94.
10. Hu FZ, Eutsey R, Ahmed A, Frazao N, Powell E, Hiller NL, et al. In Vivo Capsular Switch in Streptococcus pneumoniae - Analysis by Whole Genome Sequencing. PLoS ONE. 2012;7(11).
11. Assaad, U., El-Masri, I., Porhomayon, J., & El-Solh AA. Pneumonia immunization in older adults: review of vaccine effectiveness and strategies. Clinical Interventions in Aging. 2012;7:453–461.
12. Brueggemann AB, Pai R, Crook DW, Beall B. Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. PLoS pathogens. 2007 Nov;3(11):e168.
13. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, et al. Rapid pneumococcal evolution in response to clinical interventions. Science. 2011 Jan 28;331(6016):430–4.
14. Finkelstein JA, Huang SS, Daniel J, Rifas-Shiman SL, Kleinman K, Goldmann D, et al. Antibiotic-resistant Streptococcus pneumoniae in the heptavalent pneumococcal conjugate vaccine era: predictors of carriage in a multicommunity sample. Pediatrics. 2003 Oct;112(4):862–9.
15. Frazão N, Hiller NL, Powell E, Earl J, Ahmed A, Sá-Leão R, et al. Virulence Potential and Genome-Wide Characterization of Drug Resistant Streptococcus pneumoniae Clones

Selected In Vivo by the 7-Valent Pneumococcal Conjugate Vaccine. PLoS ONE. 2013;8(9):1–16.

16. Lynch JP, Zhanel GG. Streptococcus pneumoniae: does antimicrobial resistance matter? Seminars in Respiratory and Critical Care Medicine. 2009 Apr;30(2):210–38.

17. Wainer H. MEDICAL ILLUMINATIONS Using Evidence, Visualization and Statistical Thinking to Improve Healthcare. First Edition. Oxford University Press; 2014.

18. Howard LV, Gooder H. Specificity of the autolysin of Streptococcus (Diplococcus) pneumoniae. J Bacteriol. 1974 Feb;117(2):796–804.

19. Todar K. Todar's Online Textbook of Bacteriology, In. 2018;

20. Pikis A, Campos JM, Rodriguez WJ, Keith JM. Optochin resistance in Streptococcus pneumoniae: mechanism, significance, and clinical implications. J Infect Dis. 2001 Sep 1;184(5):582–90.

21. Gámez G, Castro A, Gómez-Mejia A, Gallego M, Bedoya A, Camargo M, et al. The variome of pneumococcal virulence factors and regulators. BMC Genomics [Internet]. 2018 Jan 3 [cited 2019 Oct 28];19. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5753484/

22. Croucher NJ, Finkelstein JA, Pelton SI, Mitchell PK, Lee GM, Parkhill J, et al. Population genomics of post-vaccine changes in pneumococcal epidemiology. Nat Genet. 2013 Jun;45(6):656–63.

23. Donati C, Hiller NL, Tettelin H, Muzzi A, Croucher NJ, Angiuoli SV, et al. Structure and dynamics of the pan-genome of Streptococcus pneumoniae and closely related species. Genome Biol. 2010;11(10):R107.

24. Gladstone RA, Jefferies JM, Tocheva AS, Beard KR, Garley D, Chong WW, et al. Five winters of pneumococcal serotype replacement in UK carriage following PCV introduction. Vaccine. 2015 Apr 21;33(17):2015–21.

25. Tonder AJ van, Bray JE, Jolley KA, Quirk SJ, Haraldsson G, Maiden MCJ, et al. Heterogeneity Among Estimates Of The Core Genome And Pan-Genome In Different Pneumococcal Populations. bioRxiv. 2017 May 6;133991.

26. Forbes ML, Horsey E, Hiller NL, Buchinsky FJ, Hayes JD, Compliment JM, et al. Strain-specific virulence phenotypes of Streptococcus pneumoniae assessed using the Chinchilla laniger model of otitis media. PLoS ONE. 2008;3(4):e1969.

27. Salvadori G, Junges R, Morrison DA, Petersen FC. Competence in Streptococcus pneumoniae and Close Commensal Relatives: Mechanisms and Implications. Front Cell Infect Microbiol. 2019;9:94.

28. Tai SS. Streptococcus pneumoniae protein vaccine candidates: properties, activities and animal studies. Crit Rev Microbiol. 2006;32(3):139–53.

29. Hathaway LJ, Brugger SD, Morand B, Bangert M, Rotzetter JU, Hauser C, et al. Capsule type of Streptococcus pneumoniae determines growth phenotype. PLoS Pathog. 2012;8(3):e1002574.

30. Larson TR, Yother J. Streptococcus pneumoniae capsular polysaccharide is linked to peptidoglycan via a direct glycosidic bond to β-D-N-acetylglucosamine. Proc Natl Acad Sci USA. 2017 30;114(22):5695–700.

31. Domon H, Oda M, Maekawa T, Nagai K, Takeda W, Terao Y. Streptococcus pneumoniae disrupts pulmonary immune defence via elastase release following pneumolysin-dependent neutrophil lysis. Sci Rep. 2016 28;6:38013.

32. Berry AM, Lock RA, Hansman D, Paton JC. Contribution of autolysin to virulence of Streptococcus pneumoniae. Infect Immun. 1989 Aug;57(8):2324–30.

33. Jensch I, Gámez G, Rothe M, Ebert S, Fulde M, Somplatzki D, et al. PavB is a surface-exposed adhesin of Streptococcus pneumoniae contributing to nasopharyngeal colonization and airways infections. Mol Microbiol. 2010 Jul 1;77(1):22–43.

34. Pracht D, Elm C, Gerber J, Bergmann S, Rohde M, Seiler M, et al. PavA of Streptococcus pneumoniae modulates adherence, invasion, and meningeal inflammation. Infect Immun. 2005 May;73(5):2680–9.

35. Antikainen J, Kuparinen V, Lähteenmäki K, Korhonen TK. Enolases from Gram-positive bacterial pathogens and commensal lactobacilli share functional similarity in virulence-associated traits. FEMS Immunol Med Microbiol. 2007 Dec;51(3):526–34.

36. Bergmann S, Rohde M, Chhatwal GS, Hammerschmidt S. alpha-Enolase of Streptococcus pneumoniae is a plasmin(ogen)-binding protein displayed on the bacterial cell surface. Mol Microbiol. 2001 Jun;40(6):1273–87.

37. Limoli DH, Sladek JA, Fuller LA, Singh AK, King SJ. BgaA acts as an adhesin to mediate attachment of some pneumococcal strains to human epithelial cells. Microbiology (Reading, Engl). 2011 Aug;157(Pt 8):2369–81.

38. Henriques-Normark B, Tuomanen EI. The Pneumococcus: Epidemiology, Microbiology, and Pathogenesis. Cold Spring Harb Perspect Med [Internet]. 2013 Jul [cited 2016 Oct 5];3(7). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3685878/

39. CDC. Antibiotic Resistance Threats in the United States [Internet]. 2013. Available from: https://www.cdc.gov/drugresistance/biggest_threats.html

40. World Health Organization. Global immunization data. World Health Organization [Internet]. 2014. Available from: http://www.who.int/immunization/monitoring_surveillance/global_immunization_data.pdf?ua=1

41. World Health Organization. Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. 2017; Available from: https://www.who.int/medicines/publications/global-priority-list-antibiotic-resistant-bacteria/en/

42. Hamaluba M, Kandasamy R, Ndimah S, Morton R, Caccamo M, Robinson H, et al. A cross-sectional observational study of pneumococcal carriage in children, their parents, and older adults following the introduction of the 7-valent pneumococcal conjugate vaccine. Medicine (Baltimore). 2015 Jan;94(1):e335.

43. O'Brien KL, Wolfson LJ, Watt JP, Henkle E, Deloria-Knoll M, McCall N, et al. Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates. The Lancet. 2009 Sep;374(9693):893–902.

44. Weycker D, Strutton D, Edelsberg J, Sato R, Jackson LA. Clinical and economic burden of pneumococcal disease in older US adults. Vaccine. 2010 Jul 12;28(31):4955–60.

45. Wroe PC, Finkelstein JA, Ray GT, Linder JA, Johnson KM, Rifas-Shiman S, et al. Aging population and future burden of pneumococcal pneumonia in the United States. J Infect Dis. 2012 May 15;205(10):1589–92.

46. Willyard C. The drug-resistant bacteria that pose the greatest health threats. Nature. 2017 28;543(7643):15.

47. CDC. Pneumococcal Vaccination | What You Should Know | CDC [Internet]. 2019 [cited 2019 Nov 9]. Available from: https://www.cdc.gov/vaccines/vpd/pneumo/public/index.html

48. Kim L, McGee L, Tomczyk S, Beall B. Biological and Epidemiological Features of Antibiotic-Resistant Streptococcus pneumoniae in Pre- and Post-Conjugate Vaccine Eras: a United States Perspective. Clin Microbiol Rev. 2016 Jul;29(3):525–52.

49. Navalkele P, Özgönenel B, McGrath E, Lephart P, Sarnaik S. Invasive Pneumococcal Disease in Patients With Sickle Cell Disease. J Pediatr Hematol Oncol. 2017;39(5):341–4.

50. Oligbu G, Fallaha M, Pay L, Ladhani S. Risk of invasive pneumococcal disease in children with sickle cell disease in the era of conjugate vaccines: a systematic review of the literature. Br J Haematol. 2019 May;185(4):743–51.

51. Nobbs AH, Lamont RJ, Jenkinson HF. Streptococcus adherence and colonization. Microbiol Mol Biol Rev. 2009 Sep;73(3):407–50, Table of Contents.

52. Hansman D, Andrews G. Hospital infection with pneumococci resistant to tetracycline. Med J Aust. 1967 Mar 11;1(10):498–501.

53. Low DE. Changing trends in antimicrobial-resistant pneumococci: it's not all bad news. Clin Infect Dis. 2005 Aug 15;41 Suppl 4:S228-233.

54. Jacobs MR, Bajaksouzian S, Zilles A, Lin G, Pankuch GA, Appelbaum PC. Susceptibilities of Streptococcus pneumoniae and Haemophilus influenzae to 10 oral antimicrobial agents based on pharmacodynamic parameters: 1997 U.S. Surveillance study. Antimicrob Agents Chemother. 1999 Aug;43(8):1901–8.

55. Song J-H, Jung S-I, Ko KS, Kim NY, Son JS, Chang H-H, et al. High prevalence of antimicrobial resistance among clinical Streptococcus pneumoniae isolates in Asia (an ANSORP study). Antimicrob Agents Chemother. 2004 Jun;48(6):2101–7.

56. Gillis LM, White HD, Whitehurst A, Sullivan DC. Vancomycin-tolerance among clinical isolates of Streptococcus pneumoniae in Mississippi during 1999-2001. Am J Med Sci. 2005 Aug;330(2):65–8.

57. Lencastre H De, Tomasz A. From ecological reservoir to disease : the nasopharynx , day-care centres and drug-resistant clones of Streptococcus pneumoniae. Journal of Antimicrobial Chemotherapy. 2002;50:75–81.

58. Domenech A, Ardanuy C, Grau I, Calatayud L, Pallares R, Fenoll A, et al. Evolution and genetic diversity of the Spain. J Antimicrob Chemother. 2014;69:924–31.

59. Tomasz A, Corso A, Severina EP, Echániz-Aviles G, Brandileone MC, Camou T, et al. Molecular epidemiologic characterization of penicillin-resistant Streptococcus pneumoniae invasive pediatric isolates recovered in six Latin-American countries: an overview. PAHO/Rockefeller University Workshop. Pan American Health Organization. Microb Drug Resist. 1998;4(3):195–207.

60. Adegbola RA, DeAntonio R, Hill PC, Roca A, Usuf E, Hoet B, et al. Carriage of Streptococcus pneumoniae and other respiratory bacterial pathogens in low and lower-

middle income countries: a systematic review and meta-analysis. PLoS ONE. 2014;9(8):e103293.

61. Usuf E, Bottomley C, Adegbola RA, Hall A. Pneumococcal carriage in sub-Saharan Africa--a systematic review. PLoS ONE. 2014;9(1):e85001.

62. Hammitt LL, Etyang AO, Morpeth SC, Ojal J, Mutuku A, Mturi N, et al. Effect of ten-valent pneumococcal conjugate vaccine on invasive pneumococcal disease and nasopharyngeal carriage in Kenya: a longitudinal surveillance study. Lancet. 2019 May 25;393(10186):2146–54.

63. Hammitt LL, Campbell JC, Borys D, Weatherholtz RC, Reid R, Goklish N, et al. Efficacy, safety and immunogenicity of a pneumococcal protein-based vaccine co-administered with 13-valent pneumococcal conjugate vaccine against acute otitis media in young children: A phase IIb randomized study. Vaccine. 2019 Oct 16;

64. Huang SS, Finkelstein JA, Rifas-Shiman SL, Kleinman K, Platt R. Community-level predictors of pneumococcal carriage and resistance in young children. Am J Epidemiol. 2004 Apr 1;159(7):645–54.

65. Lipsky BA, Boyko EJ, Inui TS, Koepsell TD. Risk factors for acquiring pneumococcal infections. Arch Intern Med. 1986 Nov;146(11):2179–85.

66. Nuorti JP, Butler JC, Farley MM, Harrison LH, McGeer A, Kolczak MS, et al. Cigarette smoking and invasive pneumococcal disease. Active Bacterial Core Surveillance Team. N Engl J Med. 2000 Mar 9;342(10):681–9.

67. Xiang L, Zhou TJ, Zhou LL, Luo J, Qin Z, You JZ, et al. Influenza a virus and Streptococcus pneumonia coinfection potentially promotes bacterial colonization and enhances B lymphocyte depression and reduction. J Biol Regul Homeost Agents. 2019 Oct;33(5):1437–49.

68. De Lencastre H, Tomasz A. From ecological reservoir to disease: the nasopharynx, day-care centres and drug-resistant clones of Streptococcus pneumoniae. J Antimicrob Chemother. 2002 Dec;50 Suppl S2:75–81.

69. Hava DL, LeMieux J, Camilli A. From nose to lung: the regulation behind Streptococcus pneumoniae virulence factors. Mol Microbiol. 2003 Nov;50(4):1103–10.

70. Nelson AL, Roche AM, Gould JM, Chim K, Ratner AJ, Weiser JN. Capsule Enhances Pneumococcal Colonization by Limiting Mucus-Mediated Clearance. Infect Immun. 2007 Jan;75(1):83–90.

71. World Health Organization. Pneumonia [Internet]. 2019 [cited 2019 Nov 9]. Available from: https://www.who.int/news-room/fact-sheets/detail/pneumonia

72. Bergmann S, Hammerschmidt S. Versatility of pneumococcal surface proteins. Microbiology. 2006 Feb 1;152(2):295–303.

73. García B, Merayo-Lloves J, Rodríguez D, Alcalde I, García-Suárez O, Alfonso JF, et al. Different Use of Cell Surface Glycosaminoglycans As Adherence Receptors to Corneal Cells by Gram Positive and Gram Negative Pathogens. Front Cell Infect Microbiol. 2016;6:173.

74. Paterson GK, Orihuela CJ. Pneumococcal microbial surface components recognizing adhesive matrix molecules targeting of the extracellular matrix. Mol Microbiol. 2010 Jul 1;77(1):1–5.

75. Papasergi S, Garibaldi M, Tuscano G, Signorino G, Ricci S, Peppoloni S, et al. Plasminogen- and fibronectin-binding protein B is involved in the adherence of Streptococcus pneumoniae to human epithelial cells. J Biol Chem. 2010 Mar 5;285(10):7517–24.

76. Yamaguchi M, Terao Y, Mori Y, Hamada S, Kawabata S. PfbA, a novel plasmin- and fibronectin-binding protein of Streptococcus pneumoniae, contributes to fibronectin-dependent adhesion and antiphagocytosis. J Biol Chem. 2008 Dec 26;283(52):36272–9.

77. King SJ, Hippe KR, Weiser JN. Deglycosylation of human glycoconjugates by the sequential activities of exoglycosidases expressed by Streptococcus pneumoniae. Mol Microbiol. 2006 Feb;59(3):961–74.

78. Chang AB, Redding GJ, Everard ML. Chronic wet cough: Protracted bronchitis, chronic suppurative lung disease and bronchiectasis. Pediatr Pulmonol. 2008 Jun;43(6):519–31.

79. Hare KM, Smith-Vaughan HC, Chang AB, Pizzutto S, Petsky HL, McCallum GB, et al. Propensity of pneumococcal carriage serotypes to infect the lower airways of children with chronic endobronchial infections. Vaccine. 2017 01;35(5):747–56.

80. Barbeito-Castiñeiras G, Guinda-Giménez M, Cores-Calvo O, Hernández-Blanco M, Pardo-Sánchez F. [Pneumococcal arthritis in paediatric population]. Rev Esp Quimioter. 2017 Apr;30(2):118–22.

81. Moro-Lago I, Talavera G, Moraleda L, González-Morán G. Clinical presentation and treatment of septic arthritis in children. Rev Esp Cir Ortop Traumatol. 2017 Jun;61(3):170–5.

82. Erdem G, Singh AK, Brusnahan AJ, Moore AN, Barson WJ, Leber A, et al. Pneumococcal colonization among tracheostomy tube dependent children. PLoS ONE. 2018;13(10):e0206305.

83. Olarte L, Lin PL, Barson WJ, Romero JR, Tan TQ, Givner LB, et al. Invasive pneumococcal infections in children following transplantation in the pneumococcal conjugate vaccine era. Transpl Infect Dis. 2017 Feb;19(1).

84. McCullers JA. The co-pathogenesis of influenza viruses with bacteria in the lung. Nat Rev Microbiol. 2014 Apr;12(4):252–62.

85. Falsey AR, Becker KL, Swinburne AJ, Nylen ES, Formica MA, Hennessey PA, et al. Bacterial complications of respiratory tract viral illness: a comprehensive evaluation. J Infect Dis. 2013 Aug 1;208(3):432–41.

86. Musher DM, Abers MS, Bartlett JG. Evolving Understanding of the Causes of Pneumonia in Adults, With Special Attention to the Role of Pneumococcus. Clin Infect Dis. 2017 Oct 30;65(10):1736–44.

87. Metzger DW, Sun K. Immune dysfunction and bacterial coinfections following influenza. J Immunol. 2013 Sep 1;191(5):2047–52.

88. Ngo CC, Massa HM, Thornton RB, Cripps AW. Predominant Bacteria Detected from the Middle Ear Fluid of Children Experiencing Otitis Media: A Systematic Review. PLoS ONE. 2016;11(3):e0150949.

89. Donati C, Hiller NL, Tettelin H, Muzzi A, Croucher NJ, Angiuoli S V, et al. Structure and dynamics of the pan-genome of Streptococcus pneumoniae and closely related species. Genome biology. 2010;11(10):R107.

90. Qureishi A, Lee Y, Belfield K, Birchall JP, Daniel M. Update on otitis media - prevention and treatment. Infect Drug Resist. 2014 Jan 10;7:15–24.

91. Schilder AGM, Chonmaitree T, Cripps AW, Rosenfeld RM, Casselbrant ML, Haggard MP, et al. Otitis media. Nat Rev Dis Primers. 2016 08;2:16063.

92. Nistico L, Kreft R, Gieseke A, Coticchia JM, Burrows A, Khampang P, et al. Adenoid reservoir for pathogenic biofilm bacteria. Journal of Clinical Microbiology. 2011;49(4):1411–20.

93. Gordon KA, Jiwani S, Papsin BC. Benefits and detriments of unilateral cochlear implant use on bilateral auditory development in children who are deaf. Front Psychol [Internet]. 2013 Oct 16 [cited 2019 Apr 17];4. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3797443/

94. Vila PM, Ghogomu NT, Odom-John AR, Hullar TE, Hirose K. Infectious complications of pediatric cochlear implants are highly influenced by otitis media. Int J Pediatr Otorhinolaryngol. 2017 Jun;97:76–82.

95. Hall-stoodley L, Hu FZ, Gieseke A, Nistico L, Hayes J, Forbes M, et al. Direct Detection of Bacterial Biofilms on the Middle-Ear Mucosa of Children With Chronic Otitis Media. JAMA. 2006;296(2):202–11.

96. Millar B, Moore J. Emerging Issues in Infective Endocarditis. Emerging Infectious Diseases. 2004;10(4):1110–6.

97. Baddour LM, Wilson WR, Bayer AS, Fowler VG, Tleyjeh IM, Rybak MJ, et al. Infective endocarditis in adults: Diagnosis, antimicrobial therapy, and management of complications: A scientific statement for healthcare professionals from the American Heart Association. Vol. 132, Circulation. 2015. 1435–1486 p.

98. Werdan K, Dietz S, Löffler B, Niemann S, Bushnaq H, Silber R-E, et al. Mechanisms of infective endocarditis: pathogen–host interaction and risk states. Nature Reviews Cardiology. 2013;11(1):35–50.

99. Musher DM, Rueda AM, Kaka AS, Mapara SM. The Association between Pneumococcal Pneumonia and Acute Cardiac Events. Clinical Infectious Diseases. 2007;45(2):158–65.

100. Thomsen RW, Kasatpibal N, Riis A, Nørgaard M, Sørensen HT. The impact of pre-existing heart failure on pneumonia prognosis: Population-based cohort study. Journal of General Internal Medicine. 2008;23(9):1407–13.

101. Corrales-Medina VF, Musher DM, Wells GA, Chirinos JA, Chen L, Fine MJ. Cardiac complications in patients with community-acquired pneumonia incidence, timing, risk factors, and association with short-term mortality. Circulation. 2012;125(6):773–81.

102. Cunningham R, Sinha L. Recurrent Streptococcus pneumoniae Endocarditis. Eur J Clin Microbiol Infect Dis. 1995;14:526–8.

103. Shrestha S, Chintanaboina J, Pancholy S. Case Report Recurrent Mitral Valve Endocarditis Caused by Streptococcus pneumoniae in a Splenectomized Host. 2013;2013.

104. Lindberg J, Fangel S. Recurrent Endocarditis Caused by Streptococcus pneumoniae. Scandinavian Journal of Infectious Diseases. 1999;31(4):409–10.

105. Alhamdi Y, Neill DR, Abrams ST, Malak HA, Yahya R, Barrett-Jolley R, et al. Circulating Pneumolysin Is a Potent Inducer of Cardiac Injury during Pneumococcal Infection. PLOS Pathogens. 2015;11(5):e1004836.

106. Aronin SI, Mukherjee SK, West JC, Cooney EL. Review of pneumococcal endocarditis in adults in the penicillin era. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 1998;26(1):165–71.

107. Okada T, Yoshitomi H, Harada Y, Ito S, Nakamura T, Adachi T, et al. Pneumococcal endocarditis complicating meningitis and arthritis in a previously healthy woman: A case report. Journal of Cardiology Cases. 2015;11(3):96–9.

108. Siegel M, Timpone J. Penicillin-resistant Streptococcus pneumoniae endocarditis: a case report and review. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2001;32(6):972–4.

109. Natsheh A, Vidberg M, Friedmann R, Ben-chetrit E, Yinnon AM, Zevin S. Prosthetic valve endocarditis due to Streptococcus pneumoniae. 2014;(October 2016):0–3.

110. Fefer P, Raveh D, Rudensky B, Schlesinger Y, Yinnon A. Changing epidemiology of infective endocarditis: A retrospective survey of 108 cases, 1990-1999. European Journal of Clinical Microbiology and Infectious Diseases. 2002;21(6):432–7.

111. Kan B, Ries J, Normark BH, Chang FY, Feldman C, Ko WC, et al. Endocarditis and pericarditis complicating pneumococcal bacteraemia, with special reference to the adhesive abilities of pneumococci: Results from a prospective study. Clinical Microbiology and Infection. 2006;12(4):338–44.

112. Hall-Stoodley L, Costerton JW, Stoodley P. Bacterial biofilms: from the Natural environment to infectious diseases. Nat Rev Micro. 2004 Feb;2(2):95–108.

113. Bashore TM, Cabell C, Fowler V. Update on infective endocarditis. Current Problems in Cardiology. 2006;31(4):265–352.

114. Moreillon P, Que Y, Bayer A. Pathogenesis of streptococcal and staphylococcal endocarditis. Infect Dis Clin North Am. 2002;16(2):297–318.

115. Que Y-A, Moreillon P. Infective endocarditis. Nat Rev Cardiol. 2011;8:322–36.

116. Miller JJ, Scott IU, Flynn HW, Smiddy WE, Corey RP, Miller D. Endophthalmitis caused by Streptococcus pneumoniae. Am J Ophthalmol. 2004 Aug;138(2):231–6.

117. Marquart ME, O'Callaghan RJ. Infectious Keratitis: Secreted Bacterial Proteins That Mediate Corneal Damage. Journal of Ophthalmology. 2013 Jan 8;2013:e369094.

118. Croucher NJ, Coupland PG, Stevenson AE, Callendrello A, Bentley SD, Hanage WP. Diversification of bacterial genome content through distinct mechanisms over different timescales. Nat Commun [Internet]. 2014 Nov 19 [cited 2015 Feb 1];5. Available from: http://www.nature.com/ncomms/2014/141119/ncomms6471/abs/ncomms6471.html

119. Hilty M, Wüthrich D, Salter SJ, Engel H, Campbell S, Sá-Leão R, et al. Global phylogenomic analysis of nonencapsulated Streptococcus pneumoniae reveals a deep-branching classic lineage that is distinct from multiple sporadic lineages. Genome Biol Evol. 2014 Dec 4;6(12):3281–94.

120. Valentino MD, McGuire AM, Rosch JW, Bispo PJM, Burnham C, Sanfilippo CM, et al. Unencapsulated Streptococcus pneumoniae from conjunctivitis encode variant traits and belong to a distinct phylogenetic cluster. Nat Commun. 2014;5:5411.

121. Ertugrul N, Rodriguez-Barradas MC, Musher DM, Ryan MA, Agin CS, Murphy SJ, et al. BOX-polymerase chain reaction-based DNA analysis of nonserotypeable Streptococcus pneumoniae implicated in outbreaks of conjunctivitis. J Infect Dis. 1997 Nov;176(5):1401–5.

122. Shayegani M, Parsons LM, Gibbons WE, Campbell D. Characterization of nontypable Streptococcus pneumoniae-like organisms isolated from outbreaks of conjunctivitis. J Clin Microbiol. 1982 Jul;16(1):8–14.

123. Tarabishy AB, Jeng BH. Bacterial conjunctivitis: a review for internists. Cleve Clin J Med. 2008 Jul;75(7):507–12.

124. Henrichsen J. Six newly recognized types of Streptococcus pneumoniae. J Clin Microbiol. 1995 Oct;33(10):2759–62.

125. Kadioglu A, Weiser JN, Paton JC, Andrew PW. The role of Streptococcus pneumoniae virulence factors in host respiratory colonization and disease. Nat Rev Micro. 2008 Apr;6(4):288–301.

126. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabbinowitsch E, Collins M, et al. Genetic Analysis of the Capsular Biosynthetic Locus from All 90 Pneumococcal Serotypes. PLoS Genet. 2006 Mar 10;2(3):e31.

127. Green C, Moore CA, Mahajan A, Bajaj K. A Simple Approach to Pneumococcal Vaccination in Adults. J Glob Infect Dis. 2018 Sep;10(3):159–62.

128. Kong Y, Zhang W, Jiang Z, Wang L, Li C, Li Y, et al. Immunogenicity and safety of a 23-valent pneumococcal polysaccharide vaccine in Chinese healthy population aged >2 years: A randomized, double-blinded, active control, phase III trial. Hum Vaccin Immunother. 2015 Jun 17;11(10):2425–33.

129. Clarke E, Kampmann B, Goldblatt D. Maternal and neonatal pneumococcal vaccination - where are we now? Expert Rev Vaccines. 2016;15(10):1305–17.

130. Poehling KA, Talbot TR, Griffin MR, Craig AS, Whitney CG, Zell E, et al. Invasive pneumococcal disease among infants before and after introduction of pneumococcal conjugate vaccine. JAMA. 2006 Apr 12;295(14):1668–74.

131. CDC. Pneumococcal Disease | Symptoms and Complications | CDC [Internet]. 2016 [cited 2016 Sep 15]. Available from: http://www.cdc.gov/pneumococcal/about/symptoms-complications.html

132. CDC. Pneumococcal Disease: Fast Facts About Pneumococcal Pneumonia - CDC [Internet]. 2014 [cited 2014 Oct 9]. Available from: http://www.cdc.gov/pneumococcal/about/facts.html

133. Bogaert D, Hermans PWM, Adrian PV, Rümke HC, de Groot R. Pneumococcal vaccines: an update on current strategies. Vaccine. 2004 Jun 2;22(17–18):2209–20.

134. Wilson R, Cohen JM, Reglinski M, Jose RJ, Chan WY, Marshall H, et al. Naturally Acquired Human Immunity to Pneumococcus Is Dependent on Antibody to Protein Antigens. PLoS Pathog. 2017;13(1):e1006137.

135. Chan W-Y, Entwisle C, Ercoli G, Ramos-Sevillano E, McIlgorm A, Cecchini P, et al. A Novel, Multiple-Antigen Pneumococcal Vaccine Protects against Lethal Streptococcus pneumoniae Challenge. Infect Immun. 2019;87(3).

136. Kirkham L-AS, Wiertsema SP, Corscadden KJ, Mateus T, Mullaney GL, Zhang G, et al. Otitis-Prone Children Produce Functional Antibodies to Pneumolysin and Pneumococcal Polysaccharides. Clin Vaccine Immunol. 2017 Mar;24(3).

137. Hermand P, Vandercammen A, Mertens E, Di Paolo E, Verlant V, Denoël P, et al. Preclinical evaluation of a chemically detoxified pneumolysin as pneumococcal vaccine antigen. Hum Vaccin Immunother. 2017 02;13(1):220–8.

138. Girard MP, Cherian T, Pervikov Y, Kieny MP. A review of vaccine research and development: human acute respiratory infections. Vaccine. 2005 Dec 30;23(50):5708–24.

139. Brown JS, Ogunniyi AD, Woodrow MC, Holden DW, Paton JC. Immunization with components of two iron uptake ABC transporters protects mice against systemic Streptococcus pneumoniae infection. Infect Immun. 2001 Nov;69(11):6702–6.

140. Jomaa M, Yuste J, Paton JC, Jones C, Dougan G, Brown JS. Antibodies to the iron uptake ABC transporter lipoproteins PiaA and PiuA promote opsonophagocytosis of Streptococcus pneumoniae. Infect Immun. 2005 Oct;73(10):6852–9.

141. Ogunniyi AD, Mahdi LK, Trappetti C, Verhoeven N, Mermans D, Van der Hoek MB, et al. Identification of Genes That Contribute to the Pathogenesis of Invasive Pneumococcal Disease by In Vivo Transcriptomic Analysis. Infect Immun. 2012 Sep;80(9):3268–78.

142. Brooks-Walter A, Briles DE, Hollingshead SK. The pspC gene of Streptococcus pneumoniae encodes a polymorphic protein, PspC, which elicits cross-reactive antibodies to PspA and provides immunity to pneumococcal bacteremia. Infect Immun. 1999 Dec;67(12):6533–42.

143. Ogunniyi AD, Woodrow MC, Poolman JT, Paton JC. Protection against Streptococcus pneumoniae elicited by immunization with pneumolysin and CbpA. Infect Immun. 2001 Oct;69(10):5997–6003.

144. Long JP, Tong HH, DeMaria TF. Immunization with native or recombinant Streptococcus pneumoniae neuraminidase affords protection in the chinchilla otitis media model. Infect Immun. 2004 Jul;72(7):4309–13.

145. Tong HH, Li D, Chen S, Long JP, DeMaria TF. Immunization with recombinant Streptococcus pneumoniae neuraminidase NanA protects chinchillas against nasopharyngeal colonization. Infect Immun. 2005 Nov;73(11):7775–8.

146. Adrian PV, Bogaert D, Oprins M, Rapola S, Lahdenkari M, Kilpi T, et al. Development of antibodies against pneumococcal proteins alpha-enolase, immunoglobulin A1 protease, streptococcal lipoprotein rotamase A, and putative proteinase maturation protein A in relation to pneumococcal carriage and Otitis Media. Vaccine. 2004 Jul 29;22(21–22):2737–42.

147. Audouy SAL, van Selm S, van Roosmalen ML, Post E, Kanninga R, Neef J, et al. Development of lactococcal GEM-based pneumococcal vaccines. Vaccine. 2007 Mar 22;25(13):2497–506.

148. Janoff EN, Rubins JB, Fasching C, Charboneau D, Rahkola JT, Plaut AG, et al. Pneumococcal IgA1 protease subverts specific protection by human IgA1. Mucosal Immunol. 2014 Mar;7(2):249–56.

149. Reglinski M, Ercoli G, Plumptre C, Kay E, Petersen FC, Paton JC, et al. A recombinant conjugated pneumococcal vaccine that protects against murine infections with a similar efficacy to Prevnar-13. NPJ Vaccines. 2018;3:53.

150. Borg MA, Tiemersma E, Scicluna E, van de Sande-Bruinsma N, de Kraker M, Monen J, et al. Prevalence of penicillin and erythromycin resistance among invasive Streptococcus pneumoniae isolates reported by laboratories in the southern and eastern Mediterranean region. Clin Microbiol Infect. 2009 Mar;15(3):232–7.

151. Felmingham D, Cantón R, Jenkins SG. Regional trends in beta-lactam, macrolide, fluoroquinolone and telithromycin resistance among Streptococcus pneumoniae isolates 2001-2004. J Infect. 2007 Aug;55(2):111–8.

152. Geno KA, Gilbert GL, Song JY, Skovsted IC, Klugman KP, Jones C, et al. Pneumococcal Capsules and Their Types: Past, Present, and Future. Clin Microbiol Rev. 2015 Jul;28(3):871–99.

153. Hiller NL, Eutsey RA, Powell E, Earl JP, Janto B, Martin DP, et al. Differences in genotype and virulence among four multidrug-resistant streptococcus pneumoniae isolates belonging to the PMEN1 clone. PLoS ONE. 2011;6(12):e28850.

154. Wyres KL, Lambertsen LM, Croucher NJ, McGee L, von Gottberg A, Liñares J, et al. The multidrug-resistant PMEN1 pneumococcus is a paradigm for genetic success. Genome Biol. 2012 Nov 16;13(11):R103.

155. Wyres KL, Lambertsen LM, Croucher NJ, McGee L, von Gottberg A, Liñares J, et al. Pneumococcal capsular switching: a historical perspective. J Infect Dis. 2013 Feb 1;207(3):439–49.

156. Zivich PN, Grabenstein JD, Becker-Dreps SI, Weber DJ. Streptococcus pneumoniae outbreaks and implications for transmission and control: a systematic review. Pneumonia (Nathan) [Internet]. 2018 Nov 5 [cited 2019 Nov 2];10. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6217781/

157. Hiller NL, Eutsey RA, Powell E, Earl JP, Janto B, Martin DP, et al. Differences in Genotype and Virulence among Four Multidrug-Resistant Streptococcus pneumoniae Isolates Belonging to the PMEN1 Clone. PLoS ONE. 2011 Dec 19;6(12):e28850.

158. Martin C, Sibold C, Hakenbeck R. Relatedness of penicillin-binding protein 1a genes from different clones of penicillin-resistant Streptococcus pneumoniae isolated in South Africa and Spain. EMBO J. 1992 Nov;11(11):3831–6.

159. McDougal LK, Facklam R, Reeves M, Hunter S, Swenson JM, Hill BC, et al. Analysis of multiply antimicrobial-resistant isolates of Streptococcus pneumoniae from the United States. Antimicrob Agents Chemother. 1992 Oct;36(10):2176–84.

160. McGee L, McDougal L, Zhou J, Spratt BG, Tenover FC, George R, et al. Nomenclature of major antimicrobial-resistant clones of Streptococcus pneumoniae defined by the pneumococcal molecular epidemiology network. J Clin Microbiol. 2001 Jul;39(7):2565–71.

161. PMEN Consortium. PMEN Public Release, 2006 [Internet]. 2006 [cited 2019 Apr 12]. Available from: http://web1.sph.emory.edu/PMEN/pmen_table1.html

162. Everett DB, Mukaka M, Denis B, Gordon SB, Carrol ED, Joep J, et al. Ten Years of Surveillance for Invasive Streptococcus pneumoniae during the Era of Antiretroviral Scale-Up and Cotrimoxazole Prophylaxis in Malawi. PlosOne. 2011;6(3):1–9.

163. Hiller NL, Janto B, Hogg JS, Boissy R, Yu S, Powell E, et al. Comparative genomic analyses of seventeen Streptococcus pneumoniae strains: insights into the pneumococcal supragenome. J Bacteriol. 2007 Nov;189(22):8186–95.

164. Donati C, Hiller NL, Tettelin H, Muzzi A, Croucher NJ, Angiuoli SV, et al. Structure and dynamics of the pan-genome of Streptococcus pneumoniae and closely related species. Genome Biol. 2010;11(10):R107.

165. Eutsey RA, Powell E, Dordel J, Salter SJ, Clark TA, Korlach J, et al. Genetic Stabilization of the Drug-Resistant PMEN1 Pneumococcus Lineage by Its Distinctive DpnIII Restriction-Modification System. mBio. 2015 Jul 1;6(3):e00173-15.

166. Tettelin H, Masignani V, Cieslewicz MJ, Eisen JA, Peterson S, Wessels MR, et al. Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V Streptococcus agalactiae. Proc Natl Acad Sci USA. 2002 Sep 17;99(19):12391–6.

167. Aaberge IS, Eng J, Lermark G, Løvik M. Virulence of Streptococcus pneumoniae in mice: a standardized method for preparation and frozen storage of the experimental bacterial inoculum. Microb Pathog. 1995 Feb;18(2):141–52.

168. Bricker AL, Camilli A. Transformation of a type 4 encapsulated strain of Streptococcus pneumoniae. FEMS Microbiol Lett. 1999 Mar 15;172(2):131–5.

169. Hakenbeck R, Balmelle N, Weber B, Gardès C, Keck W, de Saizieu A. Mosaic genes and mosaic chromosomes: intra- and interspecies genomic variation of Streptococcus pneumoniae. Infect Immun. 2001 Apr;69(4):2477–86.

170. de Saizieu A, Gardès C, Flint N, Wagner C, Kamber M, Mitchell TJ, et al. Microarray-based identification of a novel Streptococcus pneumoniae regulon controlled by an autoinduced peptide. J Bacteriol. 2000 Sep;182(17):4696–703.

171. Wizemann TM, Heinrichs JH, Adamou JE, Erwin AL, Kunsch C, Choi GH, et al. Use of a whole genome approach to identify vaccine molecules affording protection against Streptococcus pneumoniae infection. Infect Immun. 2001 Mar;69(3):1593–8.

172. Armbruster C, Hong W, Pang B, Weimer K, Juneau R, Turner J, et al. Indirect Pathogenicity of Haemophilus influenzae and Moraxella catarrhalis in Polymicrobial Otitis Media Occurs via Interspecies. mBio. 2010;1(3):1–9.

173. Clementi CF, Murphy TF. Non-Typeable Haemophilus influenzae Invasion and Persistence in the Human Respiratory Tract. Frontiers in Cellular and Infection Microbiology. 2011;1(November):1–9.

174. Hall-Stoodley L, Hu FZ, Gieseke A, Nistico L, Nguyen D, Hayes J, et al. Direct Detection of Bacterial Biofilms on the Middle-Ear Mucosa of Children With Chronic Otitis Media. JAMA. 2006 Jul 12;296(2):202–11.

175. Perez AC, Pang B, King LB, Tan L, Murrah KA, Reimche JL, et al. Residence of Streptococcus pneumoniae and Moraxella catarrhalis within polymicrobial biofilm promotes antibiotic resistance and bacterial persistence in vivo. 2014;280–8.

176. Weiser JN, Austrian R, Sreenivasan PK, Masure HR. Phase variation in pneumococcal opacity: relationship between colonial morphology and nasopharyngeal colonization. Infect Immun. 1994 Jun;62(6):2582–9.

177. Cundell DR, Weiser JN, Shen J, Young A, Tuomanen EI. Relationship between colonial morphology and adherence of Streptococcus pneumoniae. Infect Immun. 1995 Mar;63(3):757–61.

178. Dahl null, Mygind null. Anatomy, physiology and function of the nasal cavities in health and disease. Adv Drug Deliv Rev. 1998 Jan 5;29(1–2):3–12.

179. Feldman C, Anderson R, Cockeran R, Mitchell T, Cole P, Wilson R. The effects of pneumolysin and hydrogen peroxide, alone and in combination, on human ciliated epithelium in vitro. Respir Med. 2002 Aug;96(8):580–5.

180. Luo P, Li H, Morrison DA. ComX is a unique link between multiple quorum sensing outputs and competence in Streptococcus pneumoniae. Molecular Microbiology. 2003 Oct;50(2):623–33.

181. Alloing G, Martin B, Granadel C, Claverys JP. Development of competence in Streptococcus pneumonaie: pheromone autoinduction and control of quorum sensing by the oligopeptide permease. Molecular Microbiology. 1998 Jul;29(1):75–83.

182. Johnston C, Martin B, Fichant G, Polard P, Claverys J-P. Bacterial transformation: distribution, shared mechanisms and divergent control. Nature Reviews Microbiology. 2014 Mar;12(3):181–96.

183. Zafar MA, Kono M, Wang Y, Zangari T, Weiser JN. Infant Mouse Model for the Study of Shedding and Transmission during Streptococcus pneumoniae Monoinfection. Infect Immun. 2016;84(9):2714–22.

184. Zafar MA, Hamaguchi S, Zangari T, Cammer M, Weiser JN. Capsule Type and Amount Affect Shedding and Transmission of Streptococcus pneumoniae. MBio. 2017 22;8(4).

185. Muñoz-Elías EJ, Marcano J, Camilli A. Isolation of Streptococcus pneumoniae biofilm mutants and their characterization during nasopharyngeal colonization. Infect Immun. 2008 Nov;76(11):5049–61.

186. Tong HH, Blue LE, James MA, DeMaria TF. Evaluation of the virulence of a Streptococcus pneumoniae neuraminidase-deficient mutant in nasopharyngeal colonization and development of otitis media in the chinchilla model. Infect Immun. 2000 Feb;68(2):921–4.

187. Bidossi A, Mulas L, Decorosi F, Colomba L, Ricci S, Pozzi G, et al. A functional genomics approach to establish the complement of carbohydrate transporters in Streptococcus pneumoniae. PLoS ONE. 2012;7(3):e33320.

188. Buckwalter CM, King SJ. Pneumococcal carbohydrate transport: food for thought. Trends Microbiol. 2012 Nov;20(11):517–22.

189. Marks LR, Reddinger RM, Hakansson AP. High Levels of Genetic Recombination during Nasopharyngeal Carriage and Biofilm Formation in Streptococcus pneumoniae. mBio. 2012 Nov 1;3(5):e00200-12.

190. Chao Y, Marks LR, Pettigrew MM, Hakansson AP. Streptococcus pneumoniae biofilm formation and dispersion during colonization and disease. Front Cell Infect Microbiol. 2014;4:194.

191. Croucher NJ, Harris SR, Barquist L, Parkhill J, Bentley SD. A high-resolution view of genome-wide pneumococcal transformation. PLoS Pathog. 2012;8(6):e1002745.

192. Tonnaer EL, Mylanus EA, Mulder JJ, Curfs JH. Detection of bacteria in healthy middle ears during cochlear implantation. Arch Otolaryngol Head Neck Surg. 2009 Mar;135(3):232–7.

193. Trappetti C, Potter AJ, Paton AW, Oggioni MR, Paton JC. LuxS Mediates Iron-Dependent Biofilm Formation, Competence, and Fratricide in Streptococcus pneumoniae ▿. Infect Immun. 2011 Nov;79(11):4550–8.

194. Chao Y, Marks LR, Pettigrew MM, Hakansson AP. Streptococcus pneumoniae biofilm formation and dispersion during colonization and disease. Front Cell Infect Microbiol. 2014;4:194.

195. Collins AM, Rylance J, Wootton DG, Wright AD, Wright AKA, Fullerton DG, et al. Bronchoalveolar lavage (BAL) for research; obtaining adequate sample yield. J Vis Exp. 2014 Mar 24;(85).

196. Weiser JN, Bae D, Fasching C, Scamurra RW, Ratner AJ, Janoff EN. Antibody-enhanced pneumococcal adherence requires IgA1 protease. Proc Natl Acad Sci USA. 2003 Apr 1;100(7):4215–20.

197. Beiter K, Wartha F, Albiger B, Normark S, Zychlinsky A, Henriques-Normark B. An endonuclease allows Streptococcus pneumoniae to escape from neutrophil extracellular traps. Curr Biol. 2006 Feb 21;16(4):401–7.

198. Arnold RR, Cole MF, McGhee JR. A bactericidal effect for human lactoferrin. Science. 1977 Jul 15;197(4300):263–5.

199. Bellamy W, Takase M, Wakabayashi H, Kawase K, Tomita M. Antibacterial spectrum of lactoferricin B, a potent bactericidal peptide derived from the N-terminal region of bovine lactoferrin. J Appl Bacteriol. 1992 Dec;73(6):472–9.

200. Yamauchi K, Tomita M, Giehl TJ, Ellison RT. Antibacterial activity of lactoferrin and a pepsin-derived lactoferrin peptide fragment. Infect Immun. 1993 Feb;61(2):719–28.

201. Hood MI, Skaar EP. Nutritional immunity: transition metals at the pathogen-host interface. Nat Rev Microbiol [Internet]. 2012 Jul 16 [cited 2019 Nov 13];10(8). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3875331/

202. Núñez G, Sakamoto K, Soares MP. Innate Nutritional Immunity. J Immunol. 2018 01;201(1):11–8.

203. Dave S, Carmicle S, Hammerschmidt S, Pangburn MK, McDaniel LS. Dual roles of PspC, a surface protein of Streptococcus pneumoniae, in binding human secretory IgA and factor H. J Immunol. 2004 Jul 1;173(1):471–7.

204. Duthy TG, Ormsby RJ, Giannakis E, Ogunniyi AD, Stroeher UH, Paton JC, et al. The human complement regulator factor H binds pneumococcal surface protein PspC via short consensus repeats 13 to 15. Infect Immun. 2002 Oct;70(10):5604–11.

205. Lu L, Ma Y, Zhang J-R. Streptococcus pneumoniae recruits complement factor H through the amino terminus of CbpA. J Biol Chem. 2006 Jun 2;281(22):15464–74.

206. Saleh M, Bartual SG, Abdullah MR, Jensch I, Asmat TM, Petruschka L, et al. Molecular architecture of Streptococcus pneumoniae surface thioredoxin-fold lipoproteins crucial for extracellular oxidative stress resistance and maintenance of virulence. EMBO Molecular Medicine. 2013 Dec 2;5(12):1852–70.

207. Yesilkaya H, Andisi VF, Andrew PW, Bijlsma JJE. Streptococcus pneumoniae and reactive oxygen species: an unusual approach to living with radicals. Trends Microbiol. 2013 Apr;21(4):187–95.

208. Agarwal V, Asmat TM, Dierdorf NI, Hauck CR, Hammerschmidt S. Polymeric immunoglobulin receptor-mediated invasion of Streptococcus pneumoniae into host cells requires a coordinate signaling of SRC family of protein-tyrosine kinases, ERK, and c-Jun N-terminal kinase. J Biol Chem. 2010 Nov 12;285(46):35615–23.

209. Asmat TM, Agarwal V, Räth S, Hildebrandt J-P, Hammerschmidt S. Streptococcus pneumoniae Infection of Host Epithelial Cells via Polymeric Immunoglobulin Receptor Transiently Induces Calcium Release from Intracellular Stores. J Biol Chem. 2011 May 20;286(20):17861–9.

210. Asmat TM, Agarwal V, Saleh M, Hammerschmidt S. Endocytosis of Streptococcus pneumoniae via the polymeric immunoglobulin receptor of epithelial cells relies on clathrin and caveolin dependent mechanisms. Int J Med Microbiol. 2014 Nov;304(8):1233–46.

211. Fillon S, Soulis K, Rajasekaran S, Benedict-Hamilton H, Radin JN, Orihuela CJ, et al. Platelet-activating factor receptor and innate immunity: uptake of gram-positive bacterial cell wall into host cells and cell-specific pathophysiology. J Immunol. 2006 Nov 1;177(9):6182–91.

212. Zhang JR, Mostov KE, Lamm ME, Nanno M, Shimida S, Ohwaki M, et al. The polymeric immunoglobulin receptor translocates pneumococci across human nasopharyngeal epithelial cells. Cell. 2000 Sep 15;102(6):827–37.

213. Tilley SJ, Orlova EV, Gilbert RJC, Andrew PW, Saibil HR. Structural basis of pore formation by the bacterial toxin pneumolysin. Cell. 2005 Apr 22;121(2):247–56.

214. Bättig P, Mühlemann K. Influence of the spxB Gene on Competence in Streptococcus pneumoniae. J Bacteriol. 2008 Feb;190(4):1184–9.

215. Li J, Glover DT, Szalai AJ, Hollingshead SK, Briles DE. PspA and PspC minimize immune adherence and transfer of pneumococci from erythrocytes to macrophages through their effects on complement activation. Infect Immun. 2007 Dec;75(12):5877–85.

216. Voss S, Hallström T, Saleh M, Burchhardt G, Pribyl T, Singh B, et al. The choline-binding protein PspC of Streptococcus pneumoniae interacts with the C-terminal heparin-binding domain of vitronectin. J Biol Chem. 2013 May 31;288(22):15614–27.

217. Flemming H-C, Wuertz S. Bacteria and archaea on Earth and their abundance in biofilms. Nat Rev Microbiol. 2019 Apr;17(4):247–60.

218. Hoa M, Syamal M, Sachdeva L, Berk R, Coticchia J. Demonstration of nasopharyngeal and middle ear mucosal biofilms in an animal model of acute otitis media. Ann Otol Rhinol Laryngol. 2009 Apr;118(4):292–8.

219. Post JC, Hiller NL, Nistico L, Stoodley P, Ehrlich GD. The role of biofilms in otolaryngologic infections: update 2007. Curr Opin Otolaryngol Head Neck Surg. 2007 Oct;15(5):347–51.

220. Massa HM, Cripps AW, Lehmann D. Otitis media: viruses, bacteria, biofilms and vaccines. Med J Aust. 2009 02;191(S9):S44-49.

221. Stewart PS, Costerton JW. Antibiotic resistance of bacteria in biofilms. Lancet. 2001 Jul 14;358(9276):135–8.

222. Budhani RK, Struthers JK. The use of Sorbarod biofilms to study the antimicrobial susceptibility of a strain of Streptococcus pneumoniae. J Antimicrob Chemother. 1997 Oct;40(4):601–2.

223. Moscoso M, García E, López R. Biofilm Formation by Streptococcus pneumoniae: Role of Choline, Extracellular DNA, and Capsular Polysaccharide in Microbial Accretion. J Bacteriol. 2006 Nov 15;188(22):7785–95.

224. Donlan RM, Piede JA, Heyes CD, Sanii L, Murga R, Edmonds P, et al. Model system for growing and quantifying Streptococcus pneumoniae biofilms in situ and in real time. Appl Environ Microbiol. 2004 Aug;70(8):4980–8.

225. Goeres DM, Loetterle LR, Hamilton MA, Murga R, Kirby DW, Donlan RM. Statistical assessment of a laboratory method for growing biofilms. Microbiology (Reading, Engl). 2005 Mar;151(Pt 3):757–62.

226. Sanchez CJ, Hurtgen BJ, Lizcano A, Shivshankar P, Cole GT, Orihuela CJ. Biofilm and planktonic pneumococci demonstrate disparate immunoreactivity to human convalescent sera. BMC Microbiol. 2011;11:245.

227. Chao Y, Marks LR, Pettigrew MM, Hakansson AP. Streptococcus pneumoniae biofilm formation and dispersion during colonization and disease. Frontiers in cellular and infection microbiology. 2015;4(January):194.

228. Marks LR, Davidson BA, Knight PR, Hakansson AP. Interkingdom Signaling Induces Streptococcus pneumoniae Biofilm Dispersion and Transition from Asymptomatic Colonization to Disease. mBio. 2013;4(4):e00438-13.

229. Antic I, Brothers KM, Stolzer M, Lai H, Powell E, Eutsey R, et al. Gene Acquisition by a Distinct Phyletic Group within Streptococcus pneumoniae Promotes Adhesion to the Ocular Epithelium. mSphere. 2017 Oct 25;2(5):e00213-17.

230. Aprianto R, Slager J, Holsappel S, Veening J-W. High-resolution analysis of the pneumococcal transcriptome under a wide range of infection-relevant conditions. Nucleic Acids Res. 2018 02;46(19):9990–10006.

231. Cuevas RA, Ebrahimi E, Gazioglu O, Yesilkaya H, Hiller NL. Pneumococcal attachment to epithelial cells is enhanced by the secreted peptide VP1 via its control of hyaluronic acid processing. bioRxiv. 2019 Oct 1;788430.

232. Sutherland IW. The biofilm matrix--an immobilized but dynamic microbial environment. Trends Microbiol. 2001 May;9(5):222–7.

233. Flemming H-C, Wingender J. The biofilm matrix. Nat Rev Microbiol. 2010 Sep;8(9):623–33.

234. Hall-Stoodley L, Nistico L, Sambanthamoorthy K, Dice B, Nguyen D, Mershon WJ, et al. Characterization of biofilm matrix, degradation by DNase treatment and evidence of capsule downregulation in Streptococcus pneumoniae clinical isolates. BMC Microbiol. 2008;8:173.

235. Guiral S, Mitchell TJ, Martin B, Claverys J-P. Competence-programmed predation of noncompetent cells in the human pathogen Streptococcus pneumoniae: genetic requirements. Proc Natl Acad Sci USA. 2005 Jun 14;102(24):8710–5.

236. Eldholm V, Gutt B, Johnsborg O, Brückner R, Maurer P, Hakenbeck R, et al. The Pneumococcal Cell Envelope Stress-Sensing System LiaFSR Is Activated by Murein Hydrolases and Lipid II-Interacting Antibiotics. J Bacteriol. 2010 Apr 1;192(7):1761–73.

237. Johnsborg O, Eldholm V, Bjørnstad ML, Håvarstein LS. A predatory mechanism dramatically increases the efficiency of lateral gene transfer in Streptococcus pneumoniae and related commensal species. Mol Microbiol. 2008 Jul;69(1):245–53.

238. Yadav MK, Kwon SK, Cho CG, Park S-W, Chae S-W, Song J-J. Gene expression profile of early in vitro biofilms of Streptococcus pneumoniae. Microbiol Immunol. 2012 Sep;56(9):621–9.

239. Yadav MK, Go YY, Kim SH, Chae S-W, Song J-J. Antimicrobial and Antibiofilm Effects of Human Amniotic/Chorionic Membrane Extract on Streptococcus pneumoniae. Front Microbiol. 2017;8:1948.

240. Cundell DR, Gerard NP, Gerard C, Idanpaan-Heikkila I, Tuomanen EI. Streptococcus pneumoniae anchor to activated human cells by the receptor for platelet-activating factor. Nature. 1995 Oct 5;377(6548):435–8.

241. Jedrzejas MJ. Pneumococcal Virulence Factors: Structure and Function. Microbiol Mol Biol Rev. 2001 Jun;65(2):187–207.

242. Ren B, Li J, Genschmer K, Hollingshead SK, Briles DE. The absence of PspA or presence of antibody to PspA facilitates the complement-dependent phagocytosis of pneumococci in vitro. Clin Vaccine Immunol. 2012 Oct;19(10):1574–82.

243. Farnaud S, Evans RW. Lactoferrin--a multifunctional protein with antimicrobial properties. Mol Immunol. 2003 Nov;40(7):395–405.

244. Mirza S, Wilson L, Benjamin WH, Novak J, Barnes S, Hollingshead SK, et al. Serine protease PrtA from Streptococcus pneumoniae plays a role in the killing of S. pneumoniae by apolactoferrin. Infect Immun. 2011 Jun;79(6):2440–50.

245. Bethe G, Nau R, Wellmer A, Hakenbeck R, Reinert RR, Heinz HP, et al. The cell wall-associated serine protease PrtA: a highly conserved virulence factor of Streptococcus pneumoniae. FEMS Microbiol Lett. 2001 Nov 27;205(1):99–104.

246. Elm C, Braathen R, Bergmann S, Frank R, Vaerman J-P, Kaetzel CS, et al. Ectodomains 3 and 4 of human polymeric Immunoglobulin receptor (hpIgR) mediate invasion of Streptococcus pneumoniae into the epithelium. J Biol Chem. 2004 Feb 20;279(8):6296–304.

247. Elm C, Rohde M, Vaerman J-P, Chhatwal GS, Hammerschmidt S. Characterization of the interaction of the pneumococcal surface protein SpsA with the human polymeric immunoglobulin receptor (hpIgR). Indian J Med Res. 2004 May;119 Suppl:61–5.

248. Hammerschmidt S, Agarwal V, Kunert A, Haelbich S, Skerka C, Zipfel PF. The host immune regulator factor H interacts via two contact sites with the PspC protein of Streptococcus pneumoniae and mediates adhesion to host epithelial cells. J Immunol. 2007 May 1;178(9):5848–58.

249. Kohler S, Hallström T, Singh B, Riesbeck K, Spartà G, Zipfel PF, et al. Binding of vitronectin and Factor H to Hic contributes to immune evasion of Streptococcus pneumoniae serotype 3. Thromb Haemost. 2015 Jan;113(1):125–42.

250. Asmat TM, Klingbeil K, Jensch I, Burchhardt G, Hammerschmidt S. Heterologous expression of pneumococcal virulence factor PspC on the surface of Lactococcus lactis confers adhesive properties. Microbiology (Reading, Engl). 2012 Mar;158(Pt 3):771–80.

251. Rosenow C, Ryan P, Weiser JN, Johnson S, Fontan P, Ortqvist A, et al. Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of Streptococcus pneumoniae. Mol Microbiol. 1997 Sep;25(5):819–29.

252. Iannelli F, Chiavolini D, Ricci S, Oggioni MR, Pozzi G. Pneumococcal surface protein C contributes to sepsis caused by Streptococcus pneumoniae in mice. Infect Immun. 2004 May;72(5):3077–80.

253. Laitinen H, Tomasz A. Changes in composition of peptidoglycan during maturation of the cell wall in pneumococci. J Bacteriol. 1990 Oct;172(10):5961–7.

254. Berry AM, Paton JC. Additive attenuation of virulence of Streptococcus pneumoniae by mutation of the genes encoding pneumolysin and other putative pneumococcal virulence proteins. Infect Immun. 2000 Jan;68(1):133–40.

255. Orihuela CJ, Gao G, Francis KP, Yu J, Tuomanen EI. Tissue-Specific Contributions of Pneumococcal Virulence Factors to Pathogenesis. J Infect Dis. 2004 Nov 1;190(9):1661–9.

256. Obregón V, García P, García E, Fenoll A, López R, García JL. Molecular peculiarities of the lytA gene isolated from clinical pneumococcal strains that are bile insoluble. J Clin Microbiol. 2002 Jul;40(7):2545–54.

257. García P, González MP, García E, López R, García JL. LytB, a novel pneumococcal murein hydrolase essential for cell separation. Mol Microbiol. 1999 Feb;31(4):1275–81.

258. De Las Rivas B, García JL, López R, García P. Purification and polar localization of pneumococcal LytB, a putative endo-beta-N-acetylglucosaminidase: the chain-dispersing murein hydrolase. J Bacteriol. 2002 Sep;184(18):4988–5000.

259. López R, González MP, García E, García JL, García P. Biological roles of two new murein hydrolases of Streptococcus pneumoniae representing examples of module shuffling. Res Microbiol. 2000 Aug;151(6):437–43.

260. Berry AM, Paton JC. Sequence heterogeneity of PsaA, a 37-kilodalton putative adhesin essential for virulence of Streptococcus pneumoniae. Infect Immun. 1996 Dec;64(12):5255–62.

261. McAllister LJ, Tseng H-J, Ogunniyi AD, Jennings MP, McEwan AG, Paton JC. Molecular analysis of the psa permease complex of Streptococcus pneumoniae. Mol Microbiol. 2004 Aug;53(3):889–901.

262. Dintilhac A, Alloing G, Granadel C, Claverys JP. Competence and virulence of Streptococcus pneumoniae: Adc and PsaA mutants exhibit a requirement for Zn and Mn resulting from inactivation of putative ABC metal permeases. Mol Microbiol. 1997 Aug;25(4):727–39.

263. Tseng H-J, McEwan AG, Paton JC, Jennings MP. Virulence of Streptococcus pneumoniae: PsaA Mutants Are Hypersensitive to Oxidative Stress. Infect Immun. 2002 Mar;70(3):1635–9.

264. Jomaa M, Terry S, Hale C, Jones C, Dougan G, Brown J. Immunization with the iron uptake ABC transporter proteins PiaA and PiuA prevents respiratory infection with Streptococcus pneumoniae. Vaccine. 2006 Jun 12;24(24):5133–9.

265. Xu Q, Casey JR, Newman E, Pichichero ME. Otitis-prone Children Have Immunologic Deficiencies in Naturally Acquired Nasopharyngeal Mucosal Antibody Response after Streptococcus pneumoniae Colonization. Pediatr Infect Dis J. 2016 Jan;35(1):54–60.

266. Hermans PWM, Adrian PV, Albert C, Estevão S, Hoogenboezem T, Luijendijk IHT, et al. The streptococcal lipoprotein rotamase A (SlrA) is a functional peptidyl-prolyl isomerase involved in pneumococcal colonization. J Biol Chem. 2006 Jan 13;281(2):968–76.

267. Blue CE, Mitchell TJ. Contribution of a Response Regulator to the Virulence of Streptococcus pneumoniae Is Strain Dependent. Infect Immun. 2003 Aug 1;71(8):4405–13.

268. Frolet C, Beniazza M, Roux L, Gallet B, Noirclerc-Savoye M, Vernet T, et al. New adhesin functions of surface-exposed pneumococcal proteins. BMC Microbiol. 2010 Jul 12;10:190.

269. Govindarajan B, Menon BB, Spurr-Michaud S, Rastogi K, Gilmore MS, Argüeso P, et al. A metalloproteinase secreted by Streptococcus pneumoniae removes membrane mucin MUC16 from the epithelial glycocalyx barrier. PLoS ONE. 2012;7(3):e32418.

270. Bek-Thomsen M, Poulsen K, Kilian M. Occurrence and evolution of the paralogous zinc metalloproteases IgA1 protease, ZmpB, ZmpC, and ZmpD in Streptococcus pneumoniae and related commensal species. MBio. 2012;3(5).

271. Camilli R, Pettini E, Del Grosso M, Pozzi G, Pantosti A, Oggioni MR. Zinc metalloproteinase genes in clinical isolates of Streptococcus pneumoniae: association of the full array with a clonal cluster comprising serotypes 8 and 11A. Microbiology (Reading, Engl). 2006 Feb;152(Pt 2):313–21.

272. Holmes AR, McNab R, Millsap KW, Rohde M, Hammerschmidt S, Mawdsley JL, et al. The pavA gene of Streptococcus pneumoniae encodes a fibronectin-binding protein that is essential for virulence. Mol Microbiol. 2001 Sep;41(6):1395–408.

273. Noske N, Kämmerer U, Rohde M, Hammerschmidt S. Pneumococcal interaction with human dendritic cells: phagocytosis, survival, and induced adaptive immune response are manipulated by PavA. J Immunol. 2009 Aug 1;183(3):1952–63.

274. Kadioglu A, Brewin H, Härtel T, Brittan JL, Klein M, Hammerschmidt S, et al. Pneumococcal protein PavA is important for nasopharyngeal carriage and development of sepsis. Mol Oral Microbiol. 2010 Feb;25(1):50–60.

275. Brittan JL, Buckeridge TJ, Finn A, Kadioglu A, Jenkinson HF. Pneumococcal neuraminidase A: an essential upper airway colonization factor for Streptococcus pneumoniae. Mol Oral Microbiol. 2012 Aug;27(4):270–83.

276. Pettigrew MM, Fennie KP, York MP, Daniels J, Ghaffar F. Variation in the presence of neuraminidase genes among Streptococcus pneumoniae isolates with identical sequence types. Infect Immun. 2006 Jun;74(6):3360–5.

277. King SJ, Hippe KR, Gould JM, Bae D, Peterson S, Cline RT, et al. Phase variable desialylation of host proteins that bind to Streptococcus pneumoniae in vivo and protect the airway. Mol Microbiol. 2004 Oct;54(1):159–71.

278. Parker D, Soong G, Planet P, Brower J, Ratner AJ, Prince A. The NanA neuraminidase of Streptococcus pneumoniae is involved in biofilm formation. Infect Immun. 2009 Sep;77(9):3722–30.

279. Manco S, Hernon F, Yesilkaya H, Paton JC, Andrew PW, Kadioglu A. Pneumococcal Neuraminidases A and B Both Have Essential Roles during Infection of the Respiratory Tract and Sepsis. Infect Immun. 2006 Jul 1;74(7):4014–20.

280. Paixão L, Oliveira J, Veríssimo A, Vinga S, Lourenço EC, Ventura MR, et al. Host glycan sugar-specific pathways in Streptococcus pneumonia: galactose as a key sugar in colonisation and infection. PLoS ONE. 2015;10(3):e0121042.

281. Krivan H, Roberts D, Ginsburg V. Many pulmonary pathogenic bacteria bind specifically to the carbohydrate sequence GalNAc beta 1-4Gal found in some glycolipids. - PubMed - NCBI [Internet]. [cited 2019 Nov 19]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/3413084

282. Singh AK, Pluvinage B, Higgins MA, Dalia AB, Woodiga SA, Flynn M, et al. Unravelling the multiple functions of the architecturally intricate Streptococcus pneumoniae β-galactosidase, BgaA. PLoS Pathog. 2014 Sep;10(9):e1004364.

283. Stern R, Jedrzejas MJ. Hyaluronidases: their genomics, structures, and mechanisms of action. Chem Rev. 2006 Mar;106(3):818–39.

284. Oiki S, Mikami B, Maruyama Y, Murata K, Hashimoto W. A bacterial ABC transporter enables import of mammalian host glycosaminoglycans. Sci Rep. 2017 21;7(1):1069.

285. Marion C, Stewart JM, Tazi MF, Burnaugh AM, Linke CM, Woodiga SA, et al. Streptococcus pneumoniae can utilize multiple sources of hyaluronic acid for growth. Infect Immun. 2012 Apr;80(4):1390–8.

286. Yadav MK, Chae S-W, Park K, Song J-J. Hyaluronic acid derived from other streptococci supports Streptococcus pneumoniae in vitro biofilm formation. Biomed Res Int. 2013;2013:690217.

287. Chapuy-Regaud S, Ogunniyi AD, Diallo N, Huet Y, Desnottes J-F, Paton JC, et al. RegR, a global LacI/GalR family regulator, modulates virulence and competence in Streptococcus pneumoniae. Infect Immun. 2003 May;71(5):2615–25.

288. Nandoskar M, Ferrante A, Bates EJ, Hurst N, Paton JC. Inhibition of human monocyte respiratory burst, degranulation, phospholipid methylation and bactericidal activity by pneumolysin. Immunology. 1986 Dec;59(4):515–20.

289. Kadioglu A, Coward W, Colston MJ, Hewitt CRA, Andrew PW. CD4-T-lymphocyte interactions with pneumolysin and pneumococci suggest a crucial protective role in the host response to pneumococcal infection. Infect Immun. 2004 May;72(5):2689–97.

290. Yuste J, Botto M, Paton JC, Holden DW, Brown JS. Additive inhibition of complement deposition by pneumolysin and PspA facilitates Streptococcus pneumoniae septicemia. J Immunol. 2005 Aug 1;175(3):1813–9.

291. Hirst RA, Kadioglu A, O'callaghan C, Andrew PW. The role of pneumolysin in pneumococcal pneumonia and meningitis. Clin Exp Immunol. 2004 Nov;138(2):195–201.

292. Hotomi M, Yuasa J, Briles DE, Yamanaka N. Pneumolysin plays a key role at the initial step of establishing pneumococcal nasal colonization. Folia Microbiol (Praha). 2016 Sep;61(5):375–83.

293. Rayner CF, Jackson AD, Rutman A, Dewar A, Mitchell TJ, Andrew PW, et al. Interaction of pneumolysin-sufficient and -deficient isogenic variants of Streptococcus pneumoniae with human respiratory mucosa. Infect Immun. 1995 Feb;63(2):442–7.

294. Kadioglu A, Gingles NA, Grattan K, Kerr A, Mitchell TJ, Andrew PW. Host cellular immune response to pneumococcal lung infection in mice. Infect Immun. 2000 Feb;68(2):492–501.

295. Kadioglu A, Taylor S, Iannelli F, Pozzi G, Mitchell TJ, Andrew PW. Upper and lower respiratory tract infection by Streptococcus pneumoniae is affected by pneumolysin deficiency and differences in capsule type. Infect Immun. 2002 Jun;70(6):2886–90.

296. Ehrlich GD, Hiller NL, Hu FZ. What makes pathogens pathogenic. Genome biology. 2008;9(6):225.

297. Lee MS, Morrison DA. Identification of a New Regulator in Streptococcus pneumoniae Linking Quorum Sensing to Competence for Genetic Transformation. Journal of bacteriology. 1999;181(16):5004–16.

298. Declerck N, Bouillaut L, Chaix D, Rugani N, Slamti L, Hoh F, et al. Structure of PlcR: Insights into virulence regulation and evolution of quorum sensing in Gram-positive bacteria. Proc Natl Acad Sci USA. 2007;104(47):18490–5.

299. Lyon GJ, Novick RP. Peptide signaling in Staphylococcus aureus and other Gram-positive bacteria. Peptides. 2004;25:1389–403.

300. Kuipers OP, Ruyter PGGA De, Kleerebezem M, Vos WM De. Quorum sensing-controlled gene expression in lactic acid bacteria. Journal of Biotechnology. 1998;64:15–21.

301. Bassler BL. Small Talk : Cell-to-Cell Communication in Bacteria. Cell. 2002;109:421–4.

302. Bassler BL. How bacteria talk to each other : regulation of gene expression by quorum sensing. Current Opinion in Microbiology. 1999;2:582–7.

303. Kleerebezem M, Quadri LEN, Oscar P. Quorum sensing by peptide pheromones and two-component signal-transduction systems in Gram-positive bacteria. Molecular Microbiology. 1997;24:895–904.

304. Williams P. Quorum sensing, communication and cross-kingdom signalling in the bacterial world. Microbiology. 2007;153(12):3923–38.

305. Popat R, Crusz SA, Diggle SP. The social behaviours of bacterial pathogens. Br Med Bull. 2008;87:63–75.

306. Li Y-H, Tian X. Quorum Sensing and Bacterial Social Interactions in Biofilms. Sensors (Basel). 2012;12(2):2519–2538.

307. Fleuchot B, Gitton C, Guillot A, Vidic J, Nicolas P, Besset C, et al. Rgg proteins associated with internalized small hydrophobic peptides: a new quorum-sensing mechanism in streptococci. Mol Microbiol. 2011 May;80(4):1102–19.

308. Dawid S, Roche AM, Weiser JN. The blp bacteriocins of Streptococcus pneumoniae mediate intraspecies competition both in vitro and in vivo. Infection and Immunity. 2007;75(1):443–51.

309. Håvarstein LS, Gaustad P, Nes IF, Morrison DA. Identification of the streptococcal competence-pheromone receptor. Molecular Microbiology. 1996 Aug;21(4):863–9.

310. Nes IF, Yoon S-S, D.B. D. Ribosomally Synthesiszed Antimicrobial Peptides ( Bacteriocins ) in Lactic Acid Bacteria : A Review. Food Sci Biotechnol. 2007;16(5):675–90.

311. Härtel T, Eylert E, Schulz C, Petruschka L, Gierok P, Grubmüller S, et al. Characterization of Central Carbon Metabolism of Streptococcus pneumoniae by Isotopologue Profiling. The Journal of Biological Chemistry. 2012;287(6):4260–74.

312. Hoover SE, Perez AJ, Tsui H-CT, Sinha D, Smiley DL, DiMarchi RD, et al. A new quorum-sensing system (TprA/PhrA) for Streptococcus pneumoniae D39 that regulates a lantibiotic biosynthesis gene cluster. Mol Microbiol. 2015 Apr 13;

313. Darmstadt G, Mentele L, Podbielski A, Rubens C. Role of group A streptococcal virulence factors in adherence to keratinocytes. Infection and Immunity. 2000;68(3):1215–21.

314. Gohar M, Faegri K, Perchat S, Ravnum S, Økstad O, Gominet M, et al. The PlcR virulence regulon of Bacillus cereus. Plos One. 2008;3(7):e2793.

315. Rocha-Estrada J, Aceves-diez AE. The RNPP family of quorum-sensing proteins in Gram-positive bacteria. Appl Microbiol Biotechnol. 2010;87:913–23.

316. Bortoni ME, Terra VS, Hinds J, Andrew PW, Yesilkaya H. The pneumococcal response to oxidative stress includes a role for Rgg. Microbiology (Reading, Engl). 2009 Dec;155(Pt 12):4123–34.

317. Kadam A, Janto B, Eutsey R, Earl JP, Powell E, Dahlgren ME, et al. Streptococcus pneumoniae Supragenome Hybridization Arrays for Profiling of Genetic Content and Gene Expression. Curr Protoc Microbiol. 2015;36:9D.4.1-9D.4.20.

318. Zhi X, Abdullah IT, Gazioglu O, Manzoor I, Shafeeq S, Kuipers OP, et al. Rgg-Shp regulators are important for pneumococcal colonization and invasion through their effect on mannose utilization and capsule synthesis. Sci Rep. 2018 Apr 23;8(1):6369.

319. Kadam A, Eutsey RA, Rosch J, Miao X, Longwell M, Xu W, et al. Promiscuous signaling by a regulatory system unique to the pandemic PMEN1 pneumococcal lineage. PLOS Pathogens. 2017 May 18;13(5):e1006339.

320. Motib AS, Al-Bayati FAY, Manzoor I, Shafeeq S, Kadam A, Kuipers OP, et al. TprA/PhrA Quorum Sensing System Has a Major Effect on Pneumococcal Survival in Respiratory Tract and Blood, and Its Activity Is Controlled by CcpA and GlnR. Front Cell Infect Microbiol [Internet]. 2019 Sep 13 [cited 2019 Oct 27];9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6753895/

321. Aggarwal C, Jimenez JC, Nanavati D, Federle MJ. Multiple length peptide-pheromone variants produced by Streptococcus pyogenes directly bind Rgg proteins to confer transcriptional regulation. J Biol Chem. 2014 Aug 8;289(32):22427–36.

322. Ibrahim M, Guillot A, Wessner F, Algaron F, Besset C, Courtin P, et al. Control of the transcription of a short gene encoding a cyclic peptide in Streptococcus thermophilus: a new quorum-sensing system? J Bacteriol. 2007 Dec;189(24):8844–54.

323. Pottathil M, Lazazzera B. The extracellular Phr peptide–Rap phosphatase signaling circuit of Bacillus subtilis. Front Biosci. 2003;8:32–45.

324. Dunny G. The peptide pheromone-inducible conjugation system of Enterococcus faecalis plasmid pCF10: cell–cell signaling, gene transfer, complexity and evolution. Philos Trans R Soc B. 2007;362:1185–1193.

325. Pérez-Pascual D, Gaudu P, Fleuchot B, Besset C, Rosinski-Chupin I, Guillot A, et al. RovS and Its Associated Signaling Peptide Form a Cell-To-Cell Communication System Required for Streptococcus agalactiae Pathogenesis. mBio. 2015 Feb 27;6(1):e02306-14.

326. Chang JC, Federle MJ. PptAB Exports Rgg Quorum-Sensing Peptides in Streptococcus. PLoS ONE. 2016;11(12):e0168461.

327. Cook LC, Federle MJ. Peptide pheromone signaling in Streptococcus and Enterococcus. FEMS Microbiol Rev. 2014 May;38(3):473–92.

328. Besset C, Chambellon E, Fleuchot B, Guillot A, Me C. Rgg-Associated SHP Signaling Peptides Mediate Cross- Talk in Streptococci. 2013;8(6).

329. Sanders JW, Leenhouts K, Burghoorn J, Brands JR, Venema G, Kok J. A chloride-inducible acid resistance mechanism in Lactococcus lactis and its regulation. Molecular Microbiology. 1998;27(2):299–310.

330. Qi F, Chen P, Caufield PW. Functional analyses of the promoters in the lantibiotic mutacin II biosynthetic locus in Streptococcus mutans. Applied and Environmental Microbiology. 1999;65(2):652–8.

331. Junges R, Salvadori G, Shekhar S, Åmdal HA, Periselneris JN, Chen T, et al. A Quorum-Sensing System That Regulates Streptococcus pneumoniae Biofilm Formation and Surface Polysaccharide Production. mSphere [Internet]. 2017 Sep 13;2(5). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5597970/

332. Cuevas RA, Eutsey R, Kadam A, West-Roberts JA, Woolford CA, Mitchell AP, et al. A novel streptococcal cell-cell communication peptide promotes pneumococcal virulence and biofilm formation. Mol Microbiol. 2017 Aug;105(4):554–71.

333. Chaussee MS, Somerville G a, Reitzer L, Musser JM. Rgg Coordinates Virulence Factor Synthesis and Metabolism in Streptococcus pyogenes Rgg Coordinates Virulence Factor Synthesis and Metabolism in Streptococcus pyogenes. Journal of bacteriology. 2003;185(20):6016–6024.

334. Chang J, LaSarre B, Jimenez J, Aggarwal C, Federle M. Two Group A Streptococcal Peptide Pheromones Act through Opposing Rgg Regulators to Control Biofilm Development. PLoS Pathogens. 2011;7(8):e1002190.

335. Perez-Pascual D, Monnet V, Gardan R. Bacterial Cell–Cell Communication in the Host via RRNPP Peptide-Binding Regulators. Front Microbiol [Internet]. 2016 May 20 [cited 2016 Sep 15];7. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4873490/

336. Parashar V, Aggarwal C, Federle MJ, Neiditch MB. Rgg protein structure–function and inhibition by cyclic peptide compounds. Proc Natl Acad Sci U S A. 2015 Apr 21;112(16):5177–82.

337. Lysenko ES, Lijek RS, Brown SP, Weiser JN. Within-host competition drives selection for the capsule virulence determinant of Streptococcus pneumoniae. Curr Biol. 2010 Jul 13;20(13):1222–6.

338. Havarstein LS, Coomaraswamy G, Morrison DA. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in Streptococcus pneumoniae. Proc Natl Acad Sci USA. 1995 Nov 21;92(24):11140–4.

339. Havarstein LS, Coomaraswamy G, Morrison DA. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in Streptococcus pneumoniae. Proc Natl Acad Sci U S A. 1995 Nov 21;92(24):11140–4.

340. Cozzone AJ. Protein phosphorylation in prokaryotes. Annu Rev Microbiol. 1988;42:97–125.

341. Lange R, Wagner C, de Saizieu A, Flint N, Molnos J, Stieger M, et al. Domain organization and molecular characterization of 13 two-component systems identified by genome sequencing of Streptococcus pneumoniae. Gene. 1999 Sep 3;237(1):223–34.

342. Throup JP, Koretke KK, Bryant AP, Ingraham KA, Chalker AF, Ge Y, et al. A genomic analysis of two-component signal transduction in Streptococcus pneumoniae. Mol Microbiol. 2000 Feb;35(3):566–76.

343. Gómez-Mejia A, Gámez G, Hammerschmidt S. Streptococcus pneumoniae two-component regulatory systems: The interplay of the pneumococcus with its environment. International Journal of Medical Microbiology. 2018 Aug 1;308(6):722–37.

344. Throup JP, Koretke KK, Bryant AP, Ingraham KA, Chalker AF, Ge Y, et al. A genomic analysis of two-component signal transduction in Streptococcus pneumoniae. Mol Microbiol. 2000 Feb;35(3):566–76.

345. Leer RJ, van der Vossen JM, van Giezen M, van Noort JM, Pouwels PH. Genetic analysis of acidocin B, a novel bacteriocin produced by Lactobacillus acidophilus. Microbiology (Reading, Engl). 1995 Jul;141 ( Pt 7):1629–35.

346. Worobo RW, Van Belkum MJ, Sailer M, Roy KL, Vederas JC, Stiles ME. A signal peptide secretion-dependent bacteriocin from Carnobacterium divergens. J Bacteriol. 1995 Jun;177(11):3143–9.

347. Havarstein LS, Holo H, Nes IF. The leader peptide of colicin V shares consensus sequences with leader peptides that are common among peptide bacteriocins produced by gram-positive bacteria. Microbiology (Reading, Engl). 1994 Sep;140 ( Pt 9):2383–9.

348. Havarstein LS, Diep DB, Nes IF. A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. Mol Microbiol. 1995 Apr;16(2):229–40.

349. Fontaine L, Boutry C, de Frahan MH, Delplace B, Fremaux C, Horvath P, et al. A novel pheromone quorum-sensing system controls the development of natural competence in Streptococcus thermophilus and Streptococcus salivarius. J Bacteriol. 2010 Mar;192(5):1444–54.

350. Pestova EV, Håvarstein LS, Morrison DA. Regulation of competence for genetic transformation in Streptococcus pneumoniae by an auto-induced peptide pheromone and a two-component regulatory system. Mol Microbiol. 1996 Aug;21(4):853–62.

351. Claverys J-P, Prudhomme M, Martin B. Induction of Competence Regulons as a General Response to Stress in Gram-Positive Bacteria. Annual Review of Microbiology. 2006;60(1):451–75.

352. Peterson SN, Sung CK, Cline R, Desai BV, Snesrud EC, Luo P, et al. Identification of competence pheromone responsive genes in Streptococcus pneumoniae by use of DNA microarrays. Mol Microbiol. 2004 Feb;51(4):1051–70.

353. Wang CY, Patel N, Wholey W-Y, Dawid S. ABC transporter content diversity in Streptococcus pneumoniae impacts competence regulation and bacteriocin production. Proc Natl Acad Sci USA. 2018 19;115(25):E5776–85.

354. Wholey W-Y, Abu-Khdeir M, Yu EA, Siddiqui S, Esimai O, Dawid S. Characterization of the Competitive Pneumocin Peptides of Streptococcus pneumoniae. Front Cell Infect Microbiol. 2019;9:55.

# PART II RESULTS

## CHAPTER 2 - A novel streptococcal cell-cell communication peptide promotes pneumococcal virulence and biofilm formation

This chapter focuses on the motivation that led to the discovery of the signaling peptide VP1. I will explain the bioinformatic analysis that allows us to identify and determine the key motif common to a particular type of signaling peptides in the pneumococcus. These criteria serve as a signature to classify new potential classes of signaling peptides in streptococci. Then, I will detail the criteria utilized to prioritize on a selected candidate for further analyses. Finally, I will describe how the VP1 peptide influences the pneumococcal biofilm capacity. Part of this work was published in 2017 in the Journal Molecular Microbiology, and since them, I supplemented and updated the manuscript that this chapter covers with unreleased data. The original manuscript can be accessed using the following reference:

**Cuevas RA**, Eutsey R, Kadam A, West-Roberts JA, Woolford CA, Mitchell AP, Mason KM, Hiller NL. A novel streptococcal cell-cell communication peptide promotes pneumococcal virulence and biofilm formation. Molecular microbiology. **2017**; 105(4):554-571. NIHMSID: NIHMS879801 PubMed PMID: 28557053, PMCID: PMC5550342

**Author Contributions**. Rolando A. Cuevas contributed to the conception, design, experimentation, analysis and interpretation, and manuscript preparation; Rory Eutsey contributed to bacterial construction; Dr. Anagha Kadam contributed to NanoString data interpretation. Jacob A. West-Roberts contributed to phylogenetic analysis; Dr. Carol A. Woolford contributed to sample manipulation; Dr. Aaron P. Mitchell contributed to manuscript edition; Dr. Kevin M. Mason contributed to CMEE cells; Dr. N. Luisa Hiller contributed to conception, design, analysis and interpretation, and manuscript preparation.

## 2.1 Abstract

*Streptococcus pneumoniae* (pneumococcus) is a major human pathogen. It is a common colonizer of the human respiratory tract, where it utilizes various communication systems to coordinate population-level behaviors. We hypothesized that secreted peptides that are highly expressed during infection are pivotal for virulence. Thus, we used a bioinformatic approach to define a pneumococcal transcriptome and secretome in the clinically important PMEN1 lineage. Next, we measured the resulting peptide-encoding genes expression in an animal model of otitis media infection. In this study, we characterized the virulence peptide 1 (*vp1*), a highly expressed Gly-Gly peptide-encoding gene in chinchilla middle ear effusions. The *vp1* gene is widely distributed across pneumococcus as well as encoded in related species. Further, studies in the chinchilla model of middle ear infection demonstrated that VP1 is a virulence determinant. The *vp1* gene is positively regulated by a transcription factor from the Rgg family and its cognate regulatory peptide SHP (short hydrophobic peptide). *In vitro* data indicated that VP1 promotes biofilm development by increasing the thickness and the biomass of pneumococci biofilms grown on chinchilla middle ear epithelial cells. Further, the wild-type biofilm capacity is restored with the exogenous addition of a synthetic VP1. Our results reveal for the first time how a signaling molecule influences both biofilm capacity and virulent traits in a clinical pneumococcal isolated. We conclude that VP1 is a novel streptococcal regulatory peptide that controls biofilm development and pneumococcal pathogenesis.

## 2.2 Introduction

A wide range of bacterial pathogens can cause diseases by a coordinated communication behavior across cells at a population-level (1–3). *Streptococcus pneumoniae* (pneumococcus) is a major human pathogen that engages in social behaviors. Secreted pneumococcal peptides have been shown to coordinate competence and bactericidal activity, and have been implicated in the extensive genomic plasticity and adaptability of this species (4–7). However, additional signaling molecules that control pneumococcal behaviors in the context of the human host are poorly understood.

Bacterial biofilms are recalcitrant to antibiotic treatment, and thus pose a major threat to human health (8). Pneumococcus forms biofilms on mucosal surfaces such as the nasopharynx during asymptomatic colonization, the middle ear during otitis media, and the sinus during rhinosinusitis (9–12). Pneumococcal biofilms play a central role in drug resistance, as cells in this mode of growth display decreased sensitivity to antibiotics (13). Further, the biofilm is an environment propitious for transformation. The available DNA in the matrix serves as a substrate for transformation, leading to strain evolution and the spread of vaccine-escape and drug-resistant genotypes (14–17). Pneumococcal biofilms play a key role in the virulence potential of an infection. Transcriptional changes induced during a biofilm mode of growth have been associated with pneumococcal colonization (18). Furthermore, bacteria dispersed from a biofilm display increased propensity for tissue dissemination and pathogenesis, when compared to cells from the biofilm layer or planktonic cultures (19,20). In summary, pneumococcal biofilms are of high clinical importance for drug-resistance and pathogenesis.

Bacteria orchestrate community behaviors to cause disease (21). In Gram-positive bacteria, these behaviors mostly rely on signaling cascades regulated by extracellular peptides (22). Many of the Gram-positive ribosome-synthesized peptides that initiate cell-cell communication signaling cascades, can be categorized based on sequence features (23). The pneumococcal CSP and BIP belong to the double glycine (Gly-Gly) peptide family, a group distinguished by a cleavable N-terminal domain with a

double glycine motif at the cleavage site (7). The pro-peptide is processed and exported by an ATP-binding cassette (ABC) transporter, and cleavage occurs immediately after the double glycine (24). The secreted forms are active and bind the input domain of a surface-exposed histidine kinases from a two component system (TCS) (25,26). Another distinct peptide group is represented by the pneumococcal PhrA, which belongs to the family of PlcR (phospholipase C regulator) cognate peptides, a group secreted by the Sec pathway and internalized by transporters from the conserved oligopeptidase permease family (27). Finally, pneumococci also encode peptides belonging to the family of Rgg (regulator gene glucosyltransferase) cognate peptides, specifically the short hydrophobic sequences (SHPs) (28). The SHPs fall into three categories, identified based on the nature of the charged residue and/or organization relative to the Rgg (28). Pneumococcal genomes encode SHPs from groups I and III (28). These SHPs are predicted to be processed and secreted by ABC transporters and internalized by members of the conserved oligopeptidase permease family (29). The molecular components of trafficking provide clues regarding the likelihood that signaling peptides are functional across strain and species barriers (30). The conserved features of these peptide families set the stage for *in silico* pattern searches to identify novel signaling molecules.

Secreted peptides can exert their influence via multiple different types of signal transduction systems. The best-characterized signaling pathway is the TCS. In this architecture, the secreted peptide binds a surface exposed histidine kinase triggering autophosphorylation and subsequent phosphotransfer to a cognate response regulator (31). There are thirteen TCSs in the pneumococcal core genome (32,33). However, cognate peptides have been identified for only two of these systems: CSP for ComDE and BIP for BlpDE (for the remainder, the molecule(s) that directly activate the kinases are unknown) (25,26). More recent studies have revealed a distinct signal transduction architecture, where an active form of a secreted peptide is internalized into cells and directly binds a transcription factor from the RRNPP (Rap, Rgg, NprR, PlcR, PrgX) superfamily (34). There are multiple RRNPP systems encoded in the pneumococcal pangenome (35–37), and cognate peptides have been identified for two of the

regulators (PhrA for TprA and PhrA2 for TprA2) (36,37). Thus, pneumococcal strains make use of distinct peptide-induced signal transduction pathways during infection.

Pneumococcus can asymptomatically colonize the nasopharynx. However, in a percentage of cases pneumococcal strains disseminate to other tissues and cause moderate to severe disease (38). Bacterial dissemination is often initiated as a response to extracellular stresses caused by changes in microbiome composition, immune defenses, and/or nutritional changes (39,40). Free sugars are scarce in the nasopharynx and upper respiratory tract, such that pneumococci survive by breaking down host mucins into galactose and other free complex sugars. In contrast, levels of glucose and/or free sugars are substantially higher in the lower respiratory tract and the blood leading to tissue-specific differences in pathways utilized for the uptake and processing of nutrients (41,42). In pneumococcus, CodY, GntR, and CcpA are master regulators involved in amino acid metabolism, carbohydrate metabolism, and iron uptake (43–45). Further, RRNPP regulators have been associated with metabolic control in multiple species (23,36,46–48). Transcriptomic studies demonstrate that the PlcR-type regulator TprA (spd1745) is regulated by CcpA (43), and that a Rgg-type regulator (spr0144) is regulated by CodY and by effectors of GntR (36,44,45). Thus, pneumococcal RRNPP regulators are likely to play an important role in nutritional response, and consequently in population level behaviors that promote either dissemination or asymptomatic colonization.

Regulators in the TCS and RRNPP signaling pathways can monitor the extracellular environment and thus are well positioned in the cellular circuitry to control microbial-host interactions, intra- and inter-strain interactions, biofilm development, and virulence. We hypothesized that bacterial communication peptides involved in virulence are secreted and highly expressed *in vivo*. For this study, we selected the multi-drug resistant PMEN1 lineage (multi locus sequence type 81 and capsular type 23F). This lineage represents a significant percentage of carriage clones, have high incidence recorded among OM cases, and cause a wide spectrum of invasive disease (49,50). In particular we selected the clinical strain PN4595-T23, which was isolated from the nasopharynx of a child in Portugal (51). Two research advantages made us chose to select this strain: firstly, the whole reference genome is available (GenBank:

FM211187.1) and secondly, the strain is a low passage strain, which guarantees that the strain behaves as wild-type strain. Further, as we show later, that PN4595-T23 is highly virulent in the pneumococcal OM chinchilla model.

In this section, we employed comparative genomics to identify putative secreted peptides in the pneumococcal pangenome, and prioritized candidates based on expression levels *in vivo*. In this study, we characterize the virulence peptide 1 (*vp1*). The gene encoding VP1 is highly expressed during infection and regulated by CodY and Rgg. VP1 enhances biofilm development and activates the pneumococcal virulence potential.

## 2.3 Results

### 2.3.1 Identification of Gly-Gly peptides in the pneumococcal pangenome

We reasoned that secreted peptides that are highly expressed during infection and/or highly induced relative to planktonic growth are critical in pneumococcal virulence. Thus, we undertook a two-pronged approach: a) identify secreted peptides in the pneumococcal genome and b) measure their expression in a model of infection.

To identify secreted peptides, we focused our attention on the pneumococcal peptides CSP and BIP. Both are exported by a dedicated ABC transporter with a proteolytic domain required for peptide cleavage (24). The ABC transporters recognize N-terminal sequences with conserved features, notably the presence of the previously characterized Gly-Gly leader peptide (LSXXELXXIXGG) (7). To search for additional secreted pneumococcal peptides, we used Multiple Expectation Maximization for Motif Elicitation (MEME) to generate a position dependent probability matrix that defines the Gly-Gly motif and captures its length and positional variability at each peptide position. Our input consisted of 167 homologues of CSP and BIP, obtained from a NCBI blastP searched of pneumococcal CSP and BIP against the non-redundant protein database. Next, we employed the Motif Alignment and Search Tool (MAST) to search for the MEME motif in a database of coding sequences from 60 streptococcal genomes. This database is composed of thirty-nine *S. pneumoniae* (encompassing nineteen serotypes and thirty-six MLST types), three *S. pseudopneumoniae*, nine *S. mitis*, and seven *S. oralis* strains, selected to represent much of the genomic diversity of pneumococcus and viridans, plus two *S. infantis* strains as representatives of more distantly related genomes (**Suppl. Table 2.1**). The output of this MAST search included 192 protein sequences. This expanded set of sequences was used as input to generate a MEME motif, and subsequently search the same database with MAST (**Suppl. Table 2.2**). Reiterative use of MEME and MAST returned a list of 716 proteins, of which 696 encode the motif at the N-terminus and 98% are peptides with less than 130 residues. To illustrate the domain conserved across these sequences we generated a MEME logo

using the 696 sequences (**Figure 2.1)**. This set of sequences from the MAST output corresponds to a predicted secretome.

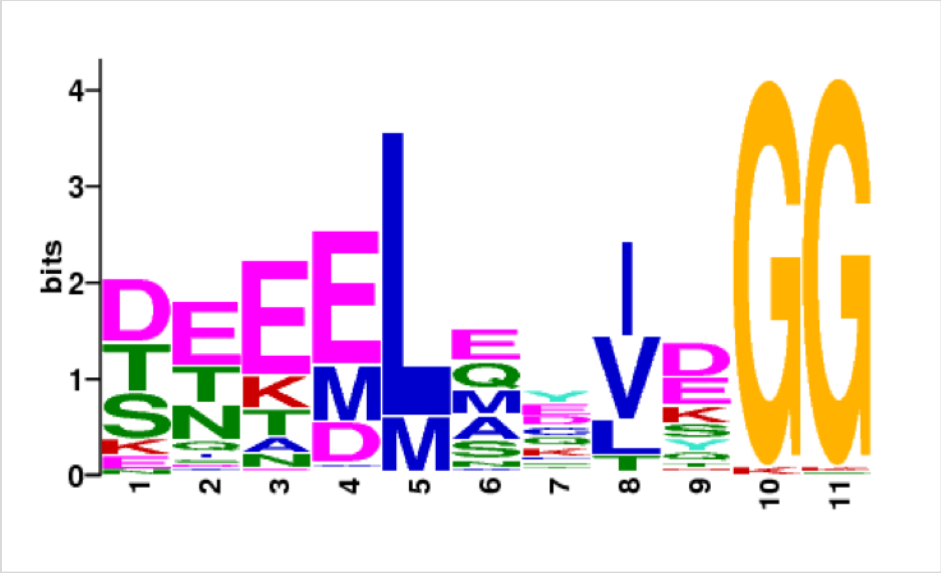



**Figure 2.1 Identification of a predicted secretome of Gly-Gly family peptides**
Sequence logo derived from 696 predicted secreted peptides. Amino acids are represented by one letter abbreviation and color coded as follows: blue, basic; red, acidic; green, polar; yellow, neutral. Height of amino acids is proportional to the fraction of the observed frequency relative to the expected frequency in the streptococcal database.

The predicted secretome was organized into 25 clusters using sequence similarity and genomic localization (**Table 2.1**). These clusters display variable distributions across strains from the same species and across species. As expected, this set of clusters contains the input molecules CSP and BIP. In addition, this set also includes previously identified peptides, such as seven peptides localized into multiple *blp loci* (52,53), the bacteriocidal molecules CibA and CibB (54), and the LanA lantibiotic (36). We also identified thirteen peptides that have not yet been functionally characterized. Twelve of these are present in some subset of pneumococcal strains as determined using blastP against the NCBI database of *Streptococcus pneumoniae*, and one is present in *S. oralis* and *S. mitis* but appears to be absent in pneumococcal strains. The pneumococcal peptides range from rare in the species to distributed in the vast majority of strains. We hypothesize that these peptides correspond to the Gly-Gly components of the pneumococcal secretome.

**Table 2.1 List of Putative Secreted Peptides**

| Blast ID | Annotation |
|---|---|
| SPN23F01520 | VP1 |
| SPN23F04030 | Hypothetical |
| CGSSp9BS68_07272 | Hypothetical |
| SPH_0224 | Hypothetical |
| SP_0924 | Hypothetical |
| SPH_0218 | Hypothetical |
| SP_0115 | Hypothetical |
| SPN23F01370 | CibB |
| SPN23F01380 | CibA |
| SPN23F19710 | LanA |
| **SPN23F22700** | **ComC (CSP2)** |
| **SPN23F04791** | **BlpC (BIP)** |
| SPN23F04830 | BlpI |
| SPN23F04840 | BlpJ |
| SPN23F04850 | BlpK |
| SPN23F04860 | BlpO |
| SPN23F00570 | BlpU |
| SP_0539 | BlpM |
| SP_0540 | BlpN |
| WP_001814148 (SPH_0227) | Hypothetical |
| smi_0057.1 | Hypothetical |
| smi_0059.1 | Hypothetical |
| smi_0061 | Hypothetical |
| HMPREF1112_0784 | Hypothetical |
| H354_08695 | Hypothetical (not present in pneumococcus) |

IDs correspond to pneumococcal strain ATCC700669. If gene is absent in this strain, the IDs from representative strains (TIGR4, BS9BS68, and Hungary19A) are provided. For peptides observed outside the selected set of pneumococcal genomes the IDs correspond to *Streptococcus mitis* B6 or *Streptococcus oralis* TAZ3a. Bold: Input for original MEME. Blue: previously characterized and known genes.

To prioritize on the pneumococcal peptides for characterization, we measured their transcription in the chinchilla model of human middle ear infection. To this end, we selected an isolate from the clinically important PMEN1 lineage (PN4595-T23). The PMEN1 lineage is widespread and multidrug-resistant, and includes vaccine-escape isolates (55–57). To measure gene expression, we employed nCounter NanoString technology, which provides an automated, highly sensitive enumeration of pathogen mRNA transcripts in infected tissues (58). Our probe set included eleven peptides identified from the MEME/MAST and encoded in PMEN1 (**Table 2.1** and **Suppl. Table 2.3**).

Then, we measured the expression *in-vivo* of all candidates in our peptide-encoding gene set in the chinchilla OM animal model. The PMEN1 strain PN4595-T23 was directly inoculated transbullarly into the middle ear of the chinchilla. RNA was extracted from effusions from the animal ears 48h post-transbullar inoculation and the expression was analyzed by NanoString technology that allows us to distinguish between bacterial and chinchilla RNA. The top expressing gene in the codeset was termed virulence peptide 1 (VP1, based on analysis presented below) and it is the most highly expressed Gly-Gly peptide-encoding gene in effusions, and appears highly induced relative to planktonic growth

The level of *vp1* mRNA was 1.3 times that of *epsA* (exopolysaccharide A) and 1.6 times that of *cbpA* (choline binding protein A), which are highly expressed during infections (59,60) (**Suppl. Table 2.3**). We also measured peptide expression in planktonic cells grown to mid-log in rich media, to identify genes differentially expressed *in vivo* relative to *in vitro*. The average number of RNA molecules for *vp1 in vivo* was 36.15 $\pm$ 1.69-fold higher than in planktonic culture (**Figure 2.2**)**.** We conclude that the levels of *vp1* during middle ear infection are comparable to levels of transcripts highly expressed *in vivo*, and thus VP1 may play a role in infection.

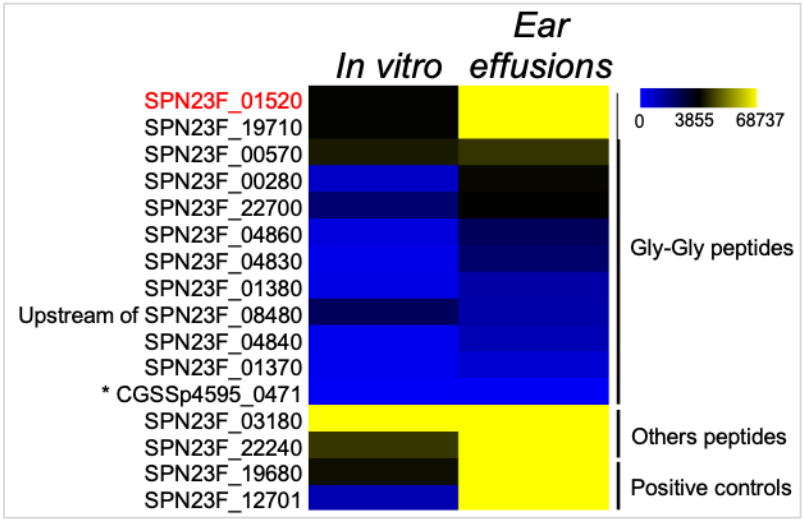



**Figure 2.2 Analysis of peptide expression during host infection**

NanoString RNA count averages for the set of predicted secreted peptides in PMEN1 strain PN4595-T23 in mid-log planktonic cultures (left column) and effusions extracted from the chinchilla middle ear 48h post-inoculation (right column). As positive controls, we utilized the quorum sensing peptide encoding gene *phrA*, and the predicted lanthiopeptides *lanA* and *lcpA* genes. RNA counts were normalized to the geometric mean of the housekeeping genes *metG* and *gyrB*, and sorted by expression level. The VP1-encoding gene is highlighted in red. Upstream refers to the relative position of the NanoString probe. ID's are represented in the model PMEN1 genome (strain ATCC 700669, GenBank FM211187), and corresponding IDs in PN4595-T23 are listed in **Suppl. Table 2.1** and **Suppl. Table 2.3**.

## 2.3.2 VP1 is a novel virulence determinant

To evaluate the virulence properties of VP1, we employed the chinchilla model of human middle ear infection. Cohorts of 10 chinchillas were injected transbularlly with either the parental wild-type strain or the isogenic *vp1* deletion mutant (Δ*vp1*). Chinchillas inoculated with Δ*vp1* displayed a significant decrease in mortality relative to animals inoculated with the wild-type strain ($p$ = 0.023) (**Figure 2.3A**). Further, relative to the wild-type, the Δ*vp1* displayed a significant reduction in dissemination to the brain ($p$ = 0.035), and although not significant, reduced dissemination to the lungs ($p$ = 0.152) (**Figure 2.3B**). These findings suggest that VP1 plays a role in pneumococcal *in vivo* survival and pathogenesis.

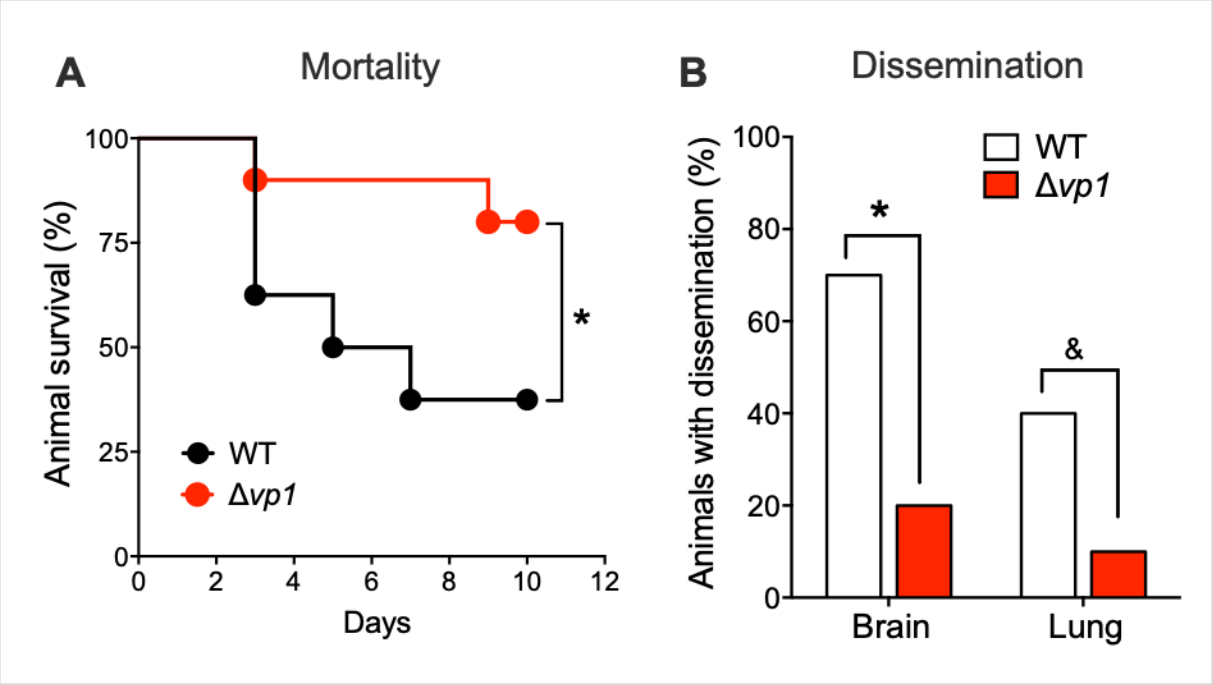



**Figure 2.3 VP1 is a virulence factor.**

Pneumococcal wild-type or Δ*vp1* strains were injected bilaterally into the tympanic bullae of chinchillas and mortality and dissemination were scored. A. Kaplan-Meier survival curves of chinchillas. B. Percent of chinchillas with bacterial dissemination to the brain or the lungs, as determined by CFU counts in these tissues at time of death. Statistical significance was calculated using paired two-tailed Student's *t*-test (* $p < 0.05$, & $p = 0.152$).

### 2.3.3 VP1 plays a role in pneumococcal biofilm development

In middle ear infections pneumococci adopt a biofilm mode of growth, thus we investigated whether *vp1* influences biofilm development (9). To this end we utilized chinchilla middle ear epithelial (CMEE) cells (61). We established that PMEN1 pneumococcal cells form a robust biofilm on CMEE substrate sixteen hours post-inoculation (**Suppl. Figure 2.1**). We compared biofilm biomass and thickness between wild-type and Δ*vp1* strains using confocal imaging (**Figure 2.4**). The Δ*vp1* displays approximately 2-fold reduction in biofilm biomass and thickness when compared to the wild-type. Furthermore, we demonstrate that the wild-type phenotype is rescued in a *vp1* complemented strain (Δ*vp1:vp1*) (**Figures 2.4A** and **2.4B**). No differences were observed between biofilms of the Δ*vp1* and Δ*vp1:vp1* strains on glass surface (**Suppl. Figure 2.2**). The difference in biofilm growth on CMEE cells is not the result of variation in growth rate, as wild-type, Δ*vp1,* and Δ*vp1:vp1* strains display the same rate of growth in the epithelial growth media used for the pneumococcal-CMEE biofilm assays (**Figure 2.4C**). This was further corroborated by growing the strains in rich liquid cultures, where no differences in growth capacity or viability were observed between wild-type the strain and the derivative strains Δ*vp1* and Δ*vp1:vp1* (**Suppl. Figure 2.3**).

Further, we determined that the *vp1* expression is about 20-times higher on biofilm growing on epithelial cells compared to planktonic growth and 4-times higher to biofilm growing on glass (**Suppl. Figure 2.4**). Finally, we did not observe a fitness advantage when the wild-type and the Δ*vp1* were co-culture (**Suppl. Figure 2.5A**) while we observed a striking fitness difference in vivo: *vp1* provides a fitness advantage solely when the bacteria was inoculated in the host (**Suppl. Figure 2.5B**).

Thus, we conclude that *vp1* contributes to the development of biofilm thickness on epithelial cells.

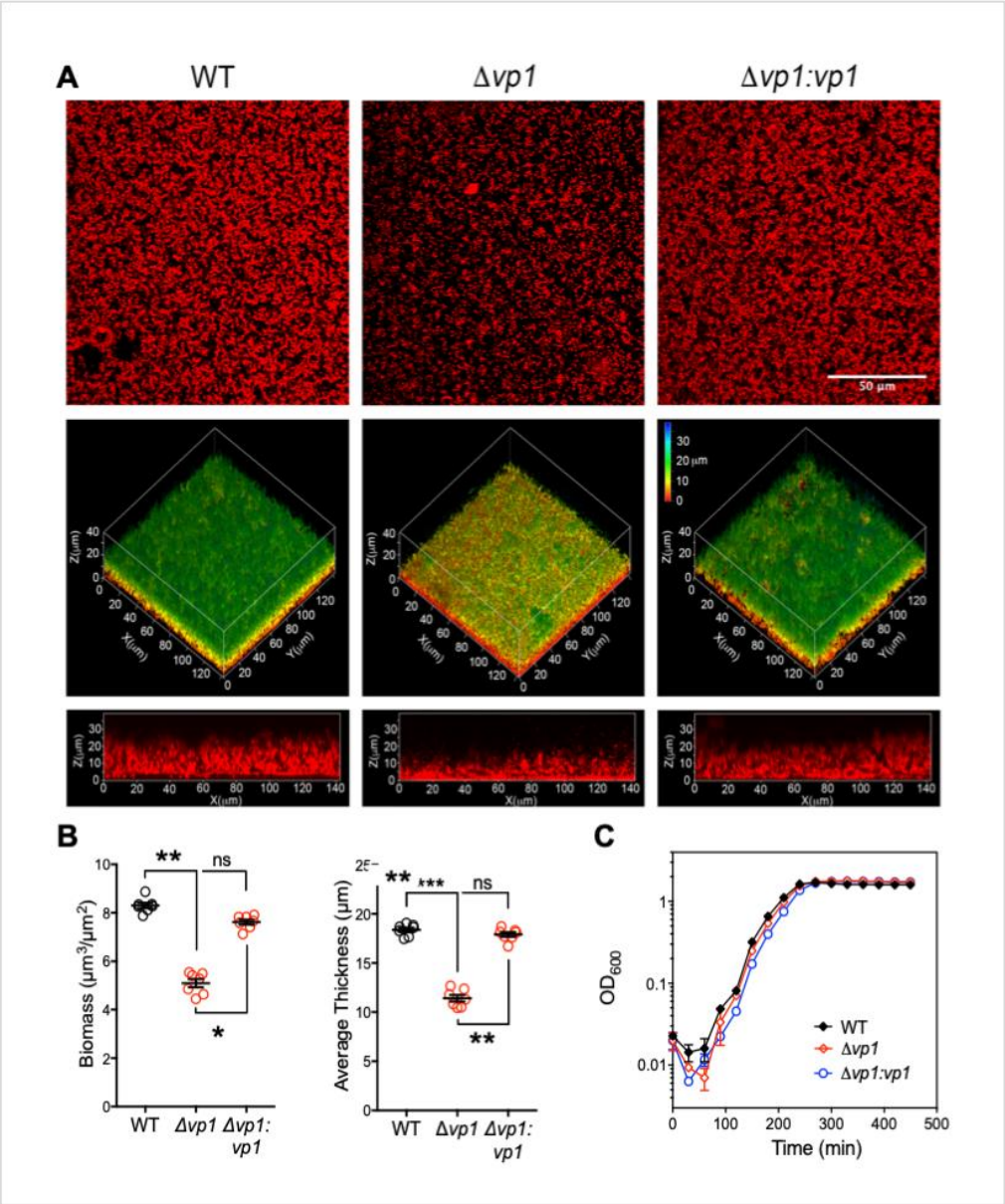



**Figure 2.4 VP1 enhances biofilm development on CMEE cells.**

**A.** Representative images assembled from confocal micrographs of biofilms of wild-type, Δ*vp1*, or Δ*vp1:vp1* grown on CMEE cells and imaged 16h post-seeding. Top panels: confocal image of the lower end of the biofilm, just above the mammalian cells. Middle panels: 3D projection pseudo-colored to highlight biofilm thickness, according to scale on the left. Lower panels: side view of a biofilm reconstruction. Axis represent distance in mm. **B.** Biofilm biomass and average thickness were computed from confocal images. Data represent the mean ± S.E.M. of at least three independent experiments. Statistical significance was calculated using 2-way ANOVA analysis with Tukey correction. *$p = 0.0039$, ** $p \leqq 0.0007$, ns = not significant. **C**. Growth curve of wild-type, Δ*vp1*, or Δ*vp1:vp1* in epithelial cell media demonstrating that differences in biofilm development are not due to differences in growth rates (n = 3). Scale bar is conserved across all figures.

### 2.3.4 Synthetic VP1 localizes to the pneumococcal surface

VP1 is part of the Gly-Gly family of peptides. Known Gly-Gly peptides are processed and secreted. Subsequently, their active form signals through kinases. Thus, we tested whether synthetic VP1 (labeled with FITC at the N-terminus) added extracellularly can bind to PMEN1 cells. As positive controls, we utilized two FITC-labeled peptides known to interact with PMEN1 cells. The first was the CSP pherotype CSP2, a Gly-Gly peptide that binds to the surface exposed histidine kinase ComD and initiates competence (7). The second was PhrA2, a cognate peptide to TprA2 (RRNPP superfamily) that is proposed to be internalized into the cytosol (37). As a negative control, we used FITC-CSP1, as this peptide does not bind to the PMEN1 ComD variant (62). The distribution of VP1 is distinct from the cytosolic localization of PhrA2 (**Figure 2.5A**). Instead, we find that VP1 distribution resembles that of CSP2, both are observed on ~95% of the cells and dispersed on the cell surface (**Figures 2.5A** and **2.5B**). Further, the VP1 signal appears internal to capsular staining and external to DNA staining (**Suppl. Figure 2.6**). These localization experiments are consistent with surface localization of VP1.

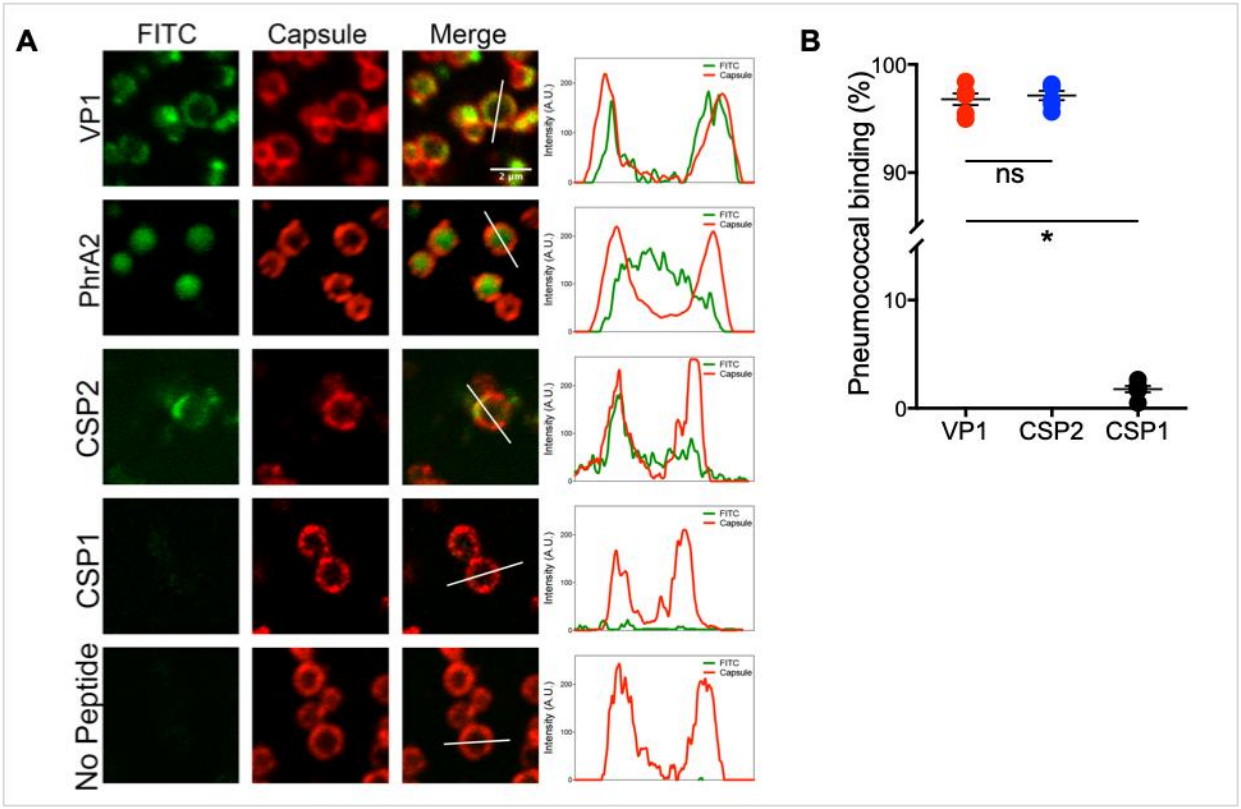
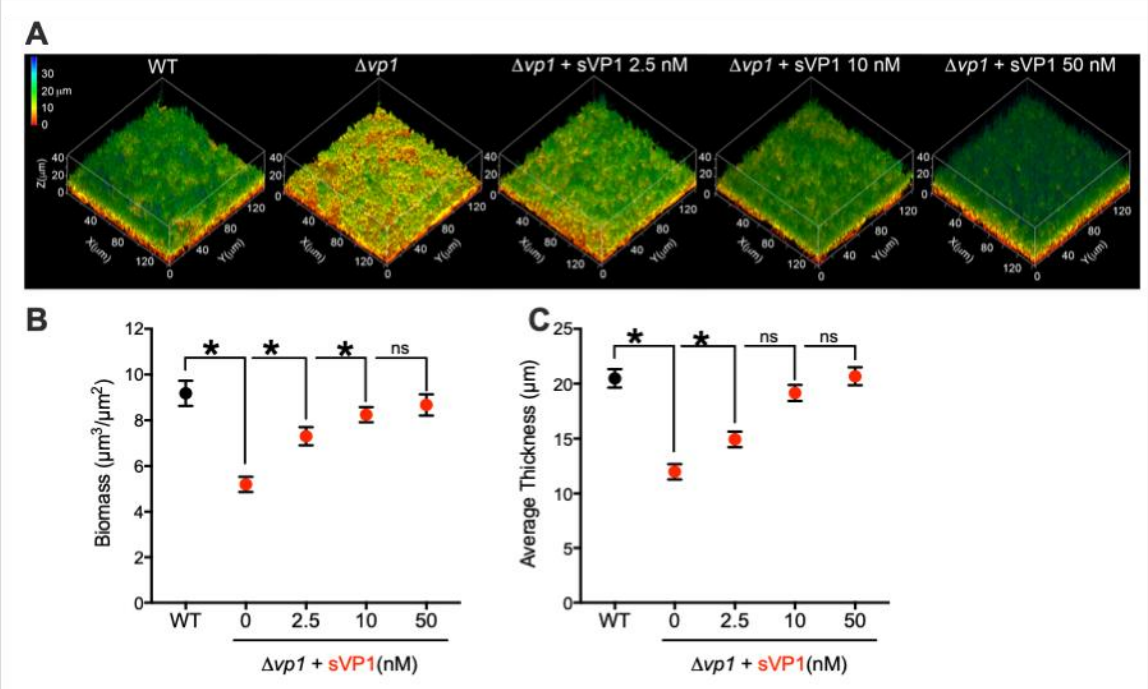



**Figure 2.5 VP1 peptide binds to the surface of pneumococcal cells.**

**A.** Confocal micrographs of early biofilms. Left Panel: wild-type cells stained with FITC-labeled VP1, PhrA2, CSP2, CSP1, or no peptide. Peptides were added to 16h biofilms for 1h prior to fixation and imaging. Middle Panel: Indirect immunofluorescence using antibodies targeting the pneumococcal capsule. Fluorescence intensity across the width of the bacterial cell (white line) was measured on representative cells and results are plotted on the right panels. **B.** Quantitation of the percentage of cells positive for FITC-VP1 and FITC-CSP2. Statistical significance was calculated using unpaired two-tailed Student's *t*-test. * $p < 0.0001$, ns = not significant.

## 2.3.5 The synthetic VP1 enhances pneumococcal biofilm

We stablished that the Δ*vp1* strain displays reduced biofilm biomass and thickness relative to the wild-type strain. We also stablished that VP1 exerts its function via binding to the pneumococcal surface to trigger a signaling cascade. These two observations allowed us to hypothesize the Δ*vp1* defect should be restored to wild-type levels when biofilm cultures of the knock-out are incubated with a synthetic VP1 peptide. To test this prediction, we grew Δ*vp1* biofilms in the presence of synthetic VP1. The synthetic peptide rescued the biofilm defect in a dose-dependent manner (**Figures 2.6A** and **2.6B**). Together, these findings suggest that VP1 is a secreted peptide, which plays a role in biofilm development and exerts its effect by binding to the surface of




producer and neighboring cells.

**Figure 2.6 Synthetic VP1 rescues the biofilm defect of the *vp1*-deficient strain.**

**A.** Representative 3D projection of biofilms on CMEE cells, pseudo-colored to highlight biofilm thickness according to the scale on the left. (i) wild-type, (ii) Δ*vp1*, (iii) Δ*vp1* + 2.5 nM sVP1, (iv) Δ*vp1* + 10 nM sVP1, (v) Δ*vp1* + 50 nM sVP1. **B.** Biomass and average thickness computed from micrographs. sVP1 stands for synthetic VP1 peptide. Data represent the mean ± S.E.M. of at

least four independent experiments. * $p = 0.0055$, * $p < 0.0001$, ns = not significant; statistical significance was calculated using 2-way ANOVA analysis with Tukey correction.

### 2.3.6 The VP1 operon is regulated by an Rgg transcriptional regulator

To gather insight on the regulation of *vp1,* we analyzed the organization of the *vp1* genomic neighborhood in strain PN4595-T23 (**Figure 2.7A**). Immediately downstream of *vp1* is a predicted protease belonging to the family of CAAX proteases and bacteriocin-processing enzymes (CPBP) and a predicted efflux pump, denoted *vpB* (*vp1* operon gene B) and *vpC* (*vp1* operon gene *C*), respectively. Another 453 bp downstream is a predicted *azlC* permease for transport of branched amino acids, denoted *vpD.* Upstream of *vp1* is a gene encoding a predicted *rgg* regulator. Computational analysis indicated that the transcription factor belongs to *S. dysgalactiae* shp pheromone Rgg2 receptor family (e-value 2.41e-61) (**Suppl. Figure 2.7**). It has been also shown that in *S. pyogenes* Rgg activity is regulated by cognates short hydrophobic peptide (SHP) (39,46,47). Computational analysis indicated the presence of a SHP-encoding short open reading frame (ORF) encoded in the complementary strand where its sequence overlaps with that of *rgg*. Initial comparative analysis shown high level of conservation with other SHP sequences for which their sequences is available (**Suppl. Figure 2.8A**). The peptide sequences display a well-defined motif organization: starting with three or four basic amino acids, followed by a distinctive hydrophobic domain interrupted by a single and conserved glutamic acid. To illustrate the conservation across SHP in different pneumococcal strains (**Supp. Table 2.4**), we generated a MEME logo (**Suppl. Figure 2.8B**).

To define the *vp1* operon structure of this region, we performed PCR analyses on cDNA. The analysis suggests that *rgg* is one transcriptional unit, and that *vp1-vpBCD* is a second transcriptional unit (**Suppl. Figure 2.9**). No product was detected for primers spanning *azlC* and downstream genes, suggesting that *azlC* demarks the 3'end of the *vp1* transcript. *In silico* analyses of the predicted promoter region for *rgg* and for the *vp1* operon revealed putative CodY binding-boxes ("AACTATCAGAATATT" and "GATTTTCTAAAATAA", respectively) (**Figure 2.7A**). This is consistent with previous work in which a deletion of *codY* led to a 2.5 fold increase in the levels of *rgg*, *vp1, vpB* and *vpC* (spd0141 to spd0145 as those experiments were performed in the D39 strain) (44). Finally, a region resembling a Rgg-box (TTGAAGGAGTGTAA) is located

upstream of *vp1*, suggesting that Rgg may regulate the *vp1* operon. This genomic organization is consistent with a role for CodY in the negative regulation of the *vp1* operon, and a regulatory link between *rgg* and *vp1*.

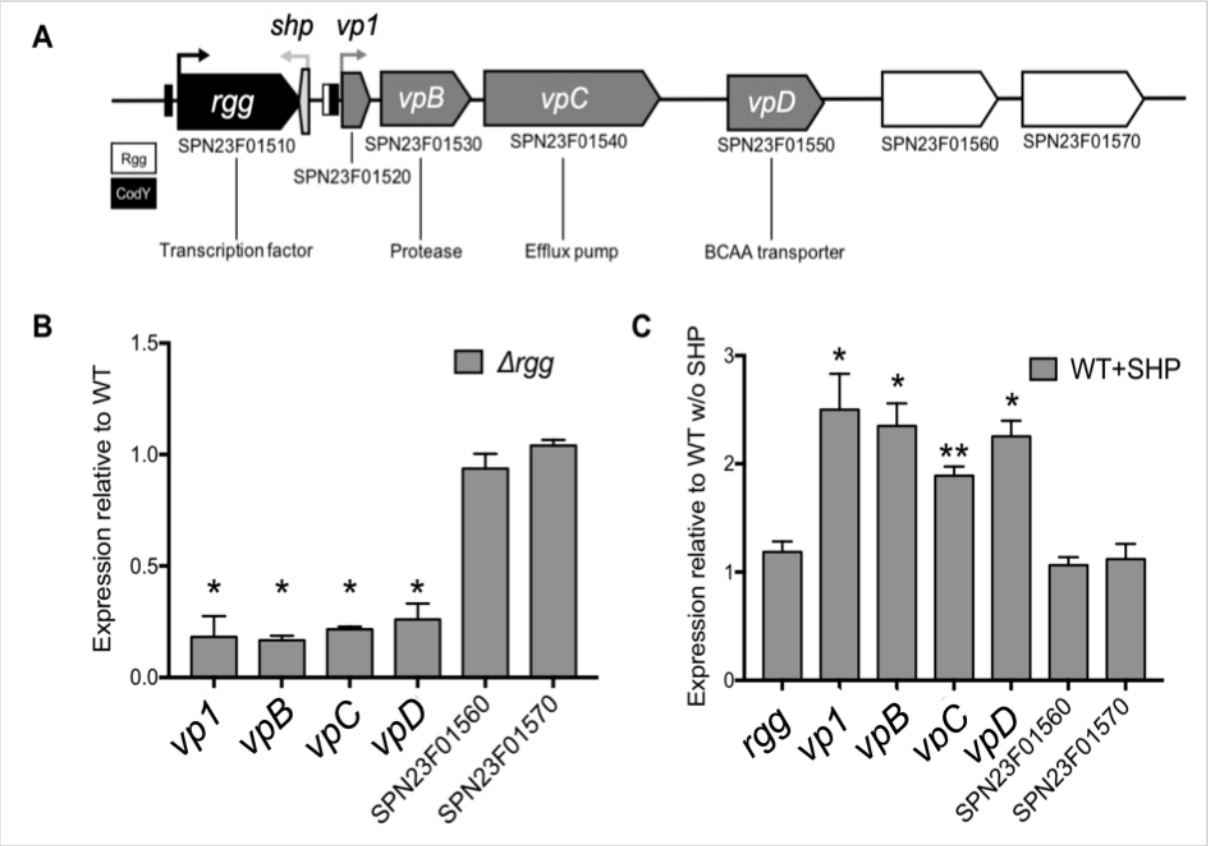



**Figure 2.7 Genomic organization and *vp1* regulation by a Rgg144/SHP**

**A.** Genomic organization of *vp1* genomic *locus*. The *locus* was adapted from a RAST annotation of the genomic region. Genes in the same color are in the same transcriptional unit, and IDs correspond to gene names in strain ATCC 700669. Thin rectangles correspond to predicted binding sites for CodY and Rgg. Predicted Rgg binding site is "TTGAAGGAGTGTAA". Putative functions are indicated below each gene. **B.** qRT-PCR analysis of transcript levels in the Δ*rgg* strain relative to wild-type strain of *vp1* and neighboring genes relative to *gapdh*. **C**. Analysis of transcript levels in wild-type and wild-type + 0.68 mM SHP synthetic peptide. * $p \leq 0.0001$ and ** $p = 0.0002$ with n = 3. Normalized to levels of *gapdh*. Statistical significance was calculated using 2-way ANOVA analysis with Bonferroni correction. The figure displays the IDs in strain ATCC 700669, the corresponding IDs in strain PN4595-T23 are: CGSSp4595_0133 (SPN23F01510), CGSSp4595_0134 (SPN23F01520), CGSSp4595_0135 (SPN23F01530), CGSSp4595_0136(SPN23F01540), CGSSp4595_0137 (SPN23F01550), CGSSp4595_0138 (SPN23F01560), and CGSSp4595_0139 (SPN23F01570).

The relative position of *rgg* and predicted *shp* suggest their gene products function as a cognate pair. In other streptococcal species, SHPs bind to their cognate Rgg activating transcription (28). In PN4595-T23, the predicted *shp* encodes a twenty-five-amino acid peptide (MKKQILTLLKIVAEIIIILPFLTNR) and is encoded in the opposite strand to and partially overlaps *rgg*. Moreover, the relative location of *vp1* and *rgg* combined with the presence of a putative *rgg*-box upstream of *vp1* are consistent with regulation of *vp1* by Rgg. From here, we called Rgg144 as later renamed by the group of Zhi et al.) (63). We hypothesized that Rgg144 (SPD_0144 locus on the strain D39) activates *vp1* expression, and that activation is SHP-dependent. To test this hypothesis we first searched for an *in vitro* condition with high levels of *vp1* expression; we tested rich media (Columbia broth) and chemically defined media supplemented with glucose (CDM-Glu) (64), at various stages of growth (**Suppl. Figure 2.10**). The highest level of *vp1* expression was observed in CDM-Glu at early-log phase. Thus, using this condition, we compared the expression levels of the genes in the *vp1* transcriptional unit between wild-type and a Δ*rgg144* strain. All four genes expressed five-fold lower in the deletion mutant relative to the wild-type, suggesting that Rgg144 is a positive regulator of the *vp1* operon (**Figure 2.7B**). Given that SHPs induce Rgg activity (46), we tested the effect of SHP on gene expression. Addition of exogenous full-length SHP to the wild-type strains leads to over two-fold induction of all the genes in the *vp1* operon (**Figure 2.7C)**. Further, our preliminary exploratory RNAseq analysis determined the Rgg144 regulates the *vp1* operon exclusively (**Suppl. Figure 2.11**). Our initial experiments were performed using a 25-amino acids SHP peptide. Then we tested a shorter SHP144 peptide, corresponding to the last 12 amino acids of the C-terminal, which have been reported as the form with the highest activity in *S. pneumoniae* (63) and in *S. pyogenes* (65). As expected, the shorter SHP-C12 form induced an acute, 10-times higher *vp1* expression when added exogenously (**Suppl. Figure 2.12A**). The peptide has little to no-effect effect over *rgg* expression, similar to the exogenously added sVP1 **(Suppl. Figure 2.12B**). These experiments show that a secreted SHP peptide is required for Rgg144-mediated *vp1* expression. Together, these data suggest that a cognate Rgg144/SHP pair positively regulates *vp1* expression.

### 2.3.7 VP1 alleles distribution in streptococcal species

The intra and inter species influence of *vp1* depends on the distribution and variation of *vp1* alleles and its receptor. Epidemiologic studies reveal that infection with multiple pneumococcal strains occur in an estimated 20-40% of infections (66–68). Furthermore, multiple streptococcal species, such as *S. mitis* and *S. oralis,* commonly colonize the human nasopharynx and there are many examples of DNA exchange between pneumococcus and these species (15,69–73). We have analyzed the distribution of *vp1* across pneumococci and related species: *vp1* is widely distributed in pneumococcal strains (**Figure 2.8**). Using thirty-five pneumococcal genomes from nineteen serotypes and thirty-six MLSTs, we determined that VP1 is encoded in thirty-three strains (**Suppl. Table 2.1**). The exceptions are strain CGSSp14, where the region encoding VP1 appears to be replaced with a lantibiotic biosynthetic locus, and strain CDC10870 where it appears to be replaced with a domain from a different pneumococcal peptide (KGI31098.1). The thirty-three sequences can be grouped into three allelic groups (**Suppl. Figure 2.13**). In all instances the alleles are syntenic, with *rgg* upstream and *vpBC* downstream. Some strains encode two adjacent copies of the protease (*vpB*). This distribution suggests that the function of VP1 is widely conserved. Further, signaling may be limited to subsets of strains based on specific interactions between each variant and its cognate receptor.

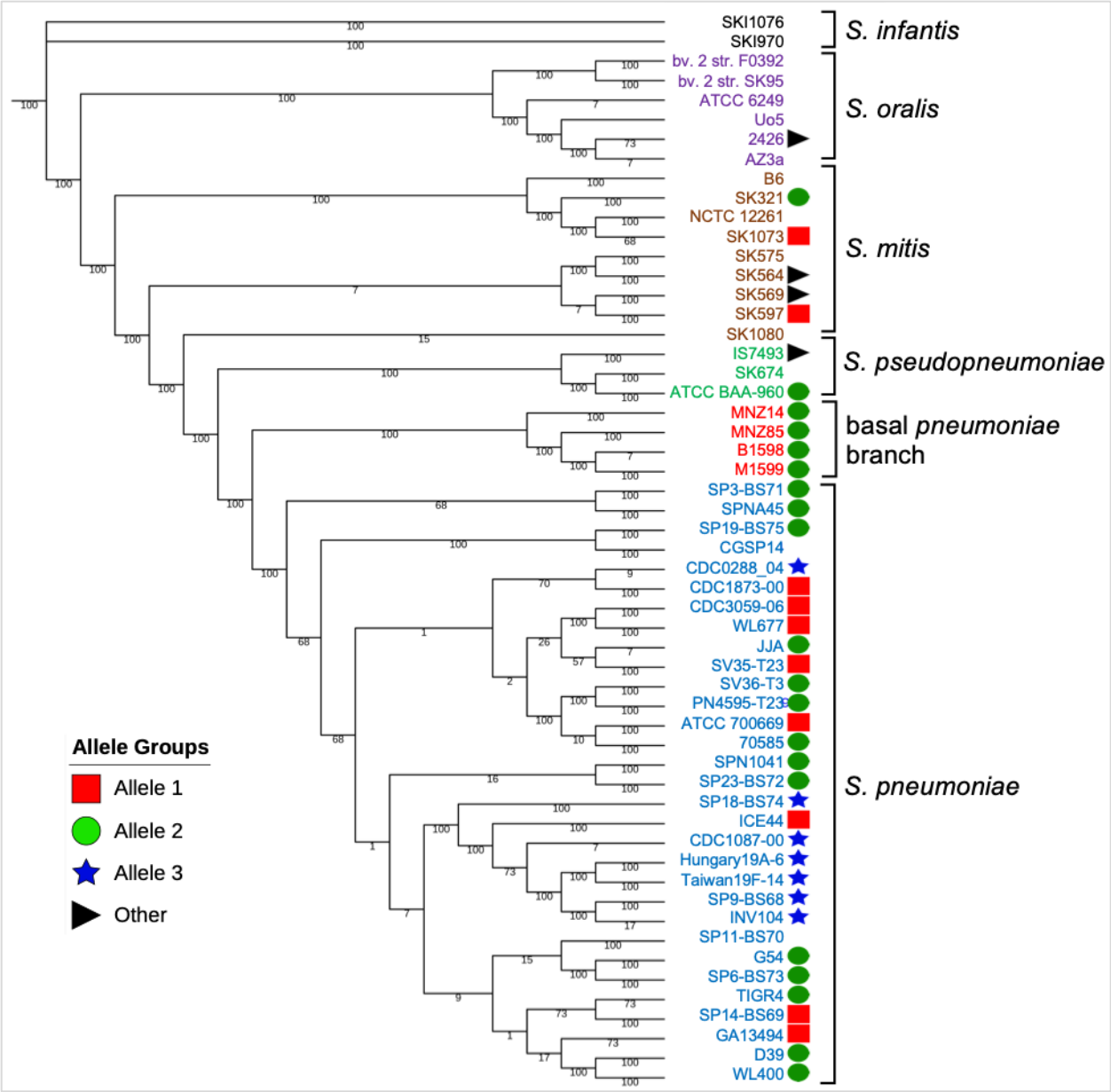



**Figure 2.8 Distribution of *vp1* alleles in *Streptococcus pneumoniae* and related species**

Specie and strain conservation of *vp1*. Maximum likelihood phylogenetic tree of streptococcal genomes was generated from the core genome with bootstrap values displayed on the branches. Green circles, red squares, and blue stars represent distinct allelic types; black triangles represent multiple allelic types identified outside of pneumococcal strains. Species are colored coded as follows: gray, *S. infantis;* beige, *S. oralis; green, S. mitis;* pink, *S. pseudopneumoniae;* light blue, distinct phyletic group of *S. pneumoniae*; blue, *S. pneumoniae.*

The distribution of *vp1* in related streptococci is highly variable (**Figure 2.8** and **Suppl. Figure 2.14**). Two out of three *S. pseudopneumoniae* genomes encodes a *vp1* allele, and one differs from those captured in the pneumococcus genomes. Approximately half of the *S. mitis* genomes encode *vp1* alleles; in this species, there are multiple alleles including some observed in pneumococcus. This gene appears less frequently in the *S. oralis* clade, where only one *S. oralis* (*S. tigurinus* subspecies) encodes a *vp1* allele. The absence of a clear correlation between allelic type and phylogeny of the species is consistent with gene transfer of the *vp1* region. The wide distribution of this gene, suggests *vp1* is a key component of pneumococcal virulence and dissemination.

## 2.4 Discussion

Our *in-silico* screen suggests that pneumococci utilize multiple cell-cell communication peptides during infection. Pneumococcal CSP and BIP control competence and bacteriocidal activities via cell-cell interactions and their role in multi-strain communities is well documented (4–7). In this study, we present peptides predicted to be part of the pneumococcal secretome; all encode a Gly-Gly secretion motif consistent with export and cell-cell interactions (**Suppl. Table 2.1**). We characterize VP1 as a signaling molecule, highly expressed *in vivo*, and with a role in biofilm formation and pathogenesis.

Multiple lines of evidence support the hypothesis that *vp1* and the genes in the *vp1* operon sense amino acid levels in a cellular community. First, *vp1* and downstream genes are negatively associated with levels of *glnA* and *glnP,* which are critical players in glutamine/glutamate metabolism (45). Second, genes in the *vp1* operon are negatively regulated by the nutritional regulator CodY, and CodY-mediated inhibition is enhanced in the presence of branched-amino acids (44). Third, *vpD* belongs to a protein family predicted to function in the transport of branched-chain amino acids (AzlC, score: $1e^{-31}$). It will be fascinating when future work reveals the extracellular signal(s) that control the high *vp1* levels we have observed *in vivo*.

Rgg144 (SPN23F01510), the positive regulator of the *vp1-vpBCD* operon, has multiple molecules controlling its transcription and activity. Analogous to the *vp1* operon, *rgg* is also negatively regulated by CodY and glutamine/glutamate metabolism (44,45). In this manner, it appears that Rgg144 amplifies the regulation of *vp1* operon. Presumably, once inhibition by these nutritional regulators is released, transcription of *rgg* is activated and serves to further enhance transcription of the *vp1* operon. While it is likely that *rgg* receives transcriptional input from other sources, these remain to be identified. Regarding the activity of Rgg144, we discovered and defined the function of and SHP-encoding ORF immediately downstream to *rgg*, in a position conserved across pneumococcal strains, and that the product of the *shp* modulates the transcriptional activity of Rgg144. We also predicted the promoter region for *shp* is on the opposite strand and overlapping with the *vp1* promoter. It is further revealing that the *vp1* promoter also encodes a putative CodY-box proximal to *vp1*. In this manner, SHP levels may also be under the control of this nutritional regulator and/or amino acid levels since CodY box is palindromic (74). Rgg144/SHP quorum-sensing systems have been extensively characterized in Group A streptococci (46,75,76), and a Rgg has been characterized in pneumococcus and implicated in response to oxidative stress (35) (**Figure 2.9**). We found that Rgg144 activity is enhanced by the peptide SHP encoded downstream *rgg*, and it is sufficient to induce *vp1* operon expression. The genome of the *S. pneumoniae PN*4595-T23 strain encodes four additional *rgg* members (SPN23F09790, SPN23F10360, SPN23F17070, SPN23F21530, and they are usually annotated as *mutR*). Our data shows that Rgg144/SHP regulated the *vp1* operon exclusively. Whether SHP binds and influence the activity of those other Rgg members, or whether the activity may by mediated via Rgg144/SHP144 complex needs to be determined (**Figure 2.9**).

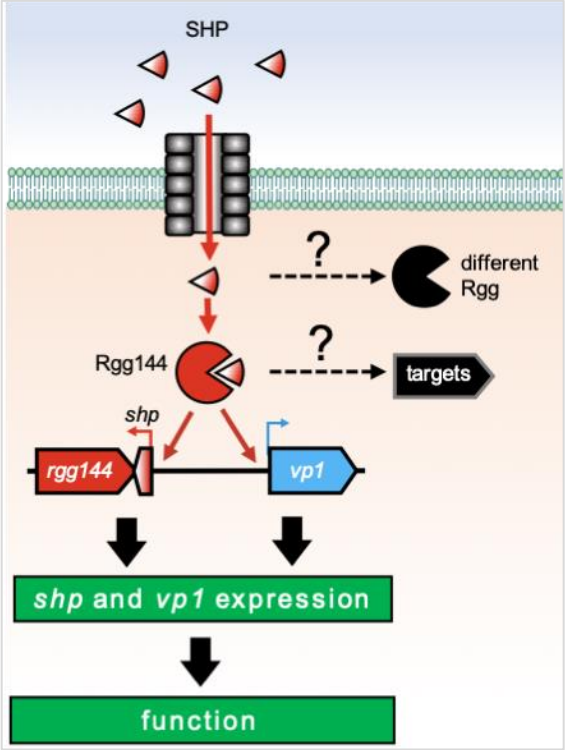



**Figure 2.9 Model for interactions between Rgg144/SHP and its targets**

SHP enhances the Rgg144 activity which in turn induce both *shp* and *vp1* expression. Alternatively, Either SHP may interact with other Rgg members and influence their activity or the par Rgg144/SHP may function over different target depending in the condition assessed.

Further, based on the GAS orthologues, we predict this SHP is processed and secreted from producer cells, accumulates in the extracellular environment in a density-dependent manner, and is imported by the AmiABCDE transporter (28). Once in the cytosol, SHP presumably binds Rgg144, stimulates this positive regulator and induces transcription of the Rgg144 regulon (**Figure 2.10**).

The organization of the *vp1* operon is conserved across strains and species, further supporting the functional association between these co-regulated genes. Encoded immediately downstream are a transmembrane protease (SPN23F01530) and a transporter (SPN23F01540) (**Figure 2.7A**). The protease is part of the Abi protease family (Pfam: Abi, PF02517, score: $7.3e^{-17}$). The presence of a protease is common to many Gly-Gly peptides (6); while their function is unknown, these proteins are presumed to play a role in self-immunity against bacteriocins. This functional annotation is consistent with a role for this protease in VP1 regulation via control of degradation. The *vpC* encodes for a transporter from the major facilitator superfamily 1 (PF07690; score: $3.6e^{-11}$). The functional link between this transporter and VP1 is still unclear. We

predict that it is not part of the VP1 import or export machinery. The Gly-Gly motif in VP1 is consistent with export via an ABC-transporter. Similarly, the peptide family combined with the surface localization of FITC-VP1 suggests VP1 is not internalized but instead binds to the cell surface. The MSF1 transporter (77) may be related to controlling molecules upstream or downstream of VP1. The role of the molecules in the VP1 locus and the identity of the VP1 receptor are currently under investigation.

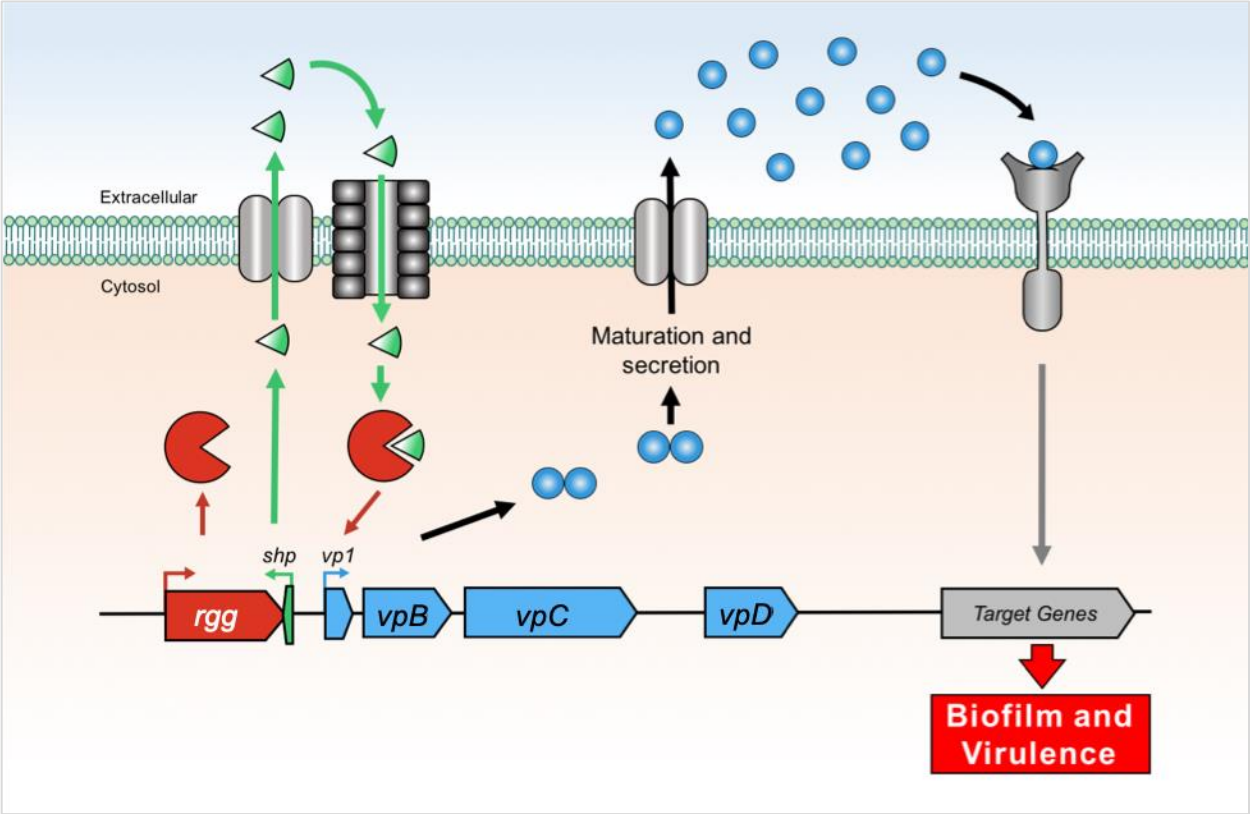



**Figure 2.10 Model for VP1 regulation and for cell-cell communication in *S. pneumoniae***

Regulation: Rgg144 is a positive regulator of the *vp1-vpBCD* operon. Rgg144 activity is induced upon binding to its cognate SHP. We predict that SHP is a quorum-sensing molecule, activated during export and functional upon internalization. Red: *rgg*, green: *shp*. Signaling: VP1 is processed and secreted into the extracellular milieu, from where it binds to the surface of producing and neighboring cells. VP1 influences biofilm and virulence. We propose VP1 binds to a surface molecule triggering a signaling cascade. Blue: genes in the *vp1* operon, and blue circle: VP1 peptide.

Comparative genomics of pneumococcal strains revealed three distinct groups of VP1 alleles, as well as additional alleles in related streptococci. The non-clustered distribution of alleles across the species tree is consistent with extensive *vp1* gene transfer across both strains and species. The presence of variants raises the question: how do these alleles interact? For signaling and bacteriocidal molecules, specificity has been documented between the signal and the receptor and between the bacteriocins and immunity system, respectively (6,78,79). Further, specificity has also been observed for bi-functional molecules that play a role in both cooperation and competition (80). Thus, it is likely that the allelic variations in *vp1* are important in multi-strain infections by increasing cooperative behaviors across strains that share the same allele, and/or increasing competitive behaviors across strains with diverse alleles.

In conclusion, in this section, we have characterized VP1 as a signaling molecule. Our findings invite exciting questions regarding VP1 regulation and function. What are the *in vivo* conditions that activate *vp1* expression? What are the downstream components required for VP1 signaling? What effector molecules are downstream of this signaling? Is VP1 a good target for anti-infective therapies? Future work on VP1 will provide answers to these questions, and in doing so should advance our understanding of pneumococcal biology in the context of human infections.

## 2.5 Materials and Methods

***In silico* screen for predicted secreted peptides.** To generate a set of predicted secreted peptides we utilized MEME to generate a position-dependent probability matrix from a set of secreted peptides and MAST to search for the MEME motif in a database of streptococcal genomes (81). The MAST database was comprised of a set of 134,662 sequences in sixty genomes: thirty-nine *S. pneumoniae*, three *S. pseudopneumoniae*, nine *S. mitis*, seven *S. oralis* strains, and two *S. infantis* strains. The thirty-nine pneumococcal genomes capture nineteen serotypes and thirty-six MLST types, and include isolates from North America, South America, Africa, Europe, and Asia (**Suppl. Table 2.1**). This set was assembled in multiple steps. First, we started from a large-scale pneumococcal pangenome study and selected at least one genome from each branch of a maximum likelihood tree build from the core genome (Donati et al., 2010). To this set of twenty-six genomes, we added two genomes (WL677 and WL400) isolated from PCV-7 immunized children to capture samples of vaccine escape strains emerging in the post-vaccine era (84). Further, recent studies have revealed that a subset of non-encapsulated isolates form a distinct phyletic group basal to other pneumococcal strains (73,85). Thus we added isolates from this group, ending up with thirty-two pneumococcal genomes (Antic et al, unpublished) (86) (light blue, **Suppl. Table 2.1**). Finally, for the MAST, we added seven genomes from isolates from eye infections and non-encapsulated strains that are distributed throughout the phylogenetic tree (Antic et al, under review). Thus, the final set of pneumococcal strains for our MAST search corresponds to thirty-nine genomes. Previous phylogenetic work on pneumococcus and viridans reveals that pneumococcus, *S. pseudopneumoniae*, and *S. mitis* form a large lineage, which is well-separated from *S. oralis* and *S. infantis* (87). Using this phylogenetic work as a guide we selected twenty-one strains distributed with these four non-pneumococcal species.

Our first MEME input consisted of 167 homologues of CSP and BIP. This set was obtained using NCBI blastP (e-value cutoff of $1e^{-4}$) to search the non-redundant protein database using query sequences of pneumococcal CSP and pneumococcal BIP. The MEME motif was normalized to the amino acid frequency of the coding sequences in

the database of sixty genomes. Next, we employed MAST to search for the MEME motif in the database of streptococcal genomes. Importantly, this method is limited to annotated coding sequences, such that peptides that have not been predicted by annotation programs will be overlooked in this analysis. The output of this MAST search consisted of 192 coding sequences. This expanded set of sequences was used as input to generate a second MEME motif, and subsequently search the same database with MAST (**Suppl**. **Table 2.2**). The reiterative use of MEME and MAST converged on a list of 716 proteins, of which 696 encode the motif at the N-terminus and 98% are coding sequences less than 130 residues in length. The strong enrichment of the motif at the N-terminus of short coding sequences is consistent with the hypothesis that the identified motif function in the secretion of small peptides.

The next step was to organize the 696 coding sequences into clusters based on sequence similarity and conservation of the genomic locus. To this end, we generated a maximum parsimony tree from this predicted secretome, using iQ-TREE and the cpREV amino acid exchange model (88). The branches were manually curated using RAST to visualize the genomic neighborhoods and to ensure that alleles in the same groups shared the same genomic location (89). One cluster was curated out of the set based on its annotation as a v-type ATP synthase subunit G. This analysis revealed 25 clusters (**Table 1**).

**Bacterial strains and culture conditions.** Wild-type *S. pneumoniae* strain PN4595-T23 (GenBank ABXO01), graciously provided by Drs. Alexander Tomasz and Herminia de Lencastre, was used as a PMEN1 representative (90). This isolate was recovered in 1996 from the nasopharynx of a child in Portugal, and corresponds to a low passage number. We avoided strain ATCC 700669, as we have evidence that our aliquot of this strain has incurred mutations during serial passage (unpublished). Strain PN4595-T23 was modified to generate a *vp1* deletion mutant (Δ*vp1*), the *vp1* complemented strain (Δ*vp1:vp1*), and a *rgg* deletion mutant (Δ*rgg*) (**Suppl. Table 2.4**). For growth on solid media, strains were streaked onto Trypticase Soy Agar II plates containing 5% sheep blood (BD BBL, New Jersey, USA). For growth in liquid culture, colonies from a frozen stock were grown overnight on TSA plates and inoculated into

Columbia broth (Remel Microbiology Products, Thermo Fisher Scientific, USA). The media was supplemented with antibiotics as needed. Cultures were incubated at 37°C and 5% CO2 without shaking. Experiments in chemically defined media (CDM) were performed utilizing previously published recipe (43), and glucose was used at a final concentration of 55mM. Growth in CDM was initiated by growing a pre-culture in Columbia broth for 2-3 hours and back dilution to $OD_{600}$ 0.1 to initiate a culture.

**Growth curves and CFU counts**. Bacteria were inoculated from overnight plates into 30 ml of liquid cultures, and samples were taken at constant intervals to evaluate bacterial growth. Liquid cultures were grown statically and monitored by optical density at 600 nm ($OD_{600}$) using Nanodrop 2000c spectrophotometer (Thermo Scientific). Bacterial counts were assessed by streaking dilution on TSA or TSA-antibiotic plates.

**Strain construction**. The Δ*vp1* and Δ*rgg* strains were generated by site-directed homologous recombination where the target region was replaced with the spectinomycin-resistance gene (*aadR*) (71,91). The spectinomycin-resistance gene was inserted on the complementary strand, to decrease the likelihood of polar effects. The Δ*vp1*:*vp1* complemented strain was generated by inserting the *vp1* gene fused to the kanamycin resistance cassette into the spectinomycin cassette of the Δ*vp1* strain. Briefly, approximately 2kb of flanking region upstream and downstream of the deletion target were amplified from the parental strain by PCR using Q5 2x Master Mix (New England Biolabs, USA) generating flanking regions, and the spectinomycin resistant gene was amplified from the plasmid pR412 (kindly provided by Dr. Donald Morrison). Assembly of the transforming cassette was achieved by either sticky-end ligation of restriction enzyme-cut PCR products or by Gibson Assembly using NEBuilder HiFi DNA Assembly Cloning Kit. The resulting constructs were transformed into PN4595-T23 and confirmed using PCR and DNA sequencing. Primers used to generate the constructs are listed in **Suppl. Table 2.6.**

**Bacterial transformations.** For all bacterial transformations, approximately 1µg of transforming DNA was added to the growing culture of the target strain at $OD_{600}$ of 0.05, supplemented with 125 µg $ml^{-1}$ of CSP2 (sequence: EMRISRIILDFLFLRKK;

purchased from GenScript, NJ, USA), and incubated at 37°C. After 4 hours, the treated cultures were plated on Columbia agar containing the appropriate concentration of antibiotic for selection, spectinomycin, 100 µg $ml^{-1}$, kanamycin 150 µg $ml^{-1}$). Resistant colonies were cultured in Columbia broth, the region of interest was amplified by PCR and the amplimer was submitted for Sanger sequencing (Genewiz, Inc., USA) to verify the sequence of the mutants. The strains generated in this study are listed in **Suppl. Table 2.5**

**Treatment with synthetic peptides.** Bacterial biofilms were treated with synthetic peptide labeled at the N-terminus with FITC and custom ordered from GenScript, (NJ, USA) at 90% purity. The peptides are as follows: VP1 (FGTPWSITNFWKKNFNDRPDFDSDRRRY), CSP1 (EMRLSKFFRDFILQRKK), CSP2 (EMRISRIILDFLFLRKK), and PhrA2 (VDLGLAD). Each peptide was independently added during biofilm seeding, and biofilms were incubated at 37°C, 5% $CO_2$ for 16 hours. For experiments where different concentration of peptides were compared in parallel, the original culture was seeded onto separate CMEE cultures, and each one was treated with the relevant peptide concentration, in addition to a no-peptide control. Using a single parent culture ensured minimal variation across comparing treatments. Planktonic cultures were treated with synthetic SHP (MKKQILTLLKIVAEIIIILPFLTNR) custom ordered from GenScript, (NJ, USA) at 90% purity for 1 hour.

**β-galactosidase activity assay and peptide treatment.** Liquid bacterial cultures were grown in Columbia broth to and $OD_{600}$ of 0.1. Then bacteria were pelleted and resuspended in CDM supplemented with a 55 mM of glucose for 1 h after in. Then, cells were pelleted and resuspended in Z-buffer (5 mM $MgCl_2$, 250 mM β-mercaptoethanol, 50 mM KCl, 300 mM $Na_2HPO_4 \cdot 7H_2O$, 200 mM $NaH_2PO_4 \cdot H_2O$) and cell density was recorded ad $OD_{600}$. Bacteria were lysed by adding 0.1% (v/v) of Triton X-100 for 10 min. Then, the samples were incubated in Z-buffer containing 4 mg/ml O-nitrophenyl-β-D-galactoside at 30°C for at least 30 min. Finally, reaction is stopped by adding 1M of $Na_2CO_3$. Color production was recorded at $OD_{420}$ and $OD_{550}$. Activity is expressed in Miller units (MU) according the following formula:

$$\mathrm{MU} = \frac{[\,1000 \times \mathrm{OD420}\,]}{[\mathrm{time} * \mathrm{volume} * \mathrm{OD600}]} \quad (92).$$

**Studies in the chinchilla OM model.** Young adult chinchillas (*Chinchilla laniger*) weighing 400–600 g were obtained from R and R Chinchilla Inc., Ohio. All chinchilla experiments were conducted with the approval of the Allegheny Singer Research Institute (ASRI) Institutional Animal Care and Use Committee (IACUC) at the Allegheny Hospital Animal Facility. Cohorts of ten chinchillas were used for each strain, and maintained in BSL2 facilities. Animals were anesthetized on day 0 by subcutaneous injection of 0.1 ml of a sedative solution (ketamine hydrochloride 100 mg $ml^{-1}$, xylazine hydrochloride 30 mg $ml^{-1}$ and acepromazine 5 mg $ml^{-1}$). For inoculation, 100 µl of pneumococcal suspension containing 100 CFUs was injected bilaterally into the tympanic bullae. Animals were maintained in observation for up to 10-days post-inoculation. During these ten days animals with severe acute infection perished and animals with severe signs of pain and illness were euthanized by administering an intra-cardiac injection of 1 ml potassium chloride after regular sedation. Animals showing prolonged signs of discomfort were administered with pain relief (Rimadyl, 0.1 ml of 50 mg $ml^{-1}$). We evaluated mortality, time to death, and spread of bacteria to the brain and the lungs. Tissue dissemination was tested by plating homogenized tissue on TSA plates with 5% sheep blood to establish pneumococcal presence. Handling, euthanasia, necropsy, and sample collection was performed in accordance with the IACUC protocols.

**Quantification of pneumococcal gene expression during infection.** For RNA extraction from *in vivo* experiments, chinchillas were euthanized 48h post-inoculation of PN4595-T23. Effusions were siphoned out from the middle ear though a small opening generated in the bulla, and flash frozen in liquid nitrogen to preserve bacterial RNA. Lysis was performed by resuspending the cells in an enzyme cocktail (2 mg $ml^{-1}$ proteinase K, 10 mg $ml^{-1}$ lysozyme and 20 µg $ml^{-1}$ mutanolysin), and bead beating with glass beads, acid-washed 425-600 µm (Sigma) and 0.5mm Disruption Beads made by Zirconia/Silica in FastPrep®-24 Instrument (MP Biomedicals, USA). These lysates were frozen for NanoString analyses. RNA extraction from *in vitro* mid-log planktonic cultures were performed as previously described (93). The RNA concentration was measured by NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, USA) and its integrity

was confirmed on gel electrophoresis, final concentrations were approximately 200 ng/µl for *in vivo* samples and 60 ng/µl for *in vitro* samples. RNA was hybridized onto the nCounter chip according to the manufacturer recommendations. RNA counts were normalized to the geometric mean of the housekeeping genes *gyrB* and *metG* (43,94) using manufacturer's software, nSolver. The *in vitro* and *in vivo* levels were compared using Student's *t*-test in the GraphPad Prism 6 tool and the heat map was generated using nSolver.

**RNA-Seq analysis.** RNA-Seq experiments were performed at the Core Facility at the Children's Hospital in Pittsburgh using Ribozero to eliminate 16S rRNA. The reads were analyzed using Bowtie to map reads on the ATCC700669 chromosome (95), followed by SamsTools to sort and index the resulting reads (96). Finally, the alignments were visualized with IGV (Integrated Genomic Viewer (97,98). All tools are available on the Galaxy Cluster Servers (99).

**RNA purification, RT-PCR and qRT-PCR.** Bacterial samples were collected at different ODs, and RNALater (Thermofisher) was used to ensure RNA preservation and quality. Next, bacterial pellets were lysed with 1x enzyme mix (lysozyme, proteinase K, and mutanolysin) in TE buffer (10 mM Tris·Cl, 1 mM EDTA, pH 8.0) for at least 10 min. Finally, total RNA was isolated using the RNeasy kit following manufacturer instructions. Contaminant DNA was removed by incubating total RNA samples with DNase (2U/µl) 37°C for at least 15 min. Any remaining DNA contamination was checked by PCR of *gapdh* (no visible band should be observable in RNA only samples). RT reaction was performed using 1 µg of total RNA using SuperscriptVILO kit for 1 h. Products were amplified using either Q5 polymerase or OneTag polymerase (New England Biolabs). For quantitative RT–PCR analysis, 5 ng of total cDNA was subjected to real-time PCR using PowerUp SYBR Green Master Mix in the ABI 7300 Real Time PCR system (Applied Biosystems) according to the manufacturer's instructions. All qRT-PCR amplification was normalized to pneumococcal *gapdh* and expressed as fold change with respect to wild-type strain and untreated bacterial cells. PCR and qPCR primers are listed in **Suppl. Table 2.6**. Primers were obtained from IDT.

**Mammalian cells.** 3T3 cells were generously obtained from Dr. Jonathan W. Jarvik (Carnegie Mellon University). Cells were propagated in DMEM media (Corning), supplemented with 5% of FBS and Glutamine 20 mM (Corning). 3T3 cells were maintained at 37º C and supplemented of 5% carbon dioxide at pH 7. 3T3-conditioned media was filtered using 0.45 µm pore size filter and store at 4ºC. Cells were propagated as previously described (61). Briefly, cells were grown on Seeding Media (DMEM/High Glucose, Ham's F12, HEPES 25 mM, Glutamine 2 mM, Hydrocortisone 0.4 µg $ml^{-1}$, Isoproterenol 2 µg $ml^{-1}$). Sixteen hours later, the media was replaced with 3T3-conditioned media containing EGF 10 ng $ml^{-1}$ (Sigma) and supplemented with penicillin/streptomycin. For biofilm assays, fresh CMEE media was utilized. Bacterial cells were cultured in Nunc Lab-Tek 4-chamber slides or glass-bottomed dishes (MatTek Corporation) coated with an organic polymeric matrix (100 µg $ml^{-1}$ BSA (Sigma), Collagen type 1 rat tail 20 µg $ml^{-1}$ (Millipore), 10 µg $ml^{-1}$ fibronectin rat tail (Biomedical Technologies), in phosphate-buffered saline (PBS) to facilitate CMEE cell adhesion. All assays involving co-cultivation with bacteria were carried in absence on antibiotics.

**Biofilm formation assays.** Starter cultures of pneumococcal strains were inoculated from frozen glycerol stocks into Columbia broth, serially diluted in the same media to an $OD_{600}$ of 0.010-0.015 to start final cultures. When cultures reached an $OD_{600}$ of 0.025, each bacterial strain was seeded in tissue culture plates containing attached, confluent CMEE cells or matek plates (glass surface). To promote biofilm growth, the plates were incubated aerobically in humidified conditions at 37º C and 5% carbon dioxide for 16 h. Then, the supernatants were carefully aspirated, and biofilms were washed with PBS to remove non-adherent and or weakly adherent bacteria. Subsequently, samples were fixed with 4% PFA (Electron Microscopy Sciences). The fixing solution was removed; samples were washed twice with PBS and prepared for confocal microscopy.

**Confocal microscopy.** Biofilms were prepared as mentioned above. Briefly, biofilms were fixed and stained with SYTO59 dye for 30 min. Biofilms were washed and kept in PBS buffer before imaging. Confocal microscopy was performed on the stage of Carl Zeiss LSM-510 META DuoScan or MP ConfoCor 3 confocal microscopes, using

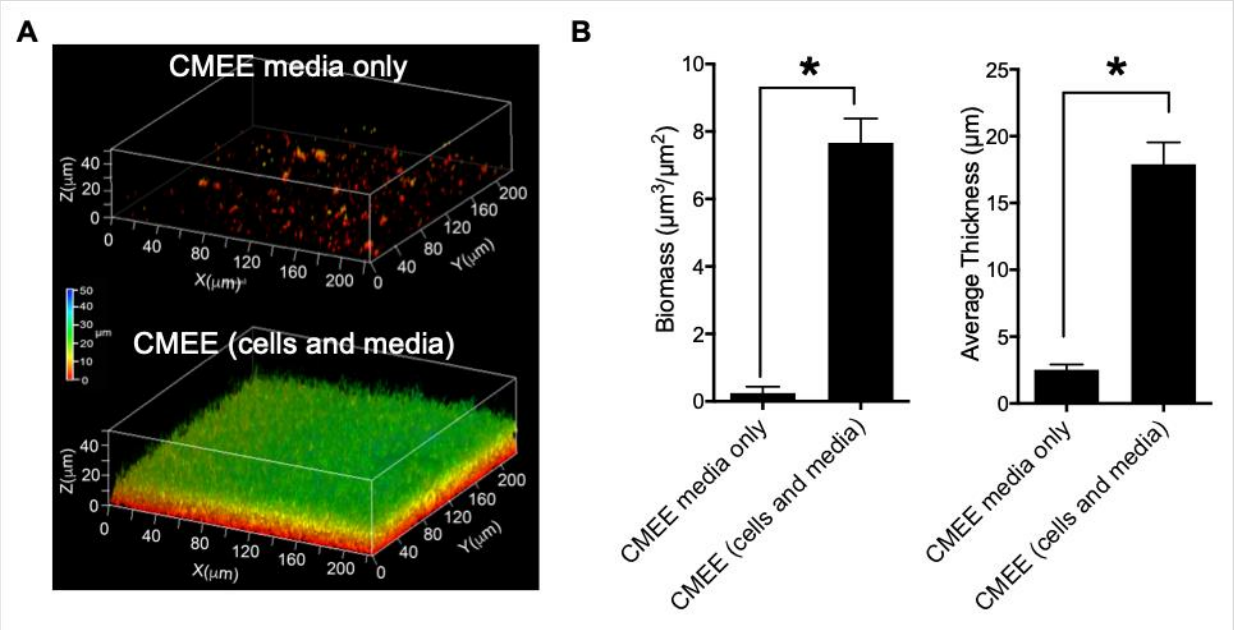
561 nm laser line for SYTO59 dye. Stacks were captured every 0.5 or 1 µm and reconstructed in Carl Zeiss Black Edition and Fiji (ImageJ. 2.0.0rc48). Laser intensity and gain were kept the same for all images. The fluorescence FITC-VP1 peptide was incubated with 16h-preformed biofilms of the wild-type strain PN4595-T23. Then biofilms were fixed and stained with rabbit anti-pneumococcal capsule antibody (AbD Serotec) and detected with goat anti-rabbit IgG (Alexa Fluor 568). Confocal microscopy was performed on the stage of Carl Zeiss LSM-880 META FCS, using 488 nm laser line for the FITC fluorescence tag and 561 nm laser line for the fluorescence Alexa tag. RGB scanning profiling was employed to validate the peptides localization relative to the capsule marker using Fiji (ImageJ. 2.0.0rc48). Laser intensity and gain were kept the same for all stacks and images.

**Statistical analysis.** Statistical significance was assessed using Student's two-tailed *t*-test for independent samples, Logrank test for animal survival experiments (Mantel-Cox method), One-tail Fisher's Exact Test for bacterial systemic dissemination, and Two-way ANOVA with Tukey multiple comparison tests. *p* values less than 0.05 were considered statistically significant.

**VP1 strain distribution and promoter analysis.** To determine the distribution of *vp1* across pneumococcal strains we used blastP to search for the VP1 sequence in a database of streptococcal genomes (**Suppl. Table 2.1**). This database includes thirty-two genomes that include nineteen serotypes and thirty-six MLST types (as detailed in "In silico screen for predicted secreted peptides**").** Further, since our work was performed in PMEN1 we expanded the set with three additional PMEN1 genomes (90). Sequences were aligned with Jalview (100) revealing three putative pherotypes in pneumococcus as well as four additional variants outside of pneumococcus. To confirm the organization of VP1 sequences, all the coding sequences were clustered using a maximum parsimony tree in iQ-TREE; it confirmed the organization into seven types (**Suppl. Figure 6.6**). The VP1 alleles are presented in the context of a maximum parsimony species tree, generated from the core region of streptococcal genomes (37). Each allelic type is colored-coded, and the visualization was generated using the Interactive Tree of Life (iTOL) (101).

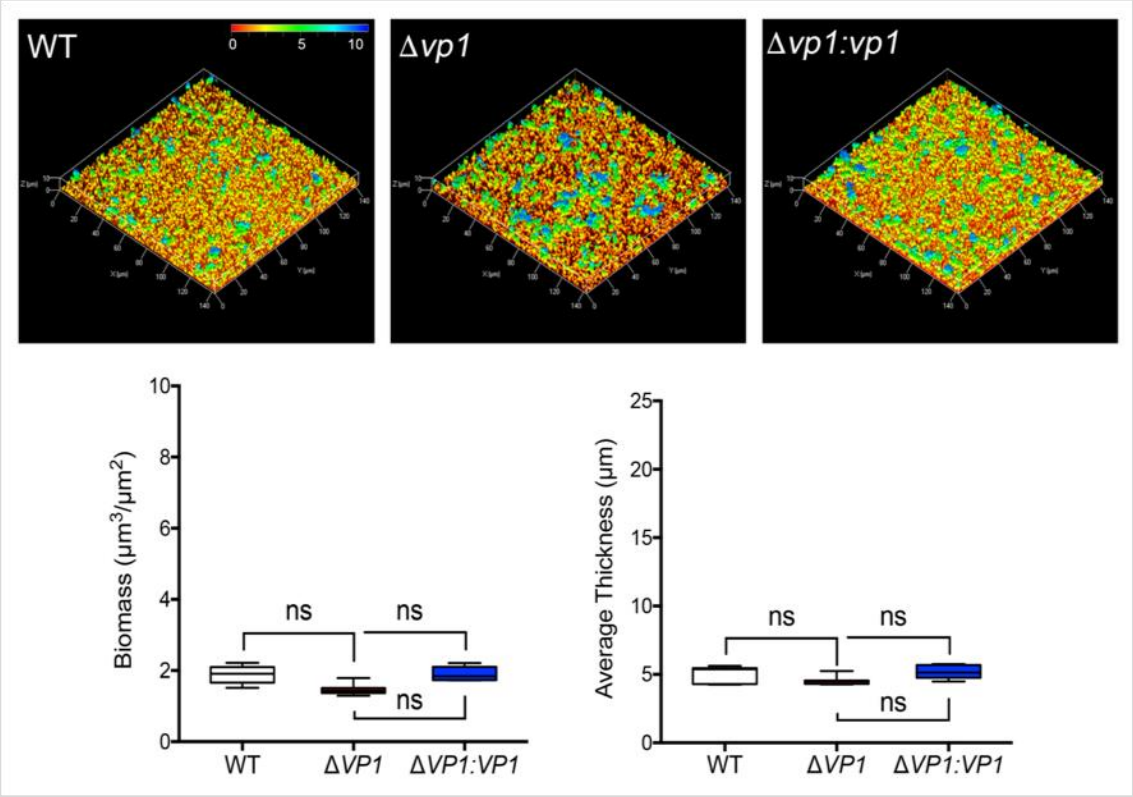
We searched *in silico* for Rgg and CodY binding regions. The putative Rgg-binding box upstream of the *vp1* operon was identified using the consensus for the Rgg-box (as presented in RegPrecise (102)). Specifically, we identified "TTGAAGGAGTGTAA" at position -17 of the *vp1* gene. Next, given that the orthologs of *rgg* and *vp1* in D39 (spd0144 and sp0145, respectively) are differentially expressed in a *codY* deletion mutant relative to the wild-type strain (44), we searched for a CodY-box upstream of these genes. In *rgg*, a CodY binding box is identified upstream of Rgg using RegPrecise on the TIGR4 genome (the representative pneumococcal genome in this database); it corresponds to the sequence "AACTATCAGAATATT" (score of 4.7) and an identical region is present at position -69 of *rgg* in PN4595-T23. In *vp1*, the sequence "GATTTTCTAAAATAA" at position –113 closely resembles the *L. lactis* CodY consensus as well as predicted CodY binding boxes in pneumococcus (103).

## 2.6 Supplementary Figures



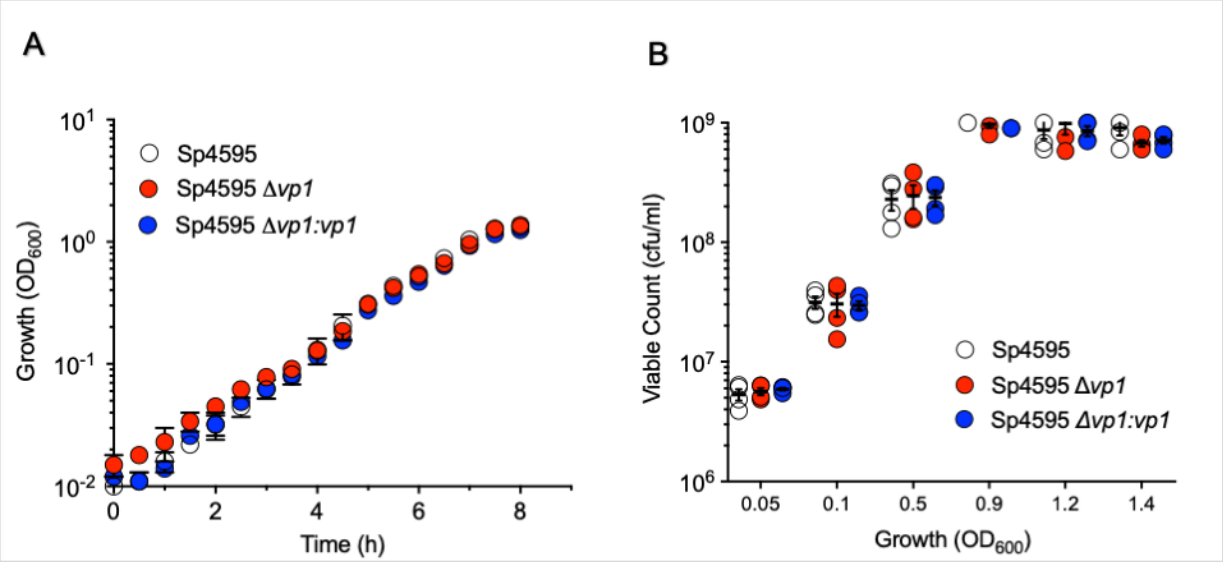

**Supplementary Figure 2.1 Effect of CMEE cells presence on biofilm capacity**

Comparison of 16h biofilms grown in CMEE media with or without pre-cultured basal layer of epithelial cells. **A.** Representative 3D images assembled from confocal images from biofilms seeded on glass or CMEE cells. The 3D projections are pseudo-colored to highlight biofilm thickness, as described in the scale. **B.** Biofilm biomass and average thickness were computed from confocal images. Data represent the mean ± S.E.M. of at least three independent experiments. Statistical significance was calculated using unpaired two-tailed Student's *t*-test. * $p < 0.0001$.



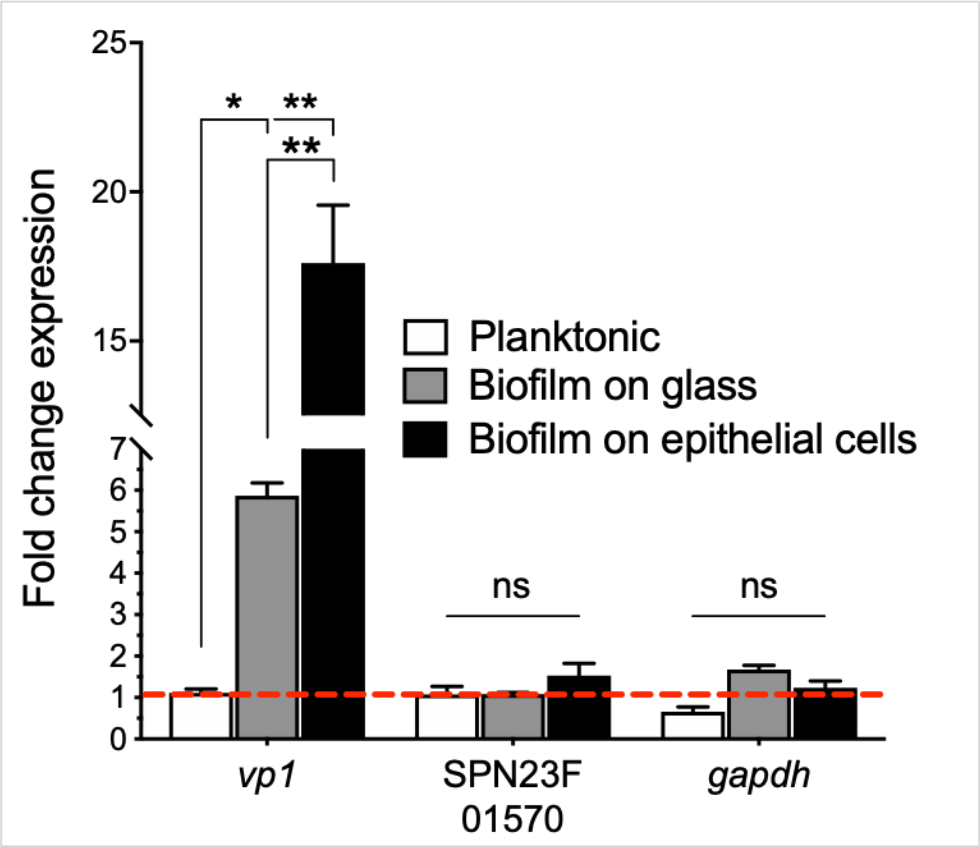

**Supplementary Figure 2.2 *vp1* does not influence biofilm capacity on abiotic surface**

Representative images assembled from confocal micrographs of biofilms of wild-type, Δ*vp1*, or Δ*vp1:vp1* grown on glass and imaged 16h post-seeding. Top panels: 3D projection was pseudo-colored to highlight biofilm thickness (in um). Bottom panels: Biofilm biomass and average thickness were computed from confocal images. Data represent the mean ± S.E.M. of at least three independent experiments. Statistical significance was calculated using 2-way ANOVA analysis with Tukey correction. ns = not significant. Scale bar is conserved across all figures.



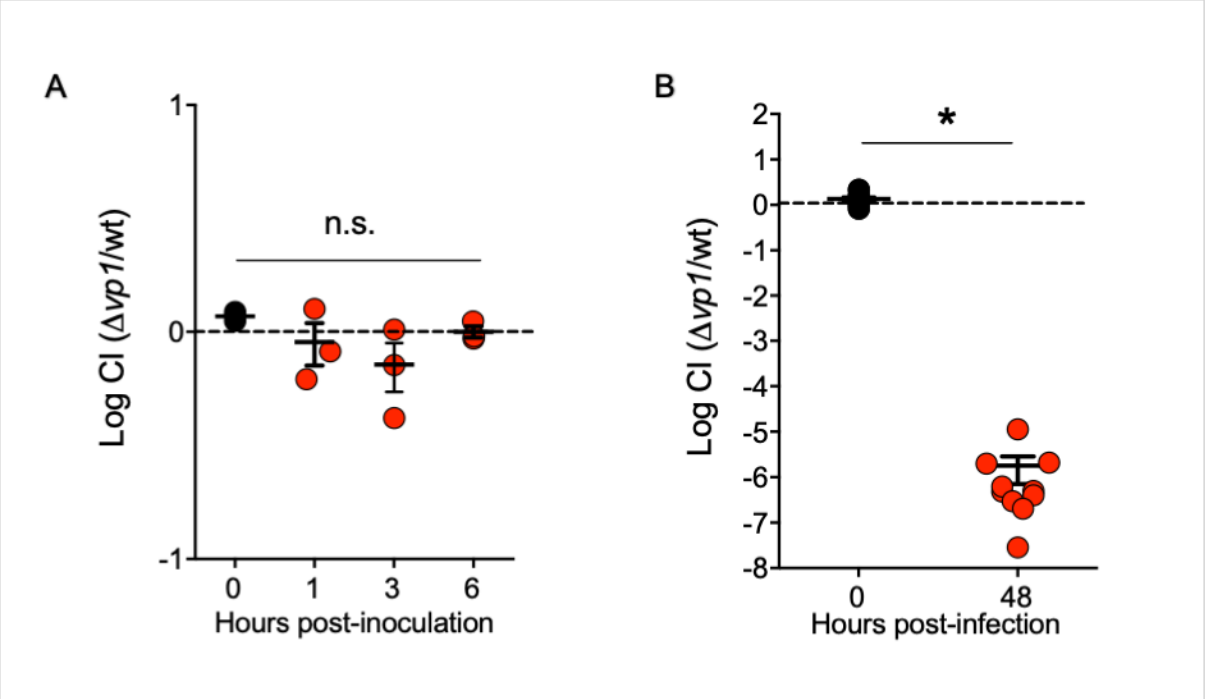

**Supplementary Figure 2.3 *vp1* provides no influence on growing or viability in rich media**

Analysis of PN4595-23T strains, wild-type, Δ*vp1* mutant, and complemented Δ*vp1:vp1*, *in vitro* during growth in rich media. **A.** Growth assay where no differences in growth is observed. **B.** Viability assay at multiple optical densities. Note that no differences are observed between strains.



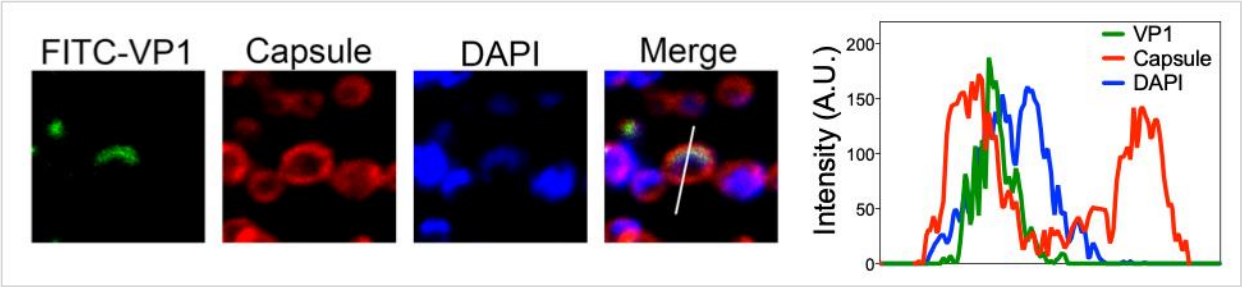

**Supplementary Figure 2.4 *vp1* expression is enhances during biofilm growth on epithelial cells**

Total RNA was isolated from bacteria growing in planktonic, biofilm on glass or biofilm on epithelial CMEE cells. Expression was assessed by RT-qPCR. Data represent the mean ± S.E.M. of at least three independent experiments. Statistical significance was calculated using unpaired two-way ANOVA analysis with Bonferroni correction, and multiple comparison between samples were performed. * $p = 0.0015$, ** $p < 0.0001$, ns = not significant relative to the planktonic. Note that SPN23F01570 correspond to a gene downstream to the *vp1* operon.



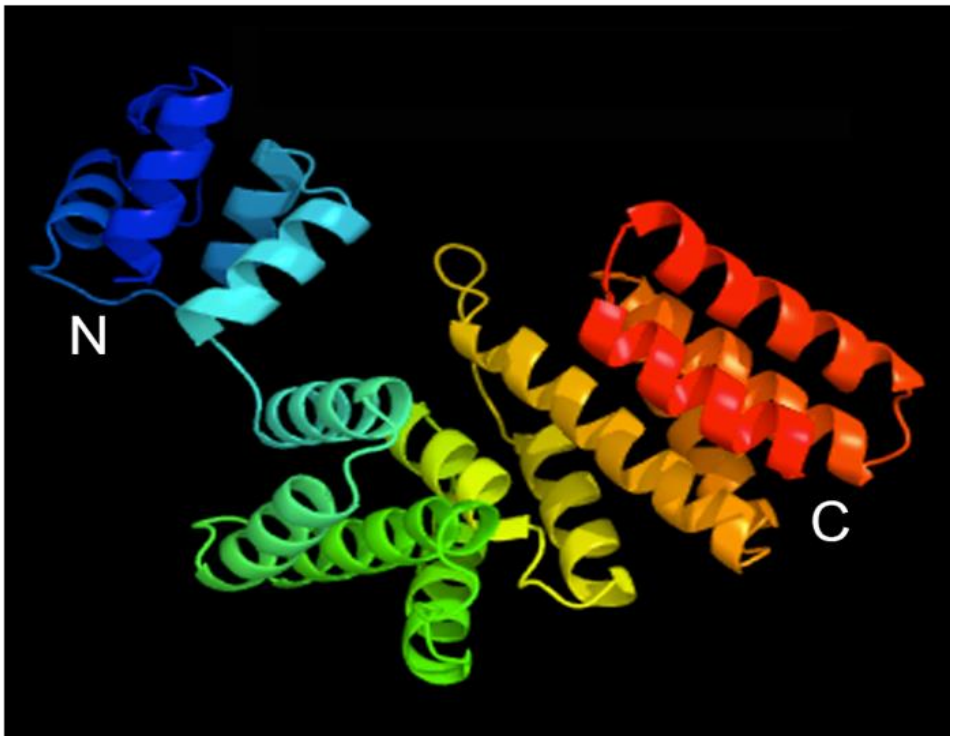

**Supplementary Figure 2.5 *vp1* provides pneumococcus with an *in vivo* fitness advantage**

**A.** *In vitro* competition assays comparing CFUs between the PN4595-23T wild-type and Δ*vp1* mutant. An equal amount of both strains was simultaneously inoculated in rich liquid culture using and samples were extracted from lag phase (1 h), exponential growth (3 h), and stationary phase (6 h). Data is expressed as competition index (CI). **B**. *In vivo* competition assays were performed comparing CFUs between PN4595-23T wt and Δ*vp1* mutant. Both strains were simultaneously injected bilaterally into the chinchilla tympanic bullae. Animals were euthanized 48h post-infection, and bacteria recovered from the animal middle ear and enumerated. Data is expressed as the mean ± S.E.M of at least three independent experiments. Data in A and B were analyzed with one-way Anova or two-tailed Student *t*-test, respectively. * $p < 0.0001$ compared to the initial inoculum for each strain.



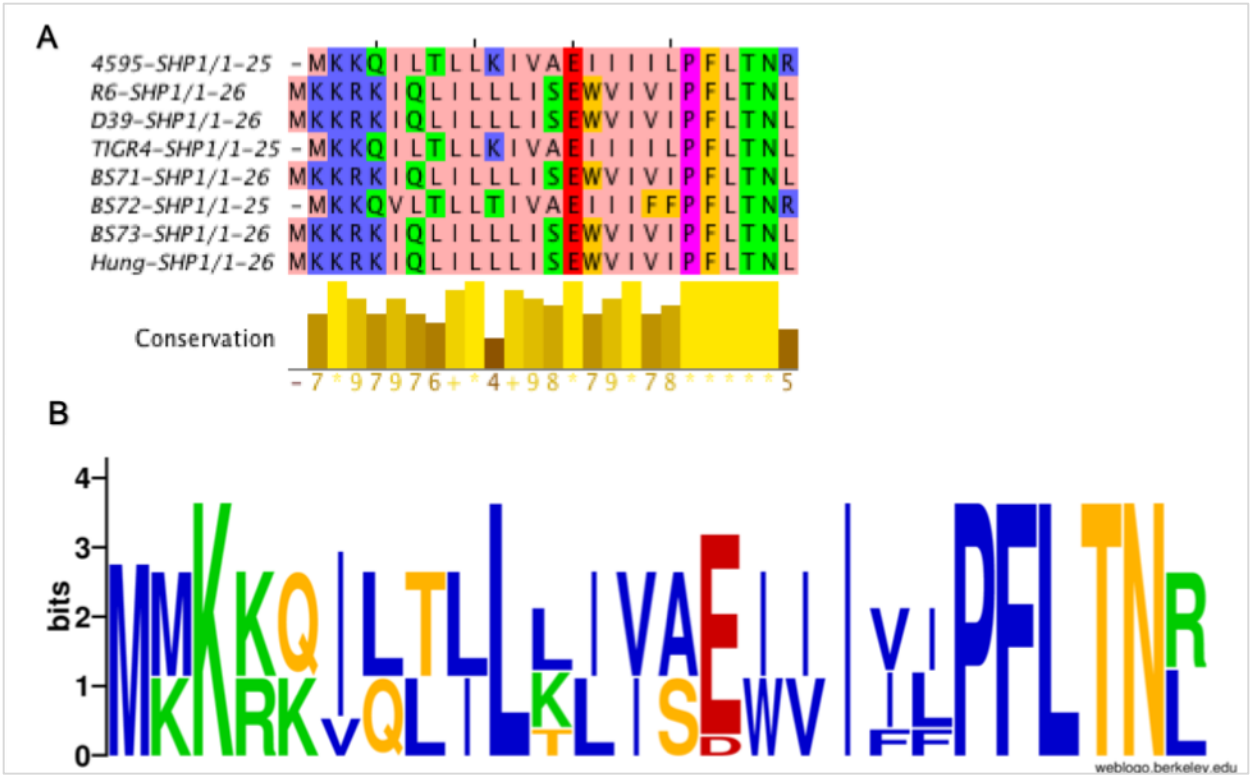

**Supplementary Figure 2.6 VP1 peptide binds to the surface of pneumococcal cells**

Fluorescence intensity across the width of the bacterial cell (white line) was measured on representative cells and compared to capsular staining (red anti-capsule) and DNA straining (blue DAPI). Red green blue (RGB) profiles (right panel) show that VP1 appears to bind internally to the capsule and externally to DNA, consistent with membrane binding.



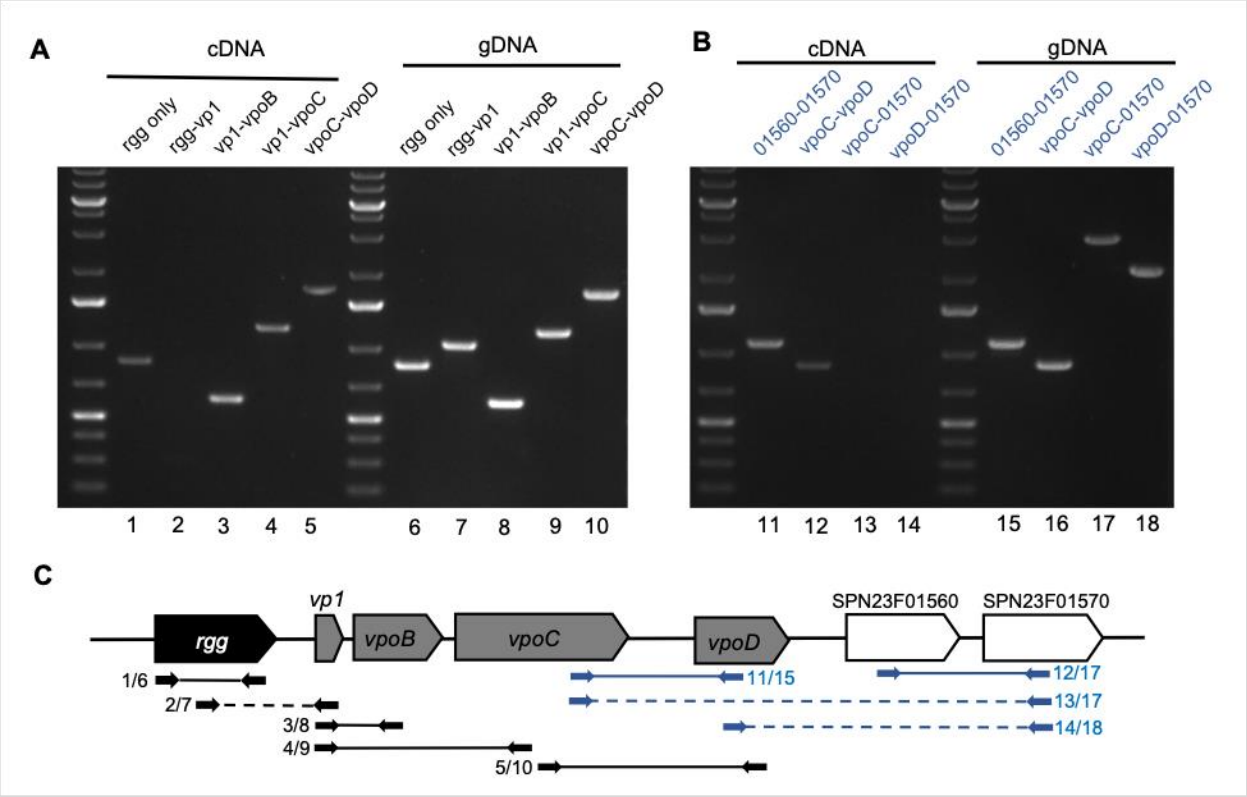

**Supplementary Figure 2.7 PHYRE2 modeling of Rgg144.**

The predicted Rgg144 structure shows high similarity to the crystal structure of Rgg2 from *S. dysgalactiae*. N- and C-termini are shown. The C-termini shows a stretch of 220 amino acids characteristics of Rgg family of transcription factors (e-value 2.41e-$_{61}$). The C-terminal show conserved sequence and an structure helix-turn-helix characteristics of DNA binding proteins (pfam e-value of 4.71e-$_{11}$). This comparative analysis utilizes the crystal structure from other Rgg proteins, including the *S. dysgalactiae* shp pheromone rgg2 receptor and *B. subtilis* PlcR. We proposed that this gene encodes a PlcR-related Rgg transcriptional factor**.**



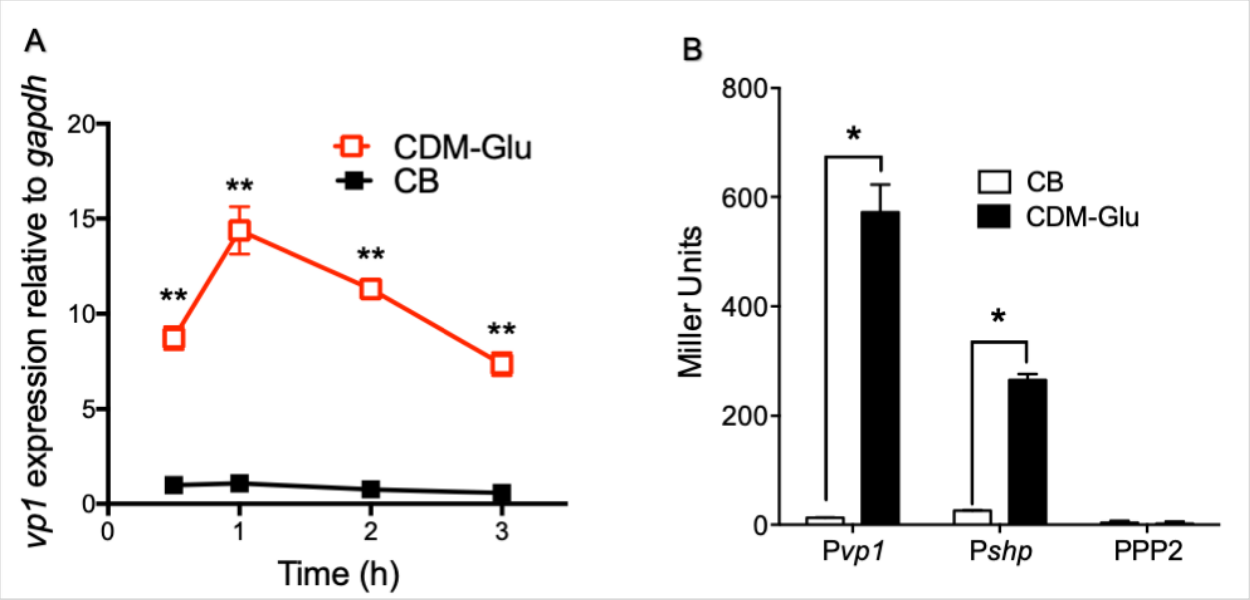

**Supplementary Figure 2.8 Identification of a novel SHP molecule in *S. pneumoniae***

**A.** Multiple SHP peptide sequence alignment translated from pneumococcal ORFs located immediately downstream of *rgg*. The hydrophobic motif starts at position +5. Glutamic acid is indicated in red. The residues are colored according to their physicochemical properties using the Zappo's color scheme. Conservation degree is indicated. **B.** Sequence logo derived from 21 predicted SHP peptides. Amino acids are represented by one letter abbreviation and color coded as follows: blue, basic; red, acidic; green, polar; yellow, neutral. Height of amino acids is proportional to the fraction of the observed frequency relative to the expected frequency in the streptococcal database.



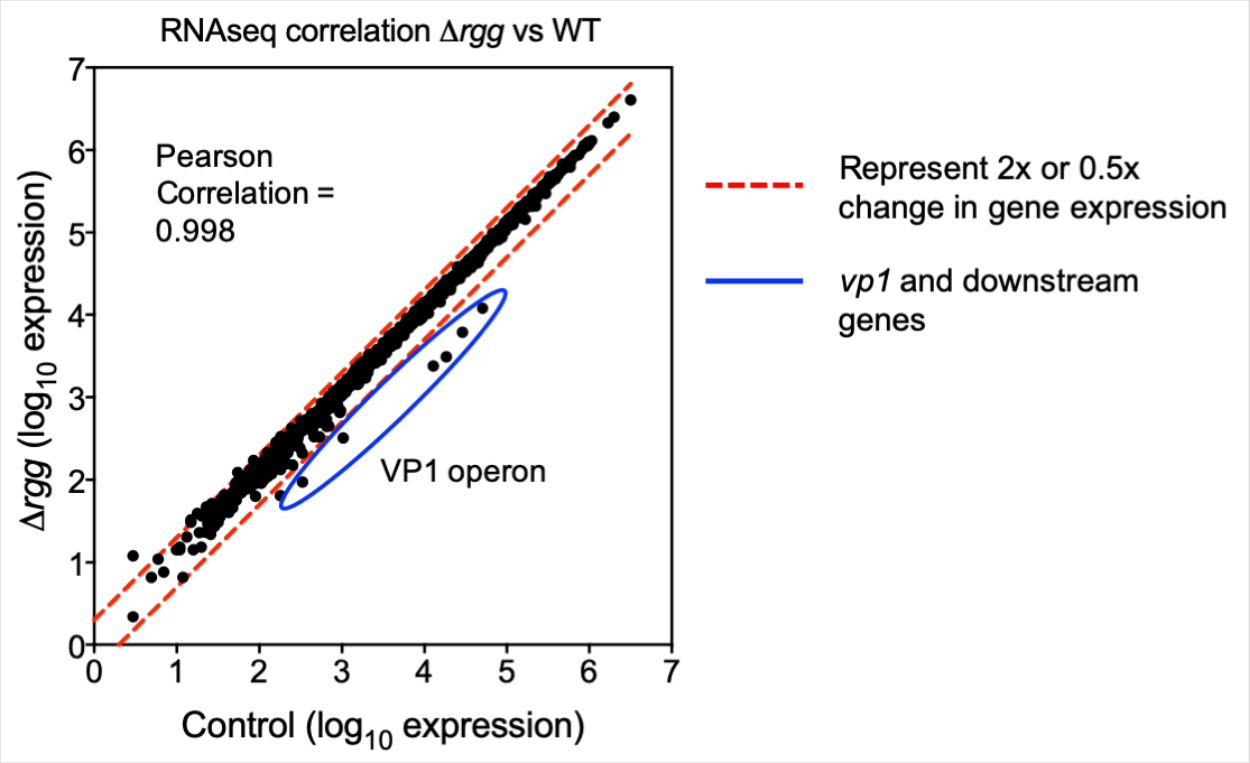

**Supplementary Figure 2.9 Demonstration of *vp1* operon transcriptional units**

Prior to cDNA synthesis, all RNA samples were DNase-treated and subjected to a PCR check using primers for *gapdh* gene to ensure total elimination of DNA. Only when no amplification was observed in the *gapdh* check PCR was the cDNA synthesized. Lanes 1-5 and 11-14 are PCRs on cDNA template, and lanes 6-10 and 15-18 on gDNA template. Primes were labeled according the *locus* they bind to. **A.** Primers are as follows: lanes 1 and 6, rggF fwd and rggF rev; lanes 2 and 7, rgg-a fwd and vp1-a rev; lanes 3 and 8, vp1-b fwd and vpB-a rev; lanes 4 and 9, vp1-b fwd and vpC-a rev; lanes 5 and 10, vpC-a fwd and vpD-a rev. **B.** Primers as follow: lanes 11 and 15, spn23f01560 fwd and spn23f01570 rev; lanes 12 and 16, vpC fwd and vpD rev; lanes 13 and 17, vpCb fwd and spn23f01570 rev; lanes 14 and 18, vpD fwd and spn23f01570 rev. **C.** Genomic *locus* schematic indicate the primer binding sites corresponding to the bands on the gel. Solid lines indicate expected amplification products. Dashed lines indicate no observed amplification product. Numbers indicates lane numbers from A and B.



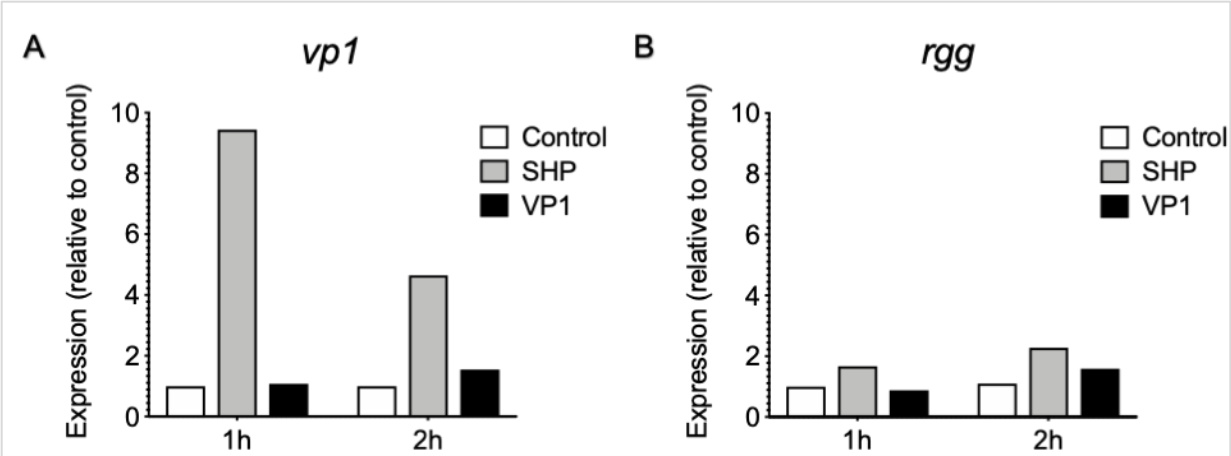

**Supplementary Figure 2.10 Experimental condition for high *vp1* expression**

**A.** Bacterial were pre-cultured in rich media (Columbia broth) and switched to CDM-Glu. RNA samples were isolated at 1h, 2h, or 3h after the switch and relative expression was measured by qRT-PCR. Expression was normalized to levels of *gapdh.* Data represent the mean ± S.E.M. of at least two independent experiments. Statistical significance was calculated using unpaired two-tailed Student's *t*-test. * $p < 0.05$ and ** $p < 0.01$. **B.** Promoter activity was measured using reporter strain in the Δ*vp1* background. Promoter activity of *vp1* (Δ*vp1-lacZ*-Δ*vp1*) and *shp* (Δ*shp-lacZ*-Δ*vp1*) was assayed in Columbia broth and CDM supplemented with 55 mM of glucose. Both promoters showed induction in CDM-glucose. PPP2 correspond to empty vector PPP2-*lacZ* used to measure background promoter activity.



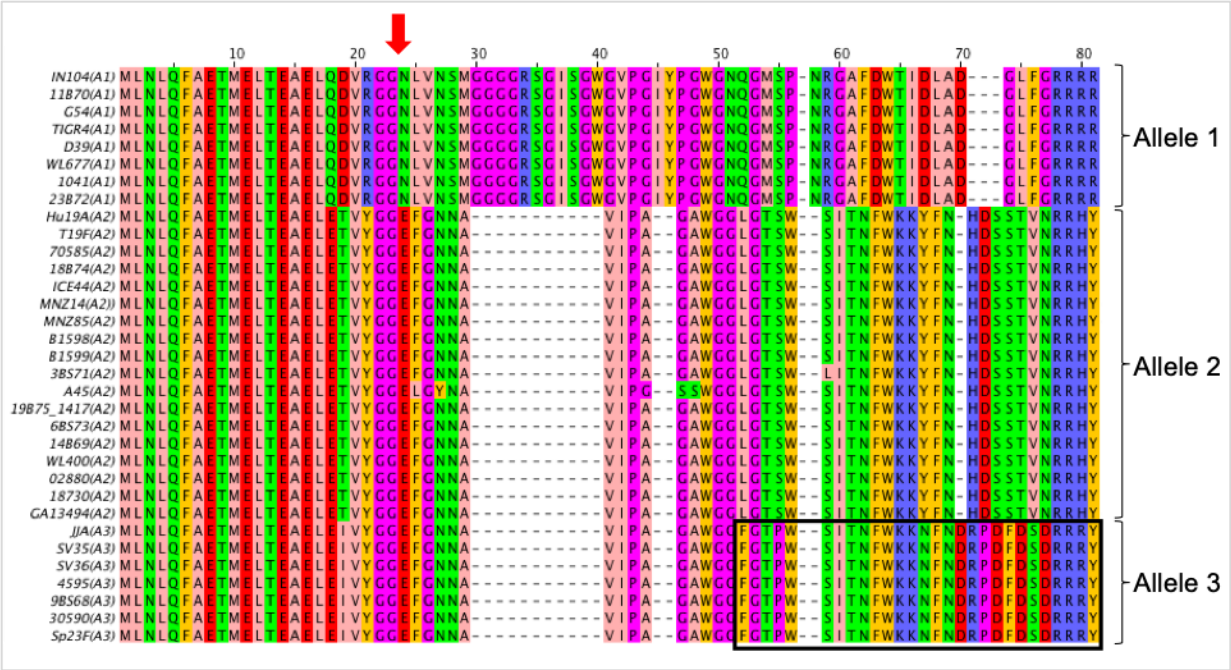

**Supplementary Figure 2.11 RNA-Seq transcriptome analysis of the *rgg*-deficient strain**

Scatterplot of global RNA expression in PN4595-T23 Δ*rgg* mutant versus the wild-type strain (n=1). Both axes show RNA expression (number of reads) on a $\log_{10}$ scale of the deletion mutants (y-axis) and the wild-type strain (x-axis). The dashed blue lines represent 2-fold up- and downregulation. Downregulation of the *vp1* operon (blue circle) was observed in the *rgg* deletion mutant, suggesting *rgg* specifically control *vp1* operon expression.



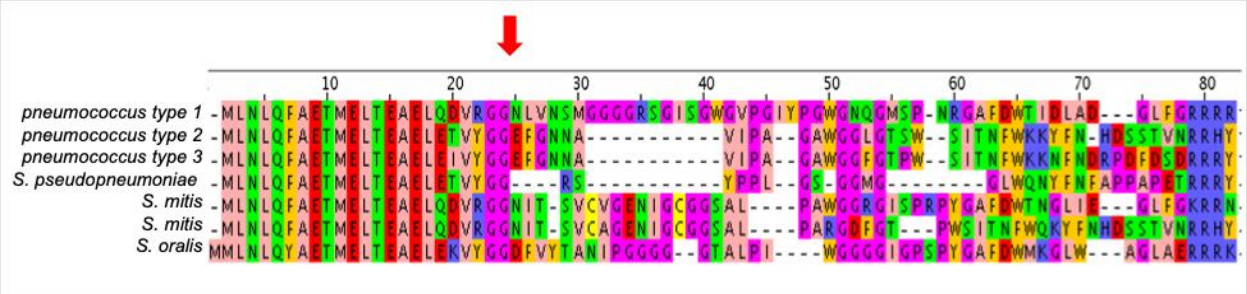

**Supplementary Figure 2.12 *vp1* expression is highly enhanced by SHP-C12 variant**

Total RNA was isolated from bacteria growing in planktonic was assessed by RT-qPCR. **A.** *vp1* expression after addition of either SHP or synthetic VP1 peptide. **B.** *rgg* expression after addition of either SHP or synthetic VP1 peptide. Data represent one independent experiment.




**Supplementary Figure 2.13 Alignments of VP1 sequences from *S. pneumoniae***

Residues are colored according to Zappo's color scheme. Black box indicates predicted and tested mature peptide. Red arrow indicates predicted cleavage site. Products of *vp1* alleles 1, 2 and 3 are shown.




**Supplementary Figure 2.14 Alignment of seven VP1 sequences from *Streptococcus sp***

All share the N-terminal motif, predicted to function in processing and export. Top three variants represent coding sequences present in pneumococcal genomes: D39, Taiwan19A, and ATCC 700669 respectively. The next four variants are from related species: *S. pseudopneumoniae* IS7493, *S. mitis* SK564 and SK569, and *S. oralis* subspecies *tigurinus* 2426). The overall conservation of the positively charged C-termini suggests this region is critical for VP1 signaling.

## 2.7 Supplementary Tables

**Supplementary Table 2.1** Strains utilized for the pangenome analysis
Bold: PMEN1 strains. Set 1: genomes in the tree. Set 2: Genomes in MAST search.

| Strain ID | Species | GenBank Accession No. | Set1 | Set 2 |
|---|---|---|---|---|
| SK1076 | *Streptococcus infantis* | AFNN00000000 | X | X |
| SK970 | *Streptococcus infantis* | AFUT00000000 | X | X |
| SK1073 | *Streptococcus mitis* | AFQT00000000 | X | X |
| B6 | *Streptococcus mitis* | FN568063 | X | X |
| NCTC 12261 | *Streptococcus mitis* | AEDX00000000 | X | X |
| SK1080 | *Streptococcus mitis* | AFQV00000000 | X | X |
| SK321 | *Streptococcus mitis* | AEDT00000000 | X | X |
| SK564 | *Streptococcus mitis* | AEDU00000000 | X | X |
| SK569 | *Streptococcus mitis* | AFUF00000000 | X | X |
| SK575 | *Streptococcus mitis* | AICU01000001 | X | X |
| SK597 | *Streptococcus mitis* | AEDV00000000 | X | X |
| bv. 2 str. F0392 | *Streptococcus mitis* (reclassified as S. mitis) | AFUO00000000 | X | X |
| ATCC 6249 | *Streptococcus mitis (reclassified as S. mitis)* | AEEN00000000 | X | X |
| bv. 2 str. SK95 | *Streptococcus mitis (reclassified as S. mitis)* | AFUB00000000 | X | X |
| Uo5 | *Streptococcus oralis* | FR720602 | X | X |
| 2426 | *Streptococcus oralis- subspecies tigurinus* | ASXA00000000 | X | X |
| AZ_3a | *Streptococcus oralis- subspecies tigurinus* | AORU00000000 | X | X |
| G54 | *Streptococcus pneumoniae* | CP001015 | X | X |
| SP18-BS74 | *Streptococcus pneumoniae* | ABAE00000000 | X | X |
| **SV35-T23** | *Streptococcus pneumoniae* | ADNN00000000 | X | |
| **PN4595-T23** | *Streptococcus pneumoniae* | ABXO00000000 | X | |
| 70585 | *Streptococcus pneumoniae* | CP000918 | X | X |
| **ATCC 700669** | *Streptococcus pneumoniae* | FM211187 | X | X |
| CDC0288_04 | *Streptococcus pneumoniae* | ABGF00000000 | X | X |
| CDC1087-00 | *Streptococcus pneumoniae* | ABFT00000000 | X | X |
| CDC1873-00 | *Streptococcus pneumoniae* | ABFS00000000 | X | X |
| CDC3059-06 | *Streptococcus pneumoniae* | ABGG00000000 | X | X |
| CGSP14 | *Streptococcus pneumoniae* | CP001033 | X | X |
| D39 | *Streptococcus pneumoniae* | CP000410 | X | X |
| GA13494 | *Streptococcus pneumoniae* | AGOZ01000001 | X | X |
| Hungary19A-6 | *Streptococcus pneumoniae* | CP000936 | X | X |
| ICE44 | *Streptococcus pneumoniae* | AUYF00000000 | X | X |
| INV104 | *Streptococcus pneumoniae* | FQ312030 | X | X |
| JJA | *Streptococcus pneumoniae* | CP000919 | X | X |
| SP11-BS70 | *Streptococcus pneumoniae* | ABAC00000000 | X | X |
| SP14-BS69 | *Streptococcus pneumoniae* | ABAD00000000 | X | X |
| SP19-BS75 | *Streptococcus pneumoniae* | ABAF00000000 | X | X |
| SP23-BS72 | *Streptococcus pneumoniae* | ABAG00000000 | X | X |
| SP3-BS71 | *Streptococcus pneumoniae* | AAZZ00000000 | X | X |
| SP6-BS73 | *Streptococcus pneumoniae* | ABAA00000000 | X | X |
| SP9-BS68 | *Streptococcus pneumoniae* | ABAB00000000 | X | X |
| SPN1041 | *Streptococcus pneumoniae* | CACE00000000 | X | X |
| SPNA45 | *Streptococcus pneumoniae* | CACG00000000 | X | X |
| **SV36-T3** | *Streptococcus pneumoniae* | ADNO00000000 | X | |
| Taiwan19F-14 | *Streptococcus pneumoniae* | CP000921 | X | X |
| TIGR4 | *Streptococcus pneumoniae* | AE005672 | X | X |
| WL400 | *Streptococcus pneumoniae* | AVFA00000000 | X | X |
| WL677 | *Streptococcus pneumoniae* | AUWZ00000000 | X | X |
| MNZ14 | *Streptococcus pneumoniae* (basal branch) | ASJO00000000 | X | X |
| MNZ85 | *Streptococcus pneumoniae* (basal branch) | ASJF00000000 | X | X |
| B1069B1599 | *Streptococcus pneumoniae* (basal branch) | JBOW00000000 | X | X |
| B1068B1598 | *Streptococcus pneumoniae* (basal branch) | JBOV00000000 | X | X |
| ATCC BAA-960 | *Streptococcus pseudopneumoniae* | AICS00000000 | X | X |
| IS7493 | *Streptococcus pseudopneumoniae* | CP002925 | X | X |
| SK674 | *Streptococcus pseudopneumoniae* | AJKE00000000 | X | X |
| E709 | *Streptococcus pneumoniae* | JBOR00000000 | | X |
| *K2521* | *Streptococcus pneumoniae* | JBOS00000000 | | X |
| *K2527* | *Streptococcus pneumoniae* | JBOT00000000 | | X |
| *K2557* | *Streptococcus pneumoniae* | JBOU00000000 | | X |
| *MNZ11b* | *Streptococcus pneumoniae* (basal branch) | ASJW00000000 | | X |
| *MNZ37* | *Streptococcus pneumoniae* (basal branch) | ASJP00000000 | | X |
| *MNZ41* | *Streptococcus pneumoniae* (basal branch) | ASJW00000000 | | X |
| *T1366* | Streptococcus oralis- subspecies tigurinus | | | X |

**Supplementary Table** 2.2 Summary of MEME/MAST analysis

| | No of sequences | |
|---|---|---|
| Iteration | MEME input | MAST output |
| 1 | 167 | 192 |
| 2 | 192 | 560 |
| 3 | 560 | 716 |
| 4 | 696* | 696* |

**Supplementary Table 2.3** RNA counts using NanoString. Red: Control Genes

| Annotation | Planktonic | Planktonic | Effusions | Effusions | Average Planktonic | Average Effusions | ID in ATCC 700669 |
|---|---|---|---|---|---|---|---|
| VP1 | 3597.27 | 4378.46 | 150910.93 | 137402.28 | 3988 | 144157 | SPN23F01520 |
| Hypothetical | 1724.59 | 1847.77 | 861.74 | 1036.74 | 1786 | 949 | upstream of SPN23F08480 |
| CSP2 | 1744.15 | 1387.14 | 4802.20 | 2791.53 | 1566 | 3797 | SPN23F22700 |
| CibA | 260.54 | 296.92 | 1158.58 | 771.81 | 279 | 965 | SPN23F_01380 |
| CibB | 155.77 | 183.81 | 557.49 | 407.14 | 170 | 482 | SPN23F01370 |
| BlpU | 10481.68 | 9485.49 | 18352.64 | 14660.52 | 9984 | 16507 | SPN23F00570 |
| BlpO | 340.87 | 410.03 | 2115.86 | 1473.10 | 375 | 1794 | SPN23F_04860 |
| BlpI | 210.95 | 273.07 | 1136.31 | 2096.47 | 242 | 1616 | SPN23F_04830 |
| BlpJ | 168.34 | 232.18 | 854.32 | 756.23 | 200 | 805 | SPN23F_04840 |
| BlpN | 64.26 | 70.69 | 104.82 | 82.99 | 67 | 94 | CGSSp4595_0471* |
| PhrA | 6061.57 | 7663.45 | 309657.33 | 78574.76 | 6863 | 194116 | SPN23F19680 |
| LanA | 3370.96 | 4458.18 | 146235.81 | 35443.71 | 3915 | 90840 | SPN23F_19710 |
| LcpA | 773.94 | 992.62 | 175073.15 | 87118.06 | 883 | 131096 | SPN23F12701 |
| EpsA | 60479.54 | 76995.80 | 109695.64 | 105276.86 | 68738 | 107486 | SPN23F03180 |
| CbpAs | 17981.46 | 17075.52 | 104449.11 | 76533.22 | 17528 | 90491 | SPN23F22240 |

**Supplementary Table 2.4** SHP Sequences

| | |
|---|---|
| PN4595 | MKKQILTLLKIVAEIIIILPFLTNR |
| D39 | MKKRKIQLILLLISEWVIVIPFLTNL |
| TIGR4 | MKKQILTLLKIVAEIIIILPFLTNL |
| R6 | MKKRKIQLILLLISEWVIVIPFLTNL |
| Hungary19A-6 | MKKRKIQLILLLISEWVIVIPFLTNL |
| INV200 | MKKRKIQLILLLISEWVIVIPFLTNL |
| 23F | MKKQILTLLKIVAEIIIILPFLTNR |
| SP23-BS72 | MKKQVLTLLTIVAEIIIFFPFLTNR |
| CGSP14 | MKKQVLTLLTIVAEIIIFFPFLTNR |
| CDC1087-00 | MKKQVLTLLTIVADIIIFFPFLTNR |
| INV104B | MKKQVLTLLTIVADIIIFFPFLTNR |
| SP14-BS69 | MKKRKIQLILLLISEWVIVIPFLTNL |
| SP6-BS73 | MKKRKIQLILLLISEWVIVIPFLTNL |
| OXC141 | MKKRKIQLILLLISEWVIVIPFLTNL |
| SP11-BS70 | MKKQILTLLKIVAEIIIILPFLTNR |
| CDC3059-06 | MKKQILTLLKIVAEIIIILPFLTNR |
| SP195 | MKKQILTLLKIVAEIIIILPFLTNR |
| CDC0288-04 | MKKQILTLLKIVAEIIIILPFLTNR |
| SP9-BS68 | MKKQILTLLKIVAEIIIILPFLTNR |
| SP18-BS74 | MKKRKIQLILLLISEWVIVIPFLTNL |
| SP3-BS71 | MKKRKIQLILLLISEWVIVIPFLTNL |

**Supplementary Table 2.5** Constructs used in this study

| Name | Parental Strain | Comments |
|---|---|---|
| PN4595-T23 | | Wild type |
| PN4595D*vp1* | PN4595-T23 | replaced *vp1* with spec cassette |
| PN4595D*vp1*:*vp1* | PN4595-T23 | inserted *vp1-kan* into native *locus* |
| PN4595D*rgg* | PN4595-T23 | replaced *rgg* with spec cassette |

**Supplementary Table 2.6** Oligonucleotides used in this study

| Primer | Nucleotide sequence (5 to 3 ) | Restriction site |
|---|---|---|
| rgg1 fwd | GCAGATTGGTAGAAAGGTTGTGGA | |
| rgg1 rev | CGACTTCCTCCAACAGTTGCTCTA | |
| vp1A fwd | AGTTTATGGAGGGGAATTTGGG | |
| vp1A rev | ATTGACCAAGGTGTCCCGAAC | |
| vpB fwd | GGAAAGAAATGGGGGCTAGA | |
| vpB rev | AATGTGCGAGACCCCAAATA | |
| vpC fwd | CGTCTTTAAACAAATACCTGAGGAG | |
| vpC rev | AAAACAACCGTAATGGCACTG | |
| vpD fwd | CTGTGGCAGCTATTGCTATGA | |
| vpD rev | CCAAAGACTGGTATGACGGAAA | |
| spn23f01560 fwd | GGACAAGCAGCTAGCAAGAA | |
| spn23f01560 rev | CGACTTCAATCTCGTAACCAGTC | |
| spn23f01570 fwd | TGTTAACCGTAGCGGTTCTG | |
| spn23f01570 rev | GTGAGTAGTCTGTGAACTCTGTAAA | |
| rgg2 fow | CGCACAATTAACTACAACTCCTAC | |
| rgg2 rev | AGCTGAGAGACAATCTGCTTAT | |
| rgg3 fow | TGGCACAGATTATCCCACTTTAC | |
| rgg3 rev | AGGGAAGTCGTCGAACTCTTA | |
| rgg4 fwd | GCCCATTCTCACAACAAACAG | |
| rgg4 rev | CCTTCCTCTGCTGTTAACTCTAC | |
| mutR Fwd | GGATTCTCTCAAGTACGCTTGG | |
| mutR rev | GCCTCTTCTTTGGTTAGTCGTAATA | |
| rggF fwd | ATGATTGAAAAAATGGAATTGGGG | |
| rggF rev | CTAATCGATAAGCTTTTTGTATTGCTG | |
| rgg-a fwd | GAGGAAGTCGAGCAAGAACAGATG | |
| vpA-a rev | CGTTAAAATTTTTCTTCCAGAAA | |
| vpA-b fwd | GCAGAAACAATGGAATTGACAGAAG | |
| vpB-a rev | TGTAGCCCCTGTAAAAGTCGGA | |
| vpC-a rev | GATGATGTTGAAAATCCGCC | |
| vpC-a fwd | ATCTCTTGCCTTGCTATCCACTTTT | |
| vpD-a rev | GGTTCCCAACCTTACCTGTCACTA | |
| Spec-a fwd | ATATATCTCGAGGGATCCCCCGTTTGATTTTTAATGGTAATG | XhoI |
| Spec-a rev | ATATATGCTAGCGCGAAAAACCCCGCCGAAGCGGGGTTTTTTGCGAATTGACGCGGAATTGGATC | NheI |
| Spec-b fwd | ATATATGCTAGCGGATCCCCCGTTTGATTTTTAATG | NheI |
| Spec-b rev | ATATATCTCGAGGCGAAAAAACCCCGCCGAAGCGGGGTTTTTTGCGTGGATCCAATTTTTTTATAAT | XhoI |
| Kan fwd | CAAAATAGTGAGGAGTCTATGCCTATTCCAGAGGAAATGG | |
| Kan rev | ATTCAATATTCTCTCCGCAAAAAACCCCGCTTC | |
| vp1AKO1 | TGTGACCTTTGATGAGTTTGGG | |
| vp1AKO2 | ATATATCTCGAGCCATTGTTTCTGCAA | XhoI |
| vp1AKO3 | ATATATGCTAGCCCGCCGTCGTTATTAATATCCAAA | NheI |
| vp1AKO4 | CTCCCAGAAAGATTTGTAGAACTTG | |
| vp1AComp1 | TGGATCGCTTTGCTCATCAAG | |
| vp1AComp2 | TGGAATAGGCATAGACTCCTCACTATTTTGATTAGTACCTATTTTATCAC | |
| vp1AComp3 | GCGGGGTTTTTTGCGGAGAGAATATTGAATGGACTAATGAAAATGTAAATTTAAC | |
| vp1AComp4 | CCAAGCGTTGAAATCCACTG | |
| rggKO1 | GGAGATGAGACAGTTGAAG | |
| rggKO2 | ATATATGCTAGCCCCAATTCCATTTTTTCAATC | NheI |
| rggKO3 | ATATATCTCGAGGCGTTTGCAGCAATACAAAAAG | XhoI |
| rggKO4 | CCCAACCGTGATTACTAATG | |

## 2.8 References


1. Bassler BL. Small Talk : Cell-to-Cell Communication in Bacteria. Cell. 2002;109:421–424.
2. Bassler BL, Losick R. Bacterially Speaking. Cell. 2006;125(2):237–246.
3. Federle MJ, Bassler BL. Interspecies communication in bacteria. J Clin Invest. 2003;112(9):1291–1299.
4. Claverys J, Prudhomme M, Martin B. Induction of Competence Regulons as a General Response to Stress in Gram-Positive Bacteria. Annu Rev Microbiol 2006. 2006;60:451–475.
5. Claverys J-P, Havarstein LS. Cannibalism and fratricide: mechanisms and raisons d'être. Nature Reviews Microbiology. 2007;5:219–229.
6. Dawid S, Roche AM, Weiser JN. The blp bacteriocins of Streptococcus pneumoniae mediate intraspecies competition both in vitro and in vivo. Infection and Immunity. 2007;75(1):443–451.
7. Havarstein LS, Coomaraswamy G, Morrison DA. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in Streptococcus pneumoniae. Proc Natl Acad Sci USA. 1995 Nov 21;92(24):11140–4.
8. Stewart PS, Costerton JW. Antibiotic resistance of bacteria in biofilms. Lancet. 2001 Jul 14;358(9276):135–8.
9. Hall-Stoodley L, Hu FZ, Gieseke A, Nistico L, Nguyen D, Hayes J, et al. Direct Detection of Bacterial Biofilms on the Middle-Ear Mucosa of Children With Chronic Otitis Media. JAMA. 2006 Jul 12;296(2):202–11.
10. Hoa M, Syamal M, Sachdeva L, Berk R, Coticchia J. Demonstration of nasopharyngeal and middle ear mucosal biofilms in an animal model of acute otitis media. Ann Otol Rhinol Laryngol. 2009 Apr;118(4):292–8.
11. Post JC, Hiller NL, Nistico L, Stoodley P, Ehrlich GD. The role of biofilms in otolaryngologic infections: update 2007. Curr Opin Otolaryngol Head Neck Surg. 2007 Oct;15(5):347–51.
12. Sanderson AR, Leid JG, Hunsaker D. Bacterial biofilms on the sinus mucosa of human subjects with chronic rhinosinusitis. Laryngoscope. 2006 Jul;116(7):1121–6.
13. Marks LR, Parameswaran GI, Hakansson AP. Pneumococcal interactions with epithelial cells are crucial for optimal biofilm formation and colonization in vitro and in vivo. Infect Immun. 2012 Aug;80(8):2744–60.
14. Chao Y, Marks LR, Pettigrew MM, Hakansson AP. Streptococcus pneumoniae biofilm formation and dispersion during colonization and disease. Front Cell Infect Microbiol. 2014;4:194.
15. Croucher NJ, Harris SR, Barquist L, Parkhill J, Bentley SD. A high-resolution view of genome-wide pneumococcal transformation. PLoS Pathog. 2012;8(6):e1002745.
16. Trappetti C, Gualdi L, Meola LD, Jain P, Korir CC, Edmonds P, et al. The impact of the competence quorum sensing system on Streptococcus pneumoniae biofilms varies depending on the experimental model. BMC Microbiology. 2011 Apr 14;11(1):75.
17. Vidal JE, Ludewick HP, Kunkel RM, Zähner D, Klugman KP. The LuxS-Dependent Quorum-Sensing System Regulates Early Biofilm Formation by Streptococcus pneumoniae Strain D39 ▿ . Infect Immun. 2011 Oct;79(10):4050–60.

18. Sanchez CJ, Hurtgen BJ, Lizcano A, Shivshankar P, Cole GT, Orihuela CJ. Biofilm and planktonic pneumococci demonstrate disparate immunoreactivity to human convalescent sera. BMC Microbiol. 2011;11:245.

19. Chao Y, Marks LR, Pettigrew MM, Hakansson AP. Streptococcus pneumoniae biofilm formation and dispersion during colonization and disease. Frontiers in cellular and infection microbiology. 2015;4(January):194.

20. Marks LR, Davidson BA, Knight PR, Hakansson AP. Interkingdom Signaling Induces Streptococcus pneumoniae Biofilm Dispersion and Transition from Asymptomatic Colonization to Disease. mBio. 2013;4(4):e00438-13.

21. Popat R, Crusz SA, Diggle SP. The social behaviours of bacterial pathogens. Br Med Bull. 2008;87:63–75.

22. Li Y-H, Tian X. Quorum Sensing and Bacterial Social Interactions in Biofilms. Sensors (Basel). 2012;12(2):2519–2538.

23. Cook LC, Federle MJ. Peptide pheromone signaling in Streptococcus and Enterococcus. FEMS Microbiol Rev. 2014 May;38(3):473–92.

24. Havarstein LS, Diep D, Nes I. A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. Molecular Microbiology. 1995 Apr 1;16(2):229–40.

25. de Saizieu A, Gardès C, Flint N, Wagner C, Kamber M, Mitchell TJ, et al. Microarray-based identification of a novel Streptococcus pneumoniae regulon controlled by an autoinduced peptide. J Bacteriol. 2000 Sep;182(17):4696–703.

26. Pestova EV, Håvarstein LS, Morrison DA. Regulation of competence for genetic transformation in Streptococcus pneumoniae by an auto-induced peptide pheromone and a two-component regulatory system. Mol Microbiol. 1996 Aug;21(4):853–62.

27. Bouillaut L, Perchat S, Arold S, Zorrilla S, Slamti L, Henry C, et al. Molecular basis for group-specific activation of the virulence regulator PlcR by PapR heptapeptides. Nucleic Acids Res. 2008 Jun;36(11):3791–801.

28. Fleuchot B, Gitton C, Guillot A, Vidic J, Nicolas P, Besset C, et al. Rgg proteins associated with internalized small hydrophobic peptides: a new quorum-sensing mechanism in streptococci. Molecular Microbiology. 2011 May 1;80(4):1102–19.

29. Chang JC, Federle MJ. PptAB Exports Rgg Quorum-Sensing Peptides in Streptococcus. PLoS ONE. 2016;11(12):e0168461.

30. Besset C, Chambellon E, Fleuchot B, Guillot A, Me C. Rgg-Associated SHP Signaling Peptides Mediate Cross- Talk in Streptococci. 2013;8(6).

31. Cozzone AJ. Protein phosphorylation in prokaryotes. Annu Rev Microbiol. 1988;42:97–125.

32. Lange R, Wagner C, de Saizieu A, Flint N, Molnos J, Stieger M, et al. Domain organization and molecular characterization of 13 two-component systems identified by genome sequencing of Streptococcus pneumoniae. Gene. 1999 Sep 3;237(1):223–34.

33. Throup JP, Koretke KK, Bryant AP, Ingraham KA, Chalker AF, Ge Y, et al. A genomic analysis of two-component signal transduction in Streptococcus pneumoniae. Mol Microbiol. 2000 Feb;35(3):566–76.

34. Rocha-Estrada J, Aceves-Diez AE, Guarneros G, de la Torre M. The RNPP family of quorum-sensing proteins in Gram-positive bacteria. Appl Microbiol Biotechnol. 2010 Jul;87(3):913–23.

35. Bortoni ME, Terra VS, Hinds J, Andrew PW, Yesilkaya H. The pneumococcal response to oxidative stress includes a role for Rgg. Microbiology (Reading, Engl). 2009 Dec;155(Pt 12):4123–34.

36. Hoover SE, Perez AJ, Tsui H-CT, Sinha D, Smiley DL, DiMarchi RD, et al. A new quorum-sensing system (TprA/PhrA) for Streptococcus pneumoniae D39 that regulates a lantibiotic biosynthesis gene cluster. Mol Microbiol. 2015 Apr 13;

37. Kadam A, Eutsey RA, Rosch J, Miao X, Longwell M, Xu W, et al. Promiscuous signaling by a regulatory system unique to the pandemic PMEN1 pneumococcal lineage. PLOS Pathogens. 2017 May 18;13(5):e1006339.

38. Kadioglu A, Weiser JN, Paton JC, Andrew PW. The role of Streptococcus pneumoniae virulence factors in host respiratory colonization and disease. Nat Rev Micro. 2008 Apr;6(4):288–301.

39. Chang JC, Jimenez JC, Federle MJ. Induction of a quorum sensing pathway by environmental signals enhances group A streptococcal resistance to lysozyme. Mol Microbiol. 2015 Sep;97(6):1097–113.

40. Dalton TL, Scott JR. CovS inactivates CovR and is required for growth under conditions of general stress in Streptococcus pyogenes. J Bacteriol. 2004 Jun;186(12):3928–37.

41. Paixão L, Oliveira J, Veríssimo A, Vinga S, Lourenço EC, Ventura MR, et al. Host glycan sugar-specific pathways in Streptococcus pneumonia: galactose as a key sugar in colonisation and infection. PLoS ONE. 2015;10(3):e0121042.

42. Terra VS, Homer KA, Rao SG, Andrew PW, Yesilkaya H. Characterization of novel beta-galactosidase activity that contributes to glycoprotein degradation and virulence in Streptococcus pneumoniae. Infect Immun. 2010 Jan;78(1):348–57.

43. Carvalho SM, Kloosterman TG, Kuipers OP, Neves AR. CcpA Ensures Optimal Metabolic Fitness of Streptococcus pneumoniae. PLoS ONE. 2011 Oct 21;6(10):e26707.

44. Hendriksen WT, Bootsma HJ, Estevão S, Hoogenboezem T, de Jong A, de Groot R, et al. CodY of Streptococcus pneumoniae: link between nutritional gene regulation and colonization. J Bacteriol. 2008 Jan;190(2):590–601.

45. Hendriksen WT, Kloosterman TG, Bootsma HJ, Estevão S, De Groot R, Kuipers OP, et al. Site-specific contributions of glutamine-dependent regulator GlnR and GlnR-regulated genes to virulence of Streptococcus pneumoniae. Infect Immun. 2008;76(3):1230–8.

46. Chang JC, LaSarre B, Jimenez JC, Aggarwal C, Federle MJ. Two group A streptococcal peptide pheromones act through opposing Rgg regulators to control biofilm development. PLoS Pathog. 2011 Aug;7(8):e1002190.

47. Cook LC, LaSarre B, Federle MJ. Interspecies communication among commensal and pathogenic streptococci. MBio. 2013;4(4).

48. Lasarre B, Aggarwal C, Federle MJ. Antagonistic Rgg Regulators Mediate Quorum Sensing via Competitive DNA Binding in Streptococcus pyogenes. mBio. 2012;3(6):1–11.

49. Nesin M, Ramirez M, Tomasz A. Capsular transformation of a multidrug-resistant Streptococcus pneumoniae in vivo. J Infect Dis. 1998 Mar;177(3):707–13.

50. Wyres KL, Lambertsen LM, Croucher NJ, McGee L, von Gottberg A, Liñares J, et al. Pneumococcal capsular switching: a historical perspective. J Infect Dis. 2013 Feb 1;207(3):439–49.

51. Hiller NL, Eutsey RA, Powell E, Earl JP, Janto B, Martin DP, et al. Differences in Genotype and Virulence among Four Multidrug-Resistant Streptococcus pneumoniae Isolates Belonging to the PMEN1 Clone. PLoS ONE. 2011 Dec 19;6(12):e28850.

52. Lux T, Nuhn M, Hakenbeck R, Reichmann P. Diversity of Bacteriocins and Activity Spectrum in Streptococcus pneumoniae. Journal of Bacteriology. 2007;189(21):7741–7751.

53. Reichmann P, Hakenbeck R. Allelic variation in a peptide-inducible two-component system of Streptococcus pneumoniae. FEMS Microbiol Lett. 2000 Sep 15;190(2):231–6.

54. Guiral S, Mitchell TJ, Martin B, Claverys J-P. Competence-programmed predation of noncompetent cells in the human pathogen Streptococcus pneumoniae: genetic requirements. Proc Natl Acad Sci USA. 2005 Jun 14;102(24):8710–5.

55. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, et al. Rapid pneumococcal evolution in response to clinical interventions. Science. 2011 Jan 28;331(6016):430–4.

56. Nesin M, Ramirez M, Tomasz A. Capsular transformation of a multidrug-resistant Streptococcus pneumoniae in vivo. The Journal of Infectious Diseases. 1998 Mar;177(3):707–713.

57. Wyres KL, Lambertsen LM, Croucher NJ, McGee L, von Gottberg A, Liñares J, et al. Pneumococcal capsular switching: a historical perspective. J Infect Dis. 2013 Feb 1;207(3):439–49.

58. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotech. 2008 Mar;26(3):317–25.

59. Ogunniyi AD, Giammarinaro P, Paton JC. The genes encoding virulence-associated proteins and the capsule of Streptococcus pneumoniae are upregulated and differentially expressed in vivo. Microbiology. 2002 Jul 1;148(7):2045–53.

60. Orihuela CJ, Radin JN, Sublett JE, Gao G, Kaushal D, Tuomanen EI. Microarray Analysis of Pneumococcal Gene Expression during Invasive Disease. Infection and Immunity. 2004;72(10):5582–5596.

61. Raffel FK, Szelestey BR, Beatty WL, Mason KM. The Haemophilus influenzae Sap transporter mediates bacterium-epithelial cell homeostasis. Infection and Immunity. 2013;81(1):43–54.

62. Pozzi G, Masala L, Iannelli F, Manganelli R, Havarstein LS, Piccoli L, et al. Competence for genetic transformation in encapsulated strains of Streptococcus pneumoniae: two allelic variants of the peptide pheromone. J Bacteriol. 1996 Oct;178(20):6087–90.

63. Zhi X, Abdullah IT, Gazioglu O, Manzoor I, Shafeeq S, Kuipers OP, et al. Rgg-Shp regulators are important for pneumococcal colonization and invasion through their effect on mannose utilization and capsule synthesis. Sci Rep. 2018 Apr 23;8(1):6369.

64. Carvalho SM, Kuipers OP, Neves AR. Environmental and Nutritional Factors That Affect Growth and Metabolism of the Pneumococcal Serotype 2 Strain D39 and Its Nonencapsulated Derivative Strain R6. PLoS ONE. 2013;8(3).

65. Aggarwal C, Jimenez JC, Nanavati D, Federle MJ. Multiple length peptide-pheromone variants produced by Streptococcus pyogenes directly bind Rgg proteins to confer transcriptional regulation. J Biol Chem. 2014 Aug 8;289(32):22427–36.

66. Everett DB, Cornick J, Denis B, Chewapreecha C, Croucher N, Harris S, et al. Genetic characterisation of Malawian pneumococci prior to the roll-out of the PCV13 vaccine using a high-throughput whole genome sequencing approach. PLoS ONE. 2012;7(9):e44250.

67. Saha S, Modak JK, Naziat H, Al-Emran HM, Chowdury M, Islam M, et al. Detection of co-colonization with Streptococcus pneumoniae by algorithmic use of conventional and molecular methods. Vaccine. 2015 Jan 29;33(5):713–8.

68. Rodrigues F, Morales-Aza B, Turner KME, Sikora P, Gould K, Hinds J, et al. Multiple Streptococcus pneumoniae Serotypes in Aural Discharge Samples from Children with Acute Otitis Media with Spontaneous Otorrhea. J Clin Microbiol. 2013 Oct 1;51(10):3409–11.

69. Chi F, Nolte O, Bergmann C, Ip M, Hakenbeck R. Crossing the barrier: evolution and spread of a major class of mosaic pbp2x in Streptococcus pneumoniae, S. mitis and S. oralis. Int J Med Microbiol. 2007 Nov;297(7–8):503–12.

70. Dowson CG, Coffey TJ, Kell C, Whiley RA. Evolution of penicillin resistance in Streptococcus pneumoniae; the role of Streptococcus mitis in the formation of a low affinity PBP2B in S. pneumoniae. Mol Microbiol. 1993 Aug;9(3):635–43.

71. Eutsey RA, Powell E, Dordel J, Salter SJ, Clark TA, Korlach J, et al. Genetic Stabilization of the Drug-Resistant PMEN1 Pneumococcus Lineage by Its Distinctive DpnIII Restriction-Modification System. mBio. 2015 Jul 1;6(3):e00173-15.

72. Johnston C, Polard P, Claverys J-P. The DpnI/DpnII pneumococcal system, defense against foreign attack without compromising genetic exchange. Mob Genet Elements. 2013 Jul 1;3(4):e25582.

73. Valentino MD, McGuire AM, Rosch JW, Bispo PJM, Burnham C, Sanfilippo CM, et al. Unencapsulated Streptococcus pneumoniae from conjunctivitis encode variant traits and belong to a distinct phylogenetic cluster. Nat Commun. 2014;5:5411.

74. Stenz L, Francois P, Whiteson K, Wolz C, Linder P, Schrenzel J. The CodY pleiotropic repressor controls virulence in gram-positive pathogens. FEMS Immunol Med Microbiol. 2011 Jul;62(2):123–39.

75. Aggarwal C, Jimenez JC, Lee H, Chlipala GE, Ratia K, Federle MJ. Identification of Quorum-Sensing Inhibitors Disrupting Signaling between Rgg and Short Hydrophobic Peptides in Streptococci. mBio. 2015 Jul 1;6(3):e00393-15.

76. Aggarwal C, Jimenez JC, Nanavati D, Federle MJ. Multiple length peptide-pheromone variants produced by Streptococcus pyogenes directly bind Rgg proteins to confer transcriptional regulation. J Biol Chem. 2014 Aug 8;289(32):22427–36.

77. Pao SS, Paulsen IT, Saier MH. Major Facilitator Superfamily. Microbiol Mol Biol Rev. 1998 Mar;62(1):1–34.

78. Capra EJ, Perchuk BS, Lubin EA, Ashenberg O, Skerker JM, Laub MT. Systematic dissection and trajectory-scanning mutagenesis of the molecular interface that ensures specificity of two-component signaling pathways. PLoS Genet. 2010 Nov 24;6(11):e1001220.

79. Maricic N, Anderson ES, Opipari AE, Yu EA, Dawid S. Characterization of a Multipeptide Lantibiotic Locus in Streptococcus pneumoniae. mBio. 2016 Mar 2;7(1):e01656-15.

80. Anderson MS, Garcia EC, Cotter PA. Kind discrimination and competitive exclusion mediated by contact-dependent growth inhibition systems shape biofilm community structure. PLoS Pathog. 2014 Apr;10(4):e1004076.

81. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME Suite: tools for motif discovery and searching. Nucl Acids Res. 2009 Jul 1;37(suppl 2):W202–8.

82. Donati C, Hiller NL, Tettelin H, Muzzi A, Croucher NJ, Angiuoli SV, et al. Structure and dynamics of the pan-genome of Streptococcus pneumoniae and closely related species. Genome Biol. 2010;11(10):R107.

83. Hiller NL, Janto B, Hogg JS, Boissy R, Yu S, Powell E, et al. Comparative genomic analyses of seventeen Streptococcus pneumoniae strains: insights into the pneumococcal supragenome. J Bacteriol. 2007 Nov;189(22):8186–95.

84. Frazão N, Hiller NL, Powell E, Earl J, Ahmed A, Sá-Leão R, et al. Virulence potential and genome-wide characterization of drug resistant Streptococcus pneumoniae clones selected in vivo by the 7-valent pneumococcal conjugate vaccine. PLoS ONE. 2013;8(9):e74867.

85. Croucher NJ, Finkelstein JA, Pelton SI, Mitchell PK, Lee GM, Parkhill J, et al. Population genomics of post-vaccine changes in pneumococcal epidemiology. Nat Genet. 2013 Jun;45(6):656–63.

86. Keller LE, Thomas JC, Luo X, Nahm MH, McDaniel LS, Robinson DA. Draft Genome Sequences of Five Multilocus Sequence Types of Nonencapsulated Streptococcus pneumoniae. Genome Announc. 2013 Jul 25;1(4).

87. Kilian M, Riley DR, Jensen A, Brüggemann H, Tettelin H. Parallel Evolution of Streptococcus pneumoniae and Streptococcus mitis to Pathogenic and Mutualistic Lifestyles. mBio. 2014 Aug 29;5(4):e01490-14.

88. Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res. 2016 Jul 8;44(W1):W232-235.

89. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucl Acids Res. 2014 Jan 1;42(D1):D206–14.

90. Hiller NL, Eutsey RA, Powell E, Earl JP, Janto B, Martin DP, et al. Differences in Genotype and Virulence among Four Multidrug-Resistant Streptococcus pneumoniae Isolates Belonging to the PMEN1 Clone. PLoS ONE. 2011 Dec 19;6(12):e28850.

91. Al-Bayati FAY, Kahya HFH, Damianou A, Shafeeq S, Kuipers OP, Andrew PW, et al. Pneumococcal galactose catabolism is controlled by multiple regulators acting on pyruvate formate lyase. Sci Rep. 2017 Feb 27;7:43587.

92. Steinmoen H, Knutsen E, Håvarstein LS. 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. 2002;99(11):7681–6.

93. Kadam A, Janto B, Eutsey R, Earl JP, Powell E, Dahlgren ME, et al. Streptococcus pneumoniae Supragenome Hybridization Arrays for Profiling of Genetic Content and Gene Expression. Curr Protoc Microbiol. 2015;36:9D.4.1-9D.4.20.

94. Kim W, Park HK, Hwang W-J, Shin H-S. Simultaneous Detection of Streptococcus pneumoniae, S. mitis, and S. oralis by a Novel Multiplex PCR Assay Targeting the gyrB Gene. J Clin Microbiol. 2013 Mar;51(3):835–40.

95. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.

96. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078–9.

97. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011 Jan;29(1):24–6.

98. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinformatics. 2013 Mar;14(2):178–92.

99. Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. 2016 Jul 8;44(W1):W3–10.

100. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009 May 1;25(9):1189–91.

101. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016 Jul 8;44(W1):W242-245.

102. Novichkov PS, Kazakov AE, Ravcheev DA, Leyn SA, Kovaleva GY, Sutormin RA, et al. RegPrecise 3.0--a resource for genome-scale exploration of transcriptional regulation in bacteria. BMC Genomics. 2013 Nov 1;14:745.

103. den Hengst CD, van Hijum SAFT, Geurts JMW, Nauta A, Kok J, Kuipers OP. The Lactococcus lactis CodY regulon: identification of a conserved cis-regulatory element. J Biol Chem. 2005 Oct 7;280(40):34332–42.

# CHAPTER 3 - Pneumococcal attachment to epithelial cells is enhanced by the secreted peptide VP1 via its control of hyaluronic acid processing

This chapter focuses on the downstream effectors of *vp1* and their role in bacterial attachment. One of the initial phenotypes observed in the *vp1*-deficient strain was its reduced capacity to attach to epithelial cells in culture. This observation conducted us to determine the *vp1* regulon via a whole-genome screening analysis. What we found was revealing; *vp1* controls the expression of a whole set of genes involved in the processing and acquisition of hyaluronic acid. Then, we characterized the strength of the attachment phenotype, asking whether *vp1* and its targets participate in the attachment. I finalize this chapter addressing a long-standing question: does *vp1*, and by extension, its targets, participate in the pneumococcal colonization? To answer that, I worked in collaboration with Dr. Hasan Yesilkaya and Dr. Oscar Gazioglu from University of Leicester, United Kingdom, which explored the contribution of VP1 and PTS-HAL to disease in an animal model of pneumococcal carriage. Part of this work was submitted in the online repository BioRxiv and also submitted for journal publication. Since then, I have supplemented and updated the manuscript that this chapter covers with new data. The original manuscript can be accessed using the following reference:

**Cuevas RA**, Ebrahimi E, Gazioglu O, Yesilkaya H, N. Hiller NL. Pneumococcal attachment to epithelial cells is enhanced by the secreted peptide VP1 via its control of hyaluronic acid processing. bioRxiv 788430; doi: https://doi.org/10.1101/788430

**Author Contributions**: Rolando A. Cuevas contributed to the conception, general design, experimentation, analysis and interpretation, and manuscript preparation; Elnaz Ebrahimi contributed to construction of bacterial clones; Dr. Ozcan Gazioglu contributed to animal experiments and data analysis; Dr. Hasan Yesilkaya contributed to animal

experiments and analysis; Dr. N. Luisa Hiller contributed to design, analysis and interpretation, and manuscript preparation.

## 3.1 Abstract

The Gram-positive bacterium *Streptococcus pneumoniae* (pneumococcus) is an important human pathogen. It can either asymptomatically colonize the nasopharynx or spread to other tissues to cause mild to severe diseases. Nasopharyngeal colonization is a prerequisite for all pneumococcal diseases. We describe a molecular pathway utilized by pneumococcus to adhere to host cells and promote colonization. We demonstrate that the secreted peptide VP1 enhances pneumococcal attachment to epithelial cells. Transcriptional studies reveal that VP1 triggers the expression of operons involved in the transport and metabolism of hyaluronic acid (HA), a glycosaminoglycan present in the host extracellular matrix. Genetic experiments in the pneumococcus reveal that HA processing locus (HAL) promotes attachment. Further, overexpression of HAL genes in the Δ*vp1* background, reveal that the influence of VP1 on attachment is mediated via its effect on HA. In addition, VP1 also enhances degradation of the HA polymer, in a process that depends on the HAL genes. siRNA experiments to knockdown host HA synthesis support this conclusion. In these knockdown cells, attachment of wild-type pneumococci is decreased, and VP1 and HAL genes no longer contribute to the attachment. Finally, experiments in a murine model of colonization reveal that VP1 and HAL genes are significant contributors to colonization. Our working model, which combines our previous and current work, is that changes in nutrient availability that influence CodY and Rgg144 lead to changes in the levels of VP1. In turn, VP1 controls the expression of a genomic region involved in the transport and metabolism of HA, and these HAL genes promote adherence in an HA-dependent manner. VP1 is encoded by a core gene, which is highly induced *in vivo* and is a major contributor to host adhesion, biofilm development, colonization, and virulence. In conclusion, the VP1 peptide plays a central role in a pathway that connects nutrient availability, population-level signaling, adhesion, biofilm formation, colonization, and virulence.

## 3.2 Introduction

The Gram-positive bacterium *Streptococcus pneumoniae* (also known as the pneumococcus) is an important human pathogen. A recent global study on lower respiratory infections determined that the pneumococcus contributed to morbidity more than all other etiologies combined: it was responsible for an estimated 1.18 million deaths (1,2). This pathogen can either asymptomatically colonize the nasopharynx or spread to other tissues to cause mild to severe diseases. It can spread to the middle ear and sinus, leading to otitis media or sinus infections, and the lungs causing pneumonia (3). It can also disseminate into the bloodstream, brain or heart causing sepsis, meningitis, or heart disease, respectively (4,5). It is well established that nasopharyngeal colonization is a prerequisite for all pneumococcal diseases. In this study, we explore the mechanisms of colonization and virulence.

Pneumococcus secretes many small peptides, which influence colonization, virulence, and adaptation via effects on competence, intra-species competition, and biofilm development (6–18). The **V**irulence **P**eptide 1 (VP1) is a member of a family of peptides with a conserved N-terminal sequence characterized by a double glycine motif, which directs its export into the extracellular milieu via ABC transporters (7,19). The *vp1* gene is widely distributed across pneumococcal strains, as well as encoded in related streptococcal species. The gene encoding VP1 is a virulence determinant and is highly upregulated in the presence of host cells where it promotes biofilm development (7,20). Pneumococcal biofilms have been documented in the nasopharynx, the middle ear, and the sinus (21–24). This mode of growth enhances pneumococcal dissemination and pathogenesis (25,26). Further, biofilms promote pneumococcal colonization in multiple ways: they facilitate immune evasion, increase resistance to antimicrobials, and provide a platform for DNA acquisition (27–29,29–32). In summary, VP1 is a secreted peptide, induced in the presence of human cells, that promotes biofilm development and virulence.

Glycosaminoglycans (GAGs) are major components of the extracellular matrix. GAGs are acidic linear polysaccharides with a repeated disaccharide unit. The disaccharides consist of an amino sugar (*N*-acetylglucosamine or *N*-

acetylgalactosamine) and a uronic acid (glucuronic acid or iduronic acid) or galactose. These polymers are classified into groups based on the nature of the disaccharides, the mode of glycoside bond, and the sulfation levels (33). The main groups are heparin and heparan sulfates (HS), chondroitin and dermatan sulfate (CS), keratan sulfate, and finally hyaluronic acid (HA) (also known as hyaluronate or hyaluronan) (34). Their molecular weights range from 15 to over 100 kDa, and their synthesis takes place in the Golgi apparatus, with the exception of HA, which is synthesized at the membrane (35). GAGs participate in numerous biological processes of the host and the bacteria. In the host, they influence cell adhesion and growth, cell proliferation and differentiation, and tissue formation (for comprehensive reviews, please see Iozzo and Schaefer, 2015; Pomin and Mulloy, 2018; Taylor and Gallo, 2006). In the context of microbial infections, they serve as sources of nutrients and modulators of the immune response (38–41). In addition, many bacteria utilize GAGs to adhere to host cells (42). This includes the pneumococcus, where treatment of lung cells to decrease HS and CS levels reduces pneumococcal adherence (43). In this manner, GAGs are critical components of the bacterial environment during biofilm formation and host interactions.

HA consists of disaccharides of N-acetyl-D-glucosamine (GlcNAc) and D-glucuronic acid (GlcUA). This GAG is present on the apical surface of human bronchial epithelial cells and in the airway mucosa (44,45). It is also abundant in the synovial fluid, skin, umbilical cord, and vitreous body and is the only non-sulfated glycosaminoglycan in the lung (46–48). This unbranched polysaccharide provides mechanical support, activates immunity, and modulates cell proliferation, migration, and intracellular signaling within the host (49–51). Many streptococcal genomes encode multiple operons for HA transport and metabolism. In the pneumococcus, it has been shown that HA processing and internalization is carried out by a unique set of enzymes including a hyaluronidase, a transporter system, and a hydrolase (52,53). The current model suggests that HA is digested into disaccharides by a hyaluronidase (*hysA* or *hylA*), predicted to be covalently linked to the cell wall via an LPxTG motif. The disaccharides are imported into the bacteria and phosphorylated via a phosphoenolpyruvate-dependent phosphotransferase system (PTS) system. The PTS system is composed of general and substrate-specific components. Enzyme I (EI) and the histidine-containing

phosphocarrier protein (HPr) are cytoplasmic subunits, shared by multiple PTS systems. In contrast, Enzyme II (EII) provides specificity to the PTS via substrate recognition. EII has four enzymes: EIIA and EIIB are cytosolic, and EIIC and EIID are the transmembrane components and form a pore. During internalization via the PTS, the HA disaccharides are phosphorylated. These disaccharides are processed by an unsaturated glucuronyl hydrolase (*ugl*) into the monosaccharide components. Finally, the GlcUA can be further processed into glyceraldehyde 3-phosphate and pyruvate by a set of four genes (the isomerase *kduL*, the reductase *kduD*, the kinase *kdgK*, and the aldolase *kdgA*). Operons for the hydrolase, the PTS-EII system, and the GlcUA metabolism enzymes are all adjacent in the genome, and under the control of RegR, a negative regulator located immediately downstream (54). Note that the unsaturated glucuronyl hydrolase not only cleaves HA into GlcNAc and GlcUA, it can also cleave chondroitin sulfate into their component monosaccharides (53,55). The HA pathway has been shown to influence pneumococcal biology. Specifically, *hysA* and the genes in the PTS-EII-encoding operon (EIIB, EIIC, EIID, and *ugl*), are required for pneumococcal growth on human HA, and contribute to pneumococcal biofilm formation and virulence (54,56,57).

In this study, we show that VP1 promotes pneumococcal adherence via activation of HA transport and metabolism. Our experiments in a murine model of pneumococcal carriage demonstrate that both VP1 and HA processing promote colonization. We propose a model where the nutritional status of the host induces expression of the pneumococcal secreted peptide VP1, via CodY and Rgg144 (7,18,58). Signaling by VP1 triggers HA metabolism in the bacteria, which in turn, regulates attachment to host cells and colonization in the upper airways.

## 3.3 Results

### 3.3.1 *vp1* enhances pneumococcal attachment to epithelial cells

We have previously shown that *vp1* plays a role in biofilm development (7). Thus, we investigated whether *vp1* influences attachment to epithelial cells, an early step in biofilm formation. For these studies, we utilized strain PN4595-T23, a representative of the drug-resistant and pandemic PMEN1 pneumococcal lineage (59–62). To show that the results extend beyond one lineage, we also employed TIGR4, a clinical isolate frequently used as a model strain (63). For host cells, we utilized the human lung epithelial cells line A549. These cells have been extensively used to study pneumococcal behavior (20,64,65). Early works demonstrated that A549 cells display features of type II alveolar pneumocytes: they are capable of surfactant production (66–68), they produce cell surface-associated MUC1 and secrete MUC5AC mucin in air-liquid cultures (69,70), and they produce HA. Further, our experiments on adhesion are extended to NIH-3T3 fibroblasts as they express HA and chinchilla middle ear epithelial cells (CMEE) as these primary cells were used in the original characterization of VP1 function (7,71).

We established that the wild-type strain displays strong attachment to A549 cells after one hour of exposure (**Suppl. Figure 3.1A**). Thus, we used this time point to compare adhesion between wild-type, *vp1* deletion mutant (Δ*vp1*) and *vp1* complement strains (Δ*vp1:vp1*) by using spot plating assays and enumerating the total bound bacteria on TSA plates. The Δ*vp1* displays a reduction in attachment compared to the wild-type cells. Furthermore, the wild-type phenotype is rescued in a *vp1* complement strain (Δ*vp1:vp1*) (**Figure 3.1A**). These data suggest that VP1 promotes attachment to the lung epithelia. This phenotype was not specific to A549 cells, as *vp1* also promotes adherence to chinchilla middle ear epithelial cells (CMEE) and NIH-3T3 fibroblasts (**Suppl. Figure 3.1B** and **C**). Similarly, these results extend beyond the PMEN1 lineage, deletion of the

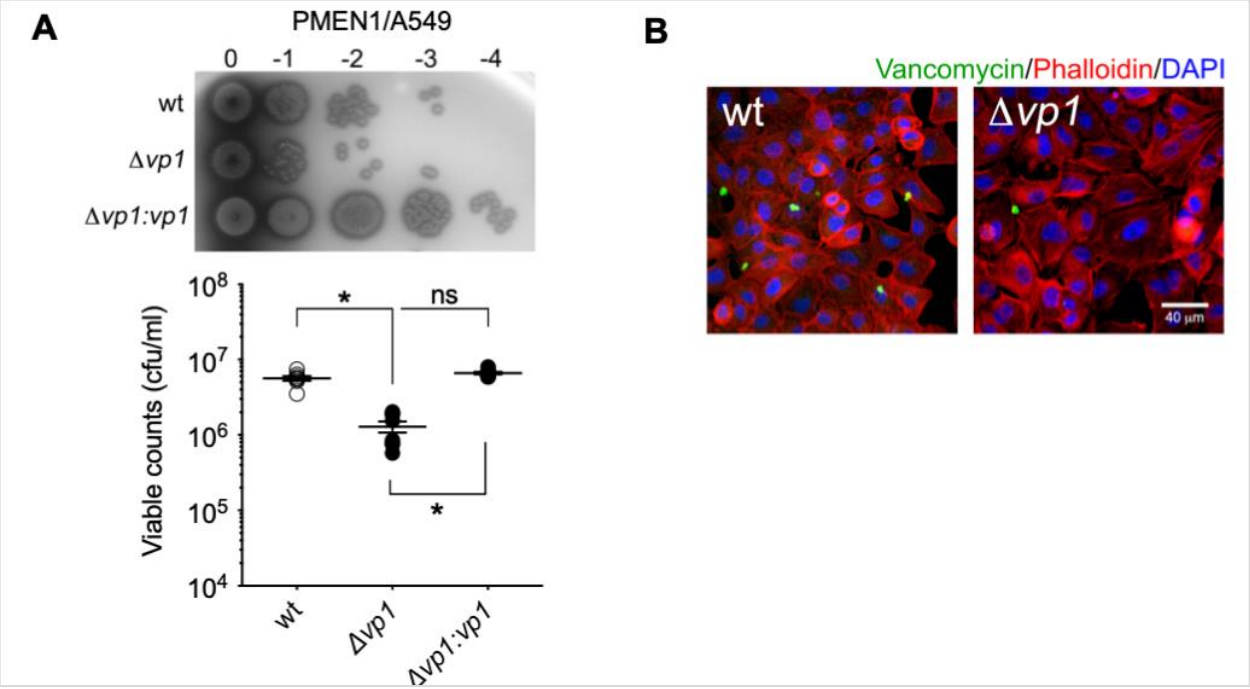



**Figure 3.1 *vp1* enhances pneumococcal attachment to epithelial cells**

**A.** A549 lung epithelial cells were exposed to the wild-type PMEN1 strain PN4595-T23, and isogenic *vp1* mutant and *vp1* complemented strains (Δ*vp1* and the Δ*vp1:vp1,* respectively) for 1h at 37 °C. After washing the unbound bacteria, the remaining cell-bound bacteria were recovered and quantified on TSA plates. Visualization by a spot assay performed with 5 µl of ten-fold dilutions (top panel). Quantification of the total number of bacteria bound to epithelial cells based on counts from TSA plates (bottom panel). **B.** Representative image of wild-type and the Δ*vp1* strains bound to A549 cells. Bacterial cells were stained with BODYPI-FL vancomycin (green), actin was visualized with TRITC-phalloidin (red), and bacterial and host DNA were visualized with DAPI (blue). Data on A represents the mean ± S.E.M of at least six independent experiments. Statistical significance was calculated using unpaired one-way ANOVA analysis with Bonferroni correction, and multiple comparison between samples were performed. * $p < 0.0001$, ns = not significant with $p = 0.080$.

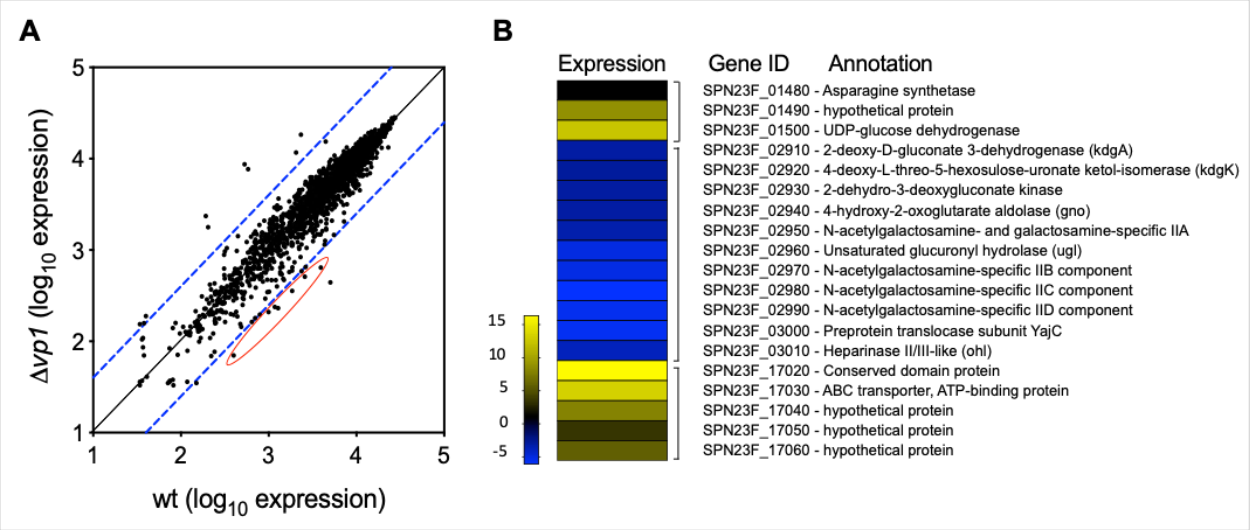



**Figure 3.2 The *vp1* enhance the expression of an operon critical to processing of HA**

**A**. Microarray transcriptome analysis comparing the expression profiles of the wild-type PMEN1 strain PN4595-T23 and *vp1* derivative strain (Δ*vp1*) on CDM supplemented with glucose. Data was analyzed for statistical significance using CyberT as per Student's T-test. The dashed blue lines represent 4-fold up- and down-regulation, respectively. The red circle highlights genes associated with hyaluronic acid metabolism. **B**. Selected genomic regions with 3-fold difference in expression between the wild-type and the Δ*vp1* strains. Targets were sorted by their genomic position on the genome. Operons are indicated by brackets. Color code illustrates expression ratio between the Δ*vp1* strain relative to wild-type strain. Gene IDs correspond to those in PMEN1 reference strain ATCC700669 (GenBank FM211187), and the predicted functions based on Pfam. Summary is listed presented in **Suppl. Table 3.1**. The complete microarray analysis is available online at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE137991

complement strain (**Suppl. Figure 3.1D**). Adherent bacteria were visualized by fluorescence microscopy of A549 cells inoculated with PMEN1 cells and stained with vancomycin-FL (**Figure 3.1B** and **Supple. Figure 3.2**). The operon encoding *vp1* (*vpo*) in a TIGR4 background also displays decreased attachment relative to the wild-type strain and this phenotype is restored to wild-type levels in a *vpo* confocal orthogonal view shows localization at the surface of A549 cells (**Suppl. Figure 3.1E** and **Supple. Figure 3.2**). Finally, the difference in attachment is not the result of variation in growth rate between strains, as growth in DMEM media is the same for wild-type, Δ*vp1,* and Δ*vp1:vp1* strains, as previously reported (7). We conclude that *vp1* promotes pneumococcal attachment to mammalian epithelial cells.

### 3.3.2 *vp1* regulates expression of genes implicated in the metabolism

The VP1 pro-peptide possesses a double-glycine amino acid sequence commonly present on secreted peptides that bind surface-bound histidine kinase receptors from streptococci two-component signal transduction system (7,72). Further, synthetic VP1 binds to the pneumococcal surface (7). Thus, we hypothesized that VP1 signals to producing and neighboring cells and induces expression of genes that participate in the attachment.

To test this hypothesis, we compared the genome-wide expression of the wild-type PN4595-T23 strain and the isogenic Δ*vp1* mutant using the pangenome pneumococcal array (13) (**Figure 3.2A**) Strains were grown in chemically defined media (CDM) supplemented with glucose (CDM-Glu), as *vp1* expression is induced in this media relative to rich media (7). In support of the hypothesis, twenty-eight genes displayed a 3-fold difference in expression between the wild-type strain and the Δ*vp1* strain (**Figure 3.2A** and **Suppl. Table 3.1**). Many of these genes are organized into three genomic regions (**Figure 3.2B**). Microarray results were validated by qRT-PCR for a subset of genes (**Suppl. Table 3.2**). Taken together, we observed a significant number of genes with transcription levels that respond to *vp1*.

Our analysis shows that two neighboring and opposite oriented operons display higher expression in the wild-type strain relative to the Δ*vp1;* these operons have been previously implicated in the processing of HA (53). Together, these two operons encode eleven genes that displayed an average 4.28-fold decrease in expression in the Δ*vp1* relative to the wild-type strain (**Figure 3.2B** and **Figure 3.3**). One operon is on the positive strand and encodes seven genes. In this study, for simplicity, we refer to this as the PTS-HAL operon. A second, single operon displays lower expression in the wild-type strain relative to the Δ*vp1*; they are downstream of another Rgg regulator (Rgg1518, nomenclature based on gene name in D39, SPD_1518), appear annotated as hypothetical genes and their function remains largely unknown.

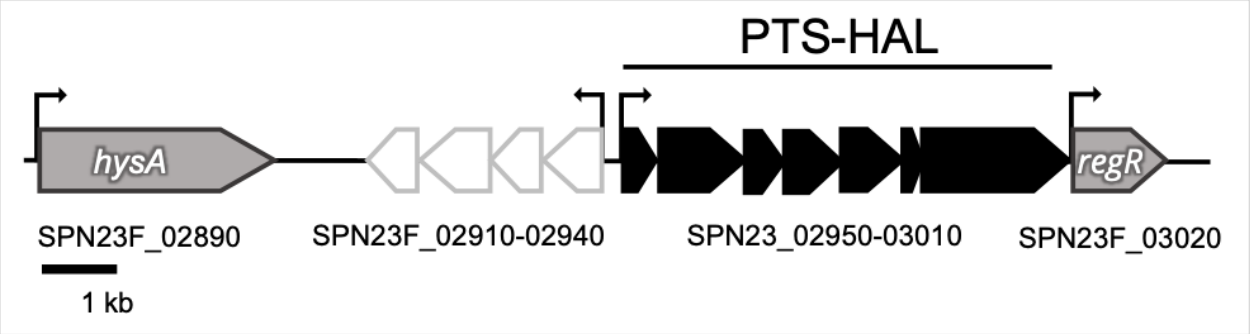



**Figure 3.3 Schematic of the genomic region implicated in the metabolism of hyaluronic acid**

Arrows correspond to predicted open reading frames. Black and white denote predicted operons. Grey corresponds to genes involved in HA metabolism, but not highly controlled by *vp1*. The IDs refer to gene names in the PMEN1 reference strain ATCC700669.

The PTS-HAL encodes the four EIIABCD enzymes that form the PTS importer (**Figure 3.3**). This importer is required for phosphorylation and internalization of HA (EII genes are: SPN23F_02950, and SPN23F_02970 to SPN23F_02990). The operon also contains the *ugl* gene that encodes an unsaturated glucuronyl hydrolase (SPN23F_02960), the preprotein translocase YajC that is a subunit of the bacterial holo-translocon (SPN23F_03000) (73), and finally a predicted heparinase (SPN23F_3010). Our analysis with Phyre2 (74), indicates that the heparinase is a glycosidase that belongs to the superfamily of heparinases type II/III; member of this family bind and degrade heparin and heparin sulfate (75). It is noteworthy, that 60 base-pairs downstream of the end of *hep*, is the negative regulator RegR (SPN23F_3020) (76). Analyses of the region upstream of RegR with the promoter identification programs BDPG and BPROM suggests the presence of an active promoter with the -35 within the *hep* sequence (77,78). The second operon regulated by VP1 is immediately upstream of the PTS-HAL operon on the negative strand, it encodes four enzymes predicted to metabolize glucuronic acid into glyceraldehyde-3 phosphate and pyruvate (SPN23F_02910 to SPN23F_02940). These transcriptional findings suggest that *vp1* triggers the expression of genes implicated in the transport and metabolism of HA.

### 3.3.3 *vp1* controls biofilm development via the HA-processing operon

We compared biofilm biomass and thickness between wild-type and ΔPTS-HAL strains using confocal imaging. The ΔPTS-HAL displayed approximately 20-fold reduction in biofilm biomass and 12.75-fold reduction in thickness when compared to the wild-type strain growing on A549 cells. This reduction is higher compared to the Δ*vp1* biofilm under the same conditions: Δ*vp1* displayed approximately 3-fold reduction in biomass and thickness compared to the wild-type strain. The double mutant Δ*vp1*ΔPTS-HAL also displayed approximately 20-fold reduction in biomass and 9.14-fold reduction in thickness compared to the wild type strain. Furthermore, we demonstrated that the wild-type phenotype is rescued in a PTS-HAL complemented strain (Δ*vp1*ΔPTS-HAL-OE PTS-HAL) (**Suppl. Figure 3.3**). The difference in biofilm growth on cells media is not the result of variation in growth rate, as we demonstrated before. Thus, we conclude that *vp1* controls biofilm capacity on epithelial cells by controlling the expression of the PTS-HAL.

### 3.3.4 *vp1* promotes degradation of the HA polymer via its control of the HA processing operon

Once we established that VP1 stimulates expression of multiple genes involved in HA metabolism, we tested whether VP1 could influence the processing of the HA polymer. To this end, we measured the hyaluronidase activity in pellets and supernatants of the PN4595-T23 strain carrying deletions of *vp1,* PTS-HAL, and various genes involved in HA processing (**Figure 3.4**). In this assay, bacterial pellets or supernatants were mixed with the HA polymer, degradation of HA was measured as a change in optical density at $OD_{400}$ and reported as hyaluronidase concentration (76). As controls, we employed the Δ*hysA*, as this gene encodes for a hyaluronidase and its deletion should result in reduced hyaluronidase relative to the wild-type strain (53,79). As predicted, this strain displayed 2.6-fold lower hyaluronidase levels ($p < 0.0001$ relative to the wild-type strain). We also employed a Δ*regR* mutant, as this gene encodes for a repressor of *hysA* and its deletion should lead to increased hyaluronidase

activity (54). The Δ*regR* strain exhibited elevated hyaluronidase activity ($p = 0.0006$). The Δ*vp1* strain displayed 1.7-fold lower hyaluronidase compared to the wild-type strain ($p < 0.0001$). This defect was rescued in the Δ*vp1:vp1* ($p < 0.0001$ relative to Δ*vp1* strain). The ΔPTS-HAL mutant and the Δ*vp1*-ΔPTS-HAL double mutant showed a 1.4- and 1.5-fold reduction in hyaluronidase relative to the wild-type strain ($p < 0.0005$). This reduction was similar to the Δ*vp1* strain compared to the wild-type strain. The majority of the hyaluronidase activity was captured within the pellets, suggesting that the majority of the degradation enzymes are associated with the cell and not secreted into the supernatant. These data suggest that *vp1* and the PTS-HAL processing and transport operon promote degradation of HA.

To determine whether the ability to break down HA is conserved beyond the PMEN1 strain PN4595-T23, we performed the same set of experiments in a TIGR4 background. The locus in TIGR4 resembles that of the PMEN1 strain, with the addition of a predicted transposon at the end of the operon (see graphic **Suppl. Figure 3.4A**). The results were similar in that the genes encoded by the *vp1* operon (*vpo*) and HA processing and transport operon (PTS-HAL) contributed to the degradation of the HA polymer. In contrast to PN4595-T23, the majority of the TIGR4 activity was localized to the supernatant (**Suppl. Figure 3.4B** and **Suppl. Figure 3.4C**), suggesting differences in the subcellular localization of the hyaluronidase activity between strains.

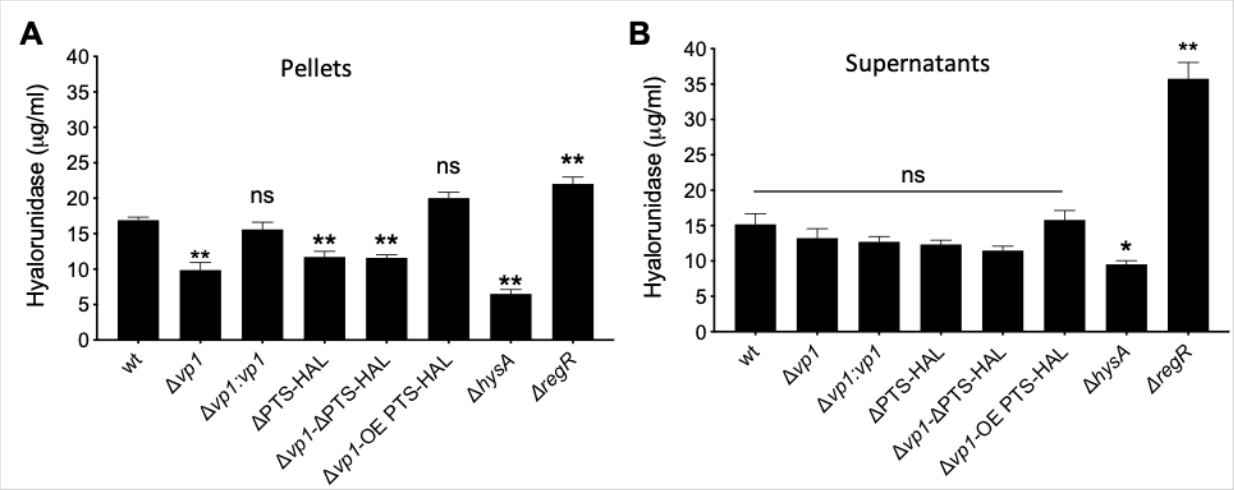



**Figure 3.4 *vp1* promotes degradation of the hyaluronan**

**A** and **B**. Measurements of breakdown of hyaluronan and isogenic deletion mutants for *vp1* (Δ*vp1*), the hyaluronic acid locus PTS operon (ΔPTS-HAL), *vp1* and the PTS operon (Δ*vp1*-ΔPTS-HAL), *hysA* (Δ*hys)*, and *regR* (Δ*regR*), and for a *vp1* complemented strains (Δ*vp1:vp1*), and a PTS overexpressor in the Δ*vp1* background (Δ*vp1*-OE PTS-HAL). Both the pellets and the supernatants of bacteria grown in liquid cultures were collected and assessed for hyaluronidase activity. Data on A and B represents the mean ± S.E.M of at least six independent experiments. Statistical significance was calculated using unpaired one-way ANOVA analysis with Bonferroni correction, and multiple comparison between samples were performed. * $p < 0.05$, ** $p < 0.001$, ns = not significant relative to the wild-type strain. Note that on B, the complemented Δ*vp1:vp1* rescues the hyaluronidase deficiency of the Δ*vp1* mutant ($p < 0.0001$) and the overexpressor Δ*vp1*-OE PTS-HAL rescues the hyaluronidase activity of the mutant Δ*vp1* ($p < 0.0001$). Further, there is no significant difference between Δ*vp1* and ΔPTS-HAL, between Δ*vp1* and Δ*vp1*-ΔPTS-HAL.

Finally, to investigate whether VP1 exerts its influence via genes encoded in the HA processing and transport operon, we tested whether overexpression of the PTS-HAL operon in a Δ*vp1* background could rescue the wild-type phenotype in both TIGR4 and PN4595-T23 strains. To this end, using the Δ*vp1* backgrounds, we replaced the native promoter of PTS-HAL with a constitutive promoter (from *amiA* gene) (strain Δ*vp1*-OE PTS-HAL) (80). These strains displayed HA degradation comparable to the isogenic wild-type strains ($p = 0.21$) (**Figure 3.4A**). These data suggest that *vp1* exerts its effect on HA metabolism via control of the gene products encoded in the PTS-HAL operon.

### 3.3.5 Growth on hyaluronic acid is dependent on *vp1*

Once we established that *vp1* promotes the degradation of HA via controlling the expression of multiple genes involved in the polymer HA metabolism, we predicted that *vp1* is required for pneumococcal growth on HA as solely source of carbon. It was previously shown that the pneumococcus can grow on HA and that genes encoded by the PTS-HAL operon are required for grow (56). To this end, we measured the TIGR4 strain growth in liquid culture on CDM supplemented with HA (5mg/ml) since our preliminary experiments indicated that PN4595-T23 wild-type strain was unable to grow in this condition (data not shown). Bacteria grown on HA showed a growth delay as previously reported due to time taken for pneumococci to adapt to growth in CDM-HA (56). We observed a pronounced growth defect on the Δ*vpo* mutant at 24 h of growth compared to the parent wild-type strain. This growth defect was partially rescued in the Δ*vpo::vpo* strain (**Suppl. Figure 3.5**). These data suggest that HA utilization is enhanced by *vp1*.

### 3.3.6 VP1-controlled attachment to host HA is mediated by molecules involved in HA processing.

We have demonstrated that the *vp1* product enhances attachment of pneumococci to lung epithelial cells (**Figure 3.1**), and promotes transcription and activity of genes in the HA processing and transport operon associated with HA metabolism (**Figures 3.2** and **3.4**). Thus, we hypothesized that *vp1* influences attachment via its effect on the PTS-HAL operon. To test this hypothesis, we performed attachment assays in a strain with a deletion in the PTS-HAL operon (ΔPTS-HAL); it displayed decreased attachment relative to the wild-type strain (**Figure 3.5A**). Moreover, the attachment defect observed in the Δ*vp1* mutant was rescued to wild-type levels in the Δ*vp1*-OE PTS-HAL strain where the genes in the PTS-HAL operon are expressed from a constitutive promoter. These data strongly suggest that the products of *vp1* and PTS-HAL promote adhesion and that the effect of *vp1* is exerted via its influence on PTS-HAL. Further, attachment levels for the double mutant Δ*vp1*ΔPTS-

HAL resembled that of the ΔPTS-HAL strain consistent with these gene products acting within the same pathway (**Suppl. Figures 3.6A** and **Suppl. Figures 3.6B**). The Δ*regR* strain, with increased levels of expression for genes encoded in the PTS-HAL operon, resembled wild-type levels. The Δ*hysA* strain displayed similar attachment to the wild-type strain, suggesting that genes in PTS-HAL are not promoting attachment via indirect effects on HysA (**Suppl. Figures 3.6B**). The attachment of the wild-type and mutant strains were also visualized by confocal microscopy, and representative images are displayed in **Figure 3.5B**. Together, these data demonstrate that the *vp1* promotes attachment via its control of the genes in the PTS-HAL operon, responsible for processing and transport of HA.

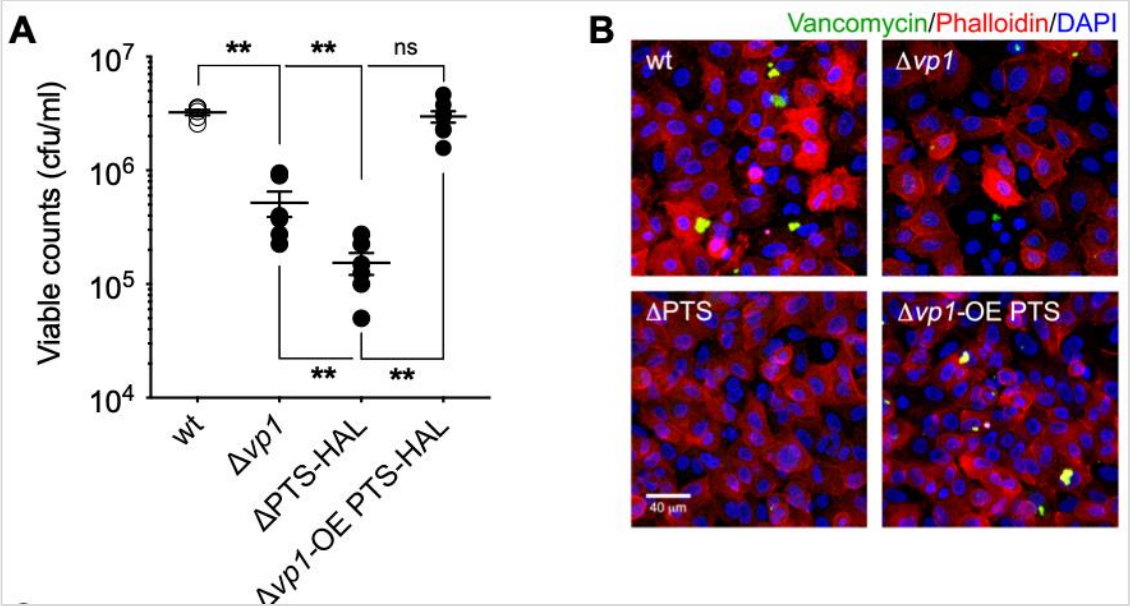



**Figure 3.5 Pneumococcus attaches to host HA in a process controlled by VP1 and mediated by molecules involved in HA processing**

**A.** The wild-type PN4595-T23, Δ*vp1*, ΔPTS-HAL, and Δ*vp1*-OE PTS-HAL strains were tested for attachment to epithelial A549 cells. Cell bound bacteria were enumerated by plating on TSA plates. **B.** Representative image of strains attached to epithelial A549 cells. Bacterial cells were stained with BODYPI-FL vancomycin (green), actin was visualized with TRITC-phalloidin (red), and DNA was visualized with DAPI (blue). Statistical significance was calculated using unpaired one-way ANOVA with Bonferroni correction, and multiple comparison between samples were performed. ** $p < 0.0001$, ns = not significant.

HA production in mammalian cells is controlled by at least three different, membrane-anchored hyaluronan synthases: *HAS1*, *HAS2*, and *HAS3* (35). Previous transcriptomic studies have reported that A549 cells exclusively express *HAS2* and *HAS3*, while *HAS1* remains undetectable (81). Our qPCR analysis corroborates these findings (**Suppl. Figure 3.6C**). To evaluate whether the HA produced by the A549 cells is implicated in pneumococcal attachment, we employed siRNA to knockdown the expression of *HAS2* and *HAS3* (independently and together) on A549 cells, and 48 h later, assessed the attachment strength of the wild-type strain. As a control for treatment, we used a scrambled construct for siRNA experiments, as well as *HAS1* because it has the low expression levels in A549 (**Suppl. Figure 3.6C** and as described by Chow et al., 2010). At the 48h time point, the siRNA treatment reduced levels of HAS2 and HAS3 to below 20% when compared to the scrambled control (**Suppl. Figure 3.6D**). Our attachment assays revealed a significant reduction in pneumococcal attachment to host cells where HAS2 or HAS3 were knockdown, with maximum reduction in the double knockdown cells (**Figure 3.6**). These data suggest that pneumococcus binding is enhanced in the presence of host HA.

If the products of the *vp1* gene and the PTS-HAL operon exert their influence on attachment via HA, a dramatic decrease in host HA levels should reduce the effect of these pneumococcal molecules on host attachment. Consistently, the Δ*vp1* strain displayed similar levels of attachment on host cells independent of the levels of HAS2 HAS3. Further, the ΔPTS-HAL strain also displayed the same levels of attachment in host cells regardless of the levels of HAS2 and HAS3 (**Figure 3.6)**. Moreover, attachment of the ΔPTS-HAL strain resembled levels of a wild-type strain on cells with reduced HAS2 and HAS3, that is, substantially lower than attachment to wild type cells. Finally, a strain overexpressing the PTS-HAL operon in the Δ*vp1* background attached better in host cells that express HA. Thus, pneumococcus binding is enhanced by host HA, in a process that depends on the products of *vp1* and PTS-HAL

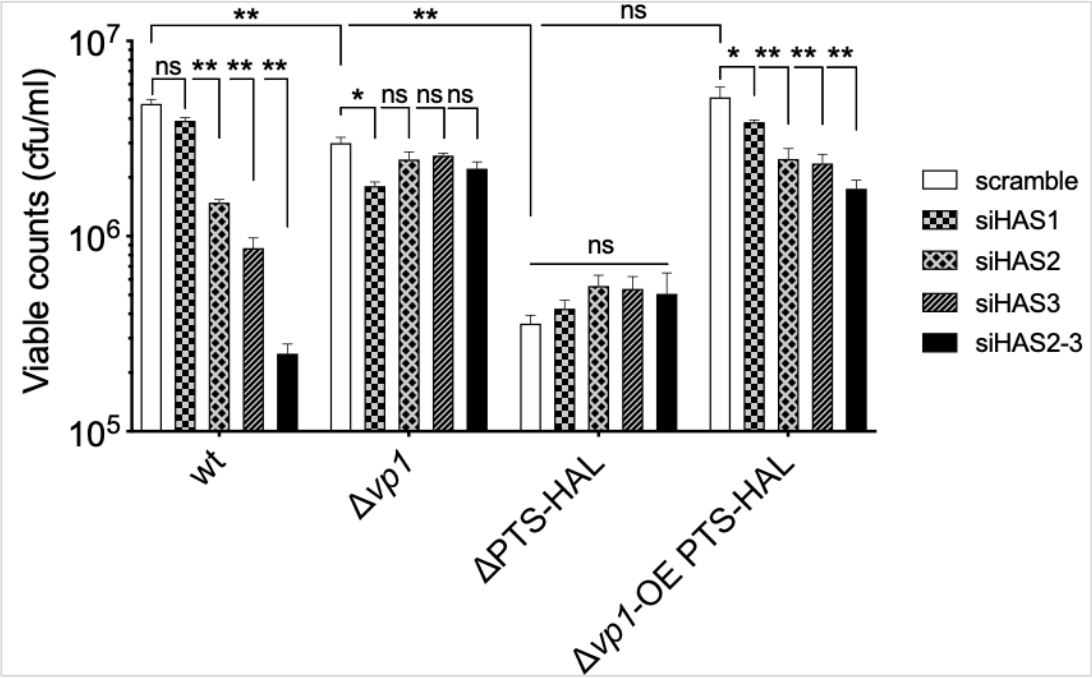



**Figure 3.6 Pneumococcus attachment is mediated via host HA**

Adherence measurements on A549 cells. *HAS1-, HAS2-,* and *HAS3*-knockdown, and the double-knockdown of *HAS2* and *HAS3* (HAS2-3) cells were exposed to wild-type and its derivative Δ*vp1*, ΔPTS-HAL, and Δ*vp1*-OE PTS-HAL strains. siRNA treatment was administered 48 h before addition of bacteria, and adherent bacteria were enumerated by plating on TSA plates. Bound bacteria were enumerated by plating cells on TSA plates. Data represent the mean ± S.E.M of at least four independent experiments. Statistical significance was calculated using two-ways ANOVA analysis with Bonferroni correction, and multiple comparison between samples were performed. * $p < 0.05$, ** $p < 0.0001$, ns = not significant.

The influence of HA in promoting pneumococcal binding is further corroborated by an experiment with exogenous HA. In one experiment, the addition of exogenous HA diminishes binding to A549, consistent with bacteria binding to detached HA (**Suppl. Figure 3.6E**). In a second experiment, we coated 6-well plates with HA from *S. equis* (**Suppl. Figure 3.7A**). A spot assay revealed that pneumococcal attachment was substantially higher on a surface coated with HA, relative to those without HA (**Suppl. Figure 3.7B**). This observation in not restricted the PN4595-T23 strains since the TIGR4 strain also show higher attachment to HA (**Suppl. Figure 3.7C**). As observed on cells, the levels of attachment for Δ*vp1* or ΔPTS-HAL strains were lower than that of the wild-type strain in HA-coated surfaces, and overexpression of the PTS-HAL operon in the Δ*vp1* restored attachment to wild-type levels (**Figure 3.7**). Moreover, these genetic changes had no significant effect on binding on surfaces that were not treated with HA. All together, these data demonstrate that the products of genes involved in the HA processing and transport promote the binding of pneumococcus to host HA in a process regulated by VP1.

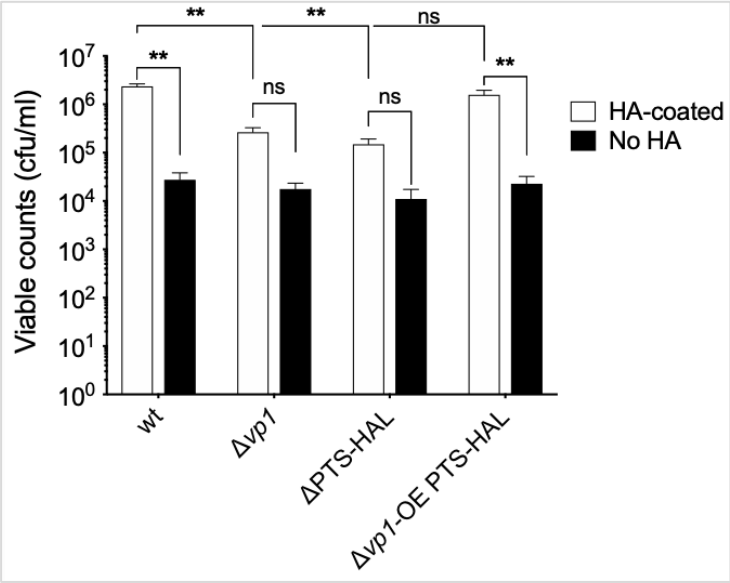



**Figure 3.7 Pneumococcus attach to *S. equis* HA**

The wild-type PN4595-T23 and its derivative Δ*vp1*, ΔPTS-HAL, and Δ*vp1*-OE PTS-HAL strains were assessed for attachment to hyaluronic acid-coated surfaces in FBS-supplemented DMEM media. The total cell-bound bacteria were recovered and quantified by growth on TSA plates. Data represent the mean ± S.E.M of at least four independent experiments. Statistical significance was calculated using two-ways ANOVA analysis with Bonferroni correction, and multiple comparison between samples were performed. ** $p < 0.0001$, ns = not significant. Note that the overexpressor Δ*vp1*-OE PTS-HAL rescues the attachment deficiency of the mutant Δ*vp1* ($p < 0.0001$).

### 3.3.7 VP1 and the HA-processing operon promote colonization

To establish the role of VP1 and of the PTS-HAL operon in pneumococcal survival in the nasopharynx, we employed the murine model of colonization. Cohorts of five mice were inoculated with single strains, using an inoculation dose of 20μl, where the wild-type strain colonizes the nasopharynx, but not the lungs (83,84). The bacterial load was measured two hours and seven days post-inoculation. Enumeration of wild-type strain in the nasopharyngeal lavages two hours post-infection did not show any difference in bacterial load among the strains, indicating the accuracy of infection and the consistency in obtaining nasopharyngeal content (**Figure 3.8A**).

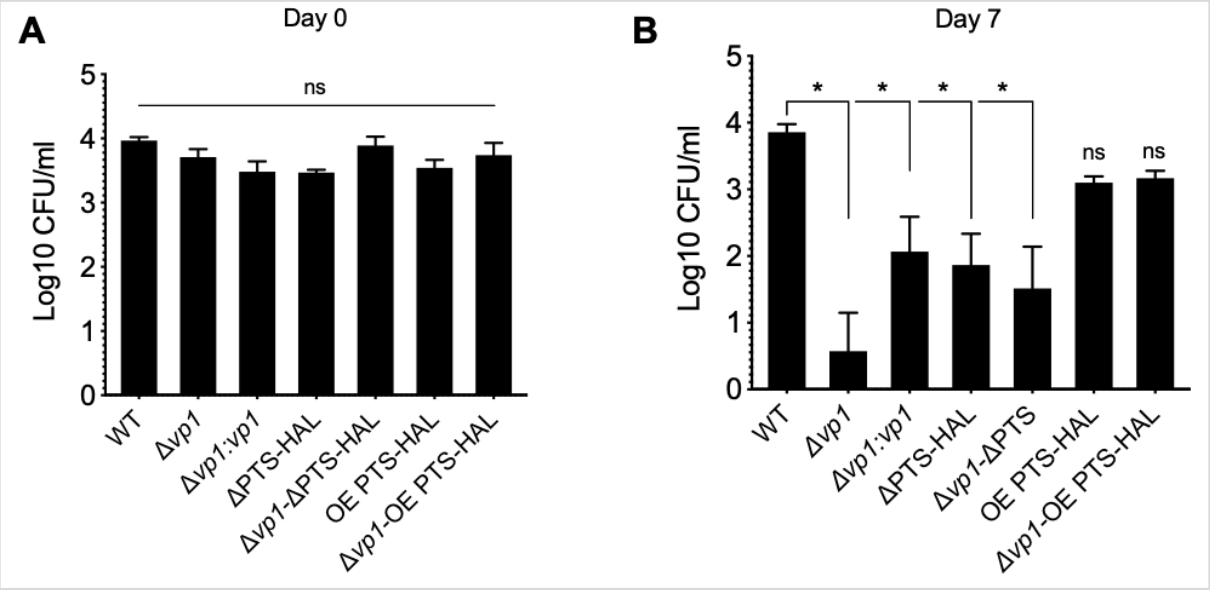



**Figure 3.8 HA processing locus is required for pneumococcal colonization**

CD1 mice were inoculated intranasally with 20 ml PBS containing approximately 5X10$_5$ CFU/mouse. At 2 hours (day 0, A), and 7-days (B) post inoculation, nasopharyngeal contents were collected and pneumococci were enumerated. Each column represents the mean of data from five mice. Error bars show the standard error of the mean. Significant differences in bacterial counts are seen comparing to the wild-type strain using two-way ANOVA and Tukey's multiple comparisons test. * $p$ < 0.0005, ns = not significant. Note that at day 7, the complemented Δ*vp1:vp1* rescues the colonization deficiency of the Δ*vp1* mutant ($p$ = 0.003). Similarly, the overexpressor OE PTS-HAL rescues the colonization deficiency of the mutants Δ*vp1* ($p$ < 0.0001). The ΔPTS-HAL ($p$ = 0.0001) and the double-mutant Δ*vp1*-ΔPTS-HAL ($p$ = 0.0001) both are significantly different from the wild-type strain and the colonization deficiency in the Δ*vp1* strain is greater to the ΔPTS strain (p = 0.0176).

It is important to notice that throughout the course of infection, the number of recovered wild-type pneumococci control remained constant (**Figures 3.8A** and **3.8B**). Samples recovered seven days post-inoculation revealed significant differences in colony counts across the pneumococcal strains (**Figures 3.8B**). Our previous work, in the chinchilla model of pneumococcal disease, established that *vp1* is a virulence determinant based on the reduced mortality and decreased dissemination associated with the Δ*vp1* strain relative to the wild-type strain (7). Akin to the chinchilla model, the Δ*vp1* strain displayed a dramatic decrease in the murine model, where four out of five animals cleared the infection ($p < 0.0001$). Complementation of *vp1* (Δ*vp1:vp1*) partially rescued the wild-type phenotype. The difference between wild-type and complement (Δ*vp1:vp1*) is likely due to changes in the dose driven by variation of the promoter. These data suggest that VP1 contributes to infection in multiple tissue types, as well as in carriage and disease.

By the seventh day post-inoculation, we also detected a significant reduction in colony counts in the ΔPTS-HAL and the double Δ*vp1*-ΔPTS-HAL mutants compared to the wild-type strain ($p < 0.0001$). Moreover, overexpression of the PTS-HAL operon in the Δ*vp1* background significantly increased the recovered pneumococci compare to Δ*vp1* ($p < 0.001$). Yet, overexpression of PTS-HAL in the wild-type background had no effect ($p = 0.542$). The observation that the PTS-HAL operon can compensate for the absence of *vp1* is consistent with the model where VP1 acts upstream of HA processing (**Figure 3.9**). These results strongly suggest that VP1 and HA transport and metabolism play a critical role in pneumococcal colonization of the upper airways.

## 3.4 Discussion

The binding of pneumococci to epithelial cells of the host mucosal surfaces is a crucial step during colonization and precedes dissemination to other tissues (3,85,86). In this study, we reveal a pneumococcal pathway that promotes adherence via host HA. We previously documented that the signaling peptide *vp1* is dramatically induced during host infection, enhances biofilm formation, and is a potent virulence determinant in a chinchilla model of middle ear infection (7). Here, we build on this work, providing mechanistic details that reveal links across signaling, adherence, and colonization. Specifically, we demonstrate that VP1 controls genes involved in the transport and metabolism of the host glycosaminoglycan, HA. Our *in vitro* studies demonstrate that genes in the PTS-HAL operon enhance attachment of the pneumococcus to epithelial cells. Our *in vivo* studies on a murine model of colonization suggest that VP1 and the operon for HA transport and metabolism play critical roles in nasopharyngeal colonization.

VP1 is part of a regulatory network involved in nutritional sensing. Previous work demonstrated that the master nutritional regulator CodY negatively regulates the level of vp1 transcripts, and positively regulated by the transcriptional regulator Rgg144 (18,87). The activity of Rgg144 (nomenclature based on gene name in D39, SPD_0144) is regulated by a cognate peptide SHP144 (short hydrophobic peptide 144), which is secreted from cells and internalized into producing and neighboring cells where it binds to Rgg144 (7,18). Rgg144 is a negative regulator of the capsule: it directly binds a predicted promoter upstream of the capsular genes, inhibits the expression of capsular genes, and leads to a reduction in the size of the type 2 capsule (18). The bacterium capsule hinders cell adhesion by concealing bacterial receptors (88). In addition, our work on murine models shows that the Rgg144-SHP144 system serves as a colonization factor in the nasopharynx and contributes to pathogenesis in pneumonia (18). These data are consistent with a model where the Rgg144-SHP144 system plays an essential role in colonization by decreasing the capsule and inducing *vp1*, and that together these events promote adhesion to epithelial cells in the nasopharynx and the lungs. Of note, adhesion experiments in this study were performed in CDM

supplemented with glucose. In this condition, levels of *rgg144* are low, and levels of *vp1* are high (7,18). Thus, while Rgg144-SHP144 likely plays a key role in multiple aspects of colonization in our animal model, it is unlikely to be a major contributor *in vitro* experiments presented in this study (Figure 3.9).

Among the gene set significatively downregulated by *vp1*, is the locus SPN23_17020 to SPN23F_017060, which is located downstream of another transcriptional regulator encoded by SPN23F_17070 (also known as Rgg1518, based on gene name in D39, SPD_1518). This locus has also been found to be negatively regulated by the *vp1* regulator Rgg144 when the strain was grown in galactose as the sole source of carbon (18). Our data suggest that Rgg144 downregulates this locus via the upregulation of *vp1*. Whether Rgg144 may regulates the expression of this locus independent of *vp1* need to be determined. We propose that VP1 is a critical component of a complex cellular response to changes in nutritional status.

Our studies in the murine model of colonization reveal a dramatic defect for the *vp1* deletion mutant, where four out of five mice cleared the infection (**Figure 3.8**). The strain with a deletion of the HA transport and metabolism operon also displayed a decrease in bacterial load, but not of the same magnitude as the Δ*vp1*. These data suggest that: (1) VP1 mediates attachment via control of the HA transport and metabolism operon, and (2) VP1 plays additional roles in colonization, which are mediated by yet unknown factors.

Many pneumococcal surface molecules serve as adhesins, mediating attachment of the bacteria to host cells (40,89,90). A subset of adhesins binds to extracellular matrix components. Pneumococcal virulence protein A and B (PavA and PavB) (91–93) and enolase (94,95) bind to fibronectin and plasminogen. Similarly, the plasmin- and fibronectin-binding proteins A and B (PfbA and PfbB) bind to the extracellular matrix and are required for adherence to lung and laryngeal epithelial cells (96,97). The choline-binding protein A (CbpA or PspC) binds vitronectin (98), and PsrP binds to keratin (99). Pneumococcus also encodes surface-exposed glycosidases, which have been shown to contribute to adhesion not only via modification of the host surface that reveal receptors but also via direct binding (100). The glycosidase NanA (64) and the β-

galactosidase BgaA (101), both process GAGs. NanA releases terminal sialic acid residues and BgaA releases terminal galactose residues from glycoconjugates. However, these enzymes promote attachment to host cells in a manner independent of their enzymatic activity. This observation is particularly evident for BgaA, where binding is mediated via a carbohydrate-binding module (102). Thus, the pneumococcal can employ glycosidases in host attachment.

The processing of host hyaluronic acid chains starts with the binding and breaking of the chain by the enzyme HysA (53). This enzymes is present in *S. oralis* (103) and *S. agalactiae* (104,105), but absent in other streptococci like *S. mitis* and *S. pseudopneumoniae* (106,107). While the function in related-species is largely unknown, in the pneumococcus it has been associated with virulence (54,56) and survival to macrophages clearance and depress proinflammatory cytokines production (108). In *S. pneumoniae* cultures, it has been reported that this surface-associated enzyme is found in both the culture and the cell-associated fractions (104). This may suggest that at least part of the enzyme is released by the bacteria to initiate the hyaluronan processing to facilitate the bacterial colonization or evade host immune surveillance. Our studies showed a differential distribution between two clinical strains; while the majority of HysA activity was detected in the supernatant of the TIGR4 strain, it remains largely associated to the pneumococcal cell in the PN4959-T23 strains. In this study, deletion of the surface exposed HysA strongly influenced HA processing (**Figure 3.4**), but not attachment (**Suppl. Figure 3.5B**). Thus, we deduce that the glycosidase HysA does not mediate the VP1-regulated defect in adhesion.

The data presented in this study strongly suggest that the transport and processing of host HA plays an essential role in adhesion. However, we have not yet established the molecular mechanism responsible for this outcome. Levels of regulators, carbohydrates, and transporters are interconnected in the bacterial cells via highly complex networks. Regulators control gene expression, including levels of transport systems and biosynthesis enzymes. Carbohydrates can function as substrates, and can positively or negatively influence the activity of regulators. Transporters not only control the import of carbohydrates, but the extent and position of phosphorylated residues on the cytoplasmic Hpr in the PTS system also controls CcpA-

dependent carbon catabolite repression (109). CcpA influences the levels of almost one-fifth of pneumococcal genes, including ABC transporters and regulators (109,110). Thus, many molecular mechanisms may explain the connections between transporters and adhesions.

The HA transport and metabolism locus manipulated in this study encodes all the PTS-EII components (EIIA-EIIB, EIIC, and EIID) that specifically mediate HA import. The locus also encodes two additional genes for a putative heparinase and YajC, a component of the machinery associated with the secretion of proteins with N-terminal signal sequences (73), Data presented in **Figure 3.5**, where host HA levels are decreased using siRNA knockdowns, strongly suggest that the PTS and VP1 mediate attachment, via a yet uncharacterized effector, in an HA-dependent manner.

Multiple lines of evidence suggest that GAGs facilitate attachment of multiple bacterial species to host cells. Several groups have explored and advanced the utilization of HA to facilitate attachment of bacteria to abiotic surfaces (49,111,112). A recent report showed that multidrug-resistant *Staphylococcus aureus* and *Pseudomonas aeruginosa* form more robust biofilms on surfaces coated with collagen and HA than on uncoated control surface (42). In addition, levels of GAGs on A549 cells were positively associated with the attachment of several pathogens, including *Staphylococcus aureus*, *Streptococcus pyogenes*, and the pneumococcus (43). Similarly, attachment of multiple bacterial species (both Gram-positive and Gram-negative) was positively associated with GAG levels on corneal epithelial cells HCE-2, well known for expressing a variety of GAGs on their surface (40). In yet another example, HA is the primary attachment site of *M. tuberculosis* in A549 cells (113). While HA has not been directly implicated in the attachment of pneumococcus, previous studies demonstrate that the pneumococcus can utilize HA as a carbon source. In particular, *hysA* and genes in the PTS-EII-encoding operon (EIIB, EIIC, EIID, and *ugl*), are required for pneumococcal growth in media where HA is the only carbohydrate (56). Furthermore, HA supports *in vitro* pneumococcal biofilm formation (57). These studies suggest that multiple bacteria may utilize host HA for attachment or nutrients and that control influences colonization and virulence.

In this study, we reveal a novel pathway involved in the attachment of the pneumococcus to host epithelial cells. Our working model is that Rgg144 and CodY monitor nutrient availability and, both of these transcription factors influence levels of the secreted peptide *vp1*. VP1, in turn, controls a genomic region involved in the transport and metabolism of HA, and these genes promote adherence in an HA-dependent manner. VP1 is encoded by a core gene, is highly expressed during infection, and is a significant contributor to both colonization and virulence. Thus, this molecule is a high-value candidate for the development of anti-pneumococcal therapies.

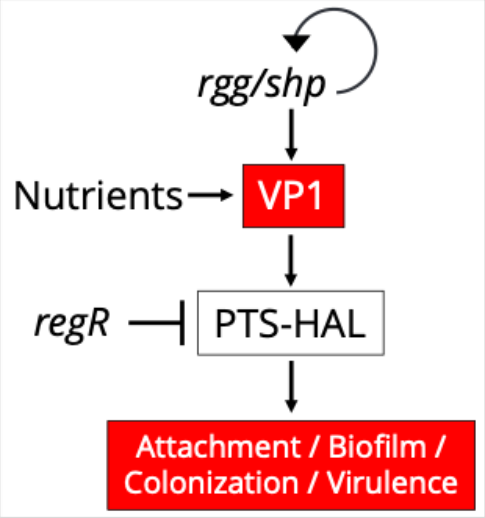



**Figure 3.9 Working model for VP1 regulation and its role in attachment in *S. pneumoniae***

The VP1 is processed and secreted into the extracellular milieu. *vp1* is highly upregulated in the presence of host cells and is induced by an upstream transcription factor, Rgg/Shp. In addition, previous work shows that levels of *vp1* respond to nutrients and to the master regulator CodY. VP1 induces changes in the expression of multiple operons, including two implicated in the processing and acquisition of hyaluronic acid. We propose that VP1 promotes attachment via its influence on the products of genes encoded in the PTS-HAL operon

## 3.5 Materials and Methods

**Bacterial strains and culture conditions.** Wild-type *S. pneumoniae* strain PN4595-T23 (GenBank ABXO01), was selected as a representative of the PMEN1 lineage (59–62). The clinical isolate TIGR4 (The Institute for Genomic Research, Aaberge et al., 1995), was graciously provided by Dr. Jason Rosch. The construction of the *vp1* deletion mutant (Δ*vp1*) and the *vp1* complemented strain (Δ*vp1:*Δ*vp1*) in the PN4595-T23 background was reported previously (7). Strain PN4595-T23 was modified to also generate the PTS-HAL operon deletion (ΔPTS-HAL), the *hysA* deletion (Δ*hysA*), and the *regR* deletion (Δ*regR*). We generated a double mutant *vp1*-PTS (Δ*vp1*-ΔPTS-HAL) and a PTS-HAL overexpressor in the Δ*vp1* and wild-type backgrounds (Δ*vp1*-OE PTS-HAL and OE PTS-HAL, respectively) (**Suppl. Table 3.5**). A similar set of strains was constructed in the TIGR4 background (**Suppl. Table 3.5**). In this set, we generated a deletion mutant strain of the operon encoding *vp1* (*vpo*) and the *vpo* complemented strain (Δ*vpo*::Δ*vpo*). For growth on solid media, strains were streaked onto Trypticase Soy Agar II plates (TSA) containing 5% sheep blood (BD BBL, New Jersey, USA). For growth in liquid culture, colonies from a frozen stock were grown overnight on TSA plates and inoculated into Columbia broth (Remel Microbiology Products, Thermo Fisher Scientific, USA). The media was supplemented with antibiotics as needed. Cultures were incubated at 37°C and 5% $CO_2$ without shaking. Experiments in chemically defined medium (CDM) were performed utilizing a published recipe (110), and glucose was used at a final concentration of 55mM. Growth in CDM was initiated by growing a pre-culture in Columbia broth for 2-3 hours and then diluted to an absorbance of 0.1 at 600 nm, this culture was then grown in CDM-Glu.

**Strain construction and transformation.** To generate the ΔPTS-HAL mutant strain we used site-directed homologous recombination to replace genes SPN23F_02950 to SPN23F_0310 with a kanamycin-resistance gene (*kan*). Likewise, the PTS-HAL operon in the TIGR4 strain (SP_0321 to SP_0327) was replaced with *kan*. The *kan* region was originally amplified from an *rpsL* cassette (80). Gene IDs for PN4595-T23 and TIGR4 are listed in **Suppl. Table 3.3**. The same strategy was used for the single gene replacements, of *regR* (SPN23F_03020) and *hysA* (SPN23F_02890).

To decrease the likelihood of polar effects in all our constructs, the strong artificial transcriptional terminator B1002 was inserted downstream of *kan* (115). Briefly, the transforming DNA was prepared by ligation of the kanamycin gene and promoter, with DNA (one to two kilobases) from the flanking regions upstream and downstream of the PTS-HAL operon. Flanking regions were amplified from the parental strains using either Q5 polymerase or OneTag polymerase (New England Biolabs). The OE PTS-HAL was generated in the Δ*vp1* and wild-type backgrounds by replacing the native PTS operon promoter with an *amiA* promoter and *kan* gene. Assembly of these transforming fragments was achieved by either sticky-end ligation of restriction enzyme-cut PCR products or by Gibson Assembly using NEBuilder HiFi DNA Assembly Cloning Kit. The resulting constructs were transformed into PN4595-T23 or TIGR4. Primers used to generate the constructs are listed in **Suppl. Table 3.4**. PN4595-T23 and TIGR4 strains were transformed with approximately 1 µg of DNA. Liquid cultures were supplemented with 125 µg $ml^{-1}$ of CSP1 (EMRLSKFFRDFILQRKK, for TIGR4) or CSP2 (EMRISRIILDFLFLRKK, for PN4595-T23), at an absorbance of 0.05 at 600 nm (GenScript, NJ, USA), and incubated at 37°C. After 2 hours, the treated cultures were plated on Columbia agar containing the appropriate concentration of antibiotic for selection, spectinomycin, 100 µg $ml^{-1}$, kanamycin 150 µg $ml^{-1}$). Resistant colonies were cultured in Columbia broth. Sequences were confirmed using DNA sequencing of the PCR amplimers (Genewiz, Inc., USA). All strains generated in this study are listed in **Suppl. Table 3.5**.

**RNA purification, reverse transcription (RT) and qPCR.** Bacterial samples were collected on RNALater (Thermofisher), and pellets were lysed with 1x lysis mix containing 2 mg $ml^{-1}$ of proteinase K, 10 mg $ml^{-1}$ of lysozyme, and 20 µg $ml^{-1}$ mutanolysin in TE buffer (10 mM Tris·Cl, 1 mM EDTA, pH 8.0) for 20 min. Total RNA was isolated using the RNeasy (Zymo Research) following manufacturer instructions. Contaminant DNA was removed by incubating total RNA samples with DNase (2U/µl) at 37°C for at least 15 min and then checked by amplification of *gapdh* (no visible band should be observable in RNA only samples). RT reaction was performed using 1 µg of total RNA using SuperscriptVILO kit for 1 h. Five ng of total cDNA was subjected to real-time PCR using PowerUp SYBR Green Master Mix in the ABI 7300 Real-Time PCR

system (Applied Biosystems) according to the manufacturer's instructions. All qRT-PCR amplification was normalized to pneumococcal *gapdh* and expressed as fold change with respect to the wild-type strain. qPCR primers are listed in S4 Table. Primers were obtained from IDT (Integrated DNA Technologies).

**Spot assays and CFU counts:** Bacteria were inoculated from overnight plates into 20 ml of Columbia broth. Liquid cultures were grown statically and monitored at an optical density of 600 nm using Nanodrop 2000c spectrophotometer (Thermo Scientific). Bacterial counts were assessed by streaking 10-fold dilutions on TSA or TSA-antibiotic containing plates. Spot assays were performed with 5 µl of ten-fold dilutions.

**Microarray Analysis:** We utilized the Pneumococcal Supragenome Hybridization Array (SpSGH) to compare gene expression between the wild-type PN4595-T23 strain and the Δ*vp1* as described previously (13). Pneumococcal cultures were collected and normalized to an absorbance of 0.3 at 600 nm. RNA extraction, cDNA preparation, and cDNA labeling were performed as described previously (116). Cyber T was used for data analysis (117). Microarray experiments were performed in triplicate. Genes with at least a 3-fold difference between strains and a Bayesian *p*-value < 0.05 were selected for further analysis. Microarray data for selected genes was confirmed by qPCR. The complete microarray analysis is available online at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE137991

**Preparation of HA-coated surfaces.** Hyaluronic acid from *Streptococcus equi* (Sigma-Aldrich) was bound to tissue culture 24-wells plates following a previously described method with modifications (49). Briefly, twenty-four well plates (VWR) were treated with 0.5 ml of concentrated sulfuric acid for 10 min at 60°C, washed extensively with distilled water, then incubated for 2 h at 37°C with 0.5 ml 5 mg $ml^{-1}$ of hyaluronic acid. Finally, the plates were rinsed with distilled water and allowed to stand for 1h at room temperature. Plates were not dried during the entire procedure. Control plates were treated solely with either sulfuric acid or HA alone. Hyaluronic acid binding was assessed with 0.5 ml of 1% of Alcian Blue 8GX (Sigma-Aldrich) for 10 min at room temperature, then washed extensively with distilled water.

**Hyaluronidase activity measurement.** Hyaluronidase activity was measured as described previously (76,118). Briefly, pneumococcal cultures were grown in Columbia broth to an absorbance of 0.7 at 600 nm. Then the bacterial pellet and supernatant were obtained by centrifugation. Prior to the reaction, bacterial pellets were washed thoroughly with PBS at room temperature then diluted in 250 □l of assay buffer while supernatant samples were diluted 1:1 in assay buffer. Hyaluronic acid isolated from *Streptococcus equi* (Sigma Aldrich) was utilized at a final concentration of 10 ug $ml^{-1}$ in assay buffer (150 mM NaCl, 200 mM sodium acetate, pH 6) for 15 min at 37°C. Stop buffer (2% NaOH, 2.5% cetrimide) was added to stop the reaction, and the absorbance at 400 nm was determined. Commercial hyaluronidase from bovine testes (Sigma-Aldrich) was used as positive control. Assay results were calculated using the standard curve from the positive controls and by comparison to that of a blank comprised of assay buffer solely or without bacterial material.

**Growth assays on CDM-HA.** Chemically defined medium CDM was prepared as described by (119) with the addition of hyaluronic acid (5 mg/ml). Experiments were performed as previously described (56). Control growth was performed on CDM supplemented with glucose (12 mM).

**Mammalian cells and media growth conditions.** Epithelial lung human carcinoma A549 and fibroblast NIH-3T3 cells were propagated and maintained at 37ºC and supplemented of 5% carbon dioxide at pH 7 in DMEM media contained 1% L-glutamine, 5% fetal bovine serum (FBS) without antibiotics. Chinchilla middle ear CMEE cells were maintained as previously reported (7,71).

**siRNA knockdown.** All transfection reagents were provided by Sigma-Aldrich. Subconfluent A549 cultures (20% to 30%) were seed in 24-wells plates and transfected with 3 pmol of a mixture containing pre-designed siRNA. Each mixture consisted of multiple siRNA targeting specifically either *HAS1*, *HAS2*, and *HAS3*. Universal control siRNA was used as a negative control. siRNA was transfected in a final volume of 100 □l of OptiMEM containing 3 μl of MISSION siRNA transfection reagent and no antibiotics. Twenty-four hours later the medium was replaced by DMEM supplemented with fetal bovine serum (FBS). Forty-eight hours after transfection, cells were used in

adherence assays. To verify the reduction in *HAS* expression, A549 cells were lysed using the RNeasy kit and relative *HAS1*, *HAS2*, *HAS3*, and *ACTINB* expression was verified by qPCR, as described above.

**Pneumococcal host cell adherence assay.** A549, NIH-3T3 or CMEE were seed in 24-wells plates or on Lab-Tek II Chamber Slides (Thermo Fisher) at a density of 1 x $10_5$, and maintained without antibiotic until confluence. Before infection, the cell monolayer medium is replaced with fresh medium and then infected with pneumococci resuspended in DMEM supplemented with 1% FBS using a multiplicity of infection (MOI) of 50 bacteria per cell. Infected cells were incubated at 37°C in 5% $CO_2$ for 1 h. Then the media of post-infected cells was aspirated and washed thoroughly three times with PBS to remove unbound bacteria. The total number of adherent bacteria was released with saponin (1% w/v) for 15 minutes at 37 °C and plated on TSA plates followed by colony formation and enumeration. The effect of exogenous HA on attachment interference was performed through the addition of HA from *Streptococcus equi* at concentrations ranging between 0.01 to 1 mg/ml to the A549 monolayers before the addition of the bacteria. Each experiment was repeated at least four times, and results were expressed as mean ± S.E.M.

**Confocal microscopy.** A549 cells were seeded on Nunc Lab-Tek 4-chamber (VWR) and inoculated with pneumococcal strains with a MOI of 50 as described above. After 1 hour of infection, cells were washed 3-times with PBS and fixed with 4% PFA, washed with PBS and stained with BODIPY FL Vancomycin (Thermo Fisher Scientific), Alexa Fluor 568-Phallodin (Thermo Fisher Scientific) following manufacturer's instructions. Finally, cells were covered with Vectashield mounting media with DAPI (Vector Laboratories). Confocal microscopy was performed on the stage of Carl Zeiss LSM-880 META FCS, using 488 nm laser line for the fluorescence tag and 561 nm laser line for the fluorescence Alexa tag. Images were analyzed using ZEN Black software (Carl Zeiss Microscopy, Thornwood, NY) and assembled in Fiji (ImageJ. 2.0.0rc48). Laser intensity and gain were kept the same for all stacks and images.

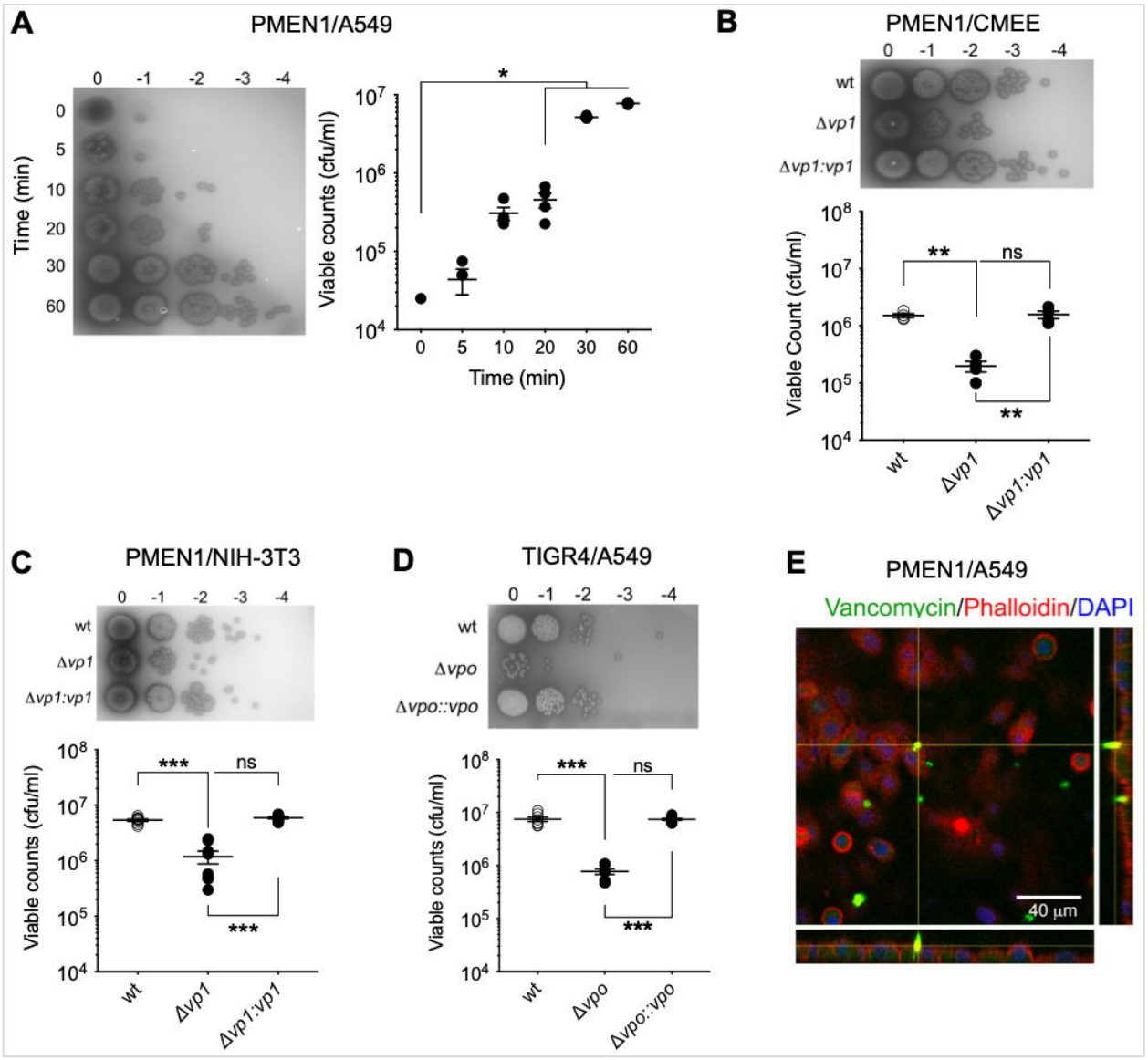
***In vivo* studies:** Colonization ability of pneumococcal strains were tested essentially as previously described (83,84). Nine-week-old CD1 mice were raised in the

pre-clinical facility of the Leicester University. Mice were anesthetized lightly with 2.5% isoflurane over oxygen (1.5 to 2 liter/min), and they were inoculated intranasally with 20 μl PBS, pH 7.0, containing approximately $5X10_5$ CFU/mouse. When inoculating, mice were kept horizontally to prevent the dissemination of inoculum into the lower respiratory tract. A group of mice (n = 5) for each cohort were killed by cervical dislocation at 2 hours and 7 days after infection. After removing the mandible, nasopharyngeal cavity was accessed using a 18G needle through the soft palate, and the content of nasopharynx was collected by injecting 500 □l PBS, pH 7.0. This process was repeated once more using the same lavage fluid in order to optimize the recovery of bacteria. Undiluted and serially diluted lavage fluid was plated on blood agar base plates and CFU/nasopharynx was calculated. The results were analyzed by two-way ANOVA followed by Tukey's multiple comparisons test.

**Ethics statement:** Mouse experiments were done under the UK Home Office approved project (permit no: P7B01C07A) and personal (permit no. I66A9D84D) licenses. The protocols used were approved by both the UK Home Office and the University of Leicester ethics committee. We used anesthetic during the procedure to minimize the suffering. Moreover, the animals were kept in individually ventilated cages, had access to water and feed *ad libido*, and their health were checked regularly in accordance with the legal and institutional requirements.

**Statistical analysis.** Statistical significance was assessed using Student's two-tailed *t*-test for independent samples, one-way ANOVA and Bonferroni's multiple comparisons test for independent samples, and two-way ANOVA and Tukey's multiple comparisons test for animal data. *p*-values less than 0.05 were considered statistically significant.

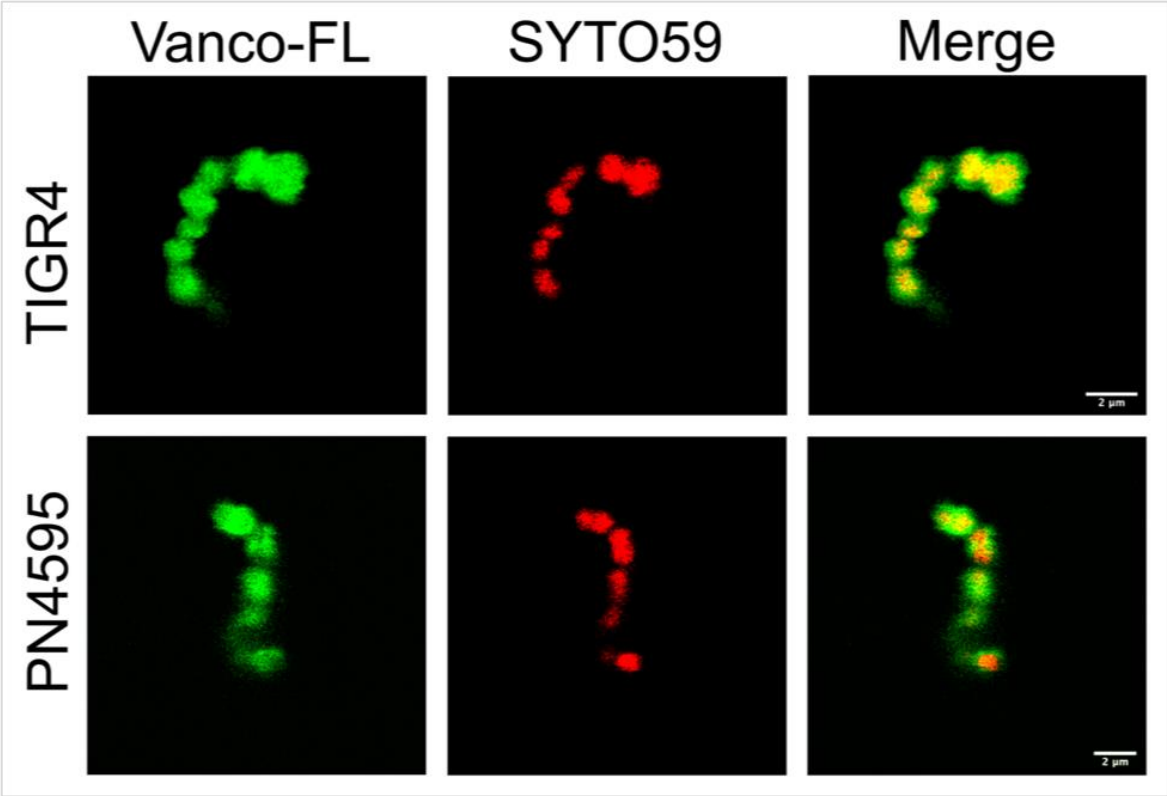
## 3.6 Supplementary Figures




**Supplementary Figure 3.1 *vp1* enhances pneumococcal attachment to epithelial cells**

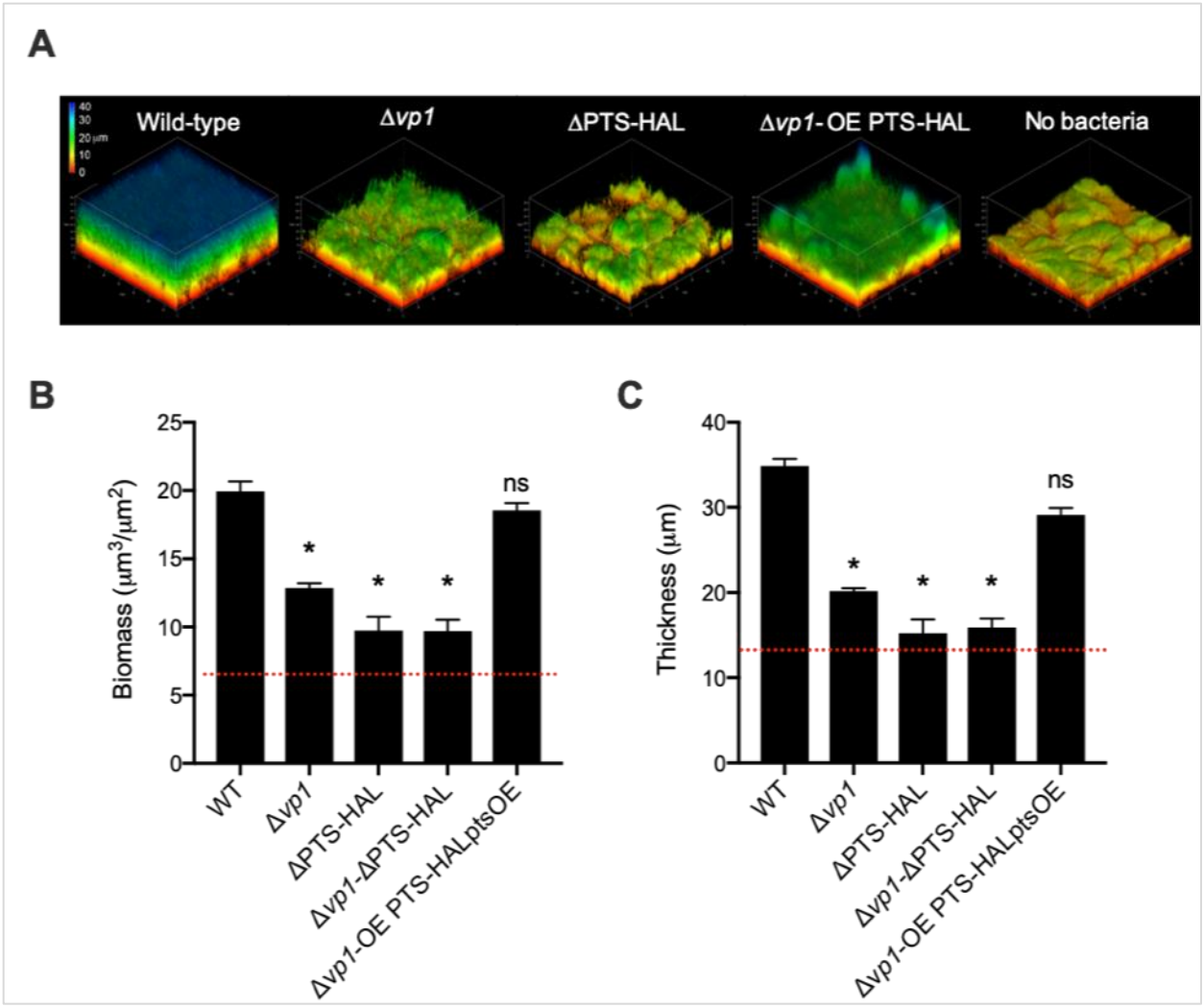
**A.** A549 cells were inoculated with the wild-type PMEN1 strain PN4595-T23. After washing the unbound bacteria, the total cell-bound bacteria were recovered. For visualization, five microliters of ten-fold dilutions were used for spot assays on TSA plates (left panel). For quantification, the total number of bacteria was determined by enumeration on TSA plates (right panel). (**B** and **C**). Attachment assay were performed on CMEE cells (**B**) and NIH-3T3 cells (**C**) inoculated with the strain wild-type PN4595-T23, the *vp1* mutant and *vp1* overexpressor strain (Δ*vp1* and the Δ*vp1:vp1,* respectively). As in A, cell-bound bacteria were recovered and enumerated on TSA plates (bottom panels). (**D**) A549 cells were inoculated with the wild-type TIGR4 strain and its derivative *vpo* mutant and *vpo* overexpressor strains (Δ*vpo* and the Δ*vpo::vpo,* respectively). The total cell-bound bacteria were recovered and enumerated as in A. (E) Representative

orthogonal view image of A549 cells inoculated with wild-type PN4595-T23. Note the extracellular *foci* indicating surface binding of the bacteria. Bacterial cells were stained with BODYPI-FL vancomycin (green), actin was visualized with TRITC-phalloidin (red), and bacterial and human DNA was visualized with DAPI (blue). Data on A, B, C, and D represents the mean ± S.E.M of at least four independent experiments. Statistical significance was calculated using unpaired one-way ANOVA analysis with Bonferroni correction, and multiple comparison between samples were performed. * $p < 0.005$, ** $p < 0.001$, *** $p < 0.0001$, ns = not significant.



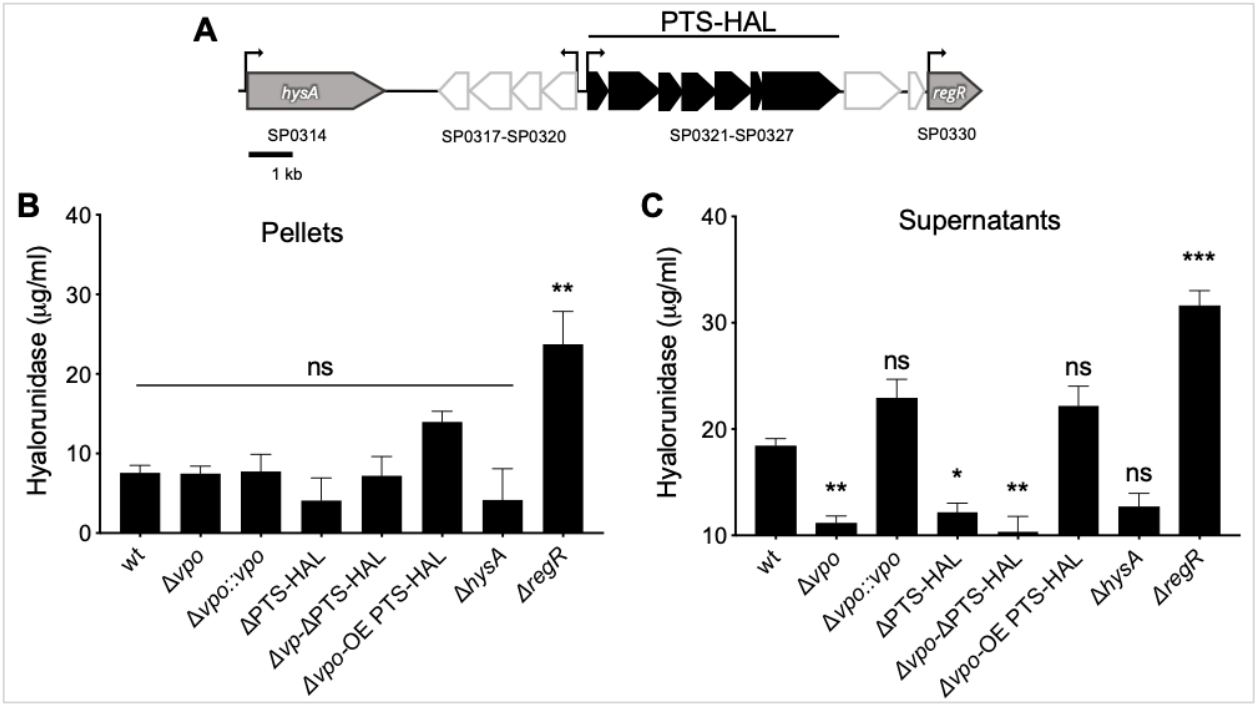

**Supplementary Figure 3.2 Vancomycin binds TIGR4 and PN4595-T23 strains**

Planktonic cultures of strains were fixed and stained with vancomycin-FL (green). Bacterial DNA was stained with SYT59 (red). Note Vancomycin-FL signal encloses the DNA signal on both strains and concentrate in the septum of the bacteria. Scale bars represent 2 μm.



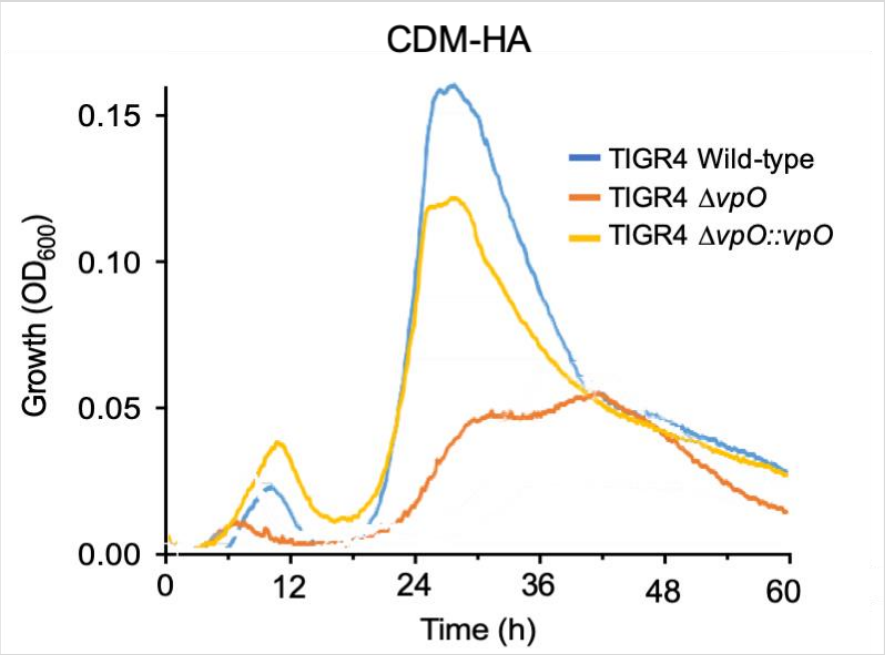

**Supplementary Figure 3.3 *vp1* rescues the PTS-HAL biofilm defect**

**A.** Representative images assembled from confocal micrographs of biofilms growing on A549 lung epithelial cells. A representative image of A549 cells is shown as reference. Axis represent distance in mm. Biofilm biomass (**B**) and average thickness (**C**) were computed from confocal images. Red line on B and C indicates the average biomass and thickness measured from uninfected A549 cells. Data on B and C represent the mean ± S.E.M. of at least three independent experiments. * $p < 0.0001$, ns = not significant.



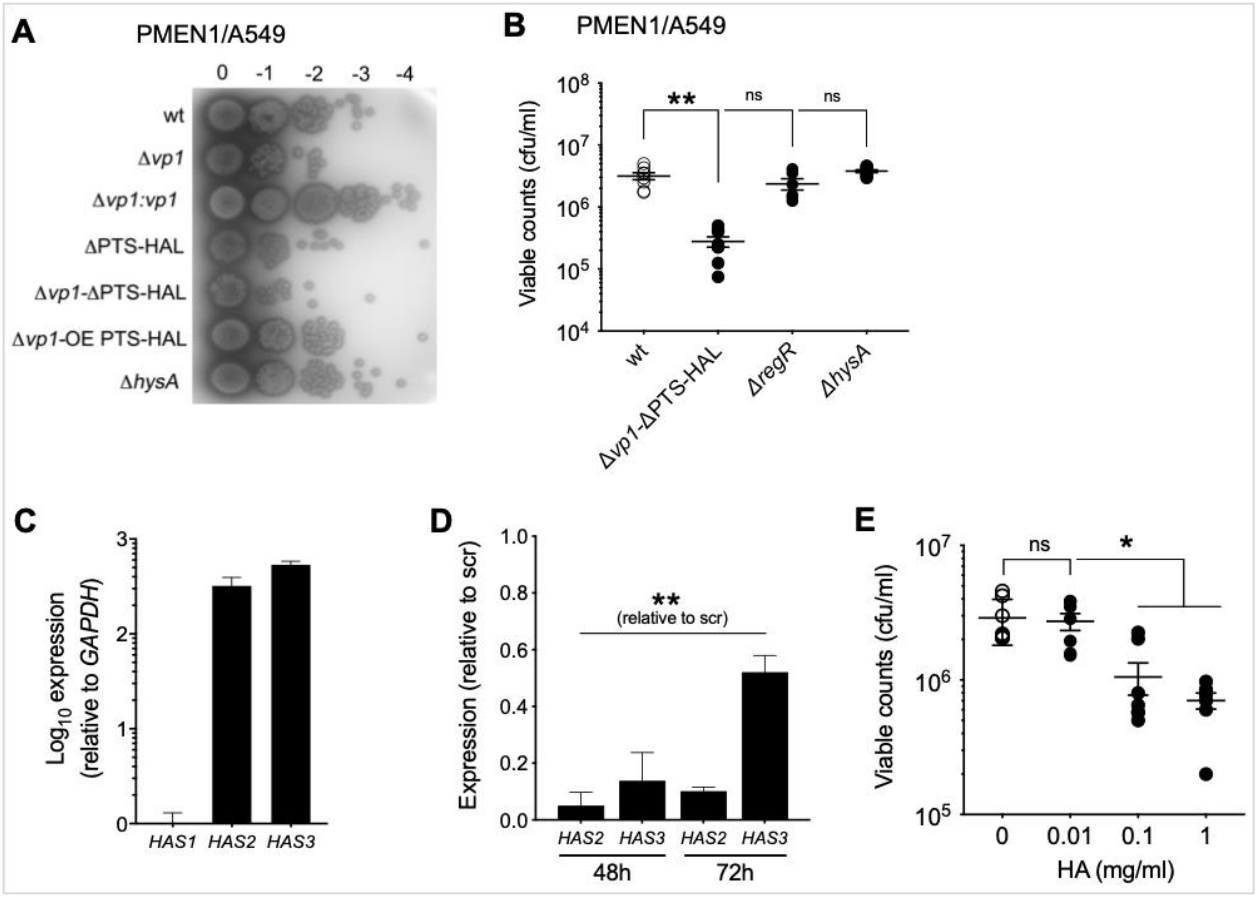

**Supplementary Figure 3.4 The *vp1* product promotes degradation of the hyaluronan polymer via its controls of the operon encoding the PTS-EII system in TIGR4**

**A.** Schematic of the genomic region implicated in the metabolism of hyaluronic acid. Arrows correspond to predicted open reading frames in TIGR4. Black and white denote predicted operons; PTS-HAL has two additional open reading frames at the end of the predicted operon (represented in white). Grey corresponds to genes involved in HA metabolism, but not highly controlled by *vp1*. The IDs refer to gene names in the TIGR4 genome. (**B** and **C**) Measurements of breakdown of hyaluronan by strain TIGR4 and isogenic mutants. Both the supernatants (**B**) and the pellets (**C**) of bacteria grown in liquid cultures were collected and assessed for hyaluronidase activity relative to the wild-type strain. Data on B and C represents the mean ± S.E.M of 8 independent experiments. Statistical significance was calculated using unpaired one-way ANOVA analysis with Bonferroni correction, and multiple comparison between samples were performed. * $p < 0.05$, ** $p < 0.05$, *** $p < 0.0001$ ns = not significant relative to the wild-type strain. Note that on C, the Δ*vpo:vpo* rescues the hyaluronidase deficiency of the Δ*vpo* mutant ($p < 0.0001$) and the overexpressor Δ*vpo*-OE PTS-HAL rescues the hyaluronidase activity of the mutant Δ*vpo* ($p < 0.0001$). Further, Δ*vpo*-OE PTS-HAL is also significantly different from the double-mutant Δ*vpo*-ΔPTS-HAL ($p < 0.0001$).



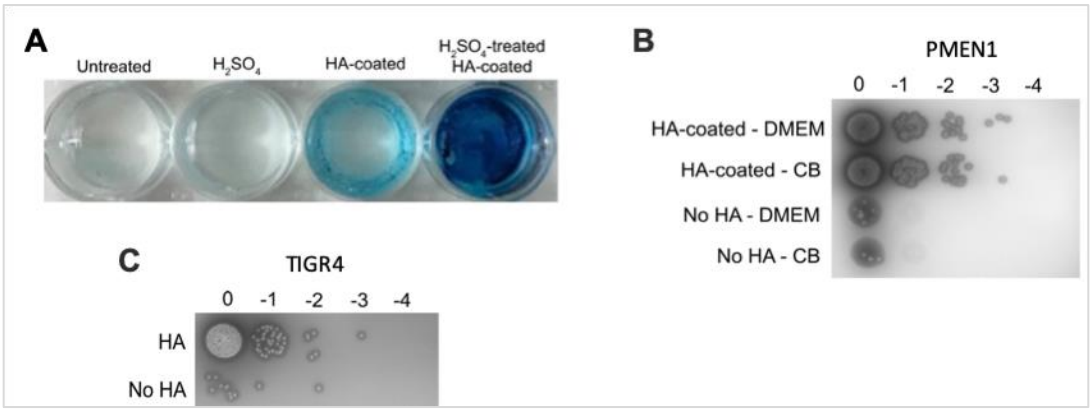

**Supplementary Figure 3.5 Growth on hyaluronic acid is dependent on *vp***

*S. pneumoniae* strain TIGR4 was grown in chemically defined medium supplemented with 5 mg/ml of hyaluronic acid. $OD_{600}$ values are the optical density and represent state of growth. Only one experiment is show in graph.




**Supplementary Figure 3.6 Pneumococci attachment and expression of HA synthases in A549 cells**

**A.** Spot assay to access attachment to host cells. The wild-type PN4595-T23, and its derivative strains Δ*vp1,* Δ*vp1*:*vp1,* ΔPTS-HAL, Δ*vp1*-ΔPTS-HAL, Δ*vp1*-OE PTS and Δ*hysA*, were assessed for attachment to A549 cells using a spot assay. After washing the unbound bacteria, the total cell-bound bacteria were recovered and plated on TSA plates. **B.** Quantification of binding, where the total number of the Δ*vp1*-ΔPTS-HAL, Δ*regR and* Δ*hysA,* bound to epithelial cells was enumerated on TSA plates. **C.** Measurement of the expression levels of hyaluronic acid synthases-encoding genes *HAS1*, *HAS2*, and *HAS3* in A549 cells at two time points where levels are compared to treatment with a scrambled control. **D.** Measurement of the expression levels of hyaluronic acid synthases-encoding genes *HAS2* and *HAS3* after treatment with specific siRNA. **E.** Quantification of bacterial binding to A549 pre-incubated with varying concentration of HA. Data on B to E represents the mean ± S.E.M of at least four independent experiments. Statistical significance was calculated using unpaired one-way ANOVA (B, D, and E) analysis with Bonferroni correction, and multiple comparison between samples were performed. * $p < 0.005$, ** $p < 0.0001$, ns = not significant.




**Supplementary Figure 3.7 *vp1* mediates attachment to hyaluronic acid of *S. equis***

**A.** Hyaluronic acid (HA) coating of 6 well plates under listed treatments, visualized by staining with Alcian Blue 8GX. **B.** The PMEN1 strain PN4595-T23 (top panel) was assessed for attachment to hyaluronic acid-coated surfaces in FBS-supplemented DMEM media (A549 growth media) or Columbia broth (bacterial growth media). **C.** The TIGR4 strain was assessed for attachment to hyaluronic acid-coated surfaces Columbia broth. On B and C, the total cell-bound bacteria were recovered and tested for growth on TSA plates by plating 5 μl of ten-fold dilutions

## 3.7 Supplementary Tables

**Supplementary Table 3.1** Microarray summary (Hits only)

| ID_REF | C1 wt | C2 wt | C3 wt | C1 Dyp1 | C2 Dyp1 | C3 Dyp1 | Mean wt | Mean Dyp1 | SD wt | SD Dyp1 | Dyp1/wt | p-value | Bonferroni | BH | Gene number | Annotation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| cluster1512 | 374 | 394 | 235 | 1075 | 1175 | 895 | 334 | 1048 | 86 | 142 | 3.13 | 0.002 | 1.00 | 0.03 | SPN23F_01480 | Asparagine synthetase |
| cluster1300 | 136 | 129 | 347 | 1719 | 2074 | 1531 | 204 | 1775 | 124 | 276 | 8.70 | 0.001 | 1.00 | 0.02 | SPN23F_01490 | hypothetical protein |
| cluster1297 | 155 | 124 | 296 | 2454 | 2874 | 1758 | 192 | 2362 | 92 | 563 | 12.31 | 0.003 | 1.00 | 0.03 | SPN23F_01490 | UDP-glucose dehydrogenase (EC 1.1.1.22) |
| cluster892 | 1812 | 1996 | 1744 | 627 | 566 | 596 | 1851 | 596 | 130 | 31 | -3.104 | 0.000 | 0.16 | 0.01 | SPN23F_02910 | 4-hydroxy-2-oxoglutarate aldolase (EC 4.1.3.16) |
| cluster892 | 1812 | 1996 | 1744 | 627 | 566 | 596 | 1851 | 596 | 130 | 31 | -3.10 | 0.000 | 0.16 | 0.01 | SPN23F_02920 | 2-dehydro-3-deoxygluconate kinase (EC 2.7.1.45) |
| cluster895 | 1380 | 1421 | 1588 | 519 | 462 | 478 | 1463 | 486 | 110 | 29 | -3.01 | 0.000 | 0.23 | 0.01 | SPN23F_02930 | predicted 4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase |
| cluster665 | 1082 | 1128 | 1518 | 373 | 405 | 422 | 1243 | 400 | 239 | 25 | -3.11 | 0.004 | 1.00 | 0.04 | SPN23F_02940 | 2-deoxy-D-gluconate 3-dehydrogenase (EC 1.1.1.125) |
| cluster664 | 591 | 587 | 284 | 135 | 137 | 161 | 487 | 144 | 176 | 14 | -3.38 | 0.028 | 1.00 | 0.11 | SPN23F_02950 | PTS system, N-acetylgalactosamine-specific IIA component (EC 2.7.1.69) |
| cluster667 | 3300 | 3059 | 1392 | 519 | 519 | 490 | 2584 | 509 | 1039 | 17 | -5.08 | 0.026 | 1.00 | 0.10 | SPN23F_02960 | Unsaturated glucuronyl hydrolase (EC 3.2.1.-) |
| cluster666 | 2393 | 2338 | 827 | 365 | 361 | 346 | 1853 | 357 | 889 | 10 | -5.19 | 0.043 | 1.00 | 0.14 | SPN23F_02970 | PTS system, N-acetylgalactosamine-specific IIB component (EC 2.7.1.69) |
| cluster668 | 4854 | 5176 | 1760 | 573 | 669 | 687 | 3930 | 643 | 1886 | 61 | -6.11 | 0.039 | 1.00 | 0.13 | SPN23F_02980 | PTS system, N-acetylgalactosamine-specific IIC component (EC 2.7.1.69) |
| cluster670 | 1579 | 1680 | 594 | 214 | 253 | 226 | 1284 | 231 | 600 | 20 | -5.57 | 0.038 | 1.00 | 0.13 | SPN23F_02990 | PTS system, N-acetylgalactosamine-specific IID component (EC 2.7.1.69) |
| cluster669 | 509 | 545 | 147 | 0 | 104 | 106 | 400 | 70 | 220 | 61 | -5.728 | 0.066 | 1.00 | 0.18 | SPN23F_03000 | Preprotein translocase subunit YajC |
| cluster770 | 803 | 863 | 438 | 148 | 199 | 195 | 701 | 180 | 230 | 28 | -3.88 | 0.018 | 1.00 | 0.09 | SPN23F_03010 | Heparinase II/III-like |
| cluster2257b | 730 | 547 | 472 | 8431 | 7704 | 6865 | 583 | 7667 | 133 | 783 | 13.15 | 0.000 | 0.20 | 0.01 | SPN23F_17020 | Conserved domain protein |
| cluster2256 | 670 | 481 | 443 | 9471 | 8475 | 8110 | 531 | 8686 | 121 | 704 | 16.35 | 0.000 | 0.07 | 0.01 | SPN23F_17030 | ABC transporter |
| cluster2259 | 3089 | 1908 | 2009 | 20195 | 17731 | 17128 | 2335 | 18351 | 655 | 1625 | 7.86 | 0.000 | 0.18 | 0.01 | SPN23F_17040 | hypothetical protein |
| cluster2258 | 5367 | 3865 | 4201 | 16341 | 15081 | 14819 | 4478 | 15414 | 788 | 813 | 3.44 | 0.000 | 0.14 | 0.01 | SPN23F_17050 | hypothetical protein |
| cluster2357 | 2522 | 1552 | 2074 | 12805 | 11480 | 10707 | 2049 | 11664 | 486 | 1061 | 5.69 | 0.000 | 0.27 | 0.01 | SPN23F_17060 | hypothetical protein |

**Supplementary Table 3.2** qPCR microarray validation (expressed in log2)

| Gene | Microarray | qPCR |
|---|---|---|
| kdgK | -1.589 | -0.492 |
| KduD | -1.531 | -0.845 |
| ptsA | -1.755 | -0.708 |
| ugl | -2.343 | -0.954 |
| ptsB | -2.375 | -0.235 |
| ptsC | -2.612 | -0.897 |
| ptsD | -2.477 | -0.833 |
| hep | -1.958 | -0.953 |

**Supplementary Table 3.3** Genes analyzed or mentioned in this study

| Gene Name | Gene ID in ATCC700669 | Gene ID in TIGR4 | Anotation |
|---|---|---|---|
| vp1 | SPN23F_01520 | SP_0141 | Signaling peptide |
| rgg | SPN23F_01510 | SP_0142 | Transcriptional regulator of vp1 |
| hysA | SPN23F_02890 | SP_0314 | LPXTG-anchored hyaluronate lyase |
| kdgA | SPN23F_02910 | SP_0317 | bifunctional 4-hydroxy-2-oxoglutarate aldolase |
| kdgK | SPN23F_02920 | SP_0318 | sugar kinase |
| KduD | SPN23F_02930 | SP_0319 | hypothetical protein |
| KduI | SPN23F_02940 | SP_0320 | gluconate 5-dehydrogenase |
| ptsA | SPN23F_02950 | SP_0321 | PTS sugar transporter subunit IIA |
| ugl | SPN23F_02960 | SP_0322 | unsaturated chondroitin disaccharide hydrolase |
| ptsB | SPN23F_02970 | SP_0323 | PTS mannose/fructose/sorbose transporter family subunit IIB |
| ptsC | SPN23F_02980 | SP_0324 | PTS mannose/fructose/sorbose/N-acetylgalactosamine transporter subunit |
| ptsD | SPN23F_02990 | SP_0325 | PTS mannose/fructose/sorbose transporter family subunit IID |
| yajC | SPN23F_03000 | SP_0326 | preprotein translocase subunit YajC |
| hep | SPN23F_03010 | SP_0327 | alginate lyase family protein |
| regR | SPN23F_03020 | SP_0330 | LacI family DNA-binding transcriptional regulator |

## **Supplementary Table 3.4** Oligonucleotides used in this study

| Primer | Nucleotidesequence(5to3) | Comments |
|---|---|---|
| Spec-a fwd | ATATATCTCGAGGGATCCCCCGTTTGATTTTTAATG | PN4595-T23 only |
| Spec-a rev | ATATATGCTAGCGCGAAAAACCCCGCCGAAGCGG | PN4595-T23 only |
| Spec-b fwd | ATATATGCTAGCGGATCCCCCGTTTGATTTTTAATG | PN4595-T23 only |
| Spec-b rev | ATATATCTCGAGGCGAAAAAACCCCGCCGAAGCG | PN4595-T23 only |
| Kan fwd | CAAAATAGTGAGGAGTCTATGCCTATTCCAGAGGA | PN4595-T23 only |
| Kan rev | ATTCAATATTCTCTCCGCAAAAAACCCCGCTTC | PN4595-T23 only |
| vp1AKO1 | TGTGACCTTTGATGAGTTTGGG | PN4595-T23 only |
| vp1AKO2 | ATATATCTCGAGCCATTGTTTCTGCAA | PN4595-T23 only |
| vp1AKO3 | ATATATGCTAGCCCGCCGTCGTTATTAATATCCAAA | PN4595-T23 only |
| vp1AKO4 | CTCCCAGAAAGATTTGTAGAACTTG | PN4595-T23 only |
| vp1AComp1 | TGGATCGCTTTGCTCATCAAG | PN4595-T23 only |
| vp1AComp2 | TGGAATAGGCATAGACTCCTCACTATTTTGATTAGT | PN4595-T23 only |
| vp1AComp3 | GCGGGGTTTTTTGCGGAGAGAATATTGAATGGACT | PN4595-T23 only |
| vp1AComp4 | CCAAGCGTTGAAATCCACTG | PN4595-T23 only |
| F1_F_hylko-kan | TTGGGAAATGGGAATGGTAGAG | Works on PN4595-T23 and TIGR4 |
| F1_R_hylko-kan | ATATATGCTAGCGGCCCTCATGACTTCACTTTAT | Works on PN4595-T23 and TIGR4 |
| KAN_F | ATATATGCTAGCTCTATGCCTATTCCAGAGGAAATG | Works on PN4595-T23 and TIGR4 |
| KAN_R-Ter | ATATATCTCGAGGCGAAAAAACCCCGCCGAAGCG | Works on PN4595-T23 and TIGR4 |
| F2_F_hylko-kan | ATATATCTCGAGAAATCCGAAGAAGGCCAAATATC | Works on PN4595-T23 and TIGR4 |
| F2_R_hylko-kan | TCACGTTGTTCAAGCCTTATTG | Works on PN4595-T23 and TIGR4 |
| F1_F_PTSko-kan | GGATATTGGCTCCACCGTATT | Works on PN4595-T23 and TIGR4 |
| F1_R_PTSko-kan | ATATATGCTAGCGGTTTACCGGTTTACTTTATATTCC | Works on PN4595-T23 and TIGR4 |
| F2_F_PTSko-kan | ATATATCTCGAGGAAATGGCTCAGACCTCGAAA | Works on PN4595-T23 and TIGR4 |
| F2_R_PTSko-kan | GCCTGTTGTCCTTCCTCAAA | Works on PN4595-T23 and TIGR4 |
| F1_F_RegRko-kan | GACTCTGGACATGTCTGCATTA | Works on PN4595-T23 and TIGR4 |
| F1_R_RegRko-kan | ATATATGCTAGCTAGCACTGCAAGCGAATCA | Works on PN4595-T23 and TIGR4 |
| F2_F_RegRko-kan | ATATATCTCGAGCTAGGTGGGATTATTTGCCTATGA | Works on PN4595-T23 and TIGR4 |
| F2_R_RegRko-kan | GCCTCCAAACTGACCTCAA | Works on PN4595-T23 and TIGR4 |
| F1_forw_vpoOE | CGGTGTTGCGATGATTGAG | Tigr4 only |
| F1_rev_vpoOE | TGGAATAGGCATAGAGGTCTTTAATGATAAAGAAG | Tigr4 only |
| F2_forw_vpoOE | TTATCATTAAAGACCTCTATGCCTATTCCAGAG | Tigr4 only |
| F2_rev_vpoOE | TTGTAAATTTAACATCTTTTGATTTACGCGGTTTG | Tigr4 only |
| vpoOE_fwd | CGCGTAAATCAAAAGATGTTAAATTTACAATTTGCA | Tigr4 only |
| vpoOE_rev_term | GCGGGGTTTTTTGCGCGGGGAATCCAAGTCACC | Tigr4 only |
| F3_forw_vpoOE_term | GACTTGGATTCCCCGCGCAAAAAACCCCGCTTC | Tigr4 only |
| F3_rev_vpoOE | CTCTACAAAATAGTATAGACAATAACACTATACAAT | Tigr4 only |
| T4-PTS-OE-F1_fwd | CGGCTTCAGCGTAGGATTT | Tigr4 only |
| T4-PTS-OE-F1_rev | TATGCTAGCGGTTTACCGGTTTACTTTATATTCCC | Tigr4 only |
| Kan-pts-OE_fwd | TATGCTAGCTCTATGCCTATTCCAGAGGAAATG | Works on PN4595-T23 and TIGR4 |
| Kan-pts-OE_rev | TATCTCGAGTCTTTGATTTACGCGGTTTGCG | Works on PN4595-T23 and TIGR4 |
| T4-PTS-OE-F2_fwd | TATCTCGAGGGAGAAAAGATGAAAATTGTACTTGT | Tigr4 only |
| T4-PTS-OE-F2_rev | CTCTAGCCTCTCCATCTCCATTTA | Tigr4 only |
| PMEN1-PTS-OE-F1_fw | CGGCATCAAGCACAGTAC | PN4595-T23 only |
| PMEN1-PTS-OE-F1_re | TAGGCATAGAGGTTTACCGGTTTACTTTATATTCC | PN4595-T23 only |
| PMEN1-PTS-OE-F2_fw | AAATCAAAAGGGAGAAAAGATGAAAATTGTAC | PN4595-T23 only |
| PMEN1-PTS-OE-F2_re | CCATTTTGACAAAAGAATGC | PN4595-T23 only |
| qPCR-ptsA-F | CGCCATTCAAGGTTTCTTCTAC | qPCR PN4595-T23 only |
| qPCR-ptsA-R | GCAAAGACTGCTTCCATTAACA | qPCR PN4595-T23 only |
| qPCR-ugl-F | GACGCCAGCTACCTTTGATAAT | qPCR PN4595-T23 only |
| qPCR-ugl-R | TCATAAGCCAACCACAGTTCTC | qPCR PN4595-T23 only |
| qPCR-ptsB-F | CAGTTGCCATGCGTTTCTTC | qPCR PN4595-T23 only |
| qPCR-ptsB-R | AGCGTCCTTCACATCCTTTAC | qPCR PN4595-T23 only |
| qPCR-ptsC-F | GGCTTATATGGGATTCGGTGTT | qPCR PN4595-T23 only |
| qPCR-ptsC-R | CCGCTTCTGGACTGACTTTAC | qPCR PN4595-T23 only |
| qPCR-ptsD-F | CCAATGTGATGTATCCTGCTTTG | qPCR PN4595-T23 only |
| qPCR-ptsD-R | CAGGAAGTGTGGGTTGGTATTA | qPCR PN4595-T23 only |
| qPCR-hep-F | GTTCCCTTGAACGCTTGATAGA | qPCR PN4595-T23 only |
| qPCR-hep-R | CATACCATCGGCTGGATTTGA | qPCR PN4595-T23 only |
| qPCR-regR-F | CCAGAATGGCTACCAGGTAATG | qPCR PN4595-T23 only |
| qPCR-regR-R | TAAAGCCGTCTACTCCCAAGA | qPCR PN4595-T23 only |
| qPCR-kdgA-F | CCGGCGCTACTATTGTCTTTA | qPCR PN4595-T23 only |
| qPCR-kdgA-R | ACAGACATATCCATGGGCTTC | qPCR PN4595-T23 only |
| qPCR-kdgK-F | CGTGGAGAAGAAGGAGAAAGTC | qPCR PN4595-T23 only |
| qPCR-kdgK-R | GCCGGTAACAACAAAGTCAAC | qPCR PN4595-T23 only |
| qPCR-kduD-F | GTCGCTGTCATTCGAGGAAATA | qPCR PN4595-T23 only |
| qPCR-kduD-R | GTCCTGCATACTGATTGGTATAGG | qPCR PN4595-T23 only |
| qPCR-kduI-F | ACCATTAGATGCCACCAGTATC | qPCR PN4595-T23 only |
| qPCR-kduI-R | TACCCAGGGCTTGCAAATTA | qPCR PN4595-T23 only |

**Supplementary Table 3.5** Construct used in this study

| Name | Parental Strain | Comments |
|---|---|---|
| PN4595 wt | | wild-type strain |
| PN4595 D*vp1* | PN4595-T23 | replaced *vp1* with spec cassette |
| PN4595 D*vp1*:*vp1* | PN4595 D*vp1* | inserted *vp1-kan* into native *locus* |
| PN4595 DPTS-HAL | PN4595-T23 | replaced PTS operon with kan cassette |
| PN4595 OE PTS-HAL | PN4595-T23 | replaced native PTS promoter with Kanamycin promoter and kan as selection mark |
| PN4595 D*vp1*-DPTS-HAL | PN4595 D*vp1* | replaced PTS operon with kan cassette |
| PN4595 D*vp1*-OE PTS-HAL | PN4595 D*vp1* | replaced native PTS promoter with Kanamycin promoter and kan as selection mark |
| PN4595 D*hysA* | PN4595-T23 | replaced *hysA* with kan cassette |
| PN4595 D*regR* | PN4595-T23 | replaced *regR* with kan cassette |
| TIGR4 wt | | wild-type strain |
| TIGR4 D*vpo* | TIGR4 | replaced *vp1* operon with spec cassette |
| TIGR4 D*vpo*::*vpo* | TIGR4 D*vpo* | inserted *vp1* operon downstream *bgaA* |
| TIGR4 DPTS-HAL | TIGR4 | replaced PTS operon with kan cassette |
| TIGR4 D*vpo*-DPTS-HAL | TIGR4 D*vpo* | replaced PTS operon with kan cassette |
| TIGR4 D*vpo*-OE PTS-HAL | TIGR4 D*vpo* | replaced native PTS promoter with Kanamycin promoter and kan as selection mark |
| TIGR4 D*hysA* | TIGR4 | replaced *hysA* with kan cassette |
| TIGR4 D*regR* | TIGR4 | replaced *regR* with kan cassette |

## 3.8 References


1. Troeger C, Blacker B, Khalil I. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Infect Dis. 2018;18(11):1191–210.
2. Murdoch DR, Howie SRC. The global burden of lower respiratory infections: making progress, but we need to do better. Lancet Infect Dis. 2018;18(11):1162–3.
3. Weiser JN, Ferreira DM, Paton JC. Streptococcus pneumoniae: transmission, colonization and invasion. Nat Rev Microbiol. 2018 Jun;16(6):355–67.
4. Brown AO, Mann B, Gao G, Hankins JS, Humann J, Giardina J, et al. Streptococcus pneumoniae Translocates into the Myocardium and Forms Unique Microlesions That Disrupt Cardiac Function. PLOS Pathogens. 2014 Sep 18;10(9):e1004383.
5. Loughran AJ, Orihuela CJ, Tuomanen EI. Streptococcus pneumoniae: Invasion and Inflammation. Microbiol Spectr. 2019;7(2).
6. Aggarwal SD, Eutsey R, West-Roberts J, Domenech A, Xu W, Abdullah IT, et al. Function of BriC peptide in the pneumococcal competence and virulence portfolio. PLoS Pathog [Internet]. 2018 Oct 11 [cited 2019 Jul 29];14(10). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6181422/
7. Cuevas RA, Eutsey R, Kadam A, West-Roberts JA, Woolford CA, Mitchell AP, et al. A novel streptococcal cell-cell communication peptide promotes pneumococcal virulence and biofilm formation. Mol Microbiol. 2017 Aug;105(4):554–71.
8. Dawid S, Roche AM, Weiser JN. The blp bacteriocins of Streptococcus pneumoniae mediate intraspecies competition both in vitro and in vivo. Infect Immun. 2007 Jan;75(1):443–51.
9. de Saizieu A, Gardès C, Flint N, Wagner C, Kamber M, Mitchell TJ, et al. Microarray-based identification of a novel Streptococcus pneumoniae regulon controlled by an autoinduced peptide. J Bacteriol. 2000 Sep;182(17):4696–703.
10. Havarstein LS, Coomaraswamy G, Morrison DA. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in Streptococcus pneumoniae. Proc Natl Acad Sci U S A. 1995 Nov 21;92(24):11140–4.
11. Hoover SE, Perez AJ, Tsui H-CT, Sinha D, Smiley DL, DiMarchi RD, et al. A new quorum-sensing system (TprA/PhrA) for Streptococcus pneumoniae D39 that regulates a lantibiotic biosynthesis gene cluster. Mol Microbiol. 2015 Apr 13;
12. Junges R, Salvadori G, Shekhar S, Åmdal HA, Periselneris JN, Chen T, et al. A Quorum-Sensing System That Regulates Streptococcus pneumoniae Biofilm Formation and Surface Polysaccharide Production. mSphere [Internet]. 2017 Sep 13;2(5). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5597970/
13. Kadam A, Eutsey RA, Rosch J, Miao X, Longwell M, Xu W, et al. Promiscuous signaling by a regulatory system unique to the pandemic PMEN1 pneumococcal lineage. PLOS Pathogens. 2017 May 18;13(5):e1006339.
14. Lux T, Nuhn M, Hakenbeck R, Reichmann P. Diversity of Bacteriocins and Activity Spectrum in Streptococcus pneumoniae. J Bacteriol. 2007 Nov;189(21):7741–51.

15. Motib Anfal, Guerreiro Antonio, Al-Bayati Firas, Piletska Elena, Manzoor Irfan, Shafeeq Sulman, et al. Modulation of Quorum Sensing in a Gram-Positive Pathogen by Linear Molecularly Imprinted Polymers with Anti-infective Properties. Angewandte Chemie. 2017 Nov 30;129(52):16782–5.

16. Motib AS, Al-Bayati FAY, Manzoor I, Shafeeq S, Kadam A, Kuipers OP, et al. TprA/PhrA Quorum Sensing System Has a Major Effect on Pneumococcal Survival in Respiratory Tract and Blood, and Its Activity Is Controlled by CcpA and GlnR. Front Cell Infect Microbiol [Internet]. 2019 [cited 2019 Sep 20];9. Available from: https://www.frontiersin.org/articles/10.3389/fcimb.2019.00326/full

17. Tomasz A. Control of the Competent State in Pneumococcus by a Hormone-Like Cell Product: An Example for a New Type of Regulatory Mechanism in Bacteria. Nature. 1965 Oct 9;208(5006):155–9.

18. Zhi X, Abdullah IT, Gazioglu O, Manzoor I, Shafeeq S, Kuipers OP, et al. Rgg-Shp regulators are important for pneumococcal colonization and invasion through their effect on mannose utilization and capsule synthesis. Sci Rep. 2018 Apr 23;8(1):6369.

19. Havarstein LS, Diep DB, Nes IF. A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. Molecular Microbiology. 1995 Apr 1;16(2):229–40.

20. Aprianto R, Slager J, Holsappel S, Veening J-W. High-resolution analysis of the pneumococcal transcriptome under a wide range of infection-relevant conditions. Nucleic Acids Res. 2018 02;46(19):9990–10006.

21. Hall-Stoodley L, Hu FZ, Gieseke A, Nistico L, Nguyen D, Hayes J, et al. Direct Detection of Bacterial Biofilms on the Middle-Ear Mucosa of Children With Chronic Otitis Media. JAMA. 2006 Jul 12;296(2):202–11.

22. Hoa M, Syamal M, Sachdeva L, Berk R, Coticchia J. Demonstration of nasopharyngeal and middle ear mucosal biofilms in an animal model of acute otitis media. Ann Otol Rhinol Laryngol. 2009 Apr;118(4):292–8.

23. Post JC, Hiller NL, Nistico L, Stoodley P, Ehrlich GD. The role of biofilms in otolaryngologic infections: update 2007. Curr Opin Otolaryngol Head Neck Surg. 2007 Oct;15(5):347–51.

24. Sanderson AR, Leid JG, Hunsaker D. Bacterial biofilms on the sinus mucosa of human subjects with chronic rhinosinusitis. Laryngoscope. 2006 Jul;116(7):1121–6.

25. Chao Y, Marks LR, Pettigrew MM, Hakansson AP. Streptococcus pneumoniae biofilm formation and dispersion during colonization and disease. Frontiers in cellular and infection microbiology. 2015;4(January):194.

26. Marks LR, Davidson BA, Knight PR, Hakansson AP. Interkingdom Signaling Induces Streptococcus pneumoniae Biofilm Dispersion and Transition from Asymptomatic Colonization to Disease. mBio. 2013;4(4):e00438-13.

27. Chao Y, Marks LR, Pettigrew MM, Hakansson AP. Streptococcus pneumoniae biofilm formation and dispersion during colonization and disease. Front Cell Infect Microbiol. 2014;4:194.

28. Croucher NJ, Harris SR, Barquist L, Parkhill J, Bentley SD. A high-resolution view of genome-wide pneumococcal transformation. PLoS Pathog. 2012;8(6):e1002745.

29. Marks LR, Parameswaran GI, Hakansson AP. Pneumococcal interactions with epithelial cells are crucial for optimal biofilm formation and colonization in vitro and in vivo. Infect Immun. 2012 Aug;80(8):2744–60.

30. Tonnaer EL, Mylanus EA, Mulder JJ, Curfs JH. Detection of bacteria in healthy middle ears during cochlear implantation. Arch Otolaryngol Head Neck Surg. 2009 Mar;135(3):232–7.

31. Trappetti C, Gualdi L, Meola LD, Jain P, Korir CC, Edmonds P, et al. The impact of the competence quorum sensing system on Streptococcus pneumoniae biofilms varies depending on the experimental model. BMC Microbiology. 2011 Apr 14;11(1):75.

32. Vidal JE, Ludewick HP, Kunkel RM, Zähner D, Klugman KP. The LuxS-Dependent Quorum-Sensing System Regulates Early Biofilm Formation by Streptococcus pneumoniae Strain D39 ▿. Infect Immun. 2011 Oct;79(10):4050–60.

33. Couchman JR, Pataki CA. An Introduction to Proteoglycans and Their Localization. J Histochem Cytochem. 2012 Dec;60(12):885–97.

34. Iozzo RV, Schaefer L. Proteoglycan form and function: A comprehensive nomenclature of proteoglycans. Matrix Biol. 2015 Mar;42:11–55.

35. Weigel PH, DeAngelis PL. Hyaluronan synthases: a decade-plus of novel glycosyltransferases. J Biol Chem. 2007 Dec 21;282(51):36777–81.

36. Pomin VH, Mulloy B. Glycosaminoglycans and Proteoglycans. Pharmaceuticals (Basel) [Internet]. 2018 Feb 27 [cited 2019 Sep 3];11(1). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5874723/

37. Taylor KR, Gallo RL. Glycosaminoglycans and their proteoglycans: host-associated molecular patterns for initiation and modulation of inflammation. FASEB J. 2006 Jan;20(1):9–22.

38. Andre GO, Converso TR, Politano WR, Ferraz LFC, Ribeiro ML, Leite LCC, et al. Role of Streptococcus pneumoniae Proteins in Evasion of Complement-Mediated Immunity. Front Microbiol. 2017;8:224.

39. Frolet C, Beniazza M, Roux L, Gallet B, Noirclerc-Savoye M, Vernet T, et al. New adhesin functions of surface-exposed pneumococcal proteins. BMC Microbiol. 2010 Jul 12;10:190.

40. García B, Merayo-Lloves J, Rodríguez D, Alcalde I, García-Suárez O, Alfonso JF, et al. Different Use of Cell Surface Glycosaminoglycans As Adherence Receptors to Corneal Cells by Gram Positive and Gram Negative Pathogens. Front Cell Infect Microbiol. 2016;6:173.

41. Hammerschmidt S. Adherence molecules of pathogenic pneumococci. Curr Opin Microbiol. 2006 Feb;9(1):12–20.

42. Birkenhauer E, Neethirajan S, Weese JS. Collagen and hyaluronan at wound sites influence early polymicrobial biofilm adhesive events. BMC Microbiol. 2014 Jul 16;14:191.

43. Rajas O, Quirós LM, Ortega M, Vazquez-Espinosa E, Merayo-Lloves J, Vazquez F, et al. Glycosaminoglycans are involved in bacterial adherence to lung cells. BMC Infect Dis. 2017 02;17(1):319.

44. Bhaskar KR, O'Sullivan DD, Opaskar-Hincman H, Reid LM, Coles SJ. Density gradient analysis of secretions produced in vitro by human and canine airway mucosa:

identification of lipids and proteoglycans in such secretions. Exp Lung Res. 1986;10(4):401–22.

45. Monzon ME, Casalino-Matsuda SM, Forteza RM. Identification of glycosaminoglycans in human airway secretions. Am J Respir Cell Mol Biol. 2006 Feb;34(2):135–41.

46. Jiang D, Liang J, Noble PW. Regulation of non-infectious lung injury, inflammation, and repair by the extracellular matrix glycosaminoglycan hyaluronan. Anat Rec (Hoboken). 2010 Jun;293(6):982–5.

47. Jiang D, Liang J, Noble PW. Hyaluronan in tissue injury and repair. Annu Rev Cell Dev Biol. 2007;23:435–61.

48. Lennon FE, Singleton PA. Role of hyaluronan and hyaluronan-binding proteins in lung pathobiology. Am J Physiol Lung Cell Mol Physiol. 2011 Aug;301(2):L137-147.

49. Kujawa MJ, Caplan AI. Hyaluronic acid bonded to cell-culture surfaces stimulates chondrogenesis in stage 24 limb mesenchyme cell cultures. Dev Biol. 1986 Apr;114(2):504–18.

50. Nusgens B-V. [Hyaluronic acid and extracellular matrix: a primitive molecule?]. Ann Dermatol Venereol. 2010 Apr;137 Suppl 1:S3-8.

51. Solis MA, Chen Y-H, Wong TY, Bittencourt VZ, Lin Y-C, Huang LLH. Hyaluronan regulates cell behavior: a potential niche matrix for stem cells. Biochem Res Int. 2012;2012:346972.

52. Bidossi A, Mulas L, Decorosi F, Colomba L, Ricci S, Pozzi G, et al. A functional genomics approach to establish the complement of carbohydrate transporters in Streptococcus pneumoniae. PLoS ONE. 2012;7(3):e33320.

53. Oiki S, Mikami B, Maruyama Y, Murata K, Hashimoto W. A bacterial ABC transporter enables import of mammalian host glycosaminoglycans. Sci Rep. 2017 21;7(1):1069.

54. Chapuy-Regaud S, Ogunniyi AD, Diallo N, Huet Y, Desnottes J-F, Paton JC, et al. RegR, a global LacI/GalR family regulator, modulates virulence and competence in Streptococcus pneumoniae. Infect Immun. 2003 May;71(5):2615–25.

55. Maruyama Y, Nakamichi Y, Itoh T, Mikami B, Hashimoto W, Murata K. Substrate specificity of streptococcal unsaturated glucuronyl hydrolases for sulfated glycosaminoglycan. J Biol Chem. 2009 Jul 3;284(27):18059–69.

56. Marion C, Stewart JM, Tazi MF, Burnaugh AM, Linke CM, Woodiga SA, et al. Streptococcus pneumoniae can utilize multiple sources of hyaluronic acid for growth. Infect Immun. 2012 Apr;80(4):1390–8.

57. Yadav MK, Chae S-W, Park K, Song J-J. Hyaluronic acid derived from other streptococci supports Streptococcus pneumoniae in vitro biofilm formation. Biomed Res Int. 2013;2013:690217.

58. Hendriksen WT, Bootsma HJ, Estevão S, Hoogenboezem T, de Jong A, de Groot R, et al. CodY of Streptococcus pneumoniae: link between nutritional gene regulation and colonization. J Bacteriol. 2008 Jan;190(2):590–601.

59. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, et al. Rapid pneumococcal evolution in response to clinical interventions. Science. 2011 Jan 28;331(6016):430–4.

60. Hiller NL, Eutsey RA, Powell E, Earl JP, Janto B, Martin DP, et al. Differences in Genotype and Virulence among Four Multidrug-Resistant Streptococcus pneumoniae Isolates Belonging to the PMEN1 Clone. PLoS ONE. 2011 Dec 19;6(12):e28850.

61. Nesin M, Ramirez M, Tomasz A. Capsular transformation of a multidrug-resistant Streptococcus pneumoniae in vivo. The Journal of Infectious Diseases. 1998 Mar;177(3):707–713.

62. Wyres KL, Lambertsen LM, Croucher NJ, McGee L, von Gottberg A, Liñares J, et al. Pneumococcal capsular switching: a historical perspective. J Infect Dis. 2013 Feb 1;207(3):439–49.

63. Tettelin H, Nelson KE, Paulsen IT, Eisen JA, Read TD, Peterson S, et al. Complete genome sequence of a virulent isolate of Streptococcus pneumoniae. Science. 2001 Jul 20;293(5529):498–506.

64. Brittan JL, Buckeridge TJ, Finn A, Kadioglu A, Jenkinson HF. Pneumococcal neuraminidase A: an essential upper airway colonization factor for Streptococcus pneumoniae. Mol Oral Microbiol. 2012 Aug;27(4):270–83.

65. Kjos M, Aprianto R, Fernandes VE, Andrew PW, van Strijp JAG, Nijland R, et al. Bright fluorescent Streptococcus pneumoniae for live-cell imaging of host-pathogen interactions. J Bacteriol. 2015 Mar;197(5):807–18.

66. Balis JU, Bumgarner SD, Paciga JE, Paterson JF, Shelley SA. Synthesis of lung surfactant-associated glycoproteins by A549 cells: description of an in vitro model for human type II cell dysfunction. Exp Lung Res. 1984;6(3–4):197–213.

67. Lieber M, Smith B, Szakal A, Nelson-Rees W, Todaro G. A continuous tumor-cell line from a human lung carcinoma with properties of type II alveolar epithelial cells. Int J Cancer. 1976 Jan 15;17(1):62–70.

68. Nardone LL, Andrews SB. Cell line A549 as a model of the type II pneumocyte. Phospholipid biosynthesis from native and organometallic precursors. Biochim Biophys Acta. 1979 May 25;573(2):276–95.

69. Berger JT, Voynow JA, Peters KW, Rose MC. Respiratory carcinoma cell lines. MUC genes and glycoconjugates. Am J Respir Cell Mol Biol. 1999 Mar;20(3):500–10.

70. Carterson AJ, Höner zu Bentrup K, Ott CM, Clarke MS, Pierson DL, Vanderburg CR, et al. A549 lung epithelial cells grown as three-dimensional aggregates: alternative tissue culture model for Pseudomonas aeruginosa pathogenesis. Infect Immun. 2005 Feb;73(2):1129–40.

71. Raffel FK, Szelestey BR, Beatty WL, Mason KM. The Haemophilus influenzae Sap transporter mediates bacterium-epithelial cell homeostasis. Infection and Immunity. 2013;81(1):43–54.

72. Cook LC, Federle MJ. Peptide pheromone signaling in Streptococcus and Enterococcus. FEMS Microbiol Rev. 2014 May;38(3):473–92.

73. Martin R, Larsen AH, Corey RA, Midtgaard SR, Frielinghaus H, Schaffitzel C, et al. Structure and Dynamics of the Central Lipid Pool and Proteins of the Bacterial Holo-Translocon. Biophys J. 2019 May 21;116(10):1931–40.

74. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015 Jun;10(6).

75. Dong W, Lu W, McKeehan WL, Luo Y, Ye S. Structural basis of heparan sulfate-specific degradation by heparinase III. Protein Cell. 2012 Dec;3(12):950–61.

76. Ma K, Cao Q, Luo S, Wang Z, Liu G, Lu C, et al. cas9 Enhances Bacterial Virulence by Repressing the regR Transcriptional Regulator in Streptococcus agalactiae. Infect Immun. 2018;86(3).

77. Reese MG. Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. Comput Chem. 2001 Dec;26(1):51–6.

78. Solovyev V, Salamov A. Automatic Annotation of Microbial Genomes and Metagenomic Sequences. In Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies. Nova Science Publisher; 2011. 61–78 p.

79. Takao A. Cloning and expression of hyaluronate lyase genes of Streptococcus intermedius and Streptococcus constellatus subsp. constellatus(1). FEMS Microbiol Lett. 2003 Feb 14;219(1):143–50.

80. Sung CK, Li H, Claverys JP, Morrison DA. An rpsL Cassette, Janus, for Gene Replacement through Negative Selection in Streptococcus pneumoniae. Appl Environ Microbiol. 2001 Nov;67(11):5190–6.

81. Wang T, Cui Y, Jin J, Guo J, Wang G, Yin X, et al. Translating mRNAs strongly correlate to proteins in a multivariate manner and their translation ratios are phenotype specific. Nucleic Acids Res. 2013 May;41(9):4743–54.

82. Chow G, Tauler J, Mulshine JL. Cytokines and growth factors stimulate hyaluronan production: role of hyaluronan in epithelial to mesenchymal-like transition in non-small cell lung cancer. J Biomed Biotechnol. 2010;2010:485468.

83. Glanville DG, Han L, Maule AF, Woodacre A, Thanki D, Abdullah IT, et al. RitR is an archetype for a novel family of redox sensors in the streptococci that has evolved from two-component response regulators and is required for pneumococcal colonization. PLoS Pathog. 2018;14(5):e1007052.

84. Kahya HF, Andrew PW, Yesilkaya H. Deacetylation of sialic acid by esterases potentiates pneumococcal neuraminidase activity for mucin utilization, colonization and virulence. PLOS Pathogens. 2017 Mar 3;13(3):e1006263.

85. Bogaert D, de Groot R, Hermans P. Streptococcus pneumoniae colonisation: the key to pneumococcal disease. The Lancet Infectious Diseases. 2004 Mar;4(3):144–54.

86. Kadioglu A, Weiser JN, Paton JC, Andrew PW. The role of Streptococcus pneumoniae virulence factors in host respiratory colonization and disease. Nat Rev Micro. 2008 Apr;6(4):288–301.

87. Hendriksen WT, Bootsma HJ, Estevão S, Hoogenboezem T, de Jong A, de Groot R, et al. CodY of Streptococcus pneumoniae: link between nutritional gene regulation and colonization. J Bacteriol. 2008 Jan;190(2):590–601.

88. Lysenko ES, Lijek RS, Brown SP, Weiser JN. Within-host competition drives selection for the capsule virulence determinant of Streptococcus pneumoniae. Curr Biol. 2010 Jul 13;20(13):1222–6.

89. Bergmann S, Hammerschmidt S. Versatility of pneumococcal surface proteins. Microbiology. 2006 Feb 1;152(2):295–303.

90. Paterson GK, Orihuela CJ. Pneumococcal microbial surface components recognizing adhesive matrix molecules targeting of the extracellular matrix. Mol Microbiol. 2010 Jul 1;77(1):1–5.

91. Jensch I, Gámez G, Rothe M, Ebert S, Fulde M, Somplatzki D, et al. PavB is a surface-exposed adhesin of Streptococcus pneumoniae contributing to nasopharyngeal colonization and airways infections. Mol Microbiol. 2010 Jul 1;77(1):22–43.

92. Kanwal S, Jensch I, Palm GJ, Brönstrup M, Rohde M, Kohler TP, et al. Mapping the recognition domains of pneumococcal fibronectin-binding proteins PavA and PavB demonstrates a common pattern of molecular interactions with fibronectin type III repeats. Mol Microbiol. 2017 Sep;105(6):839–59.

93. Pracht D, Elm C, Gerber J, Bergmann S, Rohde M, Seiler M, et al. PavA of Streptococcus pneumoniae modulates adherence, invasion, and meningeal inflammation. Infect Immun. 2005 May;73(5):2680–9.

94. Antikainen J, Kuparinen V, Lähteenmäki K, Korhonen TK. Enolases from Gram-positive bacterial pathogens and commensal lactobacilli share functional similarity in virulence-associated traits. FEMS Immunol Med Microbiol. 2007 Dec;51(3):526–34.

95. Bergmann S, Rohde M, Chhatwal GS, Hammerschmidt S. alpha-Enolase of Streptococcus pneumoniae is a plasmin(ogen)-binding protein displayed on the bacterial cell surface. Mol Microbiol. 2001 Jun;40(6):1273–87.

96. Papasergi S, Garibaldi M, Tuscano G, Signorino G, Ricci S, Peppoloni S, et al. Plasminogen- and fibronectin-binding protein B is involved in the adherence of Streptococcus pneumoniae to human epithelial cells. J Biol Chem. 2010 Mar 5;285(10):7517–24.

97. Yamaguchi M, Terao Y, Mori Y, Hamada S, Kawabata S. PfbA, a novel plasmin- and fibronectin-binding protein of Streptococcus pneumoniae, contributes to fibronectin-dependent adhesion and antiphagocytosis. J Biol Chem. 2008 Dec 26;283(52):36272–9.

98. Voss S, Hallström T, Saleh M, Burchhardt G, Pribyl T, Singh B, et al. The choline-binding protein PspC of Streptococcus pneumoniae interacts with the C-terminal heparin-binding domain of vitronectin. J Biol Chem. 2013 May 31;288(22):15614–27.

99. Shivshankar P, Sanchez C, Rose LF, Orihuela CJ. The Streptococcus pneumoniae adhesin PsrP binds to Keratin 10 on lung cells. Mol Microbiol. 2009 Aug;73(4):663–79.

100. King SJ, Hippe KR, Weiser JN. Deglycosylation of human glycoconjugates by the sequential activities of exoglycosidases expressed by Streptococcus pneumoniae. Mol Microbiol. 2006 Feb;59(3):961–74.

101. Limoli DH, Sladek JA, Fuller LA, Singh AK, King SJ. BgaA acts as an adhesin to mediate attachment of some pneumococcal strains to human epithelial cells. Microbiology (Reading, Engl). 2011 Aug;157(Pt 8):2369–81.

102. Singh AK, Pluvinage B, Higgins MA, Dalia AB, Woodiga SA, Flynn M, et al. Unravelling the multiple functions of the architecturally intricate Streptococcus pneumoniae β-galactosidase, BgaA. PLoS Pathog. 2014 Sep;10(9):e1004364.

103. Madhour A, Maurer P, Hakenbeck R. Cell surface proteins in S. pneumoniae, S. mitis and S. oralis. Iran J Microbiol. 2011 Jun;3(2):58–67.

104. Jedrzejas MJ. Pneumococcal Virulence Factors: Structure and Function. Microbiol Mol Biol Rev. 2001 Jun;65(2):187–207.

105. Jedrzejas MJ, Chantalat L. Structural studies of streptococcus agalactiae hyaluronate lyase. Acta Crystallogr D Biol Crystallogr. 2000 Apr;56(Pt 4):460–3.

106. Denapaite D, Brückner R, Nuhn M, Reichmann P, Henrich B, Maurer P, et al. The genome of Streptococcus mitis B6--what is a commensal? PLoS ONE. 2010 Feb 25;5(2):e9426.

107. Shahinas D, Thornton CS, Tamber GS, Arya G, Wong A, Jamieson FB, et al. Comparative Genomic Analyses of Streptococcus pseudopneumoniae Provide Insight into Virulence and Commensalism Dynamics. PLoS One [Internet]. 2013 Jun 19 [cited 2019 Nov 8];8(6). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3686770/

108. Wang Z, Guo C, Xu Y, Liu G, Lu C, Liu Y. Two Novel Functions of Hyaluronidase from Streptococcus agalactiae Are Enhanced Intracellular Survival and Inhibition of Proinflammatory Cytokine Expression. Infect Immun. 2014 Jun;82(6):2615–25.

109. Warner JB, Lolkema JS. CcpA-dependent carbon catabolite repression in bacteria. Microbiol Mol Biol Rev. 2003 Dec;67(4):475–90.

110. Carvalho SM, Kloosterman TG, Kuipers OP, Neves AR. CcpA Ensures Optimal Metabolic Fitness of Streptococcus pneumoniae. PLoS ONE. 2011 Oct 21;6(10):e26707.

111. Corradetti B, Taraballi F, Martinez JO, Minardi S, Basu N, Bauza G, et al. Hyaluronic acid coatings as a simple and efficient approach to improve MSC homing toward the site of inflammation. Sci Rep. 2017 11;7(1):7991.

112. Khademhosseini A, Suh KY, Yang JM, Eng G, Yeh J, Levenberg S, et al. Layer-by-layer deposition of hyaluronic acid and poly-L-lysine for patterned cell co-cultures. Biomaterials. 2004 Aug;25(17):3583–92.

113. Aoki K, Matsumoto S, Hirayama Y, Wada T, Ozeki Y, Niki M, et al. Extracellular mycobacterial DNA-binding protein 1 participates in mycobacterium-lung epithelial cell interaction through hyaluronic acid. J Biol Chem. 2004 Sep 17;279(38):39798–806.

114. Aaberge IS, Eng J, Lermark G, Løvik M. Virulence of Streptococcus pneumoniae in mice: a standardized method for preparation and frozen storage of the experimental bacterial inoculum. Microb Pathog. 1995 Feb;18(2):141–52.

115. Huang H. Design and Characterization of Artificial Transcriptional Terminators. Massachusetts Institute of Technology; 2008.

116. Kadam A, Janto B, Eutsey R, Earl JP, Powell E, Dahlgren ME, et al. Streptococcus pneumoniae Supragenome Hybridization Arrays for Profiling of Genetic Content and Gene Expression. Curr Protoc Microbiol. 2015;36:9D.4.1-9D.4.20.

117. Kayala MA, Baldi P. Cyber-T web server: differential analysis of high-throughput data. Nucleic Acids Res. 2012 Jul;40(Web Server issue):W553-559.

118. Horta CCR, Magalhães B de F, Oliveira-Mendes BBR, do Carmo AO, Duarte CG, Felicori LF, et al. Molecular, Immunological, and Biological Characterization of Tityus serrulatus Venom Hyaluronidase: New Insights into Its Role in Envenomation. PLoS Negl Trop Dis [Internet]. 2014 Feb 13 [cited 2019 Sep 20];8(2). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3923731/

119. Kloosterman TG, Bijlsma JJE, Kok J, Kuipers OP. To have neighbour's fare: extending the molecular toolbox for Streptococcus pneumoniae. Microbiology (Reading, Engl). 2006 Feb;152(Pt 2):351–9.

# Chapter 4 Genome-wide Profiling of Pneumococcal Biofilms

This chapter was initiated at the dusk of my graduated work and it focuses on pneumococcal global gene expression in biofilms. This is a timely topic, in that there are a limited number of studies characterizing biofilms from a transcriptional perspective. While several reports have established that the pneumococcus adjusts its transcriptome during colonization of host cells or during invasion of tissues, less is known regarding the transcriptional signatures associated with the stages of pneumococcal biofilm development (attachment, growth, and dispersal).

In the previous chapter, we established that VP1 controls attachment to host cells, and that this has consequences for biofilm development. These observations primed us to investigate the *vp1* regulon. Our screen revealed an expanded set of *vp1*-associated genes. It suggests a model where *vp1* connects to a group of transcription factors, and altogether, these molecules sense environmental nutrients and downregulate the capsule promoting exposure of surface receptors and biofilm development. In addition, the *vp1* screen suggests that this peptide may influence genes whose products are involved in competence. Thus, the data presented in this chapter represent work-in-progress, which I hope will lay the groundwork for future projects in the Hiller lab and elsewhere.

## 4.1 Introduction

The evolutionary cycle of the pneumococcus consists of colonization of the upper respiratory tract and transmission to another human host. During colonization of the mucosa, pneumococci adapt a biofilm mode of growth and it has been suggested that they play a crucial role in bacterial virulence during infection (1). The pneumococcus forms biofilms on the mucosal of the nasopharynx and the middle ear (2–4). These sessile and adherent structures serve as community settlements, and facilitate pneumococcal colonization in multiple ways (5). They hide the bacteria from immune surveillance, increase resistance to antimicrobials, and provide a platform for DNA acquisition (Chao et al., 2014; Croucher et al., 2012; Marks et al., 2012; Tonnaer et al., 2009; Trappetti et al., 2011). We previously determined that the product of *vp1* is a secreted peptide that features a Glycine-Glycine sequence, characteristics of secreted peptides (11,12). VP1 binds to the pneumococcal surface and controls the gene expression of a locus involved in the acquisition and processing of hyaluronic acid (PTS-HAL) (13). Finally, the gene encoding VP1 is a virulence determinant and is highly upregulated in the presence of host cells, where it promotes biofilm development (14,15).

The pneumococcus encodes a range of molecules involved in biofilm development. Using a transposon library, the group of Camilli and colleagues conducted an *in vitro* screen for biofilm-altered pneumococci mutants and identified fourth-nine genes encoding for several surface-exposed molecules and transcription factors (16). These include genes products implicated in cell-wall synthases (*murB* and *murE*), glycosidases (*nanB)*, choline-binding (*lytC*, *cbpA*, *cbpF*), proteases (*clpC*, *clpP*, and *clpX*,) oligopeptide transporter (subunit *aliB)*, and signal transduction (ciaH – the histidine kinase membrane receptor from the two-component system 5, TCS05). In agreement with the role of biofilm in carriage, several of these mutants show impaired colonization of the nasopharynx (17).

Many of the proteins associated with biofilm development exert their effect on the surface of the bacterial cells, at the interface with the human host (18). The pneumococcal surface protein A (PspA) inhibits complement activation (19,20), while

PspC (also known as CbpA) mediates attachment to epithelial cells by binding the polymeric immunoglobulin receptor (PIgR) (21–24). The glucosidases NanA and BgaA, release the terminal sialic acid and the subsequent galactose residues from glycoconjugates in host mucus, respectively (25,26). One hypothesis is that these enzymes promote attachment and biofilm development by exposing surface receptors on host cells (27). The pneumococcal serine-rich repeat protein (PsrP), which inhibits the production of the proinflammatory cytokines release during infection, is involved not only in biofilm development, but also in intraspecies aggregation (28). Pneumolysin, the proteinaceous pneumococcal toxin capable of damaging human tissue and released by the bacteria into the host milieu, greatly enhances biofilm biomass in the D39 and TIGR4 strains in a manner independent of its hemolytic activity (29). Transcriptional factors, likely related to extracellular sensing, also influence biofilms as exemplified by Rgg1952, which strongly enhance biofilm capacity in non-encapsulated pneumococcus (30). Thus, the pneumococcus toolbox for host interactions includes many classes of surface exposed molecules.

Competence is implicated in biofilm formation. In the pneumococcus, competence controls between 5-10% of the pneumococcal expression landscape, representing an extensive transcriptional network in a response to a single signal (31). Competence is controlled via the concerted function of *comAB* and *comCED*. The signaling starts with the release of CSP (*comC*) for which there are six pherotypes, and two are highly represented with the species: CSP1 (*comC1*) and CSP2 (*comC2*). Mutants carrying pherotypic variants of CSP (*comC1* and *comC2*) display a differential capacity to form in vitro biofilms: *comC1* variants form thicker biofilms than those carrying the *comC2* allele (32). Further, competence is induced in a biofilm mode of growth (10,33,34). Finally, in multiple *in vitro* models, activation of competence is essential for biofilm development (10,35). Recent work revealed that the pneumococcal peptide BriC plays a major role connecting competence to biofilm development (36). Quorum sensing via the small molecule AI-2 from the LuxS/AI-2 systems is also necessary for adherence to and colonization of the ear epithelium (37). Thus, the roles of competence and Lux in biofilm development illustrate the importance of cell-cell communication in chronic colonization.

This thesis focuses on the role of the newly discovered cell-cell communication peptide VP1. In chapter 2 we present data demonstrating that VP1 is secreted, required for bacterial attachment, and promotes biofilm formation (13,15). The product of *vp1* is a secreted peptide with a double glycine sequence characteristic of secreted peptides. Further, a tagged form of VP1 binds to the pneumococcal surface, consistent with a role in signal transduction, and a *vp1*-deficient strain exhibits a diminished biofilm capacity (15). We also established that VP1 is a signaling peptide that controls gene expression and it is associated with colonization. Finally, we determined that *vp1* controls the expression of a locus implicated in the acquisition of hyaluronic acid (PTS-HAL). The PTS-HAL is controlled by the negative regulator RegR encoded downstream of the PTS-HAL, which is not strongly transcriptionally regulated by *vp1* (13).

A common practice is to utilize microarrays to analyze gene expression in planktonic cultures, however microarray analysis in biofilms is challenging due to the low quality and yields of RNA isolated from biofilms. A common alternative is the utilization of qRT-PCR to measure expression of only a selected set of genes in biofilms (38–40). Moreover, a direct comparison of the transcriptome in biofilms to planktonic is complicated: growing conditions are different (media, temperature, presence epithelial cells), and epression profiles are vastly different. Recently, Aprianto *et al.* utilized RNA-seq to analyze the expression profile of pneumococci grown in 22 different conditions, some of media alone and other of bacteria in contact with A549 cancer lung epithelial cells, for up to four hours (14). To address the challenge of host-bacteria mRNA, where bacteria material corresponds to less than 1% of total recovered material, they develop a hybrid chimeric genome containing the concatenated human genome and *S. pneumoniae* (Aprianto et al., 2016, 2018).

To study transcription in biofilms, and specifically the role of VP1 in biofilms, we employed nCounter NanoString technology to measure gene expression in biofilms. This platform provides an automated, highly sensitive enumeration of the pathogens mRNA transcripts in infected tissue (41). Due to the selectivity, NanoString analyses allow the discrimination between both the epithelial cells RNA and the abundant rRNA (no depletion step required). Previously, we successfully measured *vp1* mRNA levels *in vivo* using the NanoString technology in the chinchilla model of otitis media (15).

Here, we measured the expression profile of the wild-type PN4595-T23 strain and the *vp1* isogenic mutant. Biofilms were grown on lung epithelial cells for 16h, we vp1 influences biofilm formation in this system and time point. Our findings suggest that *vp1* regulates the expression of the interconnected systems Rgg144 (SPD_0144 locus), and Rgg939 (SPD_0939 locus). In addition, several two-component regulatory systems including TCS04, TCS06, TCS09, TCS10, and TCS12 appeared to respond to levels of *vp1*. Further, *vp1* is associated with levels of the oligopeptide transporter AliAB, known to import an array of small peptides. In accordance with our data of planktonic cells grown in CDM-Glu, *vp1* also controls the expression of the PTS-HAL in biofilms. Our characterization of the VP1 regulon in a biofilm mode of growth provides evidence that *vp1* exerts profound effects on biofilms progression.

## 4.2 Results

### 4.2.1 *vp1* enhances pneumococcal biofilm development on lung epithelial cells

Pneumococci can adopt a biofilm mode of growth in the nasopharynx and the middle ear during infections (17,38,42,43)(2). Further, several *in vitro* models of pneumonia support the formation of biofilm in the lung using lung cancer epithelial A549 (44,45) and larynx HEp-2 cells (46,47), nonetheless direct evidence of biofilm development in the lungs (or bacterial clustering during pneumonia) is scarce (48).

In this chapter, we investigated whether *vp1* influences biofilm development on lung A549 epithelial cells. The wild-type strain PN4594-T23 forms a robust biofilm on A549 substrate. Sixteen hours post-inoculation, we observe biofilms ranging from 30 to 35 μm by confocal imaging (**Figure 4.1A**). Further, total bacterial biomass measurements support a biofilm in this model (**Suppl. Figure 1**) with average thickness similar to the biofilms observed on CMEE cells (**Chapter 2, Figure 2.4**). However, unlike the CMEE cells, in this short time frame the pneumococcal cells do not disrupt the A549 cells and the epithelial cell layer appears intact in confocal images.

A comparison of biofilm biomass and thickness between wild-type and Δ*vp1* strains revealed that *vp1* promotes biofilms in this model. The Δ*vp1* strain displays approximately 2-fold reduction in biofilm biomass and thickness when compared to the wild-type. Furthermore, we observed that the complemented strain carrying an overexpressor of *vp1* in the deletion mutant background (Δ*vp1:vp1*) rescues the wild-type biofilm phenotype (**Figure 4.1B** and **C**). Thus, we conclude that *vp1* contributes to the development of biofilm on lung epithelial cells.

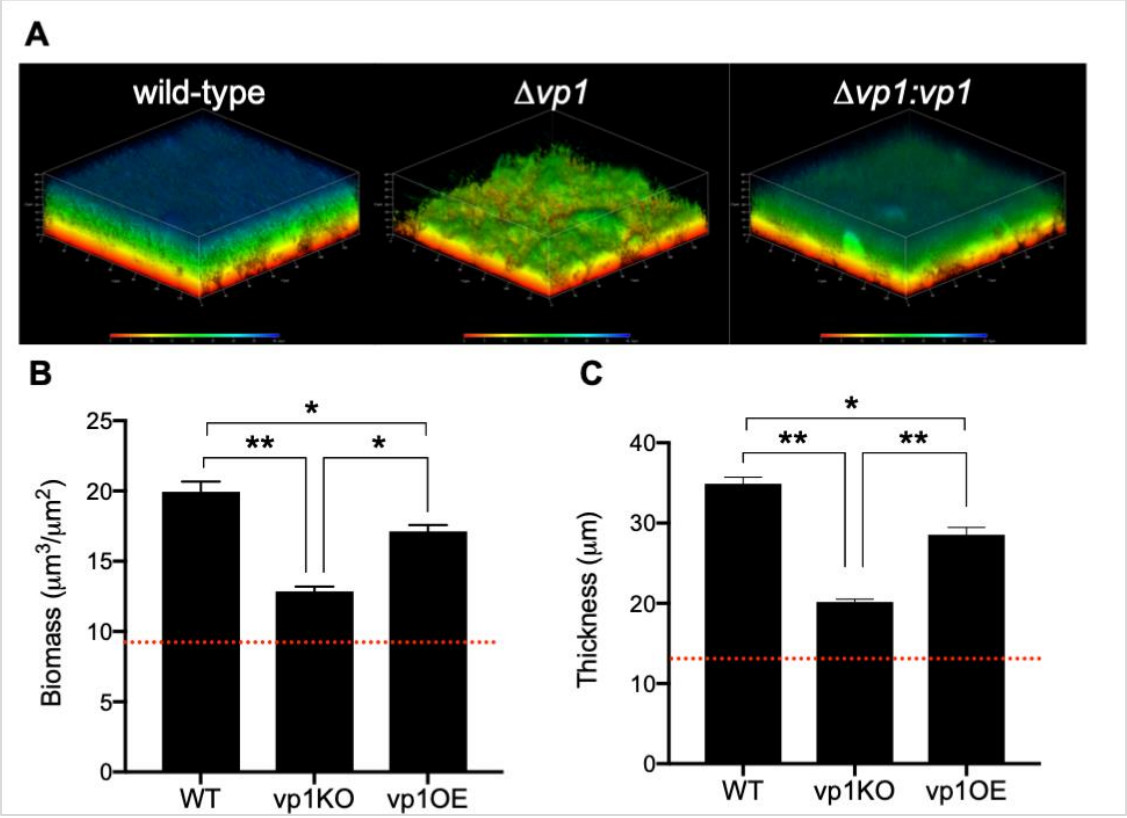



**Figure 4.1. *vp1* enhances biofilm development on lung epithelial cells**

**A.** Representative images from confocal micrographs of biofilms of wild-type, Δ*vp1*, or Δ*vp1:vp1* grown on A549 cells imaged 16h post-seeding. **B.** Biofilm biomass and **C.** average thickness were computed from confocal images. Data represent the mean ± S.E.M. of at least four independent experiments. Statistical significance was calculated using one-way ANOVA analysis with Bonferroni for multiple comparisons. *$p = 0.005$, ** $p < 0.0001$, ns = not significant.

## 4.2.2 Design of a custom nanoString array

Our laboratory has successfully utilized the nCounter analysis system from NanoString Technologies (49) to measure bacterial gene expression in the presence of host cells. This technology provides high reliability, sensitivity, and a selective enumeration of RNA transcripts. In this session we present the design, experimentation, and validation of a probe set used to access the role of *vp1* in biofilm development.

Our custom pneumococcal probe set was designed to capture the diversity in the pneumococcal genome and bypassing allelic differences between strains. To this end,

we employed the online server RAST to extract the gene sequences (50), and BLAST against multiple genomes to ensure that the probes bypassed allelic differences between strains. Due to differences in genomic content between strains, the nomenclature used for the locus identity encompasses several known strains, including PN4595-T23, D39 and TIGR4 (**Suppl. Table 4.1**). Our custom probe set captures 234 genes. These are intended to cover all the regulators encoded in the pneumococcal genome (ninety-four transcription factors). Further, it also includes genes with products predicted to be surface exposed proteins, ABC transporters, cell wall sysnthesis enzymes, adhesin molecules, and the thirteen core two-component system (TCS) (**Suppl. Figure 4.2**).

### 4.2.3 *vp1* controls gene expression during biofilm mode of growth

Our previous studies show that *vp1* is highly induced during infection (**Chapter 2, Figure 2.2**) and its expression is high in sixteen hours biofilms (**Chapter 2, Suppl. Figure 2.4** and **Figure 4.3**). Thereby, we hypothesize that *vp1* controls gene expression in biofilms. To test this hypothesis, we compared gene expression levels in the wild-type PN4595-T23 strain and its Δ*vp1* derivative when grown in biofilms on the human cancer lung epithelial cells A549 at sixteen hours post-inoculation.

Our analysis revealed multiple transcripts that are likely to be regulated by *vp1* in the biofilm mode of growth (**Figure 4.2**). Consistent with the transcriptome analysis of planktonic cells in CDM glucose, in biofilm the VP1 also positively regulates the genes involved in HA acquisition and processing (PTS-HAL). In this genomic locus, *ugl* displayed the highest degree of upregulation, followed by *gno* and the PTS subunits IIA and IID. Moreover, VP1 does not appear to regulate *regR* levels, suggesting its effect is not via the negative regulator of this region.

In these biofilm conditions, the genes most differentially regulated by *vp1* were the transcriptional repressor of the fructose-encoding operon DeoR/GlpR (peg804), followed by the fratricide bacteriocin CibA and the transcriptional regulator GntR

(peg1713). This suggests a feedback loop between VP1 and GntR, as other have shown that GntR influences levels of *vp1*.

The set also includes transcription factors Rgg144, Rgg939, and Rgg1952. Rgg 144 is a positive regulator of VP1, and there is an interdependence in the levels of *rgg144* and *rgg939*. The architecture of VP1 and Rggs remains an active area of research. Additional hits include *murE*, *murF and pbp2X*, all involved int the synthesis of the bacterial the cell wall, and the late competence genes *comX* and *comW*, involved in the DNA update and recombination. Finally, two-component system were also identified in this screen. These included the transcriptional regulators for TCS04, TCS06, and TCS09 (**Table 4.1**). We conclude that the influence of VP1 in biofilms extends beyond its role in HA-interctions, and hypothesize that it has a major role in pneumococcal virulence.

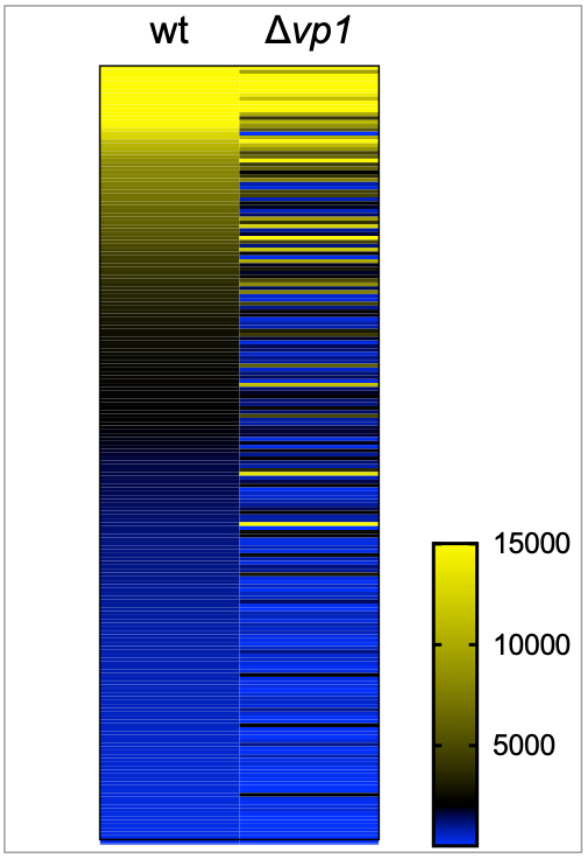



**Figure 4.2 The *vp1* product regulates the expression of multiple genes in biofilms**

NanoString transcriptome analysis comparing 234 genes the expression profiles of the wild-type strain PN4595-T23 and *vp1* derivative strain (Δ*vp1*) in biofilms on A549 cells. Data was analyzed for statistical significance using nCounter. Genes are sorted based on its levels of expression on the wild-type strain. Color code illustrates expression ratio between the Δ*vp1* strain relative to wild-type strain.

The majority of the genes in our set were positively regulated by *vp1*. However, there were also twenty genes negative regulated in the presence of *vp1* in a biofilm mode-of-growth (**Table 4.2**). This set includes the transcription factor MarR, involved in synthesis of fatty acids, the oligo peptide ABC transporter subunits AliA and AliB, the transcriptional regulator TprA2, involved in maintenance of commensalism in the pneumococcus (52), the enzyme MurM, a tRNA-dependent amino acid ligase required for pneumococcal cell wall synthesis (54), and the transcription factor Psr (part of the LytR family transcriptional regulators), involved in the attachment of teichoic acid to cell wall (55).

**Table 4.1** Genes differentially upregulated (>2-fold) in biofilms by *vp1*

| Gene | Fold | p-value | Annotation |
|---|---|---|---|
| SPD_RS04090 | 62.148 | 1.37E-06 | DeoR/GlpR transcriptional regulator fructose operon peg804 |
| peg.291 | 46.093 | 8.69E-07 | Glucuronyl hidrolase ***ugl*** |
| peg.395 | 33.220 | 1.01E-05 | fratricide two-peptide bacteriocin subunit CibA peg126 |
| SPD_RS08045 | 21.594 | 1.14E-07 | GntR family transcriptional regulator peg1713 |
| comW | 19.889 | 1.45E-03 | recombinase RecX, competence comW |
| SPR0290 | 18.755 | 3.32E-07 | phosphate-binding protein PstS 1, gluconate 5-dehydrogenase - peg289 ***gno*** |
| peg.138 | 18.440 | 2.32E-04 | UDP-glucose/GDP-mannose dehydrogenase family, central domain protein |
| SPD_RS08660 | 18.006 | 3.48E-04 | TetR/AcrR family transcriptional regulator |
| SPD_RS09875 | 17.904 | 3.03E-06 | Transcriptional regulator, MarR family |
| lclAE005672.3_cds | 16.385 | 6.90E-06 | endo-alpha-N-acetylgalactosaminidase, hypothetical |
| lclAE005672.3_cds | 13.910 | 2.84E-05 | iron chelate uptake ABC transporter family permease subunit |
| SPD_0123 | 13.788 | 7.08E-04 | vargininosuccinate lyase - argH, arginine synthesis |
| SPD_RS05070 | 13.339 | 1.16E-04 | putative transcriptional regulator |
| SP_1833 | 13.282 | 1.04E-04 | multi-ligand-binding adhesin PfbA, cell wall |
| lacZ | 11.863 | 1.32E-03 | beta-galactosidase |
| SPD_RS05570 | 11.366 | 2.25E-04 | transcription antiterminator LacT |
| peg.1721 | 10.833 | 3.08E-04 | PTS beta-glucoside transporter subunit IIBCA |
| SPD_RS06610 | 10.727 | 4.65E-06 | DNA modification methyltransferase M.XbaI, transcriptional repressor |
| SPD_RS09720 | 10.546 | 7.13E-04 | BglG family transcriptional regulator, HTHt domain containing-protein |
| SPD_RS06520 | 10.367 | 4.83E-06 | Spx/MgsR family RNA polymerase-binding regulatory protein |
| SPD_RS05550 | 10.336 | 1.12E-05 | DeoR/GlpR transcriptional regulator lactose PTS repressor |
| SPD_RS00505 | 9.815 | 1.24E-08 | padR transcriptional regulator |
| SPD_RS03380 | 9.559 | 1.53E-03 | CopY/TcrY family copper transport repressor |
| SPD_RS11310 | 9.014 | 2.39E-04 | helix-turn-helix transcriptional regulator, cI repressor |
| SPD_0293 | 8.921 | 2.96E-04 | PTS N-acetylgalactosamine transporter subunit **IIA** |
| SPD_RS03665 | 8.349 | 3.33E-07 | PadR family transcriptional regulator |
| SPD_RS09185 | 8.257 | 3.03E-03 | type II restriction-modification system regulatory protein |
| SPD_0142 | 8.247 | 2.49E-04 | N-acetylglucosaminylphosphatidylinositol deacetylase superfamily pdgA |
| SPD_RS03325 | 7.810 | 4.86E-04 | transcriptional regulator%2C TENA/THI-4 family protei |
| SPN23F_RS04990 | 7.737 | 7.05E-04 | Rgg/GadR/MutR family transcriptional regulator |
| peg.992 | 7.585 | 1.86E-04 | ABC transporter ATP-binding protein |
| comX | 7.411 | 4.36E-06 | comX2 domain protein, TF |
| SPD_RS10310 | 7.356 | 4.74E-06 | transcriptional regulator, BglG family |
| SPD_RS06720 | 7.341 | 4.56E-05 | GntR family transcriptional regulator |
| SPD_RS04350 | 7.297 | 7.39E-04 | AraC family transcriptional regulator |
| ATCC700669_171 | 7.286 | 1.78E-04 | UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase domain protein |
| SPD_1504 | 7.156 | 4.05E-05 | nanA |

**Table 4.1 Continuation.** Genes differentially upregulated (>2-fold) in biofilms by *vp1*

| | | | |
|---|---|---|---|
| SPD_RS07435 | 6.854 | 7.38E-07 | GTP-sensing transcriptional pleiotropic repressor MarR family - CodY? |
| SPD_RS01935 | 6.517 | 8.50E-04 | Transcriptional regulator MtlR |
| SPD_RS07250 | 5.885 | 1.12E-03 | M protein trans-acting positive transcriptional regulator |
| SPD_RS01225 | 5.812 | 1.33E-07 | AraC family transcriptional regulator |
| SPD_RS04880 | 5.127 | 2.72E-03 | putative transcriptional regulator |
| SPD_RS10620 | 5.074 | 3.63E-05 | CtsR family transcriptional regulator |
| SPD_RS02740 | 4.980 | 3.79E-05 | Transcriptional regulator |
| D39_436 | 4.674 | 4.04E-05 | glutamine ABC transporter substrate-binding protein |
| SPD_RS10485 | 4.500 | 8.05E-07 | DeoR/GlpR transcriptional regulator , fucose operon |
| SPD_0297 | 4.491 | 6.30E-05 | PTS system transporter subunit **IID** |
| SPD_RS10030 | 4.300 | 3.79E-03 | arginine repressor |
| ATCC700669_124 | 4.292 | 2.81E-05 | ABC transporter membrane-spanning permease-amino acid transport domain |
| SPD_1499 | 4.277 | 3.22E-03 | nanB |
| CzcD | 4.197 | 4.43E-05 | Cation efflux pump trabnstporter |
| SPD_RS06655 | 4.100 | 4.41E-03 | AraC family transcriptional regulator |
| peg.481 | 3.948 | 3.34E-04 | two-peptide bacteriocin subunit BlpN |
| SPD_RS08530 | 3.936 | 5.51E-03 | PhoU family transcriptional regulator, pho uptake |
| SP_0268 | 3.837 | 7.30E-07 | surface-anchored pullulanase spuA |
| SP_0057 | 3.693 | 8.99E-04 | beta-N-acetylhexosaminidase strH |
| SPD_RS08880 | 3.636 | 4.39E-06 | bacterial regulatory helix-turn-helix s, AraC family protein |
| SPD_0525 | 3.561 | 7.48E-06 | VncS histidine kinase **TCS10** |
| SPD_RS04965 | 3.494 | 8.64E-07 | transcriptional regulator |
| peg.1221 | 3.427 | 1.97E-04 | N-acetylmannosamine-6-phosphate 2-epimerase, sialic acid metabolis, |
| SPD_RS10090 | 3.413 | 5.79E-03 | helix-turn-helix family protein, transcriptional regulator |
| SPD_RS10505 | 3.378 | 3.58E-04 | iron dependent repressor, N-terminal DNA binding domain protein [ |
| lclAE005672.3_cds | 3.335 | 1.34E-05 | beta-galactosidase |
| lclAE005672.3_cds | 3.305 | 7.87E-03 | endo-beta-N-acetylglucosaminidase, N-acetylmuramoyl-L-alanine amidase |
| SPD_RS02290 | 3.278 | 4.88E-07 | Uncharacterised protein, potassium uptake protein |
| SPD_RS09115 | 3.266 | 3.04E-04 | Transcriptional regulator MarR family |
| SPD_RS10250 | 3.125 | 1.82E-08 | transcriptional regulator, helix-turn-helix transcriptional regulator |
| SPD_2064 | 3.027 | 1.62E-02 | putative sensor histidine kinase ComD peg2265 **TCS12** |
| SPD_RS09230 | 2.946 | 3.53E-06 | Transcriptional regulator |
| SPD_RS00340 | 2.923 | 1.70E-06 | GntR family transcriptional regulator |
| SPD_RS09665 | 2.847 | 3.01E-05 | transcriptional regulator, GntR family protein |
| peg.2139 | 2.799 | 6.13E-04 | caax amino protease family protein |
| peg.853 | 2.781 | 1.15E-04 | SP_0924 family bacteriocin-like peptide |
| ssbB | 2.716 | 6.48E-04 | single-strand DNA-binding protein, based on B.sub. |
| lclAE005672.3_cds | 2.677 | 7.92E-05 | enterochelin uptake periplasmic binding domain protein |
| SPD_RS02375 | 2.618 | 3.37E-05 | Transcriptional regulator MerR |
| SPD_RS10275 | 2.552 | 3.93E-04 | transcriptional activator, Rgg/GadR/MutR family protein **Rgg1952** |
| SP_1772 | 2.492 | 3.50E-04 | LPXTG cell wall anchor domain-containing protein |
| SPD_RS10830 | 2.478 | 3.31E-02 | DNA binding response regulator |
| SP_2201 | 2.470 | 3.21E-04 | choline binding protein D |
| SPD_RS00800 | 2.449 | 3.06E-03 | peg139 **Rgg144** |
| SPD_1726 | 2.408 | 3.15E-06 | cholesterol-dependent cytolysin pneumolysin - ply, toxin |
| SPD_RS03080 | 2.382 | 6.86E-05 | DNA bindinfg response regulator peg599 YesN **TCS09** |
| SPD_RS01680 | 2.304 | 4.57E-05 | LytR |
| SPD_RS10050 | 2.301 | 2.31E-05 | transcriptional regulatory protein PhoP peg2103 pnpR **TCS04** |
| SPD_1514 | 2.271 | 8.00E-05 | ABC Trasnporter |
| SPD_RS09605 | 2.234 | 1.63E-05 | transcription termination/antitermination protein NusG |
| SP_1948 | 2.208 | 1.15E-03 | two-peptide bacteriocin subunit PneA2 |
| ATCC700669_118 | 2.151 | 3.78E-03 | putative 5-methylcytosine restriction system,catalytic subunit |
| peg.303 | 2.137 | 4.20E-05 | penicillin-binding protein 2X |
| SPD_RS04355 | 2.120 | 7.34E-05 | LysR family transcriptional regulator |
| SPD_RS01235 | 2.085 | 1.81E-04 | DeoR family transcriptional regulator |
| D39_1359 | 2.056 | 4.06E-05 | membrane protein - murE |
| lclAE005672.3_cds | 2.045 | 2.20E-05 | metal ABC transporter substrate-binding lipoprotein/adhesin PsaA |
| SPD_RS10605 | 2.040 | 1.23E-06 | DNA-binding response regulator peg2222 **TCS06** |

**Table 4.2** Genes differentially downregulated (<2-fold) in biofilms by *vp1*

| Gene | Change | p-value | Annotation |
|---|---|---|---|
| SPD_RS02030 | -12.158 | 8.23E-10 | Transcriptional regulator MarR, fatty acid synthesis FabT peg383 |
| AliA | -8.566 | 4.13E-08 | oligo peptide ABC transporter |
| peg.1276 | -5.606 | 7.82E-06 | helix-turn-helix transcriptional regulator, peg1276 TprA2 |
| SPD_RS06345 | -5.112 | 6.91E-06 | polyisoprenyl-teichoic acid--peptidoglycan teichoic acid transferase Psr |
| SPD_0535 | -4.508 | 1.27E-05 | serine/alanine-adding enzyme MurM |
| SPD_RS08805 | -3.434 | 1.68E-05 | trehalose operon repressor |
| AliB | -2.771 | 4.00E-06 | AliB oligopeptide ABC transporter |
| peg.335 | -2.607 | 1.28E-05 | penicillin-binding protein 1A |
| lclAE005672.3_cds_AAK7455 | -2.496 | 6.95E-05 | choline binding protein F, peg359 CBPJ according to RAST |
| peg.606 | -2.491 | 1.12E-05 | Endo-beta-N-acetylglucosaminidase, lysozyme lytC or lytD |
| SPD_RS08430 | -2.403 | 1.93E-05 | helix-turn-helix transcriptional regulator, XRE familyb peg1829 |
| SPD_RS04165 | -2.346 | 2.62E-05 | arginine repressor, represssor of arg regulon peg821 |
| SPD_RS00870 | -2.311 | 1.02E-04 | resposne regulator araC peg154 |
| SPD_RS07075 | -2.289 | 1.71E-04 | transcription elongation factor GreA peg1479 |
| peg.1271 | -2.268 | 3.06E-04 | lantibiotic transport/processing ATP-binding protein LanM |
| SPD_RS09120 | -2.217 | 7.20E-05 | glucose-1-phosphate adenylyl transferase - glgC |
| SPD_0157 | -2.160 | 9.29E-04 | histidine kinase TCS07 araC, YesMN peg153 |
| SPD_RS04470 | -2.143 | 2.59E-05 | TfoX/Sxy family protein [Streptococcus pneumoniae], pe881 hypothetical |
| SP_1573 | -2.066 | 3.40E-06 | 1,4-beta-N-acetylmuramidase, lytC cell wall hydrolase |
| AdcA | -2.044 | 3.84E-06 | ABC zinc transporter ZinT peg2197 |

To validate our probe set, we compared the genome-wide expression in a PN4595-T23 derivative Δ*regR* background. It has been shown that RegR represses the PTS-HAL expression (56). Consitently, genes of the PTS-HAL displayed higher expression levels in the *regR* deletion mutant relative to the wild-type strain. These data support the current model where RegR dramatically represses the expression of the operon upstream of the PTS-HAL, which is involved in the processing of HA in the bacterial cytosol (represented by *gno*), and the surface-exposed glycosidase HysA, which is required for the release of HA from hyaluronan chains (**Figure 4.3**) (**Table 4.3**). Importantly, RegR has no effect on the levels of *vp1*. Thus, our data set is consistent with existing literature, and supports a model where, in early biofilms, RegR functions exclusively in the regulation of the genes required for the processing and internalization of HA (**Table 4.3)**.

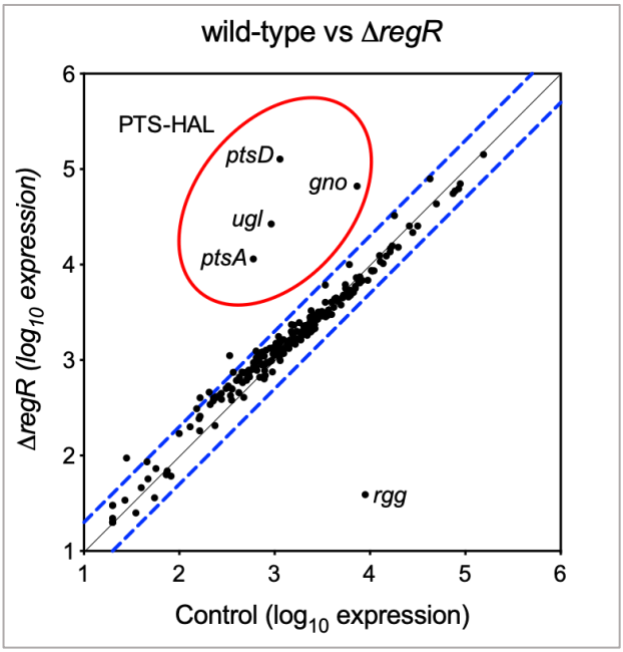



**Figure 4.3 Genes differentially expressed in the Δ*regR* mutant**

NanoString transcriptome analysis comparing the expression profiles of the wild-type strain PN4595-T23 and *regR* derivative strain (Δ*regR*) in biofilms on A549 cells. Data was analyzed for statistical significance using nCounter. The dashed blue lines represent 2-fold up- and down-regulation, respectively. The red circle highlights genes associated with hyaluronic acid processing, transport and metabolism.

**Table 4.3** Genes differentially repressed by RegR

| Locus | Fold | p-value | Function |
|---|---|---|---|
| SPD_0297 | -111.845 | 1.11E-04 | PTS system transporter subunit **IID** |
| peg.291 | -28.938 | 2.18E-04 | Glucuronyl hydrolase ***ugl*** |
| SPD_0293 | -19.202 | 3.56E-05 | PTS N-acetylgalactosamine transporter subunit **IIA** |
| SPR0290 | -9.099 | 1.62E-04 | gluconate 5-dehydrogenase - ***gno*** |
| lclAE005672.3_cds_AAK | -3.293 | 2.86E-01 | LPXTG-anchored hyaluronate lyase, ***hysA*** |

We previously reported that *vp1* positively regulates the expression of the locus PTS-HAL (13). Here, we confirm that *vp1* controls the expression of the PTS-HAL during biofilm mode of growth where the glucuronyl hydrolase-encoding gene *ugl* was one of the highest affected by *vp1*. No influence of *vp1* over *regR* or *hysA* was detected. Since no changes on RegR expression were detected, we conclude that VP1 influence on PTS-HAL is not via transcriptional regulation of *regR*, and the molecular mechanism remains an open question, which is currently being addressed by investigation of the VP1 receptor on the pneumococcus.

To test whether VP1 may influence PTS-HAL via a TCS, we took a candidate approach and tested TCS4. TCS04 was selected as there is evidence that it upregulates the RegR repressor (57). Thus, we hypothesized that TCS04 represses PTS-HAL expression via RegR. To test this hypothesis, we generated a TCS04 (Δ*pnpR*) mutant in the TIGR4 background. This mutant lacks the transcriptional regulator. We compared the expression of the genes *ugl* (second gene in the PTS-HAL operon) and SPR0289, a gene encoded in the operon upstream to PTS-HAL, as representants of the genes required for internalization and processing the HA, respectively. Experiments were performed in rich media. The TCS04 negatively regulates the expression of the PTS-HAL genes (**Figure 4.4**). However, no transcriptional changes were detected for *regR* or *vp1*. These data suggest that, in these conditions, a) TCS04 exerts its influence on PTS-HAL independent of levels of *vp1 and regR* and, b) TCS04 may influences the repression of PTS-HAL by controlling the activity of RegR. Further experiments are need to determine the mechanism by which TCS04 control on PTS-HAL.

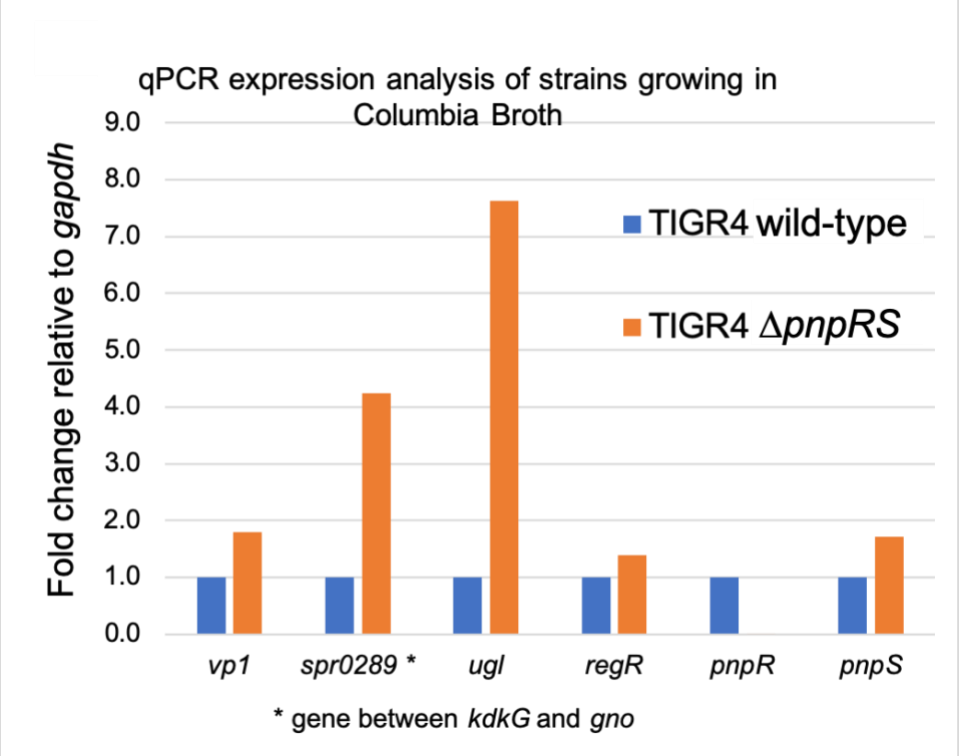



**Figure 4.4 TCS04 negatively regulates PTS-HAL**
Expression analysis of TIGR4 strains. Total RNA was isolated from the TIGR4 wild-type strains and the isogeneic *pnpR* mutant, grown in Columbia broth ($OD_{600}$ of 0.5). Expression was assessed by RT-qPCR. Note that SPR0289 correspond to the operon upstream of PTS-HAL and it is also involved in the processing of HA.

## 4.3 Discussion

During infection, the pneumococcus displays a biofilm mode of growth (58–61). Our findings indicate that *vp1* affects biofilm development on lung epithelial cells. We show that *vp1* has a profound effect on the global expression during growth as biofilms. The VP1 regulon in biofilms includes the HA processing genes previously characterized, genes involved in pneumococcal competence, as well the transcription factors Rgg144, Rgg939, and Rgg1952. We demonstrated those changes by using a custom-designed pneumococcal probe set, designed to capture the diversity in the pneumococcal processes bypassing allelic that covers the majority of the functional operon encoded in the pneumococcal genome. Our probe set was validated on the transcription factor

*regR*-deficient background that showed a high control over its cognate operons confirming previous reports (56).

Bacteria living in biofilms are phenotypically distinct from their planktonic counterparts. It has been reported that the transcriptional landscape of planktonic pneumococci differs from *in vitro* and *in vivo* pneumococcal biofilms (28,62). Further, bacterial infection initiated from dispersed biofilm displayed higher dissemination and virulence compared to those initiated from planktonic cultures, suggesting that the biofilm influences bacterial behavior and contributes to disease outcome (63,64). By studying *Streptococcus pyogenes*, Chang *et al*. demonstrated that biofilm capacity depends on extracellular peptide signaling (65). In Gram-positive bacteria, these behaviors mostly rely on signaling cascades regulated by ribosome-synthesized peptides (66,67). Many of the peptides that signal cell-to-cell can be categorized based on sequence features (68). VP1 is a member of a family of peptides with a conserved N-terminal sequence characterized by a double glycine motif, which directs its export into the extracellular milieu via ABC transporters (15,69). Further, it has been shown that VP1 controls attachment to epithelial cells and biofilm capacity (13,15)

The expression of *vp1* is controlled by the transcription factor Rgg144, which is involved in nutritional sensing (15,70). The signaling is initiated when the Rgg144 regulatory peptide SHP144 (short hydrophobic peptide 144) binds to Rgg144 activating it. Then, the pair Rgg144/SHP144 binds a predicted promoter upstream of the *vp1* operon. Our data indicate that VP1 induces Rgg144 and Rgg939, and it has also been shown that both transcription factors downregulate the expression of the genes required for the capsule synthesis (70). This observation suggests that VP1 cooperates with Rggs to downregulate the capsule, and consequently, enhance biofilm capacity: thicker bacterial capsule obstruct attachment by restricting exposure of bacterial receptors (71). However, and in stark contrast, Junges *et al*. determined that Rgg939 reduces biofilm capacity in the same strain. The group found that Rgg939 stimulates the expression operon formed by 12 genes initiated at the *shp939* gene while suggesting that this operon enhances polysaccharide synthesis and subsequently reduces biofilm formation (44). While these differences may be due to the experimental conditions (planktonic versus biofilm), more research will be required. It is clear that regulation of the Rggs is

highly condition dependent. We identified difference in Rgg144 regulation of the *vp1* between rich media and CDM supplemented with glucose (Chapter 2). Similarly, Zhi *et al*. identified differences in CDM supplemented with mannose or galactose relative to supplementation with glucose or acetyl-glucosamine (70). The connection between VP1 and Rgg939 and Rgg1952 remains to be dissected.

VP1 and Rgg144 pathways partially overlap, suggesting these molecules may be critical for initiation and maintenance of pneumococci biofilm. Our data, as well from others, indicate that CodY negatively regulates the expression of the *vp1* and its operon in response to nutrients abundance (72–75). *rgg144* expression is also negatively regulated by CodY in response to nutrients (76,77). In the absence of nutrients, CodY may release its inhibition from *vp1* and *rgg144,* enhancing its expression. While the *rgg144-vp1* interaction has been confirmed subsequently by others (70), *rgg144* and *vp1* may receive transcriptional input from other sources, and those inputs are likely determined by the environment, stage of the infection, as well as the tissue being colonizing.

Several genes related to biofilm capacity were upregulated by VP1 in our analysis. Pneumococcal biofilms express genes involved in cell-wall and isoprenoid biosynthesis, as well as purine and pyrimidine nucleotide biosynthesis (78). We detected several genes involved in the cell-wall synthesis, including *murE* and *murF* (79–81) and *pbp2X* (82), the late competence gene *comX* and *comW*, involved in the DNA update and recombination (61,83), and the bacteriocin-encoding genes *cibA* and *blpN* (84,85) and the lantibiotic-encoding gene *pneA2* (86). Genes associated with bacterial attachment were also upregulated. One of them was the cell-membrane metal-binding lipoprotein-encoding gene *psaA*, required for attachment (87), and the glycosidase-encoding gene *bgaA* as well as neuraminidase-encoding gene *nanA*, the last also required for biofilm formation (26) and colonization of the nasopharynx (88). Thus, the level of overlap between the VP1 regulon and genes required for biofilm formation is elevated, and worthy of future analyses.

Finally, we noted an association between RegR and novel genes. These interactions include a gene annotated as preprotein translocase SecY, and a

dehydrogenase located just upstream and in proximity to *rgg144* and with homology to UDP-glucose dehydrogenase involved in capsule synthesis (89). SecY is a core component of the SecYEG complex, the bacterial Sec translocon (90). It is revealing that components of the SecYEG translocon interact with the accessory proteins YidC (91) and YajC (92), the latter is also part of the PTS-HAL operon (Chapter 3). This leads to a question as to whether the RegR pathway connects with the bacterial secretion system in the pneumococcus via SecY or YajC.

In this study, we reveal a novel regulon involved in biofilm development. Our working model involved elements that sense nutrient availability and, in concert, influence the expression of the genes involved in the capsule synthesis. Our data indicate that *vp1* plays a pivotal role in pneumococcal biofilm development and behavior by orchestrating gene expression. While further studies are required to dissect the intricate interplay between VP1, its regulators, and its downstream effector, our finding shed light on this functional mechanism that contributes to pneumococcal colonization and virulence.

## 4.4 Material and Methods

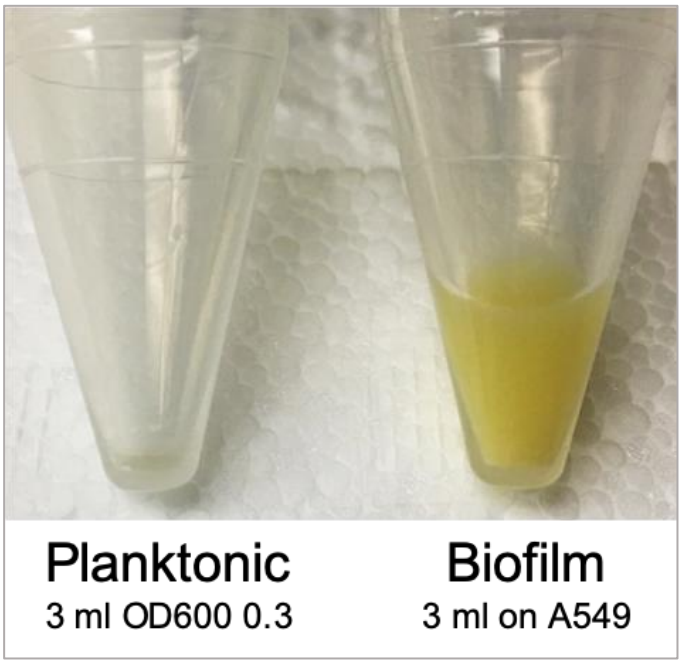
**Bacterial strains, construction, and media.** Wild-type *S. pneumoniae* strain PN4595-T23 (GenBank ABXO01), (93–96) were used to generate the mutant Δ*vp1* and the overexpressor Δ*vp1:vp1* and their construction was reported previously (15). The TIGR4(The Institute for Genomic Research, Aaberge et al., 1995) strain was used to generate the Δ*vpo mutant*, and its construction was reported previously (13). For growth on solid media, strains were streaked onto Trypticase Soy Agar II plates (TSA) containing 5% sheep blood (BD BBL, New Jersey, USA). For growth in liquid culture, colonies from a frozen stock were grown overnight on TSA plates and inoculated into Columbia broth (Remel Microbiology Products, Thermo Fisher Scientific, USA. Cultures were incubated at 37°C and 5% $CO_2$ without shaking.

**Mammalian cells and media conditions.** Epithelial lung human carcinoma A549 were propagated and maintained at 37ºC and supplemented of 5% carbon dioxide at pH 7 in DMEM media contained 1% L-glutamine, 5% fetal bovine serum (FBS) without antibiotics. Prior to infection day, fresh media was added to cells.

**RNA isolation, purification, qRT-PCR, and NanoString analysis.** Pneumococcal biofilms were collected 16 h post-inoculation by carefully aspirating the supernatants. Then, the biofilms were washed with PBS and collected on RNALater (Thermofisher), and pellets were lysed with 1x lysis mix containing 2 mg $ml^{-1}$ of proteinase K, 10 mg $ml^{-1}$ of lysozyme, and 20 µg $ml^{-1}$ mutanolysin in TE buffer (10 mM Tris·Cl, 1 mM EDTA, pH 8.0) for 20 min. Total RNA was isolated using the RNeasy (Zymo Research) following manufacturer instructions. Contaminant DNA was removed by incubating total RNA samples with DNase (2U/µl) at 37°C for at least 15 min. Approximately 80 ng of total RNA was used for peptide-only NanoString chip (small sample of 27 probes). For main NanoString analysis (Δ*vp1* biofilms versus wild-type), 200 ng of total RNA was used. A detailed protocol can be found in Chapter 7 (98), and the online version of this protocol is available at https://link.springer.com/protocol/10.1007/978-1-4939-9199-0_7

To quantify the gene expression in liquid cultures, RT reaction was performed using one µg of total RNA was retro transcribed using SuperscriptVILO kit for 1 h. Products were amplified using OneTag polymerase (New England Biolabs). For quantitative RT-PCR analysis, 5 ng of total cDNA was subjected to real-time PCR using PowerUp SYBR Green Master Mix in the ABI 7300 Real Time PCR system (Applied Biosystems) according to the manufacturer's instructions. All qRT-PCR amplification was normalized to pneumococcal *gapdh* and expressed as fold change with respect to wild-type strain and untreated bacterial cells. qPCR primers are listed in **Suppl. Table 4.2**. Primers were obtained from IDT.

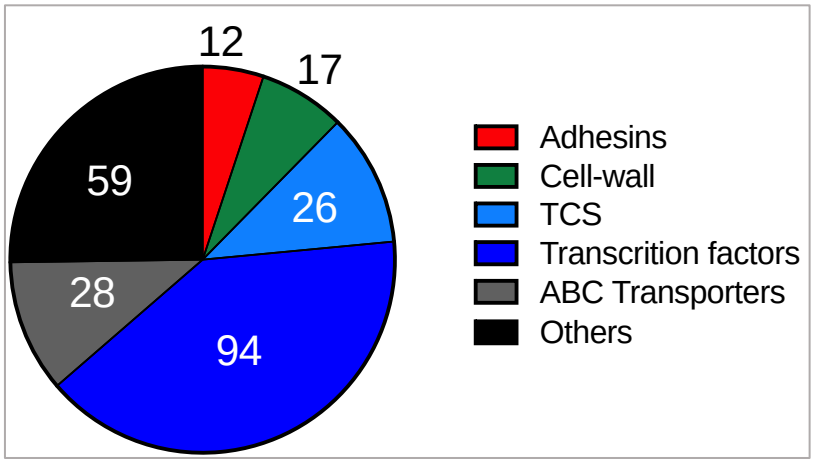
**Biofilm formation assays.** Starter cultures were started from pneumococcal strains inoculated overnight on TSA plates. Cultures were initiated on Columbia broth at an $OD_{600}$ close to 0.01. Bacterial grow was constantly monitored until it reached and $OD_{600}$ of 0.025. Then, 100 ml of each strain was inoculated directly on confluent A549 cells. To promote biofilm growth, the plates were incubated aerobically in humidified conditions at 37º C and 5% carbon dioxide for 16 h. Add the end of incubation, the supernatants were carefully aspirated, and biofilms were washed with PBS to remove non-adherent and or weakly adherent bacteria. Subsequently, samples were fixed with 4% PFA. The fixing solution was removed; samples were washed twice with PBS and prepared for confocal microscopy.




## 4.5 Supplementary Figures

**Supplementary Figure 4.1 Pie chart distribution of the Gene Ontology analysis of our custom probe set**

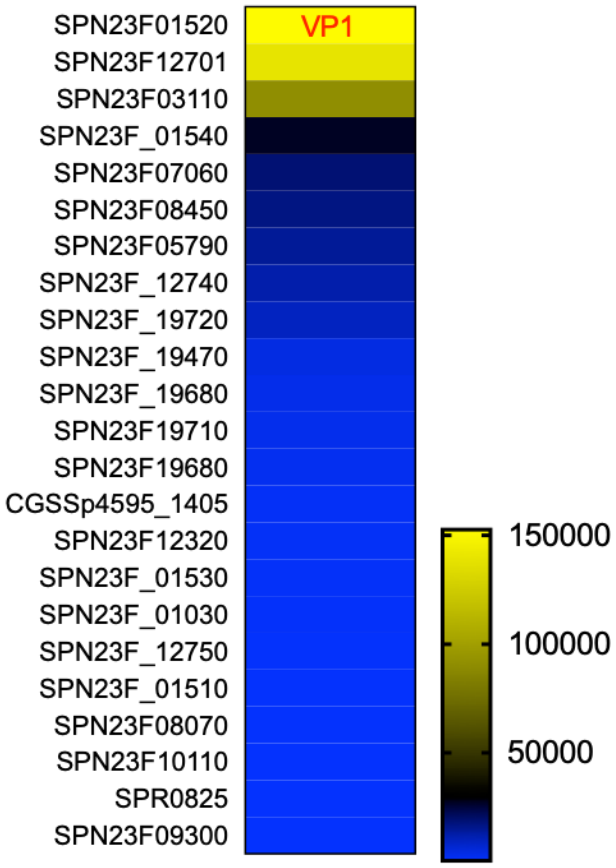
Representative image of the total pneumococcal biomass isolated from 3 ml of planktonic growth at an $OD_{600}$ of 0.3 (left) and from a biofilm on A549 epithelial cells growing in a 60 mm dish covered with 3 ml of media. Biofilm was recovered 16 h post inoculation.




**Supplementary Figure 4.2 Pie chart distribution of the Gene Ontology analysis of our custom probe set**

Pie chart distribution representation of the shows the NanoString custom probe set. Figure depict the most represent types of gene function analyzed covers 234 different genes.

| Locus | Name | Counts |
|---|---|---|
| SPN23F_01520 | VP1 | 152544.08 |
| SPN23F_12701 | SPN23F_12701 | 138419.29 |
| SPN23F_03110 | Clp_protease | 90789.04 |
| SPN23F_01540 | ABC_PMEN1 (Vp1 operon) | 26951.25 |
| SPN23F_07060 | Pep715 | 18744.92 |
| SPN23F_08450 | Arg_deiminase | 17000.46 |
| SPN23F_05790 | C5a_peptidase | 14441.62 |
| SPN23F_12740 | PhrA2 | 11937.72 |
| SPN23F_19720 | LanM_PMEN1 | 9109.80 |
| SPN23F_19470 | Ply | 4388.60 |
| SPN23F_19680 | TprA | 3250.44 |
| SPN23F_19710 | LanAcore_1969 | 2934.60 |
| SPN23F19680 | PhrA | 2297.80 |
| CGSSp4595_1405 | Pep1422 | 1358.58 |
| SPN23F_12320 | LanL_protease1231 | 1270.07 |
| SPN23F_01530 | CAAX854 (VP1 operon) | 1004.50 |
| SPN23F_01030 | MSM | 948.21 |
| SPN23F_12750 | TprA2 | 549.69 |
| SPN23F_01510 | MutR/Rgg (VP1 operon) | 510.05 |
| SPN23F_08070 | RM_815 | 418.71 |
| SPN23F_10110 | Redox_1021 | 412.84 |
| SPR0825 | peg853 | 394.89 |
| SPN23F_09300 | HTriad | 369.90 |




**Supplementary Figure 4.3 Analysis of peptide expression on biofilms on A549 cells**

NanoString RNA count averages for the set of predicted secreted peptides in PMEN1 strain PN4595-T23 in 16 h biofilms on lung cancer A549 epithelial cells (left panel) and graphic representation of the data ID's are represented in the model PMEN1 genome (strain ATCC 700669, GenBank FM211187), and corresponding working gene names are listed. Note vp1 is the highest expressed gene in the set.

## 4.6 Supplementary tables

### Supplementary Table 4.1 Biofilm NanoString analysis in the PN4595-T23 Δ*vp1*

| Locus | Locus | Fold Change | Function |
|---|---|---|---|
| peg.140 | SPN23F_RS00805 | 0.002 | VP1 |
| SPD_RS04090 | SPD_RS04090 | 0.016 | DeoR/GlpR transcriptional regulator fructose operon |
| peg.291 | SPD_RS01570 | 0.022 | ugl |
| peg.395 | SP_RS00640 | 0.030 | fratricide two-peptide bacteriocin subunit CibA |
| SPD_RS08045 | SPD_RS08045 | 0.046 | GntR family transcriptional regulator |
| comW | SPN23F_RS00140 | 0.050 | recombinase RecX |
| SPR0290 | SPD_RS01560 | 0.053 | phosphate-binding protein PstS 1, gluconate 5-dehydrogenase - gno |
| peg.138 | SPD_RS00795 | 0.054 | UDP-glucose/GDP-mannose dehydrogenase family, central domain protein |
| SPD_RS08660 | SPD_RS08660 | 0.056 | TetR/AcrR family transcriptional regulator |
| SPD_RS09875 | SPD_RS09875 | 0.056 | Transcriptional regulator, MarR family |
| lclAE005672.3_cds_ABC75807.1_282 | SP_RS01795 | 0.061 | endo-alpha-N-acetylgalactosaminidase, hypothetical |
| lclAE005672.3_cds_AAK75941.1_1006 | SP_RS09270 | 0.072 | iron chelate uptake ABC transporter family permease subunit |
| SPD_0123 | SPN23F_RS11775 | 0.073 | vargininosuccinate lyase - argH, arginine synthesis |
| SPD_RS05070 | SPD_RS05070 | 0.075 | putative transcriptional regulator |
| SP_1833 | SP_RS09095 | 0.075 | multi-ligand-binding adhesin PfbA, cell wall |
| lacZ | V00296 | 0.084 | beta-galactosidase |
| SPD_RS05570 | SPD_RS05570 | 0.088 | transcription antiterminator LacT |
| peg.1721 | SPD_RS08090 | 0.092 | PTS beta-glucoside transporter subunit IIBCA |
| SPD_RS06610 | SPD_RS06610 | 0.093 | DNA modification methyltransferase M.XbaI, transcriptional repressor, HTH binding I |
| SPD_RS09720 | SPD_RS09720 | 0.095 | BglG family transcriptional regulator, HTHt domain containing-protein |
| SPD_RS06520 | SPD_RS06520 | 0.096 | Spx/MgsR family RNA polymerase-binding regulatory protein |
| SPD_RS05550 | SPD_RS05550 | 0.097 | DeoR/GlpR transcriptional regulator lactose PTS repressor |
| SPD_RS00505 | SPD_RS00505 | 0.102 | padR transcriptional regulator |
| SPD_RS03380 | SPD_RS03380 | 0.105 | CopY/TcrY family copper transport repressor |
| SPD_1517 | SPD_RS08000 | 0.107 | hypothetical protein |
| SPD_RS11310 | SPD_RS11310 | 0.111 | helix-turn-helix transcriptional regulator, cI repressor |
| SPD_0293 | SPD_RS01565 | 0.112 | PTS N-acetylgalactosamine transporter subunit IIA |
| SPD_RS03665 | SPD_RS03665 | 0.120 | PadR family transcriptional regulator |
| SPD_RS09185 | SPD_RS09185 | 0.121 | type II restriction-modification system regulatory protein |
| SPD_0142 | SPD_RS00790 | 0.121 | N-acetylglucosaminylphosphatidylinositol deacetylase superfamily pdgA |
| ATCC700669_1499 | SPN23F_RS07350 | 0.122 | cell wall surface anchor family protein |
| SPD_RS03325 | SPD_RS03325 | 0.128 | transcriptional regulator%2C TENA/THI-4 family protei |
| SPN23F_RS04990 | SPN23F_RS04990 | 0.129 | Rgg/GadR/MutR family transcriptional regulator |
| peg.992 | SPN23F_RS05010 | 0.132 | ABC transporter ATP-binding protein |
| comX | SPD_RS09600 | 0.135 | comX2 domain protein, TF |
| SPD_RS10310 | SPD_RS10310 | 0.136 | transcriptional regulator, BglG family |
| SPD_RS06720 | SPD_RS06720 | 0.136 | GntR family transcriptional regulator |
| SPD_RS04350 | SPD_RS04350 | 0.137 | AraC family transcriptional regulator |
| ATCC700669_1717 | SPD_RS07825 | 0.137 | UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase domain protein  murF |
| SPD_1504 | SPD_RS07935 | 0.140 | nanA |
| SPD_RS07435 | SPD_RS07435 | 0.146 | GTP-sensing transcriptional pleiotropic repressor MarR family - CodY? |
| SPD_RS01935 | SPD_RS01935 | 0.153 | Transcriptional regulator MtlR |
| SPD_RS07250 | SPD_RS07250 | 0.170 | M protein trans-acting positive transcriptional regulator |
| SPD_RS01225 | SPD_RS01225 | 0.172 | AraC family transcriptional regulator |
| SPD_RS04880 | SPD_RS04880 | 0.195 | putative transcriptional regulator |
| SPD_RS10620 | SPD_RS10620 | 0.197 | CtsR family transcriptional regulator |
| SPD_RS02740 | SPD_RS02740 | 0.201 | Transcriptional regulator |
| D39_436 | SPD_RS02190 | 0.214 | glutamine ABC transporter substrate-binding protein |
| SPD_RS10485 | SPD_RS10485 | 0.222 | DeoR/GlpR transcriptional regulator , fucose operon |
| SPD_0297 | SPD_RS01585 | 0.223 | PTS system transporter subunit IID |
| SPD_RS10030 | SPD_RS10030 | 0.233 | arginine repressor |
| ATCC700669_124 | SPD_RS00560 | 0.233 | ABC transporter membrane-spanning permease-amino acid transport domain protei |
| SPD_1499 | SPD_RS07910 | 0.234 | nanB |
| CzcD | SP_RS09215 | 0.238 | Cation efflux pump trabnstporter |
| SPD_RS06655 | SPD_RS06655 | 0.244 | AraC family transcriptional regulator |
| peg.481 | SPN23F_RS02485 | 0.253 | two-peptide bacteriocin subunit BlpN |
| SPD_RS08530 | SPD_RS08530 | 0.254 | PhoU family transcriptional regulator, pho uptake |
| SP_0268 | SP_RS01315 | 0.261 | surface-anchored pullulanase spuA |
| SPD_2065 | SPD_RS10840 | 0.269 | competence-stimulating peptide ComC |
| SP_0057 | SP_RS00325 | 0.271 | beta-N-acetylhexosaminidase strH |
| VP1 | SPD_RS00805 | 0.272 | VP1 |
| SPD_RS08880 | SPD_RS08880 | 0.275 | bacterial regulatory helix-turn-helix s, AraC family protein, Msm operon regulatory pr |
| SPD_0525 | SPD_RS02815 | 0.281 | VncS histidine kinase TCS10 |
| SPD_RS04965 | SPD_RS04965 | 0.286 | transcriptional regulator |
| peg.1221 | SPN23F_RS06205 | 0.292 | N-acetylmannosamine-6-phosphate 2-epimerase, sialic acid metabolis, |
| SPD_RS10090 | SPD_RS10090 | 0.293 | helix-turn-helix family protein, transcriptional regulator |
| SPD_RS10505 | SPD_RS10505 | 0.296 | iron dependent repressor, N-terminal DNA binding domain protein [ |
| lclAE005672.3_cds_AAK74795.1_487 | SP_RS03175 | 0.300 | beta-galactosidase |
| lclAE005672.3_cds_AAK74656.1_369 | SP_RS02450 | 0.303 | endo-beta-N-acetylglucosaminidase, N-acetylmuramoyl-L-alanine amidase |

| | | | |
|---|---|---|---|
| SPD_RS02290 | SPD_RS02290 | 0.305 | Uncharacterised protein, potassium uptake protein, Trk system potassium transporte |
| SPD_RS09115 | SPD_RS09115 | 0.306 | Transcriptional regulator MarR family |
| lclAE005672.3_cds_AAK74420.1_191 | SP_RS01160 | 0.319 | phosphoglycerate mutase family protein , uncharacterized protein |
| SPD_RS10250 | SPD_RS10250 | 0.320 | transcriptional regulator, helix-turn-helix transcriptional regulator |
| SPD_2064 | SPD_RS10835 | 0.330 | putative sensor histidine kinase ComD |
| SPD_RS09230 | SPD_RS09230 | 0.339 | Transcriptional regulator |
| SPD_RS00340 | SPD_RS00340 | 0.342 | GntR family transcriptional regulator |
| lclAE005672.3_cds_AAK74544.1_289 | SP_RS01845 | 0.351 | choline-binding protein F |
| SPD_RS09665 | SPD_RS09665 | 0.351 | transcriptional regulator, GntR family protein |
| peg.2139 | SPN23F_RS11040 | 0.357 | caax amino protease family protein |
| peg.853 | SPN23F_RS04320 | 0.360 | SP_0924 family bacteriocin-like peptide |
| SPD_RS00585 | SPD_RS00585 | 0.363 | trancriptional regulator |
| ssbB | SPD_RS09040 | 0.368 | single-strand DNA-binding protein, based on B.sub. |
| lclAE005672.3_cds_AAK75944.1_1009 | SP_RS09285 | 0.374 | iron-compound ABC transporter iron-compound-binding protein |
| SPD_RS02375 | SPD_RS02375 | 0.382 | Transcriptional regulator MerR |
| SP_1992 | SP_RS10010 | 0.383 | LPXTG-motif cell wall anchor domain |
| SPD_RS10275 | SPD_RS10275 | 0.392 | transcriptional activator, Rgg/GadR/MutR family protein |
| SP_1772 | SP_RS12370 | 0.401 | LPXTG cell wall anchor domain-containing protein, comifin |
| SPD_RS10830 | SPD_RS10830 | 0.403 | DNA binding response regulator |
| SP_2201 | SP_RS11240 | 0.405 | choline binding protein D |
| SPD_RS00800 | SPD_RS00800 | 0.408 | rgg type |
| SPD_1726 | SPD_RS09125 | 0.415 | cholesterol-dependent cytolysin pneumolysin - ply, toxin |
| SPD_RS03080 | SPD_RS03080 | 0.420 | DNA bindinfg response regulator |
| SPD_RS02695 | SPD_RS02695 | 0.432 | Transcriptional antiterminaotr LicT |
| lclAE005672.3_cds_AAK75086.1_732 | SP_RS04785 | 0.432 | endo-beta-N-acetylglucosaminidase LytB, N-acetylmuramoyl-L-alanine amidase |
| SPD_RS01680 | SPD_RS01680 | 0.434 | LytR |
| SPD_RS10050 | SPD_RS10050 | 0.435 | transcriptional regulatory protein PhoP |
| SPD_1514 | SPD_RS07985 | 0.440 | ABC Trasnporter |
| SPD_RS09605 | SPD_RS09605 | 0.448 | transcription termination/antitermination protein NusG |
| SP_1948 | SP_RS09795 | 0.453 | two-peptide bacteriocin subunit PneA2 |
| ATCC700669_1180 | SPN23F_RS05810 | 0.465 | putative 5-methylcytosine restriction system,catalytic subunit |
| peg.303 | SPN23F_RS01600 | 0.468 | penicillin-binding protein 2X |
| SPD_RS04355 | SPD_RS04355 | 0.472 | LysR family transcriptional regulator |
| SPD_RS01235 | SPD_RS01235 | 0.480 | DeoR family transcriptional regulator |
| D39_1359 | SPD_RS06725 | 0.486 | membrane protein - murE |
| lclAE005672.3_cds_AAK75729.1_974 | SP_RS08145 | 0.489 | metal ABC transporter substrate-binding lipoprotein/adhesin PsaA |
| SPD_RS10605 | SPD_RS10605 | 0.490 | DNA-binding response regulator |
| SPD_RS08380 | SPD_RS08380 | 0.496 | LacI family transcriptional regulator, sucrose operon represor |
| SPD_RS08640 | SPD_RS08640 | 0.502 | LacI family transcriptional regulator |
| lclAE005672.3_cds_AAK74545.1_290 | SP_RS01850 | 0.511 | N-acetylmuramoyl-L-alanine amidase family protein Choline binding protein J |
| SPD_RS09210 | SPD_RS09210 | 0.513 | cell envelope-related, transcriptional attenuator common biofilm regulatory protein A |
| SPD_RS02810 | SPD_RS02810 | 0.531 | resposnse regulator |
| SPD_RS00895 | SPD_RS00895 | 0.531 | response regulator |
| SPD_RS05205 | SPD_RS05205 | 0.538 | redox-sensing transcriptional repressor Rex |
| SPD_1909 | SPD_RS10055 | 0.539 | histidine kinase PnpS |
| SPD_RS05750 | SPD_RS05750 | 0.541 | DNA-binding response regulator |
| SPD_RS02485 | SPD_RS02485 | 0.562 | LytR family transcriptional regulator [ |
| SPD_RS04250 | SPD_RS04250 | 0.565 | S1 RNA-binding domain-containing protein, putative transcription accessory protein |
| SPD_RS00030 | SPD_RS00030 | 0.567 | helicase conserved C-terminal domain protein, transcription repair, helicase |
| SPD_RS08005 | SPD_RS08005 | 0.569 | transcriptional activator, Rgg/GadR/MutR family protein |
| SPD_RS02255 | SPD_RS02255 | 0.571 | Transcriptional regulator ROK |
| SPD_RS08040 | SPD_RS08040 | 0.573 | NrdR family transcriptional regulator represor |
| SPD_1384 | SPD_RS07285 | 0.574 | cation effluc pump |
| SPD_RS02430 | SPD_RS02430 | 0.575 | molecular chaperone DnaK - mreB |
| SPD_0082 | SPD_RS00435 | 0.581 | sensor histidine kinase |
| SPD_RS09080 | SPD_RS09080 | 0.582 | transcriptional regulator, DNA binding protein |
| SPD_RS09440 | SPD_RS09440 | 0.584 | helix-turn-helix transcriptional regulator, activator, transposase like |
| SPD_RS08105 | SPD_RS08105 | 0.590 | transcriptional regulator |
| SPD_RS08485 | SPD_RS08485 | 0.611 | LacI family transcriptional regulator |
| SPD_RS08260 | SPD_RS08260 | 0.611 | putative MarR-family transcriptional regulator |
| SPD_RS08885 | SPD_RS08885 | 0.613 | bifunctional biotin--[acetyl-CoA-carboxylase] ligase/biotin operon repressor BirA |
| SPD_RS07615 | SPD_RS07615 | 0.615 | DNA-binding response regulator |
| peg.1672 | SPN23F_RS08440 | 0.615 | penicillin-binding protein 2 |
| SPD_RS01615 | SPD_RS01615 | 0.629 | helix-turn-helix transcriptional regulator |
| SPD_RS03450 | SPD_RS03450 | 0.640 | TetR/AcrR family transcriptional regulator |
| peg.854 | SPN23F_RS04325 | 0.644 | caax amino protease family protein |
| peg.1422 | SP_RS07145 | 0.650 | hypothetical protein |
| SPD_1445 | SPD_RS07610 | 0.654 | Histidine kinase |
| SPD_RS08650 | SPD_RS08650 | 0.654 | MerR family transcriptional regulator |
| peg.1967 | SP_RS09785 | 0.655 | Secreted hypothetical protein |
| SPD_RS09390 | SPD_RS09390 | 0.658 | pur operon repressor PurR, reprosor of biosynthetic genes |
| SPD_RS07630 | SPD_RS07630 | 0.660 | Rrf2 family transcriptional regulator, group III |
| SPD_RS09010 | SPD_RS09010 | 0.664 | RecX family transcriptional regulator |
| lclAE005672.3_cds_AAK75147.1_784 | SP_RS05120 | 0.664 | iron-compound ABC transporter, iron compound-binding protein |
| SPD_RS02555 | SPD_RS02555 | 0.669 | Transcriptional terminator NusA |
| SPD_RS08700 | SPD_RS08700 | 0.671 | MarR family transcriptional regulator |
| peg.2141 | SPN23F_RS11050 | 0.671 | hypothetical protein |
| SPD_RS05640 | SPD_RS05640 | 0.673 | transcriptional regulator of arginine metabolism expression |
| SPD_RS01600 | SPD_RS01600 | 0.690 | regR PTS |

| | | | |
|---|---|---|---|
| SPD_RS03150 | SPD_RS03150 | 0.692 | LysR tgranscriptional regulator |
| SPD_RS01805 | SPD_RS01805 | 0.711 | Cell Division Regulator GpsB |
| lclAE005672.3_cds_AAK75263.1_877 | SP_RS05695 | 0.715 | immunoglobulin A1 protease |
| Spr1707 | SPD_RS08840 | 0.717 | oligopeptide-binding protein AmiA |
| lclAE005672.3_cds_AAK74270.1_70 | SP_RS00425 | 0.719 | cell wall surface anchor family protein (pseudogene) |
| SPD_RS00600 | SPD_RS00600 | 0.728 | araC |
| LcpA | SPN23F_RS06435 | 0.760 | Phosphotransferase That Mediates Glycosylation of a Cell Wall-Anchored Protein |
| SPD_RS00430 | SPD_RS00430 | 0.774 | response regulator |
| ATCC700669_1092 | SPN23F_RS05375 | 0.776 | , IgA1 protease multipass membrane prot |
| ATCC700669_1093 | SPN23F_RS05380 | 0.781 | Select seq emb\|VJV29072.1\| zinc metalloprotease |
| bric upstream | SPD_RS02085 | 0.797 | acetyl co-enzyme A carboxylase carboxyltransferase alpha subunit |
| RitR | SPN23F_RS01795 | 0.802 | response regulator |
| ATCC700669_622 | SPN23F_RS03065 | 0.804 | ZmpB |
| SP_2190 | SP_RS11185 | 0.812 | surface protein PspC, choline-binding protein CbpA |
| SPD_0351 | SPD_RS01865 | 0.823 | histidine kinase |
| SPD_RS07640 | SPD_RS07640 | 0.824 | iron-dependent transcriptional regulator |
| SPD_RS10205 | SPD_RS10205 | 0.842 | periplasmic binding proteins and sugar binding domain of the LacI |
| SPD_RS02490 | SPD_RS02490 | 0.843 | response regulator |
| lclAE005672.3_cds_AAK74791.1_483 | SPN23F_RS02965 | 0.844 | S8 family serine peptidase |
| SPD_RS05995 | SPD_RS05995 | 0.871 | uracil phosphoribosyltransferase PyrR |
| SPD_0575 | SPD_RS03085 | 0.877 | sensor histidine kinase |
| SPD_RS01870 | SPD_RS01870 | 0.891 | response regulator |
| SPD_RS05785 | SPD_RS05785 | 0.906 | transcription repressor NadR DeoR family transcriptional regulator |
| SPD_RS03760 | SPD_RS03760 | 0.929 | zeptation ring formation EzrA |
| SPD_RS06365 | SPD_RS06365 | 0.929 | YlbF/YmcA family competence regulator AroA, EPSP |
| lclAE005672.3_cds_AAK74491.1_247 | SPD_RS01535 | 0.930 | LPXTG-anchored hyaluronate lyase, hysA on HA PTS operon |
| ATCC700669_758 | SPD_RS03740 | 0.939 | RodA |
| lclAE005672.3_cds_AAK74622.1_344 | SP_RS02280 | 0.951 | cell wall surface anchor family protein no hit 1313.153 |
| SPD_RS03430 | SPD_RS03430 | 0.991 | regulator of the efflux pump PmrA |
| SPD_RS07845 | SPD_RS07845 | 0.992 | MurR/RpiR family transcriptional regulator |
| SPD_RS05010 | SPD_RS05010 | 1.000 | Rgg/GadR/MutR family transcriptional regulator |
| SPR1403 | SPD_RS07245 | 1.000 | putative collagen-like surface-anchored protein - pclA |
| SP_1492 | SP_RS07345 | 1.000 | cell wall surface anchor family protein |
| lclAE005672.3_cds_AAK74258.1_59 | SP_RS00375 | 1.000 | N-acetylmuramoyl-L-alanine amidase family protein, choline binding protein |
| lclAE005672.3_cds_AAK74303.1_93 | SP_RS00595 | 1.000 | surface protein A - pspA |
| lclAE005672.3_cds_AAK74623.1_345 | SP_RS02285 | 1.000 | isopeptide-forming domain-containing fimbrial protein |
| lclAE005672.3_cds_AAK74624.1_346 | SP_RS02290 | 1.000 | LPXTG cell wall anchor domain-containing protein, pili surface protein |
| lclAE005672.3_cds_AAK74809.1_499 | SP_RS03260 | 1.000 | zinc metalloprotease ZmpB |
| lclAE005672.3_cds_AAK74812.1_502 | SP_RS03275 | 1.000 | peg606, Endo-beta-N-acetylglucosaminidase |
| spr1194 | SPD_RS06185 | 1.000 | oligopeptide ABC transporteroligopeptide-binding protein |
| SPD_RS03715 | SPD_RS03715 | 1.011 | DNA-binding response regulator CiaR |
| SPD_1084 | SPD_RS05745 | 1.026 | Select seq gb\|EJG55365.1\| sensory box histidine kinase YycG domain protein |
| SPD_0469 | SPD_RS02495 | 1.067 | BlpH histidine kinase TCS13 |
| SPD_RS05325 | SPD_RS05325 | 1.071 | putative DNA-binding protein Rgg/GadR/MutR family transcriptional regulator |
| SPD_1799 | SPD_RS09505 | 1.072 | histidine kinase |
| SP_1937 | SP_RS09740 | 1.160 | N-acetylmuramoyl-L-alanine amidase family protein, autolysin - lytA |
| SPD_RS06475 | SPD_RS06475 | 1.183 | phosphate transport system regulatory protein PhoU |
| bric | SPD_RS02090 | 1.188 | bric |
| SPD_RS01240 | SPD_RS01240 | 1.200 | DNA-binding transcriptional regulator |
| peg.1274 | SPN23F_RS06450 | 1.279 | lantibiotic ABC transporter ATP-binding protein |
| RSH | SPD_RS07685 | 1.343 | bifunctional (p)ppGpp synthase/hydrolase RelA domain protein - relA |
| lclAE005672.3_cds_AAK75054.1_702 | SP_RS04600 | 1.345 | choline binding protein E CbpE or cbpA in 1313.153 |
| peg.1270 | SPN23F_RS06425 | 1.364 | lantibiotic transport/processing protein SlbT ABC transporter transmembrane r |
| lclAE005672.3_cds_AAK74260.1_61 | SP_RS00380 | 1.379 | zinc metalloprotease ZmpC, YSIRK containing peptide |
| lclAE005672.3_cds_AAK74556.1_299 | SPD_RS01895 | 1.442 | choline binding protein G, serine protease, V8-like Glu-specific endopeptidase cbpG |
| SPD_RS10815 | SPD_RS10815 | 1.468 | TetR/AcrR family transcriptional regulator |
| SPD_2019 | SPD_RS10600 | 1.520 | signal transduction histidine kinase |
| peg.1275 | peg_1275 | 1.548 | Uncharacterised protein putative membrane protein, peptide, phra2 |
| SPR1995 | SPD_RS10590 | 1.580 | choline-binding protein A CbpA |
| OppA | SP_RS01790 | 1.583 | AliA |
| peg.37 | SPD_RS00240 | 1.604 | hypothetical protein |
| ATCC700669_1554 | SPN23F_RS07630 | 1.667 | collagen-like surface-anchored protein sclB? |
| SPD_0702 | SPD_RS03720 | 1.767 | sensor histidine kinase |
| SPD_RS09500 | SPD_RS09500 | 1.864 | response regulator transcription factor [Streptococcus pneumoniae] |
| AdcA | SPD_RS10490 | 2.044 | ABC zinc transporter |
| SP_1573 | SP_RS07755 | 2.066 | 1,4-beta-N-acetylmuramidase, lytC |
| SPD_RS04470 | SPD_RS04470 | 2.143 | TfoX/Sxy family protein [Streptococcus pneumoniae] |
| SPD_0157 | SPD_RS00865 | 2.160 | histidine kinase |
| SPD_RS09120 | SPD_RS09120 | 2.217 | glucose-1-phosphate adenylyl transferase - glgC, sugar metabolism |
| peg.1271 | SPN23F_RS06430 | 2.268 | lantibiotic transport/processing ATP-binding protein |
| SPD_RS07075 | SPD_RS07075 | 2.289 | transcription elongation factor GreA |
| SPD_RS00870 | SPD_RS00870 | 2.311 | resposne regulator |
| SPD_RS04165 | SPD_RS04165 | 2.346 | arginine repressor |
| SPD_RS08430 | SPD_RS08430 | 2.403 | helix-turn-helix transcriptional regulator |
| peg.606 | SPN23F_RS03080 | 2.491 | Endo-beta-N-acetylglucosaminidase, lysozyme lytC or lytD |
| lclAE005672.3_cds_AAK74557.1_300 | SP_RS01930 | 2.496 | choline binding protein F |
| peg.335 | SPN23F_RS01755 | 2.607 | penicillin-binding protein 1A |
| AliB | SPD_RS07140 | 2.771 | AliB oligopeptide ABC transporter |
| D39_623 | SPD_RS03095 | 3.122 | ZmpB |
| SPD_RS08805 | SPD_RS08805 | 3.434 | trehalose operon repressor |
| D39_1092 | SPD_RS05425 | 3.684 | Immunoglobulina A1 protease |
| SPD_0535 | SPD_RS02870 | 4.508 | serine/alanine-adding enzyme MurM |
| SPD_RS06345 | SPD_RS06345 | 5.112 | polyisoprenyl-teichoic acid--peptidoglycan teichoic acid transferase Psr |
| peg.1276 | SPN23F_RS06460 | 5.606 | helix-turn-helix transcriptional regulator |
| AliA | SPD_RS01780 | 8.566 | oligo peptide ABC transporter |
| SPD_RS02030 | SPD_RS02030 | 12.158 | Ranscriptional regulator MarR |

**Supplementary Table 4.2 Primers used in this study**

| Strain | Gene | Primer |
|---|---|---|
| TIGR4 | VP1 | TAT GGG AGG TGG TGG AAG AA |
| | | AGC TCC ACG ATT AGG ACT CA |
| TIGR4 | PnpR | GCT TTG GCA GAA ACA GAA CC |
| | | CTC TCA GCC GCT TAC AAA CT |
| TIGR4 | PnpS | CTC AGT ATG TGT CTG GCC TAA A |
| | | GGT TAG GTG CGA CAG AAG AA |
| TIGR4 | RegR | CCA GAA TGG CTA CCA GGT AAT G |
| | | TAA AGC CGT CTA CTC CCA AGA |
| TIGR4 | ugl | GGA GAA CTG TGG TTG GCT TAT |
| | | GAC ACG ATC CAG GAA AGA AAG A |
| TIGR4 | spr0289 (equivalent) | CGT GGA GAA GAA GGA GAA AGT C |
| | | GCC GGT AAC AAC AAA GTC AAC |

## 4.7 References


1. Kan B, Ries J, Normark BH, Chang FY, Feldman C, Ko WC, et al. Endocarditis and pericarditis complicating pneumococcal bacteraemia, with special reference to the adhesive abilities of pneumococci: Results from a prospective study. Clin Microbiol Infect. 2006;12(4):338–44.
2. Hall-Stoodley L, Hu FZ, Gieseke A, Nistico L, Nguyen D, Hayes J, et al. Direct Detection of Bacterial Biofilms on the Middle-Ear Mucosa of Children With Chronic Otitis Media. JAMA J Am Med Assoc. 2006 Jul 12;296(2):202–11.
3. Hoa M, Syamal M, Sachdeva L, Berk R, Coticchia J. Demonstration of nasopharyngeal and middle ear mucosal biofilms in an animal model of acute otitis media. Ann Otol Rhinol Laryngol. 2009 Apr;118(4):292–8.
4. Post JC, Hiller NL, Nistico L, Stoodley P, Ehrlich GD. The role of biofilms in otolaryngologic infections: update 2007. Curr Opin Otolaryngol Head Neck Surg. 2007 Oct;15(5):347–51.
5. Hall-Stoodley L, Costerton JW, Stoodley P. Bacterial biofilms: from the Natural environment to infectious diseases. Nat Rev Microbiol. 2004 Feb;2(2):95–108.
6. Chao Y, Marks LR, Pettigrew MM, Hakansson AP. Streptococcus pneumoniae biofilm formation and dispersion during colonization and disease. Front Cell Infect Microbiol. 2014;4:194.
7. Croucher NJ, Harris SR, Barquist L, Parkhill J, Bentley SD. A high-resolution view of genome-wide pneumococcal transformation. PLoS Pathog. 2012;8(6):e1002745.
8. Marks LR, Parameswaran GI, Hakansson AP. Pneumococcal interactions with epithelial cells are crucial for optimal biofilm formation and colonization in vitro and in vivo. Infect Immun. 2012 Aug;80(8):2744–60.
9. Tonnaer EL, Mylanus EA, Mulder JJ, Curfs JH. Detection of bacteria in healthy middle ears during cochlear implantation. Arch Otolaryngol Head Neck Surg. 2009 Mar;135(3):232–7.
10. Trappetti C, Gualdi L, Meola LD, Jain P, Korir CC, Edmonds P, et al. The impact of the competence quorum sensing system on Streptococcus pneumoniae biofilms varies depending on the experimental model. BMC Microbiol. 2011 Apr 14;11(1):75.
11. Havarstein LS, Coomaraswamy G, Morrison DA. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in Streptococcus pneumoniae. Proc Natl Acad Sci U S A. 1995 Nov 21;92(24):11140–4.
12. Havarstein LS, Diep DB, Nes IF. A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. Mol Microbiol. 1995 Apr;16(2):229–40.
13. Cuevas RA, Ebrahimi E, Gazioglu O, Yesilkaya H, Hiller NL. Pneumococcal attachment to epithelial cells is enhanced by the secreted peptide VP1 via its control of hyaluronic acid processing. bioRxiv. 2019 Oct 1;788430.
14. Aprianto R, Slager J, Holsappel S, Veening J-W. High-resolution analysis of the pneumococcal transcriptome under a wide range of infection-relevant conditions. Nucleic Acids Res. 2018 02;46(19):9990–10006.

15. Cuevas RA, Eutsey R, Kadam A, West-Roberts JA, Woolford CA, Mitchell AP, et al. A novel streptococcal cell-cell communication peptide promotes pneumococcal virulence and biofilm formation. Mol Microbiol. 2017 Aug;105(4):554–71.

16. Hava DL, Camilli A. Large-scale identification of serotype 4 Streptococcus pneumoniae virulence factors. Mol Microbiol. 2002 Sep;45(5):1389–406.

17. Muñoz-Elías EJ, Marcano J, Camilli A. Isolation of Streptococcus pneumoniae biofilm mutants and their characterization during nasopharyngeal colonization. Infect Immun. 2008 Nov;76(11):5049–61.

18. Moscoso M, García E, López R. Biofilm Formation by Streptococcus pneumoniae: Role of Choline, Extracellular DNA, and Capsular Polysaccharide in Microbial Accretion. J Bacteriol. 2006 Nov 15;188(22):7785–95.

19. Li J, Glover DT, Szalai AJ, Hollingshead SK, Briles DE. PspA and PspC minimize immune adherence and transfer of pneumococci from erythrocytes to macrophages through their effects on complement activation. Infect Immun. 2007 Dec;75(12):5877–85.

20. Ren B, Li J, Genschmer K, Hollingshead SK, Briles DE. The absence of PspA or presence of antibody to PspA facilitates the complement-dependent phagocytosis of pneumococci in vitro. Clin Vaccine Immunol CVI. 2012 Oct;19(10):1574–82.

21. Elm C, Braathen R, Bergmann S, Frank R, Vaerman J-P, Kaetzel CS, et al. Ectodomains 3 and 4 of human polymeric Immunoglobulin receptor (hpIgR) mediate invasion of Streptococcus pneumoniae into the epithelium. J Biol Chem. 2004 Feb 20;279(8):6296–304.

22. Elm C, Rohde M, Vaerman J-P, Chhatwal GS, Hammerschmidt S. Characterization of the interaction of the pneumococcal surface protein SpsA with the human polymeric immunoglobulin receptor (hpIgR). Indian J Med Res. 2004 May;119 Suppl:61–5.

23. Hammerschmidt S, Agarwal V, Kunert A, Haelbich S, Skerka C, Zipfel PF. The host immune regulator factor H interacts via two contact sites with the PspC protein of Streptococcus pneumoniae and mediates adhesion to host epithelial cells. J Immunol Baltim Md 1950. 2007 May 1;178(9):5848–58.

24. Zhang JR, Mostov KE, Lamm ME, Nanno M, Shimida S, Ohwaki M, et al. The polymeric immunoglobulin receptor translocates pneumococci across human nasopharyngeal epithelial cells. Cell. 2000 Sep 15;102(6):827–37.

25. Blanchette KA, Shenoy AT, Milner J, Gilley RP, McClure E, Hinojosa CA, et al. Neuraminidase A-Exposed Galactose Promotes Streptococcus pneumoniae Biofilm Formation during Colonization. Infect Immun. 2016;84(10):2922–32.

26. Parker D, Soong G, Planet P, Brower J, Ratner AJ, Prince A. The NanA neuraminidase of Streptococcus pneumoniae is involved in biofilm formation. Infect Immun. 2009 Sep;77(9):3722–30.

27. King SJ, Hippe KR, Weiser JN. Deglycosylation of human glycoconjugates by the sequential activities of exoglycosidases expressed by Streptococcus pneumoniae. Mol Microbiol. 2006 Feb;59(3):961–74.

28. Blanchette-Cain K, Hinojosa CA, Akula Suresh Babu R, Lizcano A, Gonzalez-Juarbe N, Munoz-Almagro C, et al. Streptococcus pneumoniae biofilm formation is strain dependent, multifactorial, and associated with reduced invasiveness and immunoreactivity during colonization. mBio. 2013;4(5):e00745-00713.

29. Shak JR, Vidal JE, Klugman KP. Influence of bacterial interactions on pneumococcal colonization of the nasopharynx. Trends Microbiol. 2013 Mar;21(3):129–35.

30. Bortoni ME, Terra VS, Hinds J, Andrew PW, Yesilkaya H. The pneumococcal response to oxidative stress includes a role for Rgg. Microbiol Read Engl. 2009 Dec;155(Pt 12):4123–34.

31. Peterson SN, Sung CK, Cline R, Desai BV, Snesrud EC, Luo P, et al. Identification of competence pheromone responsive genes in Streptococcus pneumoniae by use of DNA microarrays. Mol Microbiol. 2004 Feb;51(4):1051–70.

32. Carrolo M, Pinto FR, Melo-Cristino J, Ramirez M. Pherotype influences biofilm growth and recombination in Streptococcus pneumoniae. PloS One. 2014;9(3):e92138.

33. Aprianto R, Slager J, Holsappel S, Veening J-W. Time-resolved dual RNA-seq reveals extensive rewiring of lung epithelial and pneumococcal transcriptomes during early infection. Genome Biol. 2016 Sep 27;17:198.

34. Oggioni MR, Trappetti C, Kadioglu A, Cassone M, Iannelli F, Ricci S, et al. Switch from planktonic to sessile life: a major event in pneumococcal pathogenesis. Mol Microbiol. 2006 Sep;61(5):1196–210.

35. Vidal JE, Howery KE, Ludewick HP, Nava P, Klugman KP. Quorum-sensing systems LuxS/autoinducer 2 and Com regulate Streptococcus pneumoniae biofilms in a bioreactor with living cultures of human respiratory cells. Infect Immun. 2013 Apr;81(4):1341–53.

36. Aggarwal SD, Eutsey R, West-Roberts J, Domenech A, Xu W, Abdullah IT, et al. Function of BriC peptide in the pneumococcal competence and virulence portfolio. PLoS Pathog [Internet]. 2018 Oct 11 [cited 2019 Jul 29];14(10). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6181422/

37. Yadav MK, Vidal JE, Go YY, Kim SH, Chae S-W, Song J-J. The LuxS/AI-2 Quorum-Sensing System of Streptococcus pneumoniae Is Required to Cause Disease, and to Regulate Virulence- and Metabolism-Related Genes in a Rat Model of Middle Ear Infection. Front Cell Infect Microbiol. 2018;8:138.

38. Marks LR, Reddinger RM, Hakansson AP. High Levels of Genetic Recombination during Nasopharyngeal Carriage and Biofilm Formation in Streptococcus pneumoniae. mBio. 2012 Nov 1;3(5):e00200-12.

39. Marks LR, Davidson BA, Knight PR, Hakansson AP. Interkingdom signaling induces Streptococcus pneumoniae biofilm dispersion and transition from asymptomatic colonization to disease. mBio. 2013;4(4).

40. Sanchez CJ, Hurtgen BJ, Lizcano A, Shivshankar P, Cole GT, Orihuela CJ. Biofilm and planktonic pneumococci demonstrate disparate immunoreactivity to human convalescent sera. BMC Microbiol. 2011;11:245.

41. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol. 2008 Mar;26(3):317–25.

42. Chao Y, Marks LR, Pettigrew MM, Hakansson AP. Streptococcus pneumoniae biofilm formation and dispersion during colonization and disease. Front Cell Infect Microbiol. 2014;4:194.

43. Oggioni MR, Trappetti C, Kadioglu A, Cassone M, Iannelli F, Ricci S, et al. Switch from planktonic to sessile life: a major event in pneumococcal pathogenesis. Mol Microbiol. 2006 Sep;61(5):1196–210.

44. Junges R, Salvadori G, Shekhar S, Åmdal HA, Periselneris JN, Chen T, et al. A Quorum-Sensing System That Regulates Streptococcus pneumoniae Biofilm Formation and Surface Polysaccharide Production. mSphere [Internet]. 2017 Sep 13;2(5). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5597970/

45. Sanchez CJ, Hinojosa CA, Shivshankar P, Hyams C, Camberlein E, Brown JS, et al. Changes in Capsular Serotype Alter the Surface Exposure of Pneumococcal Adhesins and Impact Virulence. PLoS ONE. 2011 Oct 19;6(10):e26587.

46. Angulo-Zamudio UA, Vidal JE, Nazmi K, Bolscher JGM, Leon-Sicairos C, Antezana BS, et al. Lactoferrin Disaggregates Pneumococcal Biofilms and Inhibits Acquisition of Resistance Through Its DNase Activity. Front Microbiol. 2019;10:2386.

47. Vidal JE, Howery KE, Ludewick HP, Nava P, Klugman KP. Quorum-sensing systems LuxS/autoinducer 2 and Com regulate Streptococcus pneumoniae biofilms in a bioreactor with living cultures of human respiratory cells. Infect Immun. 2013 Apr;81(4):1341–53.

48. Sanchez CJ, Shivshankar P, Stol K, Trakhtenbroit S, Sullam PM, Sauer K, et al. The pneumococcal serine-rich repeat protein is an intra-species bacterial adhesin that promotes bacterial aggregation in vivo and in biofilms. PLoS Pathog. 2010 Aug 12;6(8):e1001044.

49. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol. 2008 Mar;26(3):317–25.

50. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 2014 Jan 1;42(D1):D206–14.

51. Huang MY, Woolford CA, Mitchell AP. Rapid Gene Concatenation for Genetic Rescue of Multigene Mutants in Candida albicans. mSphere [Internet]. 2018 Apr 25 [cited 2019 Aug 1];3(2). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5917427/

52. Kadam A, Eutsey RA, Rosch J, Miao X, Longwell M, Xu W, et al. Promiscuous signaling by a regulatory system unique to the pandemic PMEN1 pneumococcal lineage. PLOS Pathog. 2017 May 18;13(5):e1006339.

53. Woolford CA, Lagree K, Xu W, Aleynikov T, Adhikari H, Sanchez H, et al. Bypass of Candida albicans Filamentation/Biofilm Regulators through Diminished Expression of Protein Kinase Cak1. PLoS Genet [Internet]. 2016 Dec 9 [cited 2019 Aug 1];12(12). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5147786/

54. Filipe SR, Pinho MG, Tomasz A. Characterization of the murMN operon involved in the synthesis of branched peptidoglycan peptides in Streptococcus pneumoniae. J Biol Chem. 2000 Sep 8;275(36):27768–74.

55. Johnsborg O, Håvarstein LS. Pneumococcal LytR, a protein from the LytR-CpsA-Psr family, is essential for normal septum formation in Streptococcus pneumoniae. J Bacteriol. 2009 Sep;191(18):5859–64.

56. Chapuy-Regaud S, Ogunniyi AD, Diallo N, Huet Y, Desnottes J-F, Paton JC, et al. RegR, a global LacI/GalR family regulator, modulates virulence and competence in Streptococcus pneumoniae. Infect Immun. 2003 May;71(5):2615–25.

57. McCluskey J, Hinds J, Husain S, Witney A, Mitchell TJ. A two-component system that controls the expression of pneumococcal surface antigen A (PsaA) and regulates virulence and resistance to oxidative stress in Streptococcus pneumoniae. Mol Microbiol. 2004 Mar 1;51(6):1661–75.

58. Alloing G, Martin B, Granadel C, Claverys JP. Development of competence in Streptococcus pneumonaie: pheromone autoinduction and control of quorum sensing by the oligopeptide permease. Mol Microbiol. 1998 Jul;29(1):75–83.

59. Hall-Stoodley L, Hu FZ, Gieseke A, Nistico L, Nguyen D, Hayes J, et al. Direct Detection of Bacterial Biofilms on the Middle-Ear Mucosa of Children With Chronic Otitis Media. JAMA J Am Med Assoc. 2006 Jul 12;296(2):202–11.

60. Johnston C, Martin B, Fichant G, Polard P, Claverys J-P. Bacterial transformation: distribution, shared mechanisms and divergent control. Nat Rev Microbiol. 2014 Mar;12(3):181–96.

61. Luo P, Li H, Morrison DA. ComX is a unique link between multiple quorum sensing outputs and competence in Streptococcus pneumoniae. Mol Microbiol. 2003 Oct 1;50(2):623–33.

62. Sanchez CJ, Hurtgen BJ, Lizcano A, Shivshankar P, Cole GT, Orihuela CJ. Biofilm and planktonic pneumococci demonstrate disparate immunoreactivity to human convalescent sera. BMC Microbiol. 2011;11:245.

63. Chao Y, Marks LR, Pettigrew MM, Hakansson AP. Streptococcus pneumoniae biofilm formation and dispersion during colonization and disease. Front Cell Infect Microbiol. 2015;4(January):194.

64. Marks LR, Davidson BA, Knight PR, Hakansson AP. Interkingdom Signaling Induces Streptococcus pneumoniae Biofilm Dispersion and Transition from Asymptomatic Colonization to Disease. mBio. 2013;4(4):e00438-13.

65. Chang JC, LaSarre B, Jimenez JC, Aggarwal C, Federle MJ. Two group A streptococcal peptide pheromones act through opposing Rgg regulators to control biofilm development. PLoS Pathog. 2011 Aug;7(8):e1002190.

66. Cook LC, Federle MJ. Peptide pheromone signaling in Streptococcus and Enterococcus. FEMS Microbiol Rev. 2014 May;38(3):473–92.

67. Li Y-H, Tian X. Quorum Sensing and Bacterial Social Interactions in Biofilms. Sens Basel. 2012;12(2):2519–2538.

68. Cook LC, Federle MJ. Peptide pheromone signaling in Streptococcus and Enterococcus. FEMS Microbiol Rev. 2014 May;38(3):473–92.

69. Havarstein LS, Diep DB, Nes IF. A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. Mol Microbiol. 1995 Apr 1;16(2):229–40.

70. Zhi X, Abdullah IT, Gazioglu O, Manzoor I, Shafeeq S, Kuipers OP, et al. Rgg-Shp regulators are important for pneumococcal colonization and invasion through their effect on mannose utilization and capsule synthesis. Sci Rep. 2018 Apr 23;8(1):6369.

71. Lysenko ES, Lijek RS, Brown SP, Weiser JN. Within-host competition drives selection for the capsule virulence determinant of Streptococcus pneumoniae. Curr Biol CB. 2010 Jul 13;20(13):1222–6.

72. Barbieri G, Voigt B, Albrecht D, Hecker M, Albertini AM, Sonenshein AL, et al. CodY Regulates Expression of the Bacillus subtilis Extracellular Proteases Vpr and Mpr. J Bacteriol. 2015;197(8):1423–32.

73. Hendriksen WT, Bootsma HJ, Diepen A Van, Kuipers OP, Groot R De, Hermans PWM. Strain-specific impact of PsaR of Streptococcus pneumoniae on global gene expression and virulence. Microbiology. 2015;(2009):1569–79.

74. Hendriksen WT, Bootsma HJ, Esteva S, Hoogenboezem T, Jong A De, Groot R De, et al. CodY of Streptococcus pneumoniae : Link between Nutritional Gene. J Bacteriol. 2008;190(2):590–601.

75. Hendriksen WT, Kloosterman TG, Bootsma HJ, Estevão S, de Groot R, Kuipers OP, et al. Site-specific contributions of glutamine-dependent regulator GlnR and GlnR-regulated genes to virulence of *Streptococcus pneumoniae*. Infect Immun. 2008;76(3):1230–8.

76. Hendriksen WT, Bootsma HJ, Estevão S, Hoogenboezem T, de Jong A, de Groot R, et al. CodY of Streptococcus pneumoniae: link between nutritional gene regulation and colonization. J Bacteriol. 2008 Jan;190(2):590–601.

77. Hendriksen WT, Kloosterman TG, Bootsma HJ, Estevão S, De Groot R, Kuipers OP, et al. Site-specific contributions of glutamine-dependent regulator GlnR and GlnR-regulated genes to virulence of Streptococcus pneumoniae. Infect Immun. 2008;76(3):1230–8.

78. Yadav MK, Kwon SK, Cho CG, Park S-W, Chae S-W, Song J-J. Gene expression profile of early in vitro biofilms of Streptococcus pneumoniae. Microbiol Immunol. 2012 Sep;56(9):621–9.

79. Duong-Ly KC, Woo HN, Dunn CA, Xu W, Babič A, Bessman MJ, et al. A UDP-X diphosphatase from Streptococcus pneumoniae hydrolyzes precursors of peptidoglycan biosynthesis. PloS One. 2013;8(5):e64241.

80. Blewett AM, Lloyd AJ, Echalier A, Fülöp V, Dowson CG, Bugg TDH, et al. Expression, purification, crystallization and preliminary characterization of uridine 5'-diphospho-N-acetylmuramoyl L-alanyl-D-glutamate:lysine ligase (MurE) from Streptococcus pneumoniae 110K/70. Acta Crystallogr D Biol Crystallogr. 2004 Feb;60(Pt 2):359–61.

81. Massidda O, Anderluzzi D, Friedli L, Feger G. Unconventional organization of the division and cell wall gene cluster of Streptococcus pneumoniae. Microbiol Read Engl. 1998 Nov;144 ( Pt 11):3069–78.

82. Jamin M, Hakenbeck R, Frere JM. Penicillin binding protein 2x as a major contributor to intrinsic beta-lactam resistance of Streptococcus pneumoniae. FEBS Lett. 1993 Sep 27;331(1–2):101–4.

83. Luo P, Li H, Morrison DA. Identification of ComW as a new component in the regulation of genetic transformation in Streptococcus pneumoniae. Mol Microbiol. 2004;54(1):172–83.

84. Dawid S, Roche AM, Weiser JN. The blp bacteriocins of Streptococcus pneumoniae mediate intraspecies competition both in vitro and in vivo. Infect Immun. 2007 Jan;75(1):443–51.

85. Guiral S, Mitchell TJ, Martin B, Claverys J-P. Competence-programmed predation of noncompetent cells in the human pathogen Streptococcus pneumoniae: genetic requirements. Proc Natl Acad Sci U S A. 2005 Jun 14;102(24):8710–5.

86. Majchrzykiewicz JA, Lubelski J, Moll GN, Kuipers A, Bijlsma JJE, Kuipers OP, et al. Production of a class II two-component lantibiotic of Streptococcus pneumoniae using the

class I nisin synthetic machinery and leader sequence. Antimicrob Agents Chemother. 2010 Apr;54(4):1498–505.

87. Berry AM, Paton JC. Sequence heterogeneity of PsaA, a 37-kilodalton putative adhesin essential for virulence of Streptococcus pneumoniae. Infect Immun. 1996 Dec;64(12):5255–62.

88. Brittan JL, Buckeridge TJ, Finn A, Kadioglu A, Jenkinson HF. Pneumococcal neuraminidase A: an essential upper airway colonization factor for Streptococcus pneumoniae. Mol Oral Microbiol. 2012 Aug;27(4):270–83.

89. Domenech M, García E, Moscoso M. Versatility of the capsular genes during biofilm formation by Streptococcus pneumoniae. Environ Microbiol. 2009 Oct;11(10):2542–55.

90. Dajkovic A, Hinde E, MacKichan C, Carballido-Lopez R. Dynamic Organization of SecA and SecY Secretion Complexes in the B. subtilis Membrane. PloS One. 2016;11(6):e0157899.

91. Scotti PA, Urbanus ML, Brunner J, de Gier J-WL, von Heijne G, van der Does C, et al. YidC, the Escherichia coli homologue of mitochondrial Oxa1p, is a component of the Sec translocase. EMBO J. 2000 Feb 15;19(4):542–9.

92. Martin R, Larsen AH, Corey RA, Midtgaard SR, Frielinghaus H, Schaffitzel C, et al. Structure and Dynamics of the Central Lipid Pool and Proteins of the Bacterial Holo-Translocon. Biophys J. 2019 May 21;116(10):1931–40.

93. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, et al. Rapid pneumococcal evolution in response to clinical interventions. Science. 2011 Jan 28;331(6016):430–4.

94. Hiller NL, Eutsey RA, Powell E, Earl JP, Janto B, Martin DP, et al. Differences in Genotype and Virulence among Four Multidrug-Resistant Streptococcus pneumoniae Isolates Belonging to the PMEN1 Clone. PLoS ONE. 2011 Dec 19;6(12):e28850.

95. Nesin M, Ramirez M, Tomasz A. Capsular transformation of a multidrug-resistant Streptococcus pneumoniae in vivo. J Infect Dis. 1998 Mar;177(3):707–713.

96. Wyres KL, Lambertsen LM, Croucher NJ, McGee L, von Gottberg A, Liñares J, et al. Pneumococcal capsular switching: a historical perspective. J Infect Dis. 2013 Feb 1;207(3):439–49.

97. Aaberge IS, Eng J, Lermark G, Løvik M. Virulence of Streptococcus pneumoniae in mice: a standardized method for preparation and frozen storage of the experimental bacterial inoculum. Microb Pathog. 1995 Feb;18(2):141–52.

98. Eutsey RA, Woolford CA, Aggarwal SD, Cuevas RA, Hiller NL. Gene Expression Analysis in the Pneumococcus. Methods Mol Biol Clifton NJ. 2019;1968:79–88.

# PART III - DISCUSSION

## Chapter 5 Significance and Future Directions

To cause disease, bacteria orchestrate highly hierarchical and coordinated population-level behaviors inside the complex environment of the human body (1). Among the several strategies that the bacteria utilize to colonize and propagate in the host is the establishment of organized settlements called biofilms, which have been shown to play a pivotal role in bacterial virulence during infection (2).

What has intrigued and captivated the attention of many researchers is how this opportunistic pathogen switches between commensal and pathogenic lifestyles. What are the clues that influence the pathogen to remain as an asymptomatic colonizer or disseminate to cause disease? Over the past decades, many pneumococcal virulence determinants have been identified, yet the driving factors that differentiate commensal and pathogenic states remain unclear and are the subject of active investigation.

This dissertation has expanded the understanding of the diverse toolset utilized by the pneumococcus during human infection. It has introduced a new signaling molecule, VP1, that controls several virulent traits. In this chapter, I discuss the data presented in Chapters 2 to 4 in the context of current research in the pneumococcal field. Additionally, I propose future directions and strategies to answer how pneumococcal virulence may be regulated.

## 5.1 VP1 as a secreted signaling peptide

Several signal transduction systems have been described in Gram positive bacteria (3,4). The best-characterized signal transduction system is the TCSs, in which a surface-exposed histidine kinase receptor binds to a cognate secreted peptide to initiate a signal. The binding triggers the autophosphorylation of the receptor, which then transfers the phosphate group to an intracellular transcription factor, ultimately leading to changes in gene expression levels (5). The pneumococcal genome encodes thirteen core TCSs (TCS01 to TCS13) (6,7). A distinctive feature of these systems is that they are found grouped in operons with either *hk-rr* or *rr-hk* configuration (8,9). However, only two cognate peptides have been identified for the TCSs; these are CSP for ComDE (TCS12) and BIP for BlpDE (TCS13) (10,11). Both peptides present a conserved feature in their sequence: a double glycine that is required for export (12,13). Using CSP and BIP as templates, we refined a N-terminal motif associated with export via ABC transporters, and utilized the motif to identify the virulent peptide VP1 (Chapter 2). VP1 bears a double glycine sequence in its N-terminal, which is predicted to be proteolytically cleaved, rendering a mature C-terminal peptide. By using a synthetic VP1, we determined that the activity of VP1 is present on the C-terminal.

The presence of a protease and transporters is common in the operons of multiple secreted peptides (14). These features are also present in the *vp1* operon, albeit their role in peptide function remains an open question. Downstream of *vp1* lays *vpB,* which encodes a predicted protease belonging to the family of CAAX proteases and bacteriocin-processing enzymes. Further downstream is *vpC,* which encodes a transporter from the major facilitator superfamily 1. The other double glycine peptides, CSP and BIP, are secreted by ABC transporters with a peptidase domain. These transporters are located in the same genomic region as their cognate peptides. However, their secretion is not entirely transporter specific: ComAB can secrete CSP and BIP peptides. This discovery results in a functional crosstalk between CSP and BIP signaling systems (15). The nature of the VP1 transporter remains unknown.

VP1 is part of the Gly-Gly family of peptides that shares similarities with both CSP and BIP, which are known to signal through the histidine kinases receptor ComD

and BlpH, respectively (12,15–17). This observation led us to hypothesize that VP1 signals through a kinase. We demonstrated that VP1 binds to the surface of the bacteria, it is not internalized, and controls the expression of several genes, including genes involved in the acquisition and processing of hyaluronic acid. Moreover, our transcriptional screen suggests that *vp1* influences the expression of several histidine kinase receptors, including TCS04, TCS06, TCS09, TCS10, and TCS12 (*pnpRS*, *cbpSR*, *zmpSR*, *vncRS*, *comDE*, respectively). Thus it is tempting to speculate that any of these TCSs may correspond to a VP1 receptor.

CSP and BIP control competence and bacteriocidal activities via cell to cell signaling, important mechanisms that guarantee bacterial survival in the host (12,14,18,19). CSP also modulates metabolism and biofilm (20,21). Similarly, our characterization of VP1 indicates that this signaling molecule plays a role in bacterial attachment and biofilm capacity, key features for virulence. Further, it is highly expressed *in vivo* and plays a role in colonization (22,23). Thus, VP1 is an important virulence factor.

All in all, we provide substantial evidence demonstrating that VP1 is a signaling peptide that controls attachment and biofilm capacity, two features strongly associated with colonization and virulence. Future work on VP1 signaling will provide answers regarding its transporter, surface receptor, and regulon as well as its role(s) in colonization and virulence.

## 5.2 VP1 as a virulence determinant

Multiple virulence determinants contribute to pneumococcal infection. In Chapter 2 we reported that *vp1* is highly expressed *in vivo*, and strongly determines virulence and pathogenesis in an animal model of otitis media (22,23).

Peptide-mediated virulence has been previously reported. While primary studied as an inducer of competence, the signaling peptide CSP is also involved in pneumococci virulence by its control of LytA-mediated autolysis through the activation of ComE (24,25). It has been shown that the competence pathway is active in animal

models of systemic infection (26,27). In this line, *comE* mutants exhibit reduced respiratory tract colonization and virulence in an animal model of lung infection (28). The competence pathway is also active during, and required for, biofilms development (29,30). Similarly, CSP also plays a role in the formation of biofilms in *Streptococcus mutans*. These bacteria form dental plaques (biofilms) on the tooth surface. It has been shown that CSP-induced bacterial lysis is required for the formation of dental plaques. Therefore, CSP contributes to virulence across multiple streptococcal species (31).

In pneumococci, the QS system TprA/PhrA controls the expression of a lantibiotic gene cluster (32). In this system, the mature peptide, encoded by *phrA*, is produced as an immature peptide, which is simultaneously cleaved and transported into the extracellular milieu. The active peptide is then internalized via an oligopeptide permease and binds to the transcriptional factor TprA, inhibiting its transcriptional repression. While it was initially reported that this system does not participate in the bacterial virulence (32), subsequent reports indicate that this QS system has indeed a major effect on colonization of the respiratory tract and bacterial survival in the blood in animal models of lung infection and sepsis, respectively (33,34). Our group has recently characterized the pneumococcal signaling system TprA2/PhrA2, which is involved in the commensalism/dissemination decision-checkpoint during infection. In this QS system, the peptide PhrA2 controls virulence via its control of the downstream virulence determinant, LcpA (35).

Peptide-mediated virulence is not exclusive to streptococci. In *Staphylococcus aureus*, accessory gene regulator (Agr)-type peptides influence virulence. This peptides control cell density during the invasive phase of infection and by controlling the expression of several bacterial adhesins, including fibronectin and fibrinogen binding proteins (36,37), and secreted enzymes associated with both local and hematogenous infections (38,39). In the *agr* architecture, a cyclic peptide, encoded by *agrD*, is exported and processed via the transporter AgrB (40,41), which upon secretion, cycles the peptide to generate the active AgrD peptide (42). Subsequently, signaling occurs after binding of AgrD to a surface-exposed TCS, which transmits the signal to the transcription factor AgrA (43). The role of the *agr* architecture in pneumonia has been investigated using pneumonia animal models of infection, in which strains deficient in

*agr* displayed attenuated virulence in neonatal and adult animal models of infection (44,45). Furthermore, *agr* also controls many cellular functions, such as nutrient transport and amino acid metabolism (46). Similarly, in *Enterococcus faecalis*, virulence is controlled by a peptide known as GBAP. This signaling system is homolog to the *S. aureus agr* system and GBAP is also cycled during export (47). This architecture is heavily associated with the virulence of *E. faecalis*: animal deficient in the GBAP-encoding *fsr* locus exhibits reduced biofilm capacity (48) and displays less pathogenesis in animal models of endophthalmitis and endocarditis (49,50). Finally, in *Streptococcus agalactiae*, the Rgg-type transcription factor RovS regulates the expression of several virulent determinants as well as the bacterial attachment by controlling the expression of the adhesin fibrinogen-binding protein FbsA (51,52).

All these examples illustrate how intricated, diverse, and multifactorial is the virulence regulation in bacteria. We provide sufficient data that parallels VP1 with most of the signaling peptides architectures, as mentioned above, and in particular, with the double glycine peptide CSP. We demonstrated that VP1 controls the expression of a genomic region involved in the processing and acquisition of hyaluronic acid, and we proposed that elements encoded in this region directly participate in colonization under the control of VP1. We also determined that VP1 controls the expression of other virulence determinants, including the pneumococcal Rgg regulatory network, genes involved in the cell wall homeostasis, and competence. While the molecular links between all these molecules is under investigation, it is clear that VP1 plays a critical role in virulence.

## 5.3 Role of VP1 in nutritional sensing

A major limitation of bacterial growth is the availability of nutrients. During colonization and invasive disease, the pneumococcus faces different environmental conditions in different host tissues. For example, glucose in the nasopharynx is virtually undetectable in healthy individuals. However, in conditions such as hyperglycemia or epithelial inflammation, a considerable increase in glucose concentration occurs (53). Further, other sugars, such as galactose and mannose, greatly vary between host

structure (54–56). If pneumococcus encounters these environmental conditions, or when the nutrients become limiting, then the bacteria adapt its global genomic expression in response to those changes (55,57). In Chapter 2 and 3, we reported that *vp1* adjust its expression in response to nutrients.

It has been reported that the pneumococcus is auxotrophic for certain amino acids, including the branched-chain amino acids (BCAA) valine, leucine, and isoleucine. It has been further shown that depleting any of the BCAAs may have deleterious effects on pneumococcus growth (58). First described *in Bacillus subtilis* (59), CodY is a highly conserved, global transcriptional regulator of amino acids and iron metabolism, which respond to changes in amino acids concentration, in particular BCAA (60,61). In the pneumococcus, targets of the CodY regulon include operons involved in amino acid metabolism (58), iron metabolism (62) as well as permeases, amino acid transporters, metabolic enzymes, and proteases (63). In several pathogenic bacteria, CodY regulates major virulence determinants, including the expression of extracellular proteases (*Lactococcus lactis*), adhesins (*Listeria monocytogenes*), toxins (*Clostridium difficile*), and peptide transporters (*Enterococcus faecalis*), among others (62,64–66). In *Streptococcus pyogenes*, the stress response, quorum sensing, and virulence determinants such as globulin-binding proteins are induced by a CodY-mediated amino acid starvation (67).

Results from global scale high throughput transcriptomics and proteomics analyses, indicate that the expression of the *vp1* and its operon is regulated by CodY in response to nutrients abundance (63,68–70). Further evidence supports the hypothesis that *vp1* and the genes in the *vp1* operon sense amino acid levels: a) *vp1* and its downstream genes are negatively associated with levels of *glnA* and *glnP,* which are critical players in glutamine/glutamate metabolism (71); b) CodY and a GlnR binding sites are located just upstream of *vp1* (72,73)*;* and *c) vpD* encodes a protein predicted to function in the transport of BCAA. Metabolism and nutrients are linked to bacterial function. A study about the metabolism of nitrogen in *S. agalactiae* linked nutrients sensing to bacterial attachment (74). Finally, metal cations such as iron, magnesium, zinc, and manganese, to enumerate a few, are essential for bacterial survival because they are required as co-factors for several enzymatic reactions. The manganese

regulator and virulent determinant PsaR have been related to virulence in pneumococci (75,76) and also activates *vp1* expression (69). Thus, we conclude that there is a tight association between VP1 and nutrient availability.

We demonstrated in Chapter 2 that *vp1* is a strong virulent determinant in an animal model of otitis media and sepsis. It has been shown that amino acids concentration in the middle ear may exhibit steep gradients between compartments (77–79). The physiological conditions in the middle ear may then trigger the expression of *vp1* to induce virulence. Furthermore, the whole *vp1* operon was upregulated when pneumococci were recovered from blood in a mice model of sepsis (80). Thus, given the high *vp1* expression during infection and the fact that *vp1* expression changes in response to nutrients (71,81), then *vp1* responds to the host environmental conditions.

Rgg144, the positive regulator of the *vp1* operon, is also negatively regulated by CodY and GlnR in response to nutrients. In the absence of certain nutrients, CodY and GlnR may release their inhibition from *rgg144,* enhancing its expression. In turn, *rgg144* expression results in *vp1* upregulation. In this manner, it appears that Rgg144 amplifies the regulation of *vp1* operon. While it is likely that *rgg144* receives transcriptional input from additional sources, these interactions remain to be identified.

In conclusion VP1 is part of the CodY and GlnR regulons, which are involved in sensing of environmental nutrients. Additionally, *vp1* expression is directly controlled by the transcriptional regulator, Rgg144, which also responds to nutritional sensing. Our working model is that Rgg144, GlnR, and CodY, monitor nutrient availability and then these transcription factors influence levels of the secreted peptide *vp1,* ultimately influencing biofilm development and virulence (**Figure 5.1**)

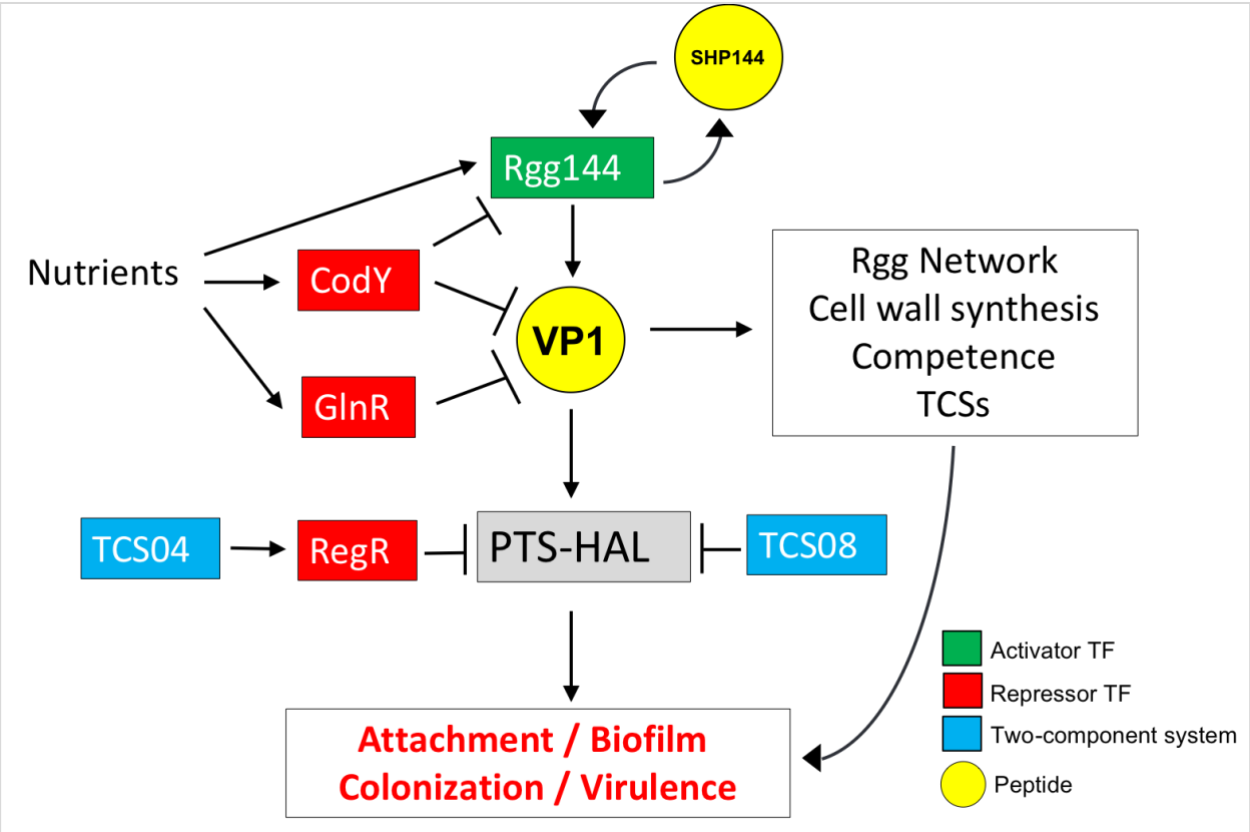



**Figure 5.1 Working model for VP1 regulation and its role in cell-cell communication and virulence**

*vp1* is highly upregulated in the presence of host cells and is induced by an upstream transcription factor, Rgg144/SHP144. In presence of nutrients, CodY and GlnR repression is removed and *rgg144* and *vp1* expression is enhanced. VP1 induces changes in the expression of multiple genes, including a genomic locus implicated in the acquisition and processing of hyaluronic acid. This genomic region is negatively regulated by the repressor RegR. Previous works and related the two-component system TSC04 and TCS08 to the repression of PTS-HAL. TCS04 inhibits PTS-HAL presumably by enhancing the expression of RegR. Our transcriptomic analysis indicates that VP1 influences multiples process important for bacterial survival. We propose that VP1 promotes functional attachment and biofilm formation via its influence on the products of genes encoded in the PTS-HAL operon, to ultimately controls colonization and virulence. TF: transcription factors. TCS: two-component system.

## 5.4 A novel Rgg/SHP locus in *S. pneumoniae*

RRNPP systems have been extensively investigated in *B. subtilis*, *Enteroccocus faecalis*, and the *B. cereus* group; Rgg proteins are widespread in the *Firmicute* species, including the *Streptococcaceae*, *Lactobacillales*, *Listeriaceae*, and *Enterococcaceae* where, in all species, RNPP systems play regulatory functions (82). The genomic organization of this group follows a common but distinct pattern: the transcriptional regulators and the activator peptide are encoded in close proximity in the bacterial chromosome. The maturation of the activator and signaling peptide also follow a common pattern. The mature peptide is derived from a pro-peptide which is enzymatically cleaved by a cytosolic protease, by membrane-anchored proteases, or by ABC transporters which also may function as a protease. In Chapter 2 we reported that upstream and adjacent to *vp1* operon, lays the transcriptional regulator Rgg144. This transcription factor appears usually, and incorrectly, annotated as *mutR* or *mutA* due to the mapping of unannotated raw sequence data to reference genomes. The original *mutR* annotation belongs to the transcriptional activator involved in the biosynthesis of the lantibiotic mutacin II (AF007761) in *S. mutans* (83). Our preliminary bioinformatic search confirmed the incorrect annotation: five different genes appear annotated as *mutR* in our pneumococcal strain model. We provided evidence that *mutR* is a transcription factor belonging to the Rgg family (84).

In the Rgg family, the transcriptional activity depends on binding to the signaling peptide (85). Signaling peptides bind to tetratricopeptide repeats in the C-terminal domain of Rgg, causing its activation (86,87). Regarding the activity of Rgg144, we discovered and characterized the function of a SHP-encoding ORF located immediately downstream to *rgg*, in a position conserved across pneumococcal strains, and that the product of the *shp* modulates the transcriptional activity of Rgg144. The predicted promoter region of *shp* lays just upstream and overlapping to *vp1* on the complementary strand. We also showed that the *vp1* promoter encodes a putative, palindromic CodY-box proximal to *vp1* (88). In this manner, SHP levels may also be under the control of CodY. Following our collaboration with Dr. Hasan Yesilkaya, we renamed Rgg and SHP to Rgg144 and SHP144, respectively, based on gene numbers in the strain D39.

We further characterized the activity of the SHP144 peptide. Our initial experiments showed that exogenously added full-length SHP (27 amino acid) has a modest influence on *vp1* expression. The length of most of the known active SHP peptides ranges between nine and twelve amino acids. For example, when SHP of various lengths where tested for Rgg1358 of *S. thermophilus*, the highest transcriptional activity for SHP1358 was observed when a 9-amino acids peptide was utilized (SHP1358-C9) (89). Additionally, in *S. pyogenes* transcriptional activity is still observed for peptides no longer than 8-amino acids (90). In this line, the SHP144-C12 is the most active peptide for Rgg144 in an *in vitro* model of Rgg activity (91), and SHP144-C12 highly induces the expression of *vp1* (Chapter 2). Mass spectroscopy experiments of the supernatant are required to establish the naturally occurring form(s) of SHP144.

Rgg/SHP QS systems have been extensively characterized in Group A streptococci (92–94). Further, an Rgg characterized in pneumococcus has been implicated in the response to oxidative stress (95). The group of Yesilkaya determined that Rgg144 monitors sugars commonly found in the respiratory tract, such as galactose and mannose, (91). Further, we observed that Rgg144 also controls *vp1* operon on planktonic cultures grown in CDM supplemented with glucose (Chapter 2). The group of Yesilkaya also showed that Rgg144 plays a key role in colonization and virulence. Thus, we propose that VP1 contributes to the overall effect of Rgg144 on colonization. We suggest complementation experiments to determine whether overexpression of *vp1* can rescue the deficiency observed in the *rgg144* mutant.

In pneumococci, peptide uptake is mediated by the general permease Ami/Ali, which is encoded by the *aliA* and *aliB*, part of the *amiACDEF* operon (58,96). It has been shown that *aliA* and *aliB* are required for successful nasopharynx colonization in mice (97) as well as adherence and virulence (98). These data also highlight the importance of communication for bacterial virulence. Does the permease AliA/B participate in the SHP144 internalization? Does AliA/B function as a bridge between Rgg144-associated virulence and VP1-associated virulence? These questions can be answered by evaluating the activity of SHP in combination with reporters constructs that take advantage of the transcriptional activity of Rgg144, utilized in the background of mutants deficient of the permeases.

Finally, it has been suggested that biofilms play a pivotal role in bacterial virulence during infection (2). Moreover, bacteria living in biofilms are phenotypically distinct from their planktonic counterparts. It has been shown that during attachment and formation of biofilms, pneumococci downregulates the genes required for capsule synthesis and sheds off the remaining shielding polysaccharides (99,100), thus, exposing adhesins that allow it to interact and attach to epithelial cells. In Chapter 4 we showed that VP1 controls the expression of Rgg144 and Rgg939 in biofilms growing on lung epithelial cells. Consistently, it has been shown that Rgg144/SHP144 and Rgg939/SHP939, reduce transcription of a gene cluster encoding capsule genes (91) in planktonic growth, presumably to enhance the biofilm capacity. Furthermore, in a murine model of nasopharynx colonization, both QS systems, SHP144/Rgg144 and SHP939/Rgg939, played a significant role in the survival of the bacteria. However, Junges *et al.* suggested, in contradiction, that Rgg939 inhibits biofilm formation in pneumococci (101). While these differences may be attributed to the experimental conditions (planktonic versus biofilm), more research addressing these differences is required.

In conclusion, we characterized an Rgg transcription factor and its cognate SHP. In concert, they control the expression of *vp1*, and potentially, other genes in a broad and interconnected circuitry, to ultimately regulate pneumococcal virulence.

## 5.5 A PTS involved in adhesion and regulation

Pneumococcal adherence is a complicated and intricated process that involves multiple molecules. Given the importance of bacterial attachment to colonization, dispersion and invasion, it is not surprising that *S. pneumoniae* uses different and overlapping mechanism to guarantee adhesion (102–106). Furthermore, different organs and structures in the host vary in terms of expression of their surface receptors (55,57). Thus, adhesion is a complex interplay between the pathogen and the host. In Chapter 3 we demonstrated that a genomic locus involved in the acquisition of HA contributed significantly to adherence, and the overexpression of such locus rescued

biofilm and attachment deficiency of *vp1*. We demonstrated this rescue in two clinical isolates.

Pneumococci utilize surface-exposed molecules as adhesins that mediate attachment of the bacteria to the host cell (107–109). They are distinguished by classical features, like the C-terminal amino-acid sequence LPxTG or a lipobox motif that serves as recognition sequences for lipid modifications (8). Although there are several proteins that do not possess those classical features, they are described primarily as cytoplasmic enzymes that relocated to the surface of the bacteria. It is now clear that these molecules function as virulence factors and participate in interaction with the host, in a manner independent of their enzymic activity (107,110). Examples of non-classical adhesin molecules are PavA, the glycolytic enzymes a-enolase (Eno), glyceraldehyde 3-phosphate dehydrogenase (GAPDH), and phosphoglycerate kinase (PGK). PavA is a surface-associated protein that lacks surface trafficking mitofs. It has been shown that PavA serves as a pneumococcal adhesin for fibronectin and facilitate adherence to epithelial and endothelial cells (111). Eno is a glycolytic, plasminogen-binding enzyme expressed on the surface of several bacteria, including *S. pneumoniae*, *S. pyogenes*, and *S. aureus* as well as of the lactic acid bacteria*, L crispatus* and *L. johnsonii*, (105). Similar to Eno, GAPDH and PGK are cytosolic molecules and surface exposed enzymes capable of binding plasmin and plasminogen (112–115). The pneumococcus genome encodes several glycosidases. A group of these enzymes that are exposed on the surface of the bacteria, have been shown to contribute to adhesion, not only via modification of the host surface that reveals receptors, but also via direct binding (116). The glycosidase β-galactosidase BgaA process GAGs (117) and promotes attachment to host cells (118), and it is required for colonization in an animal model of nasopharynx colonization (119). In BgaA, the binding is mediated via its carbohydrate-binding module (118). Thus, the finding that BgaA acts as a functional adhesin supports the hypothesis that pneumococcal non-classical adhesions may contribute to colonization.

The data presented in this dissertation strongly suggest that the molecules involved in the transport of host HA play an essential role in adhesion. The processing of host hyaluronan starts with the binding and breaking of the chain by the enzyme

HysA (120). Then components of the PTS-EII system internalize the HA disaccharide into the cell. In particular, EIIC and EIID function as transmembrane transporter. Inside the cell, HA is hydrolyzed to its constitutive components, glucuronic acid, and N-acetyl-D-glucosamine (**Figure 1.5**). Here, we demonstrated that molecules involved in this process, with the exception of HysA and RegR, are implicated in attachment and colonization. While this phenotype has not been previously reported, previous studies demonstrate that pneumococci can utilize HA as a carbon source. In particular, *hysA* and genes in the PTS-EII-encoding operon (EIIB, EIIC, EIID, and *ugl*) are required for pneumococcal growth in media where HA is the only carbohydrate (121). There is scarce evidence linking molecules involved in HA processing to virulence. RegR, is encoded downstream and in proximity to PTS-HAL. Our analysis suggests that RegR is not part of the PTS-HAL operon, but instead that a promoter is localized immediately upstream of the coding sequence. Further, RegR functions as a repressor of PTS-HAL and *hysA* expression (120). RegR and HysA have been associated with pneumococcal virulence. However, a *hysA* deficient strain behaved very similarly when compared to its parental wild-type strain in an animal infection model (122,123). Furthermore, it has been shown that HA supports *in vitro* pneumococcal biofilm formation and the molecules encoded by the PTS-HAL are upregulated in this condition (124).

Our data suggest that one or more gene products encoded in the operon that includes EIIC and EIID, the HA transporter (**Figure 1.5**), are implicated in the bacterial attachment. However, we have not yet established the molecular mechanism of attachment. One hypothesis is that the transporter binds hyaluronan and function directly as an anchor of the bacteria. A second hypothesis is that one of the two additional, uncharacterized, genes in this operon play a role in adhesion. These genes are a putative heparinase II/III-like, sometimes annotated as *hep,* and a putative gene annotated as *yajC*, predicted to be a component of the Sec machinery (associated with the protein secretion) (125). Whether heparinase or *yajC* account for the bacterial attachment to epithelial cells and the colonization deficiency observed in our animal models of nasopharynx infection, needs to be determined. Altogether, our data strongly suggest that the PTS-HAL and VP1 mediate the bacterial attachment, via a yet uncharacterized effector.

## 5.6 Regulation of the PTS system required for HA processing

Carbohydrate acquisition plays an essential role in pneumococcal pathogenesis; carbohydrates need to be extracted from the host to guarantee survival. The findings that a PTS involved in HA acquisition participates in attachment are evidence of how bacteria adapt to their environment. In chapter 3 we showed that genes encoded in the transporter are required for bacterial attachment to epithelial cells, presumably by recognizing and attaching directly to ECM (23). While we showed that HysA is not required for attachment it has been shown that HysA is required for survival to macrophages by reducing the production of proinflammatory cytokines (126). In *S. pneumoniae* cultures, it has been reported that this surface-associated enzyme is found in both the culture and the cell-associated fractions (127). This observation opens the question of how is HysA, and by extension, the components of the PTS-HAL regulated.

Levels of regulators, carbohydrates, and transporters are interconnected in the bacterial cells via highly complex networks (55). Transporters not only control the import of carbohydrates, but they dock cytosolic components that can be further regulated. For instance, cytoplasmic Hpr in the PTS system (**Figure 1.5**), also controls the carbon catabolite repression mediated by the master regulator CcpA (128), which influences the expression of nearly one-fifth of pneumococcal genes, including ABC transporters and other regulators (128,129). We showed that RegR, the PTS-HAL regulator, repress the expression of targets during biofilm mode of growth, thus expanding what was known about this locus repression (123). We found that *vp1* controls the expression of PTS-HAL, especially during the formation of biofilm on epithelial cells, suggesting an environment-dependent influence on this control. Orihuela *et al*. showed that pneumococci vary its genomic expression in a tissue-specific manner and that this differs from expression in *in vitro* conditions (130,131). The *vp1* levels are particularly high in conditions that mimic pneumococcal infection and *in vivo* (22,29) (and Chapters 2 and 3). This suggests that while RegR repressed PTS-HAL expression in liquid culture and biofilms, *vp1* enhance its expression in a RegR-independent manner. However, we have not yet established the mechanism of *vp1*-mediated expression.

Finally, we reported in Chapter 4 that TCS04 represses the expression of PTS-HAL in liquid conditions, while TCS04 has been previously connected with the repressor RegR (132). We hypothesize that TCS04 represses PTS-HAL expression via enhancing the expression of RegR. PTS-HAL regulation seems to be more complicated. A recent study by Gomez-Mejia *et al*. showed that *vp1* operon and the PTS-HAL are repressed by TCS08 (SaeRS) (9). Thus, the regulation of this locus is complex; it receives multiple signaling inputs. VP1 enhances its expression while TCS04, TCS08, and RegR repress it (**Figure 5.1**). Understanding virulence will require a detailed view of the regulatory circuitry that links nutrition, VP1, adhesins, metabolism, and biofilms.

## 5.7 Concluding remarks

Virulence in *Streptococcus pneumoniae* is the result of a multi-genic process that required tight regulation. Throughout the length of this project, we have gathered substantial evidence that supports the hypothesis that VP1 is a secreted signaling molecule that provides a robust and unique virulent phenotype. We presented substantial data on the characterization of VP1 and its influence on gene expression, cell-to-cell communication, colonization, and virulence. We also identified additional genes involved in the VP1 signaling, which are themselves virulent determinants. While several questions were addressed throughout the length of this dissertation, more questions arose regarding VP1 function and regulation. These new questions reveal exciting research projects. The work presented in this dissertation greatly deepens our current understanding of pneumococcal biology, and in doing so, advances our understanding of pneumococcal biology in the context of bacterial communication and human infections.

## 5.8 References


1. Popat R, Crusz SA, Diggle SP. The social behaviours of bacterial pathogens. Br Med Bull. 2008;87:63–75.

2. Hall-Stoodley L, Costerton JW, Stoodley P. Bacterial biofilms: from the Natural environment to infectious diseases. Nat Rev Micro. 2004 Feb;2(2):95–108.

3. Lange R, Wagner C, de Saizieu A, Flint N, Molnos J, Stieger M, et al. Domain organization and molecular characterization of 13 two-component systems identified by genome sequencing of Streptococcus pneumoniae. Gene. 1999 Sep 3;237(1):223–34.

4. Throup JP, Koretke KK, Bryant AP, Ingraham KA, Chalker AF, Ge Y, et al. A genomic analysis of two-component signal transduction in Streptococcus pneumoniae. Mol Microbiol. 2000 Feb;35(3):566–76.

5. Cozzone AJ. Protein phosphorylation in prokaryotes. Annu Rev Microbiol. 1988;42:97–125.

6. Lange R, Wagner C, de Saizieu A, Flint N, Molnos J, Stieger M, et al. Domain organization and molecular characterization of 13 two-component systems identified by genome sequencing of Streptococcus pneumoniae. Gene. 1999 Sep 3;237(1):223–34.

7. Throup JP, Koretke KK, Bryant AP, Ingraham KA, Chalker AF, Ge Y, et al. A genomic analysis of two-component signal transduction in Streptococcus pneumoniae. Mol Microbiol. 2000 Feb;35(3):566–76.

8. Gámez G, Castro A, Gómez-Mejia A, Gallego M, Bedoya A, Camargo M, et al. The variome of pneumococcal virulence factors and regulators. BMC Genomics [Internet]. 2018 Jan 3 [cited 2019 Oct 28];19. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5753484/

9. Gómez-Mejia A, Gámez G, Hammerschmidt S. Streptococcus pneumoniae two-component regulatory systems: The interplay of the pneumococcus with its environment. International Journal of Medical Microbiology. 2018 Aug 1;308(6):722–37.

10. de Saizieu A, Gardès C, Flint N, Wagner C, Kamber M, Mitchell TJ, et al. Microarray-based identification of a novel Streptococcus pneumoniae regulon controlled by an autoinduced peptide. J Bacteriol. 2000 Sep;182(17):4696–703.

11. Pestova EV, Håvarstein LS, Morrison DA. Regulation of competence for genetic transformation in Streptococcus pneumoniae by an auto-induced peptide pheromone and a two-component regulatory system. Mol Microbiol. 1996 Aug;21(4):853–62.

12. Havarstein LS, Coomaraswamy G, Morrison DA. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in Streptococcus pneumoniae. Proc Natl Acad Sci USA. 1995 Nov 21;92(24):11140–4.

13. Havarstein LS, Diep D, Nes I. A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. Molecular Microbiology. 1995 Apr 1;16(2):229–40.

14. Dawid S, Roche AM, Weiser JN. The blp bacteriocins of Streptococcus pneumoniae mediate intraspecies competition both in vitro and in vivo. Infection and Immunity. 2007;75(1):443–451.

15. Wang CY, Patel N, Wholey W-Y, Dawid S. ABC transporter content diversity in Streptococcus pneumoniae impacts competence regulation and bacteriocin production. Proc Natl Acad Sci USA. 2018 19;115(25):E5776–85.

16. Pestova EV, Håvarstein LS, Morrison DA. Regulation of competence for genetic transformation in Streptococcus pneumoniae by an auto-induced peptide pheromone and a two-component regulatory system. Mol Microbiol. 1996 Aug;21(4):853–62.

17. Wholey W-Y, Abu-Khdeir M, Yu EA, Siddiqui S, Esimai O, Dawid S. Characterization of the Competitive Pneumocin Peptides of Streptococcus pneumoniae. Front Cell Infect Microbiol. 2019;9:55.

18. Junges R, Salvadori G, Shekhar S, Åmdal HA, Periselneris JN, Chen T, et al. A Quorum-Sensing System That Regulates Streptococcus pneumoniae Biofilm Formation and Surface Polysaccharide Production. mSphere [Internet]. 2017 Sep 13;2(5). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5597970/

19. Zhi X, Abdullah IT, Gazioglu O, Manzoor I, Shafeeq S, Kuipers OP, et al. Rgg-Shp regulators are important for pneumococcal colonization and invasion through their effect on mannose utilization and capsule synthesis. Sci Rep. 2018 Apr 23;8(1):6369.

20. Duong-Ly KC, Woo HN, Dunn CA, Xu W, Babič A, Bessman MJ, et al. A UDP-X diphosphatase from Streptococcus pneumoniae hydrolyzes precursors of peptidoglycan biosynthesis. PLoS ONE. 2013;8(5):e64241.

21. Blewett AM, Lloyd AJ, Echalier A, Fülöp V, Dowson CG, Bugg TDH, et al. Expression, purification, crystallization and preliminary characterization of uridine 5'-diphospho-N-acetylmuramoyl L-alanyl-D-glutamate:lysine ligase (MurE) from Streptococcus pneumoniae 110K/70. Acta Crystallogr D Biol Crystallogr. 2004 Feb;60(Pt 2):359–61.

22. Massidda O, Anderluzzi D, Friedli L, Feger G. Unconventional organization of the division and cell wall gene cluster of Streptococcus pneumoniae. Microbiology (Reading, Engl). 1998 Nov;144 ( Pt 11):3069–78.

23. Jamin M, Hakenbeck R, Frere JM. Penicillin binding protein 2x as a major contributor to intrinsic beta-lactam resistance of Streptococcus pneumoniae. FEBS Lett. 1993 Sep 27;331(1–2):101–4.

24. Luo P, Li H, Morrison DA. ComX is a unique link between multiple quorum sensing outputs and competence in Streptococcus pneumoniae. Molecular Microbiology. 2003 Oct 1;50(2):623–33.

25. Luo P, Li H, Morrison DA. Identification of ComW as a new component in the regulation of genetic transformation in Streptococcus pneumoniae. Molecular Microbiology. 2004;54(1):172–83.

26. Claverys J, Prudhomme M, Martin B. Induction of Competence Regulons as a General Response to Stress in Gram-Positive Bacteria. Annu Rev Microbiol 2006. 2006;60:451–475.

27. Claverys J-P, Havarstein LS. Cannibalism and fratricide: mechanisms and raisons d'être. Nature Reviews Microbiology. 2007;5:219–229.

28. Claverys J-P, Prudhomme M, Martin B. Induction of Competence Regulons as a General Response to Stress in Gram-Positive Bacteria. Annual Review of Microbiology. 2006;60(1):451–75.

29. Peterson SN, Sung CK, Cline R, Desai BV, Snesrud EC, Luo P, et al. Identification of competence pheromone responsive genes in Streptococcus pneumoniae by use of DNA microarrays. Mol Microbiol. 2004 Feb;51(4):1051–70.

30. Cuevas RA, Eutsey R, Kadam A, West-Roberts JA, Woolford CA, Mitchell AP, et al. A novel streptococcal cell-cell communication peptide promotes pneumococcal virulence and biofilm formation. Mol Microbiol. 2017 Aug;105(4):554–71.

31. Cuevas RA, Ebrahimi E, Gazioglu O, Yesilkaya H, Hiller NL. Pneumococcal attachment to epithelial cells is enhanced by the secreted peptide VP1 via its control of hyaluronic acid processing. bioRxiv. 2019 Oct 1;788430.

32. Steinmoen H, Knutsen E, Håvarstein LS. Induction of natural competence in Streptococcus pneumoniae triggers lysis and DNA release from a subfraction of the cell population. Proc Natl Acad Sci USA. 2002 May 28;99(11):7681–6.

33. Steinmoen H, Teigen A, Håvarstein LS. Competence-induced cells of Streptococcus pneumoniae lyse competence-deficient cells of the same strain during cocultivation. J Bacteriol. 2003 Dec;185(24):7176–83.

34. Trappetti C, Gualdi L, Meola LD, Jain P, Korir CC, Edmonds P, et al. The impact of the competence quorum sensing system on Streptococcus pneumoniae biofilms varies depending on the experimental model. BMC Microbiology. 2011 Apr 14;11(1):75.

35. Oggioni MR, Trappetti C, Kadioglu A, Cassone M, Iannelli F, Ricci S, et al. Switch from planktonic to sessile life: a major event in pneumococcal pathogenesis. Mol Microbiol. 2006 Sep;61(5):1196–210.

36. Kowalko JE, Sebert ME. The Streptococcus pneumoniae competence regulatory system influences respiratory tract colonization. Infect Immun. 2008 Jul;76(7):3131–40.

37. Aprianto R, Slager J, Holsappel S, Veening J-W. High-resolution analysis of the pneumococcal transcriptome under a wide range of infection-relevant conditions. Nucleic Acids Res. 2018 02;46(19):9990–10006.

38. Vidal JE, Ludewick HP, Kunkel RM, Zähner D, Klugman KP. The LuxS-Dependent Quorum-Sensing System Regulates Early Biofilm Formation by Streptococcus pneumoniae Strain D39 ▿. Infect Immun. 2011 Oct;79(10):4050–60.

39. Perry JA, Cvitkovitch DG, Lévesque CM. Cell death in Streptococcus mutans biofilms: a link between CSP and extracellular DNA. FEMS Microbiol Lett. 2009 Oct;299(2):261–6.

40. Hoover SE, Perez AJ, Tsui H-CT, Sinha D, Smiley DL, DiMarchi RD, et al. A new quorum-sensing system (TprA/PhrA) for Streptococcus pneumoniae D39 that regulates a lantibiotic biosynthesis gene cluster. Mol Microbiol. 2015 Apr 13;

41. Motib AS, Al-Bayati FAY, Manzoor I, Shafeeq S, Kadam A, Kuipers OP, et al. TprA/PhrA Quorum Sensing System Has a Major Effect on Pneumococcal Survival in Respiratory Tract and Blood, and Its Activity Is Controlled by CcpA and GlnR. Front Cell Infect Microbiol [Internet]. 2019 Sep 13 [cited 2019 Oct 27];9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6753895/

42. Motib A, Guerreiro A, Al-Bayati F, Piletska E, Manzoor I, Shafeeq S, et al. Modulation of Quorum Sensing in a Gram-Positive Pathogen by Linear Molecularly Imprinted Polymers with Anti-infective Properties. Angew Chem Int Ed Engl. 2017 22;56(52):16555–8.

43. Kadam A, Janto B, Eutsey R, Earl JP, Powell E, Dahlgren ME, et al. Streptococcus pneumoniae Supragenome Hybridization Arrays for Profiling of Genetic Content and Gene Expression. Curr Protoc Microbiol. 2015;36:9D.4.1-9D.4.20.

44. Hienz SA, Schennings T, Heimdahl A, Flock JI. Collagen binding of Staphylococcus aureus is a virulence factor in experimental endocarditis. J Infect Dis. 1996 Jul;174(1):83–8.

45. Greene C, McDevitt D, Francois P, Vaudaux PE, Lew DP, Foster TJ. Adhesion properties of mutants of Staphylococcus aureus defective in fibronectin-binding proteins and studies on the expression of fnb genes. Mol Microbiol. 1995 Sep;17(6):1143–52.

46. Wang JE, Jørgensen PF, Almlöf M, Thiemermann C, Foster SJ, Aasen AO, et al. Peptidoglycan and lipoteichoic acid from Staphylococcus aureus induce tumor necrosis factor alpha, interleukin 6 (IL-6), and IL-10 production in both T cells and monocytes in a human whole blood model. Infect Immun. 2000 Jul;68(7):3965–70.

47. Soell M, Diab M, Haan-Archipoff G, Beretz A, Herbelin C, Poutrel B, et al. Capsular polysaccharide types 5 and 8 of Staphylococcus aureus bind specifically to human epithelial (KB) cells, endothelial cells, and monocytes and induce release of cytokines. Infect Immun. 1995 Apr;63(4):1380–6.

48. Nakayama J, Cao Y, Horii T, Sakuda S, Akkermans AD, de Vos WM, et al. Gelatinase biosynthesis-activating pheromone: a peptide lactone that mediates a quorum sensing in Enterococcus faecalis. Mol Microbiol. 2001 Jul;41(1):145–54.

49. Saenz HL, Augsburger V, Vuong C, Jack RW, Götz F, Otto M. Inducible expression and cellular location of AgrB, a protein involved in the maturation of the staphylococcal quorum-sensing pheromone. Arch Microbiol. 2000 Dec;174(6):452–5.

50. Qiu R, Pei W, Zhang L, Lin J, Ji G. Identification of the putative staphylococcal AgrB catalytic residues involving the proteolytic cleavage of AgrD to generate autoinducing peptide. J Biol Chem. 2005 Apr 29;280(17):16695–704.

51. Sturme MHJ, Kleerebezem M, Nakayama J, Akkermans ADL, Vaugha EE, de Vos WM. Cell to cell communication by autoinducing peptides in gram-positive bacteria. Antonie Van Leeuwenhoek. 2002 Aug;81(1–4):233–43.

52. Bubeck Wardenburg J, Patel RJ, Schneewind O. Surface proteins and exotoxins are required for the pathogenesis of Staphylococcus aureus pneumonia. Infect Immun. 2007 Feb;75(2):1040–4.

53. Heyer G, Saba S, Adamo R, Rush W, Soong G, Cheung A, et al. Staphylococcus aureus agr and sarA functions are required for invasive infection but not inflammatory responses in the lung. Infect Immun. 2002 Jan;70(1):127–33.

54. Saïd-Salim B, Dunman PM, McAleese FM, Macapagal D, Murphy E, McNamara PJ, et al. Global regulation of Staphylococcus aureus genes by Rot. J Bacteriol. 2003 Jan;185(2):610–9.

55. Qin X, Singh KV, Weinstock GM, Murray BE. Effects of Enterococcus faecalis fsr genes on production of gelatinase and a serine protease and virulence. Infect Immun. 2000 May;68(5):2579–86.

56. Hancock LE, Perego M. The Enterococcus faecalis fsr two-component system controls biofilm development through production of gelatinase. J Bacteriol. 2004 Sep;186(17):5629–39.

57. Thurlow LR, Thomas VC, Narayanan S, Olson S, Fleming SD, Hancock LE. Gelatinase contributes to the pathogenesis of endocarditis caused by Enterococcus faecalis. Infect Immun. 2010 Nov;78(11):4936–43.

58. Mylonakis E, Engelbert M, Qin X, Sifri CD, Murray BE, Ausubel FM, et al. The Enterococcus faecalis fsrB gene, a key component of the fsr quorum-sensing system, is associated with virulence in the rabbit endophthalmitis model. Infect Immun. 2002 Aug;70(8):4678–81.

59. Ibrahim M, Guillot A, Wessner F, Algaron F, Besset C, Courtin P, et al. Control of the transcription of a short gene encoding a cyclic peptide in Streptococcus thermophilus: a new quorum-sensing system? J Bacteriol. 2007 Dec;189(24):8844–54.

60. Samen UM, Eikmanns BJ, Reinscheid DJ. The transcriptional regulator RovS controls the attachment of Streptococcus agalactiae to human epithelial cells and the expression of virulence genes. Infect Immun. 2006 Oct;74(10):5625–35.

61. Philips BJ, Meguer J-X, Redman J, Baker EH. Factors determining the appearance of glucose in upper and lower respiratory tract secretions. Intensive Care Med. 2003 Dec;29(12):2204–10.

62. Blanchette KA, Shenoy AT, Milner J, Gilley RP, McClure E, Hinojosa CA, et al. Neuraminidase A-Exposed Galactose Promotes Streptococcus pneumoniae Biofilm Formation during Colonization. Infect Immun. 2016;84(10):2922–32.

63. Paixão L, Oliveira J, Veríssimo A, Vinga S, Lourenço EC, Ventura MR, et al. Host glycan sugar-specific pathways in Streptococcus pneumonia: galactose as a key sugar in colonisation and infection. PLoS ONE. 2015;10(3):e0121042.

64. Kobayashi K, Morimoto K, Kataura A, Akino T. Sugar components of glycoprotein fractions in middle ear effusions. Arch Otorhinolaryngol. 1985;242(2):177–82.

65. Verhagen LM, de Jonge MI, Burghout P, Schraa K, Spagnuolo L, Mennens S, et al. Genome-wide identification of genes essential for the survival of Streptococcus pneumoniae in human saliva. PLoS ONE. 2014;9(2):e89541.

66. Härtel T, Eylert E, Schulz C, Petruschka L, Gierok P, Grubmüller S, et al. Characterization of Central Carbon Metabolism of Streptococcus pneumoniae by Isotopologue Profiling. The Journal of Biological Chemistry. 2012;287(6):4260–74.

67. Slack F, Serror P, Joyce E, Sonenshein AL. A gene required for nutritional repression of the Bacillus subtilis dipeptide permease operon. Mol Microbiology. 1995;15:689–702.

68. Molle V, Nakaura Y, Shivers RP, Yamaguchi H, Losick R, Fujita Y, et al. Additional targets of the Bacillus subtilis global regulator CodY identified by chromatin immunoprecipitation and genome-wide transcript analysis. J Bacteriol. 2003 Mar;185(6):1911–22.

69. Sonenshein AL. CodY, a global regulator of stationary phase and virulence in Gram-positive bacteria. Curr Opin Microbiol. 2005 Apr;8(2):203–7.

70. Caymaris S, Bootsma HJ, Martin B, Hermans PWM, Prudhomme M, Claverys J. The global nutritional regulator CodY is an essential protein in the human pathogen Streptococcus pneumoniae. Molecular Microbiology. 2010;78(2):344–60.

71. Hendriksen WT, Bootsma HJ, Esteva S, Hoogenboezem T, Jong A De, Groot R De, et al. CodY of Streptococcus pneumoniae : Link between Nutritional Gene. Journal of Bacteriology. 2008;190(2):590–601.

72. Sonenshein AL. CodY, a global regulator of stationary phase and virulence in Gram-positive bacteria. Current Opinion in Microbiology. 2005;8:203–207.

73. Claverys J, Prudhomme M, Martin B. Induction of Competence Regulons as a General Response to Stress in Gram-Positive Bacteria. Annu Rev Microbiol 2006. 2006;60:451–75.

74. Sonenshein AL. Control of key metabolic intersections. Nature Reviews. 2007;5:917–28.

75. Steiner K, Malke H. relA-independent amino acid starvation response network of Streptococcus pyogenes. J Bacteriology. 2001;183:7354–64.

76. Barbieri G, Voigt B, Albrecht D, Hecker M, Albertini AM, Sonenshein AL, et al. CodY Regulates Expression of the Bacillus subtilis Extracellular Proteases Vpr and Mpr. Journal of Bacteriology. 2015;197(8):1423–32.

77. Hendriksen WT, Bootsma HJ, Diepen A Van, Kuipers OP, Groot R De, Hermans PWM. Strain-specific impact of PsaR of Streptococcus pneumoniae on global gene expression and virulence. Microbiology. 2015;(2009):1569–79.

78. Hendriksen WT, Kloosterman TG, Bootsma HJ, Estevão S, de Groot R, Kuipers OP, et al. Site-specific contributions of glutamine-dependent regulator GlnR and GlnR-regulated

genes to virulence of *Streptococcus pneumoniae*. Infection and Immunity. 2008;76(3):1230–8.

79. Hendriksen WT, Kloosterman TG, Bootsma HJ, Estevão S, De Groot R, Kuipers OP, et al. Site-specific contributions of glutamine-dependent regulator GlnR and GlnR-regulated genes to virulence of Streptococcus pneumoniae. Infect Immun. 2008;76(3):1230–8.

80. Kok J, Hermans PWM, Kuipers OP. Regulation of Glutamine and Glutamate Metabolism by GlnR and GlnA in Streptococcus pneumoniae. Journal of Biological Chemistry. 2006;281(35):25097–109.

81. Belitsky BR, Sonenshein AL. Genome-wide identification of Bacillus subtilis CodY-binding sites at single-nucleotide resolution. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(17):7026–31.

82. Tamura GS, Nittayajarn A, Schoentag DL. A Glutamine Transport Gene , glnQ , Is Required for Fibronectin Adherence and Virulence of Group B Streptococci. Infection and Immunity. 2002;70(6):2877–85.

83. Johnston J, Briles D, Myers L, Hollingshead S. Mn2+ dependent regulation of multiple genes in Streptococcus pneumoniae through PsaR and the resultant impact on virulence. Infect Immun. 2006;74:1171–80.

84. Kloosterman T, Witwicki R, van der Kooi-Pol M, Bijlsma J, Kuipers O. Opposite effects of Mn 2+ and Zn 2+ on PsaR-mediated expression of the virulence genes pcpA, prtA, and psaBCA of Streptococcus pneumoniae. Journal of Bacteriology. 2008;190:5382–93.

85. Kambayashi J, Kobayashi T, Marcus NY, DeMott JE, Thalmann I, Thalmann R. Minimal concentrations of metabolic substrates capable of supporting cochlear potentials. Hear Res. 1982 May;7(1):105–14.

86. Thalmann R, Comegys TH, Thalmann I. Amino acid profiles in inner ear fluids and cerebrospinal fluid. Laryngoscope. 1982 Mar;92(3):321–8.

87. Thalmann R. Gradients of organic substances between fluid compartments of the cochlea. Acta Otolaryngol. 1985 Apr;99(3–4):469–77.

88. Hall-Stoodley L, Hu FZ, Gieseke A, Nistico L, Nguyen D, Hayes J, et al. Direct Detection of Bacterial Biofilms on the Middle-Ear Mucosa of Children With Chronic Otitis Media. JAMA. 2006 Jul 12;296(2):202–11.

89. Hoa M, Syamal M, Sachdeva L, Berk R, Coticchia J. Demonstration of nasopharyngeal and middle ear mucosal biofilms in an animal model of acute otitis media. Ann Otol Rhinol Laryngol. 2009 Apr;118(4):292–8.

90. Post JC, Hiller NL, Nistico L, Stoodley P, Ehrlich GD. The role of biofilms in otolaryngologic infections: update 2007. Curr Opin Otolaryngol Head Neck Surg. 2007 Oct;15(5):347–51.

91. Orihuela CJ, Radin JN, Sublett JE, Gao G, Kaushal D, Tuomanen EI. Microarray Analysis of Pneumococcal Gene Expression during Invasive Disease. Infection and Immunity. 2004;72(10):5582–96.

92. Hendriksen WT, Bootsma HJ, Estevão S, Hoogenboezem T, de Jong A, de Groot R, et al. CodY of Streptococcus pneumoniae: link between nutritional gene regulation and colonization. J Bacteriol. 2008 Jan;190(2):590–601.

93. Fleuchot B, Gitton C, Guillot A, Vidic J, Nicolas P, Besset C, et al. Rgg proteins associated with internalized small hydrophobic peptides : a new quorum-sensing mechanism in streptococci. 2011;80(March):1102–19.

94. Chen P, Qi F, Novak JAN, Caufield PW. The Specific Genes for Lantibiotic Mutacin II Biosynthesis in Streptococcus mutans T8 Are Clustered and Can Be Transferred En Bloc. APPLIED AND ENVIRONMENTAL MICROBIOLOGY. 1999;65(3):1356–60.

95. Mashburn-Warren L, Morrison D, Federle M. A novel double-tryptophan peptide pheromone controls competence in Streptococcus spp. via an Rgg regulator. Mol Microbiology. 2010;78(3):589–606.

96. Chang J, LaSarre B, Jimenez J, Aggarwal C, Federle M. Two Group A Streptococcal Peptide Pheromones Act through Opposing Rgg Regulators to Control Biofilm Development. PLoS Pathogens. 2011;7(8):e1002190.

97. Declerck N, Bouillaut L, Chaix D, Rugani N, Slamti L, Hoh F, et al. Structure of PlcR: Insights into virulence regulation and evolution of quorum sensing in Gram-positive bacteria. Proc Natl Acad Sci USA. 2007;104(47):18490–5.

98. Rocha-Estrada J, Aceves-diez AE. The RNPP family of quorum-sensing proteins in Gram-positive bacteria. Appl Microbiol Biotechnol. 2010;87:913–23.

99. Stenz L, Francois P, Whiteson K, Wolz C, Linder P, Schrenzel J. The CodY pleiotropic repressor controls virulence in gram-positive pathogens. FEMS Immunol Med Microbiol. 2011 Jul;62(2):123–39.

100. Fleuchot B, Gitton C, Guillot A, Vidic J, Nicolas P, Besset C, et al. Rgg proteins associated with internalized small hydrophobic peptides: a new quorum-sensing mechanism in streptococci. Mol Microbiol. 2011 May;80(4):1102–19.

101. Chang JC, LaSarre B, Jimenez JC, Aggarwal C, Federle MJ. Two group A streptococcal peptide pheromones act through opposing Rgg regulators to control biofilm development. PLoS Pathog. 2011 Aug;7(8):e1002190.

102. Aggarwal C, Jimenez JC, Lee H, Chlipala GE, Ratia K, Federle MJ. Identification of Quorum-Sensing Inhibitors Disrupting Signaling between Rgg and Short Hydrophobic Peptides in Streptococci. mBio. 2015 Jul 1;6(3):e00393-15.

103. Aggarwal C, Jimenez JC, Nanavati D, Federle MJ. Multiple length peptide-pheromone variants produced by Streptococcus pyogenes directly bind Rgg proteins to confer transcriptional regulation. J Biol Chem. 2014 Aug 8;289(32):22427–36.

104. Chang JC, LaSarre B, Jimenez JC, Aggarwal C, Federle MJ. Two group A streptococcal peptide pheromones act through opposing Rgg regulators to control biofilm development. PLoS Pathog. 2011 Aug;7(8):e1002190.

105. Bortoni ME, Terra VS, Hinds J, Andrew PW, Yesilkaya H. The pneumococcal response to oxidative stress includes a role for Rgg. Microbiology (Reading, Engl). 2009 Dec;155(Pt 12):4123–34.

106. Hoover SE, Perez AJ, Tsui H-CT, Sinha D, Smiley DL, DiMarchi RD, et al. A new quorum sensing system (TprA/PhrA) for Streptococcus pneumoniae D39 that regulates a lantibiotic biosynthesis gene cluster. Molecular Microbiology. 2015;97(2):229–43.

107. Kerr AR, Adrian P V, Esteva S, Groot R De, Alloing G, Claverys J. The Ami-AliA / AliB Permease of Streptococcus pneumoniae Is Involved in Nasopharyngeal Colonization but Not in Invasive Disease. Infection and Immunity. 2004;72(7):3902–6.

108. Cundell DR, Pearce BJ, Sandros J, Naughton AM, Masure HR. Peptide Permeases from Streptococcus pneumoniae Affect Adherence to Eucaryotic Cells. infection and Immunity. 1995;63(7):2493–8.

109. Zafar MA, Kono M, Wang Y, Zangari T, Weiser JN. Infant Mouse Model for the Study of Shedding and Transmission during Streptococcus pneumoniae Monoinfection. Infect Immun. 2016;84(9):2714–22.

110. Zafar MA, Hamaguchi S, Zangari T, Cammer M, Weiser JN. Capsule Type and Amount Affect Shedding and Transmission of Streptococcus pneumoniae. MBio. 2017 22;8(4).

111. Jensch I, Gámez G, Rothe M, Ebert S, Fulde M, Somplatzki D, et al. PavB is a surface-exposed adhesin of Streptococcus pneumoniae contributing to nasopharyngeal colonization and airways infections. Mol Microbiol. 2010 Jul 1;77(1):22–43.

112. Kanwal S, Jensch I, Palm GJ, Brönstrup M, Rohde M, Kohler TP, et al. Mapping the recognition domains of pneumococcal fibronectin-binding proteins PavA and PavB demonstrates a common pattern of molecular interactions with fibronectin type III repeats. Mol Microbiol. 2017 Sep;105(6):839–59.

113. Pracht D, Elm C, Gerber J, Bergmann S, Rohde M, Seiler M, et al. PavA of Streptococcus pneumoniae modulates adherence, invasion, and meningeal inflammation. Infect Immun. 2005 May;73(5):2680–9.

114. Antikainen J, Kuparinen V, Lähteenmäki K, Korhonen TK. Enolases from Gram-positive bacterial pathogens and commensal lactobacilli share functional similarity in virulence-associated traits. FEMS Immunol Med Microbiol. 2007 Dec;51(3):526–34.

115. Bergmann S, Rohde M, Chhatwal GS, Hammerschmidt S. alpha-Enolase of Streptococcus pneumoniae is a plasmin(ogen)-binding protein displayed on the bacterial cell surface. Mol Microbiol. 2001 Jun;40(6):1273–87.

116. Bergmann S, Hammerschmidt S. Versatility of pneumococcal surface proteins. Microbiology. 2006 Feb 1;152(2):295–303.

117. García B, Merayo-Lloves J, Rodríguez D, Alcalde I, García-Suárez O, Alfonso JF, et al. Different Use of Cell Surface Glycosaminoglycans As Adherence Receptors to Corneal Cells by Gram Positive and Gram Negative Pathogens. Front Cell Infect Microbiol. 2016;6:173.

118. Paterson GK, Orihuela CJ. Pneumococcal microbial surface components recognizing adhesive matrix molecules targeting of the extracellular matrix. Mol Microbiol. 2010 Jul 1;77(1):1–5.

119. Chhatwal GS. Anchorless adhesins and invasins of Gram-positive bacteria: a new class of virulence factors. Trends Microbiol. 2002 May;10(5):205–8.

120. Holmes AR, McNab R, Millsap KW, Rohde M, Hammerschmidt S, Mawdsley JL, et al. The pavA gene of Streptococcus pneumoniae encodes a fibronectin-binding protein that is essential for virulence. Mol Microbiol. 2001 Sep;41(6):1395–408.

121. Seifert KN, McArthur WP, Bleiweis AS, Brady LJ. Characterization of group B streptococcal glyceraldehyde-3-phosphate dehydrogenase: surface localization, enzymatic activity, and protein-protein interactions. Can J Microbiol. 2003 May;49(5):350–6.

122. Kinnby B, Booth NA, Svensäter G. Plasminogen binding by oral streptococci from dental plaque and inflammatory lesions. Microbiology (Reading, Engl). 2008 Mar;154(Pt 3):924–31.

123. Gase K, Gase A, Schirmer H, Malke H. Cloning, sequencing and functional overexpression of the Streptococcus equisimilis H46A gapC gene encoding a glyceraldehyde-3-phosphate dehydrogenase that also functions as a plasmin(ogen)-binding protein. Purification and biochemical characterization of the protein. Eur J Biochem. 1996 Jul 1;239(1):42–51.

124. Bergmann S, Rohde M, Hammerschmidt S. Glyceraldehyde-3-Phosphate Dehydrogenase of Streptococcus pneumoniae Is a Surface-Displayed Plasminogen-Binding Protein. Infect Immun. 2004 Apr;72(4):2416–9.

125. King SJ, Hippe KR, Weiser JN. Deglycosylation of human glycoconjugates by the sequential activities of exoglycosidases expressed by Streptococcus pneumoniae. Mol Microbiol. 2006 Feb;59(3):961–74.

126. Limoli DH, Sladek JA, Fuller LA, Singh AK, King SJ. BgaA acts as an adhesin to mediate attachment of some pneumococcal strains to human epithelial cells. Microbiology (Reading, Engl). 2011 Aug;157(Pt 8):2369–81.

127. Singh AK, Pluvinage B, Higgins MA, Dalia AB, Woodiga SA, Flynn M, et al. Unravelling the multiple functions of the architecturally intricate Streptococcus pneumoniae β-galactosidase, BgaA. PLoS Pathog. 2014 Sep;10(9):e1004364.

128. Brittan JL, Buckeridge TJ, Finn A, Kadioglu A, Jenkinson HF. Pneumococcal neuraminidase A: an essential upper airway colonization factor for Streptococcus pneumoniae. Mol Oral Microbiol. 2012 Aug;27(4):270–83.

129. Oiki S, Mikami B, Maruyama Y, Murata K, Hashimoto W. A bacterial ABC transporter enables import of mammalian host glycosaminoglycans. Sci Rep. 2017 21;7(1):1069.

130. Marion C, Stewart JM, Tazi MF, Burnaugh AM, Linke CM, Woodiga SA, et al. Streptococcus pneumoniae can utilize multiple sources of hyaluronic acid for growth. Infect Immun. 2012 Apr;80(4):1390–8.

131. Berry AM, Paton JC. Additive attenuation of virulence of Streptococcus pneumoniae by mutation of the genes encoding pneumolysin and other putative pneumococcal virulence proteins. Infect Immun. 2000 Jan;68(1):133–40.

132. Chapuy-Regaud S, Ogunniyi AD, Diallo N, Huet Y, Desnottes J-F, Paton JC, et al. RegR, a global LacI/GalR family regulator, modulates virulence and competence in Streptococcus pneumoniae. Infect Immun. 2003 May;71(5):2615–25.

133. Yadav MK, Chae S-W, Park K, Song J-J. Hyaluronic acid derived from other streptococci supports Streptococcus pneumoniae in vitro biofilm formation. Biomed Res Int. 2013;2013:690217.

134. Martin R, Larsen AH, Corey RA, Midtgaard SR, Frielinghaus H, Schaffitzel C, et al. Structure and Dynamics of the Central Lipid Pool and Proteins of the Bacterial Holo-Translocon. Biophys J. 2019 May 21;116(10):1931–40.

135. Wang Z, Guo C, Xu Y, Liu G, Lu C, Liu Y. Two Novel Functions of Hyaluronidase from Streptococcus agalactiae Are Enhanced Intracellular Survival and Inhibition of Proinflammatory Cytokine Expression. Infect Immun. 2014 Jun;82(6):2615–25.

136. Jedrzejas MJ. Pneumococcal Virulence Factors: Structure and Function. Microbiol Mol Biol Rev. 2001 Jun;65(2):187–207.

137. Warner JB, Lolkema JS. CcpA-dependent carbon catabolite repression in bacteria. Microbiol Mol Biol Rev. 2003 Dec;67(4):475–90.

138. Carvalho SM, Kloosterman TG, Kuipers OP, Neves AR. CcpA Ensures Optimal Metabolic Fitness of Streptococcus pneumoniae. PLoS ONE. 2011 Oct 21;6(10):e26707.

139. Orihuela CJ, Radin JN, Sublett JE, Gao G, Kaushal D, Tuomanen EI. Microarray Analysis of Pneumococcal Gene Expression during Invasive Disease. Infect Immun. 2004 Oct;72(10):5582–96.

140. Orihuela CJ, Gao G, Francis KP, Yu J, Tuomanen EI. Tissue-Specific Contributions of Pneumococcal Virulence Factors to Pathogenesis. J Infect Dis. 2004 Nov 1;190(9):1661–9.

141. McCluskey J, Hinds J, Husain S, Witney A, Mitchell TJ. A two-component system that controls the expression of pneumococcal surface antigen A (PsaA) and regulates virulence and resistance to oxidative stress in Streptococcus pneumoniae. Molecular Microbiology. 2004 Mar 1;51(6):1661–75.

.

# PART IV - APPENDIX

## CHAPTER 6 - Gene Expression Analysis in the Pneumococcus

This chapter collects a series of protocols that aim to improve the analysis of the gene expression in pneumococci. These protocols cover sample acquisition and processing, as well as experimental design and analysis They were designed with the objective of analyzing the pneumococcal gene expression on conditions as diverse as planktonic culture, biofilm, and *in vivo*. These protocols are meant to be used with NanoString Expression analysis, a tool that has been proved accurate in term of detection and resolution. The protocols here cover methods of sample isolation and preservation, and its benefits might be useful in combination with other techniques, such as RNA-seq and microarray. While these protocols were initiated a while back, they have been perfected over the years. In particular, I contributed with the protocol for extraction and quantification of pneumococcal RNA isolated from biofilms, a condition accepted as convoluted due to the poor sample quality and yields. By using and optimizing the protocols described in this section, I was able to quantify with great precision pneumococcal gene expression in biofilms.

This collection of protocols was published as a part for the collaborative online publication “Streptococcus pneumoniae Methods and Protocols” by Federico Iovino (1).

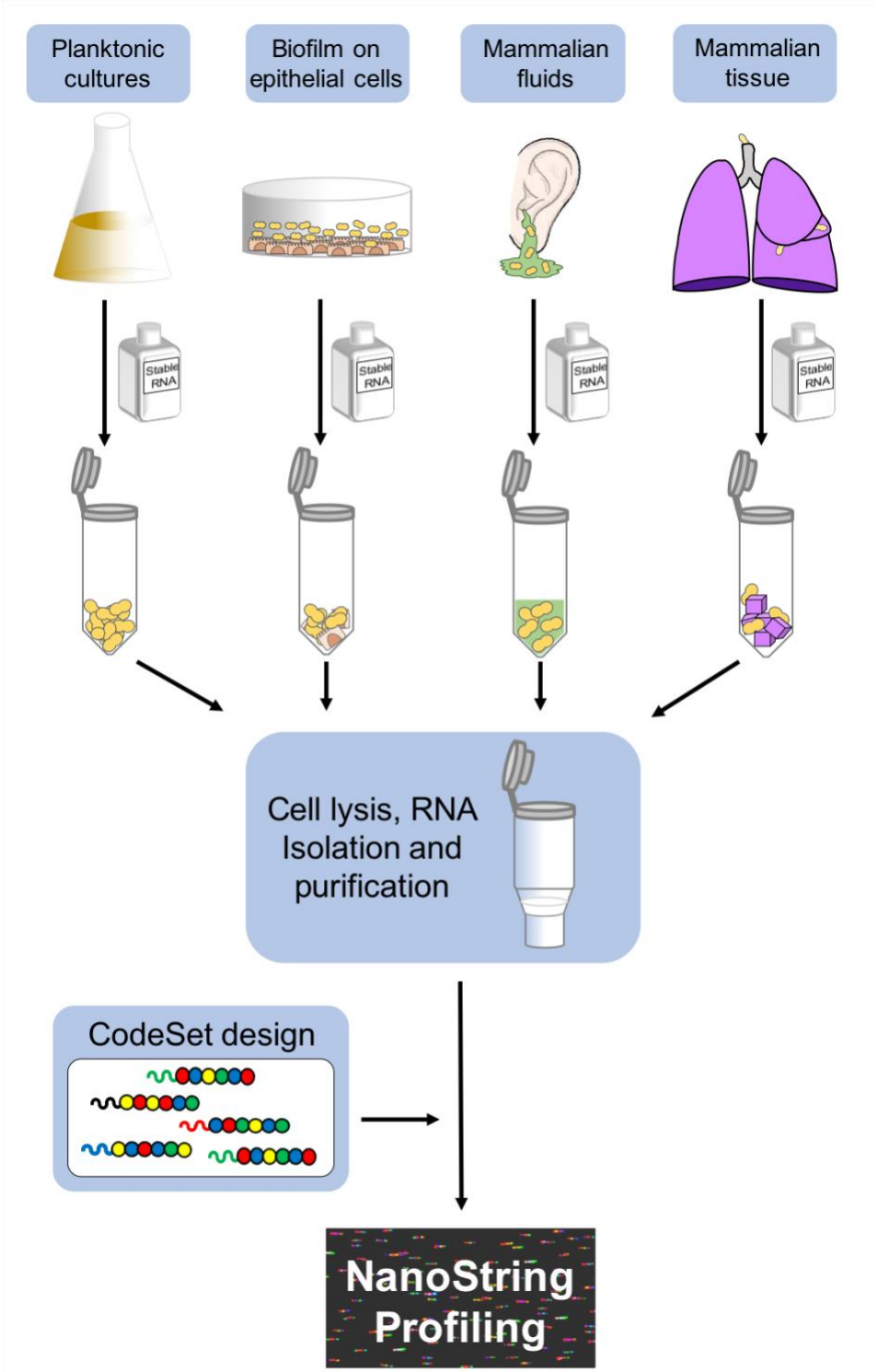
**Citation:** Eutsey R.A., Woolford C.A., Aggarwal S.D., Cuevas R.A., Hiller N.L. (2019) Gene Expression Analysis in the Pneumococcus. In: Iovino F. (eds) Streptococcus pneumoniae. Methods in Molecular Biology, vol 1968. Humana Press, New York, NY

**Disclaimer:** The compilation of protocols described in this section was developed for the analysis of the pneumococcal expression. Protocols cover sample collection, processing, probe design, data analysis, and interpretation.

**Personal contribution:** I contributed to experimental design, sample acquisition and processing, images design, manuscript writing, editing and proofreading.

## 6.1 Abstract

Bacterial cells modify their gene expression profiles throughout different stages of growth and in response to environmental cues. Analyses of gene expression across conditions reveal both conserved and condition-specific gene responses of bacteria to adapt to these dynamic conditions. In this chapter, we present a guide to pneumococcal RNA extraction for use in the NanoString nCounter platform. The nCounter is a highly effective method to measure gene expression of bacteria not only in a planktonic mode of growth, but also in the presence of host cells where the RNA of interest represents only a small portion of the total material. Here we focus on sample preparation, pneumococcal cell lysis, RNA extraction and purification, and NanoString profiling




## 6.2 Graphical abstract

**Figure 6.1 Schematic of RNA Analysis using the NanoString**

## 6.3 Introduction

Gene expression profiling has long been performed to understand the biology and pathogenesis of the human pathogen *Streptococcus pneumoniae*. Numerous studies have applied qRT-PCR to dissect the expression of small sets of genes (2–6). Further, microarrays covering the whole genome of model strains or the majority of the genes in the pneumococcal pangenome have been widely applied for a comprehensive view of gene expression (7–13). These transcriptomic studies have greatly enhanced our understanding of gene regulation in the pneumococcus. The current challenge is to measure gene expression of pneumococci grown *in vivo*. RNAseq and NanoString are technologies that address this gap (12–16). Here, we present protocols for RNA extraction and processing using the NanoString nCounter (17). Our protocols address planktonic cells, *in vitro* biofilms grown on epithelial cells, middle-ear effusion from the chinchilla model of otitis media, and lungs from the murine model of pneumonia (**Figure 6.1**).

## 6.4 Materials and Methods

### 6.4.1 Sample preservation

1. Liquid Nitrogen.
2. RNAprotect Cell Reagent (Qiagen) or RNAlater Stabilization Solution (Thermo).
3. In-house RNA stabilization solution: 5.3M Ammonium Sulfate, 20mM EDTA, 25mM Sodium Citrate in distilled water.

### 6.4.2 Lysis of pneumococcal cells

1. Lysozyme: dissolved in distilled water to 100 mg/ml (Sigma-Aldrich).
2. Mutanolysin: dissolved in distilled water to 200 μg/ml (Sigma-Aldrich).
3. Proteinase K: dissolved in distilled water to 20 mg/ml (VWR).
4. Zirconia/Silica beads 0.5mm in diameter (RPI).
5. GentleMACS dissociator (Miltenyi Biotec).
6. GentleMACS M tubes (Miltenyi Biotec).
7. Acid-Phenol:Chloroform 5:1 at pH 4.5 (Sigma-Aldrich).
8. Tris EDTA buffer (TE).
9. Nuclease-free water.

### 6.4.3 RNA extractions and purification

1. Column-based RNA extraction kit.

i. Quick-RNA MiniPrep kit (Zymo Research).

ii. RNeasy Mini kit (Qiagen).

iii. PureLink RNA Mini kit (Invitrogen).

2. Turbo DNase (Invitrogen).

### 6.4.4 RNA quantification

1. Spectrophotometer (Nanodrop).

## 6.5 Methods

### 6.5.1 Sample Preservation

*(a) planktonic cultures*

1. Grow pneumococcal cells in a liquid culture in a conical tube (see note 1 for details on growth conditions).

We routinely use $10^8$-$10^9$ cells, which provide an estimated 10-100 μg of RNA. The stage of growth and volume required to collect this number of cells will depend on experimental goals.

2. To the culture, add either 2 volumes of RNA Protect Bacteria reagent or 1 volume of in-house RNA stabilization solution.
3. Centrifuge samples at 4000×*g* for 10 minutes to pellet cells.
4. Discard the supernatant.
5. Freeze cell pellet at -80°C in the conical tubes.

*(b) in vitro biofilms on epithelial cells*

1. Grow biofilms in cell culture flasks (see notes on growth conditions).
2. Remove media from cell culture flask, using care to avoid dispersion of the biofilm.
3. Add RNAlater bacteria reagent or in-house RNA stabilization solution until the cells on the bottom of the flask are completely immersed.
4. Scrape the bottom of the flask with a cell sterile cell scraper to release the bacterial cells, use a pipette to remove all material from the flask and collect the biofilm material into a conical tube.
5. Centrifuge at 4000×*g* for 10 minutes to pellet the cells.

6. Discard the supernatant.
7. Freeze cells pellet at -80°C in the conical tube.

*(c) infected mammalian tissue (soft tissue)*[1]

1. Collect the tissue of interest from the animal, place in a microfuge tube (without any liquid).
2. *Immediately* flash-freeze the sample by placing the microfuge tube in liquid nitrogen.
3. Store sample at -80°C.

*(d) infected mammalian fluids (effusions)*[1]

1. Collect effusion fluids from chinchilla middle ears and place in a microfuge tube (see notes for details).
2. *Immediately* flash-freeze the sample by placing the microfuge tube in liquid nitrogen.
3. Store sample at -80°C.

### 6.5.2 Sample Lysis[2]

(*a,b) planktonic cultures and in vitro biofilms on epithelial culture*

1. Prepare the lysis solution: For each sample, you'll need 200μl of lysis solution (128μl TE buffer, 27μl lysozyme, 27μl mutanolysin, and 18μl proteinase K).
2. Remove sample from freezer and thaw at RT.
3. Add 200μl lysis solution to the cell pellet.
4. Resuspend the pellet and incubate at room temperature for 20 minutes, while vortexing every 5 minutes.
5. Proceed to 'RNA Extraction and Purification'.

*(c) infected mammalian tissue samples (this protocol has been used for murine lungs)*

1. Prepare the following: gentleMACS M tube, tissue lysis solution, and 15ml conical tube.
2. Remove sample from freezer and thaw at room temperature.
3. Move the tissue sample, weighing no more than 150mg, to a gentleMACS M tube. Add 800μl lysis solution (512μl TE buffer, 108μl lysozyme, 108μl mutanolysin, and 72μl proteinase K).
4. To homogenize the tissue, place the tube in the gentleMACS dissociator and run the preset RNA protocol RNA_02.01 (see note 3 for an alternative).
5. Incubate the material in the M tube at room temperature for 20 minutes to allow lysis of the bacterial cell wall. At this point, the solution will be cloudy with a reddish color.
6. Remove all liquid (~800μl) from the M tube and place into a 15ml conical tube.
7. Proceed to 'RNA Extraction and Purification'.

*(d) infected mammalian fluids (this protocol has been used for chinchilla middle ear effusions)*

1. Prepare the following: 1.5ml screw cap microcentrifuge tube with 0.5mm Zirconia/Silica beads, cell lysis solution, 1.5ml microcentrifuge tube.
2. Remove sample from freezer and thaw at room temperature.
3. Mix 200μl of effusion with 14 μl of TE, 14 μl lysozyme, 14 μl of mutanolysin, and 9 μl proteinase K in a 1.5ml screw cap microcentrifuge tube with an estimated 100μl of 0.5mm Zirconia/Silica beads.
4. Incubate at room temperature for 10 minutes.
5. Bead beat twice, for 30 seconds each.
6. Incubate at room temperature for 10 minutes.
7. Remove the liquid (leaving beads behind, as they settle to the bottom) and transfer to a 1.5ml microcentrifuge tube.

8. Proceed to ‘RNA Extraction and Purification’.

### 6.5.3 RNA Extraction and Purification

This step is the same for all samples described above.

1. Perform RNA extraction and purification using a column-based RNA mini prep kit following the specific manufacturer’s instructions. We use the Quick-RNA MiniKit from Zymo (however other kits should also work). Scale the prep for the amount of starting liquid material. Specifically: 200μl for planktonic cultures, 200μl for biofilms, 800μl for solid tissue samples, and 200μl for body fluids.
2. For NanoString analysis, samples do not need to be DNA-free. NanoString probes only hybridize to single-stranded targets, and the sample hybridization and processing steps do not result in DNA denaturation. Thus, there is no need for DNA removal beyond the one provided by the kit.
3. Quantify the RNA. To this end, one can use a spectrophotometer, such as a Nanodrop. The purity of the sample can be estimated by calculating the 260/280 absorbance ratio, where pure RNA will have a ratio around 2.0.

### 6.5.4 CodeSet Design

The design of the CodeSet is performed in conjunction with the NanoString team. CodeSet reagents and cartridges are designed for 12 samples to be run at one time. Considerations are listed below.

1. First select the RNA transcripts of interest for your experiment. Currently CodeSets can be designed to detect from 20 to 800 transcripts. Each probe is 100bp long.
2. Include multiple normalization controls in the CodeSet. We recommend at least two housekeeping genes that have little to no variability in expression across all treatment conditions. Optimally, genes used for normalization controls should

have levels of abundance in the range of the experimental genes. Normalization is performed on the geometric mean of these controls. We consistently use *gyrB* (DNA gyrase, B subunit, SPD_0709) and *metG* (methionyl-tRNA synthetase, SPD_0689). For experiments within related conditions we also use *gapdh* (glyceraldehyde-3-phosphate dehydrogenase, SPD_1004), however, we do not recommend this control to compare across diverse modes of growth (such as planktonic versus biofilms or *in vitro* versus *in vivo*). Do not use 16S rRNA, as the high abundance of this gene will cover a high proportion of the field of view, resulting in uninterpretable or suboptimal data (section 3.5).

3. We consistently capture a range of probe signals of five orders of magnitude. That is, RNA reads range from single digits to a few hundred thousand counts for a single molecule. However, if interested in very low abundance transcripts, consider designing a CodeSet with only low abundance transcripts.
4. Include positive controls using genes induced in the condition of interest. For *in vivo* conditions, we use *ply* (pneumolysin, SPD_1726) and *vp1* (virulence peptide 1, SPD_0145) (18).
5. Consider specificity. For studies in the presence of host cells, avoid probes that can cross-react to host molecules.
6. Consider strain variation. If multiple pneumococcal strains will be tested, design probes to conserved regions of the transcript, so that they capture allelic variants. In general, 95% identity to the target sequence is sufficient to ensure binding. This step can be performed in collaboration with the NanoString team and/or using BLASTn. With BLASTn, generate a database with all the alleles of interest and select probes that recognize each and every sequence with at least 95% identity over the length of the entire probe.

### 6.5.5 Estimation of RNA amount

During data acquisition, the probes-transcript complexes are immobilized on the nCounter cartridge, imaged and enumerated. For successful quantification, the binding

density on the cartridge must be within an acceptable range (varies with nCounter instrument; 0.05-2.25 for Gen1 nCounter and 0.1 to 1.8 for the nCounter SPRINT, per manufacturer's instructions). That is, there must be sufficient RNA to allow statistically relevant counts of the transcripts of interest, but not so much material that the field of view (FOV) is saturated and single molecules cannot be discerned. The transcripts of interest may correspond to a high percentage of the total RNA (in a pure planktonic culture), or a very small percentage when bacteria are mixed with mammalian tissue. Thus, the optimal starting RNA concentration is highly dependent on the percentage of the transcript of interest relative to total RNA. It will also depend on the number of probes in the CodeSet and abundance of the targets within the species of interest. In this session, we provide empirical information collected over our experiments. In all cases the final volume of RNA that can be used is set at a maximum of 5μl.

1. For pure planktonic cultures, where all the RNA is from pneumococcus, we dilute the samples to a concentration of approximately 10-15 ng/μl, for an estimated total of 50-75 ng of RNA per experiment. This has been successful for CodeSets ranging from 50-200 targets. In contrast, experiments performed with 10 times this concentration were not interpretable due to higher than optimal binding density and image saturation.
2. For biofilms: pneumococcal cells form a robust biofilm on mammalian epithelial cells (these tend to be much more robust than biofilms grown on abiotic surfaces). We recommend 50 to 150 ng of total RNA per experiment.
3. For middle ear effusions, where there is a mix of bacterial and host, we do not dilute the RNA. We use RNA at 80-200 ng/μl, for an estimated total of 400-1000ng of RNA per experiment.
4. For murine lungs, we estimate that bacteria RNA is <0.1% of total RNA. Thus, we do not dilute the RNA and sometimes concentrated the RNA 3-4 fold. In this case, the optimized concentration will depend on the infection rate of the tissue. As a rule of thumb, we homogenize one lung, and when possible use samples with at least $10^5$ CFU/mg of tissue. However, we have found that we can reach a binding density within the desired range with as little as $3x10^4$ CFU/mg of tissue,

if we concentrate the material 3-4x by lyophilization. In murine lungs, our nucleic acid concentration ranges from 1,000-3,000 ng/μl before concentration and reaches 15,000 ng/μl after concentration. In a few cases, using very high concentrations has saturated the FOV. Thus, if the CFUs are above $10^5$ CFU/mg of tissue we recommend using the RNA from purification, without any further manipulation.

### 6.5.6 NanoString Profiling

This is the hybridization protocol for the nCounter XT CodeSet gene expression assays.

1. Remove a tube of reporter CodeSet and Capture ProbeSet reagents from -80°C freezer and thaw at room temperature. Flick several times and briefly spin down reagents.
2. Create a master mix by adding 70μl hybridization buffer to the Reporter CodeSet tube. Invert repeatedly and spin down gently.
3. Add 8μl master mix to each tube in a 0.2ml thin-wall strip of 12 tubes.
4. Add 5μl RNA sample or RNA + water to bring up to 5 μl volume.
5. Add 2μl Capture ProbeSet to each tube, flick briefly, and spin down gently.
6. Immediately place tubes in preheated 65°C thermocycler and incubate reactions for at least 16h.

When used in an nCounter SPRINT Profiler, the samples are loaded into a cartridge per manufacturer's instructions and placed in the instrument for sample purification and data collection. Consumable reagents required include cartridges and the nCounter SPRINT Profiler Reagent Pack. Other instrumentation is also available from NanoString for gene expression analysis.

nSolver Analysis Software is available for data analysis following manufactures instructions. Software automatically checks data quality and flags samples that are out of the normal range. One can perform technical adjustments such as normalization and

determine ratios and differential expression using built-in functions or export data to software of choice.

## 6.6 Notes

1. Pneumococcal Growth.

(a) planktonic cultures: for growth in liquid culture, colonies from a frozen stock are grown overnight on TSA plates. The next morning colonies are inoculated into media of choice and incubated at 37°C and 5% $CO_2$ without shaking.

(b) biofilms on epithelial cells: we employ human type II lung epithelial cell line A549. These are cultured with DMEM 1x with 4.5 g/L glucose, 8mM L-glutamine and 1mM sodium pyruvate and supplemented with 10% fetal bovine serum containing 20 units/mL of penicillin and 20 µg/mL of streptomycin. Cell culture is maintained on sterile, tissue-culture treated plastic at 37°C with humidified 5% carbon dioxide atmosphere. Epithelial cells are seeded at a confluent density of 1 $x10_6$ cells·$cm_{-2}$. Confluence is confirmed next day by light microscopy. Then, the cells are thoroughly washed with PBS to remove any trace of antibiotic prior to bacterial inoculation. Starter cultures of the selected pneumococcal strain are grown to an $OD_{600}$ of 0.025 and then 100 µl of bacterial cells are inoculated into the tissue culture plates (estimated $3x10_6$ bacteria in total). Plates are incubated aerobically in humidified conditions at 37°C and 5% carbon dioxide for selected time (we use16 h). Before collection, the supernatants are carefully aspirated to remove debris and biofilms are washed with PBS to remove non-adherent and weakly adherent bacteria. Following this step guaranty low experimental variation as we have able to determine (**Figure 6.2**).

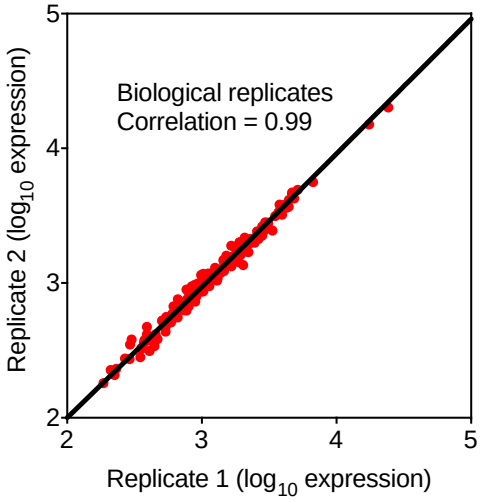



**Figure 6.2 Biofilm expression quantified by Nanostring shows strong reproducibility.**

Graph show strong correlation between biological replicates. Pearson correlation was used between sample data.

2. Collection of middle ear effusion from the chinchilla middle ear. Once the animal has been euthanized, remove the bulla from the chinchilla (19). Using small surgical scissors, make a minor incision (5mm diameter) at the top of the bulla. Check for presence of effusion by moving the bulla to sense and view liquid. If there is liquid present, insert a 0.5mL sterile disposable plastic Pasteur pipette and collect all the effusion (the volume ranges from 50 to 400 μl in a strain and host-dependent manner). If liquid is absent or estimated to be below 50 μl, add 400 μl of PBS, pipette up and down to wash bulla, and collect all liquid.

3. Alternative for tissue homogenizing. You can homogenize the tissue sample using a bead-beating apparatus. Add the tissue to a 1.5 ml screw cap tube filled with 100 μl of 0.5 mm Zirconia/Silica beads.

### 6.6.1 Footnotes.

1. In our initial RNA extractions from murine tissue, we employed RNAlater to preserve RNA (Life Technologies) and observed very low yields. Our yields improved with flash freezing the tissue immediately after removal from the animal.

2. We lyse cells at room temperature, and have found that we achieved high quality RNA (as measured by an Agilent Bioanalyzer). However, many laboratories maintain material on ice while extracting RNA. We elected room temperature, to ensure full activity of the cell wall lytic enzymes.

## 6.7 References.


1. Iovino F. Streptococcus pneuminiae, Methods and Protocols. New York, NY: Humana Press; 2019. (Methods in Molecular Biology).
2. Ogunniyi AD, Giammarinaro P, Paton JC. The genes encoding virulence-associated proteins and the capsule of Streptococcus pneumoniae are upregulated and differentially expressed in vivo. Microbiology. 2002 Jul 1;148(7):2045–53.
3. Ogunniyi AD, Mahdi LK, Trappetti C, Verhoeven N, Mermans D, Van der Hoek MB, et al. Identification of Genes That Contribute to the Pathogenesis of Invasive Pneumococcal Disease by In Vivo Transcriptomic Analysis. Infect Immun. 2012 Sep;80(9):3268–78.
4. Mahdi LK, Deihimi T, Zamansani F, Fruzangohar M, Adelson DL, Paton JC, et al. A functional genomics catalogue of activated transcription factors during pathogenesis of pneumococcal disease. BMC Genomics. 2014 Sep 8;15(1):769.
5. Mahdi LK, Ogunniyi AD, LeMessurier KS, Paton JC. Pneumococcal Virulence Gene Expression and Host Cytokine Profiles during Pathogenesis of Invasive Disease. Infect Immun. 2008 Feb;76(2):646–57.
6. LeMessurier KS, Ogunniyi AD, Paton JC. Differential expression of key pneumococcal virulence genes in vivo. Microbiology (Reading, Engl). 2006 Feb;152(Pt 2):305–11.
7. Tettelin H, Nelson KE, Paulsen IT, Eisen JA, Read TD, Peterson S, et al. Complete genome sequence of a virulent isolate of Streptococcus pneumoniae. Science. 2001 Jul 20;293(5529):498–506.
8. Orihuela CJ, Radin JN, Sublett JE, Gao G, Kaushal D, Tuomanen EI. Microarray Analysis of Pneumococcal Gene Expression during Invasive Disease. Infect Immun. 2004 Oct;72(10):5582–96.
9. Sebert ME, Patel KP, Plotnick M, Weiser JN. Pneumococcal HtrA protease mediates inhibition of competence by the CiaRH two-component signaling system. J Bacteriol. 2005 Jun;187(12):3969–79.
10. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Meth. 2011 Oct;8(10):785–6.
11. Dagkessamanskaia A, Moscoso M, Hénard V, Guiral S, Overweg K, Reuter M, et al. Interconnection of competence, stress and CiaR regulons in Streptococcus pneumoniae: competence triggers stationary phase autolysis of ciaR mutant cells. Mol Microbiol. 2004 Feb;51(4):1071–86.
12. Kimaro Mlacha SZ, Romero-Steiner S, Hotopp JCD, Kumar N, Ishmael N, Riley DR, et al. Phenotypic, genomic, and transcriptional characterization of Streptococcus pneumoniae interacting with human pharyngeal cells. BMC Genomics. 2013;14:383.
13. Kadam A, Eutsey RA, Rosch J, Miao X, Longwell M, Xu W, et al. Promiscuous signaling by a regulatory system unique to the pandemic PMEN1 pneumococcal lineage. PLOS Pathogens. 2017 May 18;13(5):e1006339.
14. Cheng S, Clancy CJ, Xu W, Schneider F, Hao B, Mitchell AP, et al. Profiling of Candida albicans gene expression during intra-abdominal candidiasis identifies biologic processes involved in pathogenesis. J Infect Dis. 2013 Nov 1;208(9):1529–37.

15. Xu W, Solis NV, Ehrlich RL, Woolford CA, Filler SG, Mitchell AP. Activation and Alliance of Regulatory Pathways in C. albicans during Mammalian Infection. PLoS Biol. 2015 Feb 18;13(2):e1002076.

16. Xu W, Solis NV, Filler SG, Mitchell AP. Pathogen Gene Expression Profiling During Infection Using a Nanostring nCounter Platform. Methods Mol Biol. 2016;1361:57–65.

17. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotech. 2008 Mar;26(3):317–25.

18. Cuevas RA, Eutsey R, Kadam A, West-Roberts JA, Woolford CA, Mitchell AP, et al. A novel streptococcal cell-cell communication peptide promotes pneumococcal virulence and biofilm formation. Mol Microbiol. 2017 Aug;105(4):554–71.

19. Giebink GS. Otitis media: the chinchilla model. Microb Drug Resist. 1999;5(1):57–72.

## CHAPTER 7 - Gene Acquisition by a Distinct Phyletic Group within *Streptococcus pneumoniae* Promotes Adhesion to the Ocular Epithelium

This chapter focuses on a horizontal gene transfer event that occurred between two streptococci species. Findings from this chapter suggest that *Streptococcus pneumoniae* acquired a gene encoding for an adhesin molecule from *Streptococcus suis.* Molecular analysis suggests that the product of the transferred gene plays a role in adhesion to the ocular epithelium. This work was published in the 2017 mSphere journal. The original manuscript can be accessed using the following reference:

Antic I, Brothers KM, Stolzer M, Lai H, Powell E, Eutsey R, **Cuevas RA**, Miao X, Kowalski RP, Shanks RMQ, Durand D, Hiller NL. Gene Acquisition by a Distinct Phyletic Group within *Streptococcus pneumoniae* Promotes Adhesion to the Ocular Epithelium. mSphere. 2017; 2(5). PubMed PMID: 29085912, PMCID: PMC5656748

**Disclaimer:** The conceptualization and work in this chapter was initiated by Dr. N. Luisa Hiller at Allegheny General Hospital, Pittsburgh, PA in collaboration with Dr. Robert M. Q. Shanks at University of Pittsburgh School of Medicine Pittsburgh, PA, and Dr. Dannie Duran at Carnegie Mellon University.

**Personal Contributions**. In this project I contributed to experimental design, data acquisition and analysis, manuscript preparation, revision and editing.

## 7.1 Abstract

Tissue tropism by *Streptococcus pneumoniae* (pneumococcus) governs which cells and body sites are colonized by this human pathogen. While pneumococcal strains colonize multiple body sites and tissues, only a distinct phyletic group appears to colonize the conjunctiva. In this study, we sequenced the genomes of six pneumococcal strains isolated from eye-infections. The conjunctivitis isolates are grouped in a distinct phyletic group together with a subset of nasopharyngeal isolates. The keratitis (infection of the cornea) and endophthalmitis (infection of the vitreous body) isolates are grouped together with the remainder of pneumococcal strains. Phenotypic characterization of these isolates is consistent with morphological differences associated with the distinct phyletic group. Specifically, conjunctivitis isolates differ from other eye-associated isolates in that they form clumps in a planktonic cultures and long chains in biofilms grown on abiotic surfaces. To begin to investigate the association between genotype and epidemiology, we focused on a predicted surface exposed agglutinin receptor protein (AR1) encoded exclusively by the distinct phyletic group. Phylogenetic analysis of AR1 in the context of a streptococcal species tree suggests that *ar1* was acquired by lateral gene transfer from *Streptococcus suis*. Further, a deletion mutant of AR1 displayed decreased adherence to cultured cells from the ocular epithelium when compared to the isogenic wild-type and complemented strains. Together these findings suggest that acquisition of genes from outside the species has contributed to pneumococcal tissue tropism by enhancing the ability of a subset of strains to infect the ocular epithelium causing conjunctivitis.

## 7.2 Introduction

Streptococcal species are characterized by extensive intra-species gene diversity that play critical roles in tissue tropism and, consequently, disease outcomes (1–3). In Group A Streptococci (GAS), a significant association has been detected between gene content and the ability to colonize the skin versus the upper respiratory tract, leading to either impetigo or pharyngitis (1). In contrast, in the human pathogen *Streptococcus pneumoniae* (pneumococcus), an association between genomic background and site of infection had not been forthcoming (4). This absence was unexpected given the extensive diversity of pneumococcal genomes, and the variety of body sites, including the nasopharynx, the middle ear, lungs, heart, brain, and eyes, that pneumococcus colonizes (2,5–7). Pneumococcal conjunctivitis is a notable exception. The genomic composition of strains causing pneumococcal conjunctivitis differs from that of strains causing disease in other tissues; this difference in gene content is so extensive that this set of strains localizes to a distinct phyletic group (8–10).

The best-characterized feature shared across conjunctivitis-associated pneumococcal strains is the absence of the genes encoding the polysaccharide capsule (8–11). The capsule is the main pneumococcal virulence determinant, thus the absence of the capsule has important clinical implications to the distinct phyletic group (11). The non-encapsulated strains are less virulent than encapsulated strains and are not thought to disseminate in single-strain infections (12). Further, the capsular structure is the target of the pneumococcal vaccine, thus the non-encapsulated strains escape the vaccine. Whereas the absence of capsule is a shared feature, it is not a distinguishing property of the distinct phyletic group. Many non-encapsulated strains have been isolated from body sites other than the eye and are phylogenetically clustered with the majority of pneumococcal strains (13,14). Thus, the absence of capsule is a highly clinically relevant feature shared by conjunctivitis strains but not what drives the association between genomic background and ability to cause conjunctivitis.

Pneumococcus eye disease is not limited to conjunctivitis, where it infects the conjunctival epithelial layers. This pathogen can also infect the vitreous body inside the eye (endophthalmitis) and the cornea (keratitis) (15–17). It is unclear whether there are

morphological and genomic features shared by all isolates that infect the human eye. In this study we combine genomics, phylogenetics, and cell adhesion studies to gain insight into the genomes of isolates from multiple types of pneumococcal ocular infections and the tissue tropism associated with conjunctivitis isolates.

## 7.3 Results

### 7.3.1 Sequencing of pneumococcal strains isolated from ocular infections

To compare the genomes of pneumococcal strains isolated from multiple types of eye infections we sequenced six ocular-associated strains. One isolate is from an endophthalmitis infection (strain E709), three are from keratitis infections (strains K2521, K2527 and K2557), and two are from conjunctivitis infections (strains B1598 and B1599). All isolates are de-identified clinical samples obtained at the Charles T. Campbell Ophthalmic Laboratory at the University of Pittsburgh Eye Center from October 2012 to May 2013. The strains were sequenced using single molecule real time (SMRT) technology and the genome sequences have been deposited in GenBank (**Table 1**).

**Table 7.1 Strains used in this study**

| Strain | E709 | K2521 | K2527 | K2557 | B1598 | B1599 |
|---|---|---|---|---|---|---|
| Isolation Date | 5/13/13 | 1/28/13 | 2/21/13 | 6/19/13 | 10/19/12 | 11/3/12 |
| Isolation Site | UPMC | UPMC | UPMC | UPMC | UPMC | UPMC |
| Serotype | 23a | 15* | 19a | 17f | NE | NE |
| MLST | 338 | 3280 | 320 | 2355 | 2315 | 2315 |
| Genbank ID | JBOR00000000 | JBOS00000000 | JBOT00000000 | JBOU00000000 | JBOV00000000 | JBOW00000000 |
| BioProject | PRJNA235160 | PRJNA235267 | PRJNA235268 | PRJNA235269 | PRJNA235270 | PRJNA235271 |
| BioSample | SAMN02691894 | SAMN02691895 | SAMN02691896 | SAMN02691897 | SAMN02691898 | SAMN02691899 |
| CG% | 39.6 | 39.6 | 39.8 | 39.8 | 39.6 | 39.6 |
| Number of Contigs | 37 | 32 | 29 | 111 | 158 | 38 |
| Read Coverage | 55.22 | 73.71 | 63.13 | 30.39 | 52.07 | 49.71 |
| Sequencing Platform | SMRT | SMRT | SMRT | SMRT | SMRT | SMRT |
| Estimated Genome Size | 2155121 | 2184812 | 2180290 | 2051681 | 2902565 | 2422872 |
| Resistance | gentamicin, amikacin | tobramycin, polymyxin B | gentamicin, tobramycin, polymyxin B | gentamicin, polymyxin B, tobramycin, sulfonamide; intermediate resistance: ciprofloxacin ofloxacin | gentamicin, trimethoprim, polymyxin B, tobramycin | gentamicin, trimethoprim, polymyxin B, tobramycin |

Pneumococcal genomes are conventionally classified using multilocus sequence typing (ST) (serves as a proxy for lineages), polysaccharide capsular types and drug resistance profiles (2,18). The six strains encompass five ST types, four capsular types

as well as non-encapsulated, and display varied resistance to a standard panel of antibiotics (Table 1). The conjunctivitis strains are non-encapsulated, while the keratitis and endophthalmitis strains are encapsulated. All strains displayed some degree of drug resistance, the broadest was observed for keratitis strain K2557 that is resistant to gentamicin, polymyxin B, tobramycin, sulfonamide and has intermediate resistance to fluoroquinolone antibiotics ciprofloxacin and ofloxacin, (**Table 1**). In summary, strains isolated from the different ocular infections display diverse ST, serotype and resistance profiles.

### 7.3.2 Phylogenetics and comparative genomics of pneumococcal strains isolated from ocular infections

To establish the phylogenetic relationship of these eye-associated strains, we compared the six genomes with a diverse set of thirty-four pneumococcal genomes from strains isolated from blood, lung, and nasopharynx (**Suppl. Table 1**). This highly curated set consists of genomes used for the first large scale pneumococcal pangenome studies (2,6), genomes from PCV-7 immunized children (19), and genomes from non-encapsulated isolates (20). Together this set reflects diversity in multi-locus sequence types, serotypes, disease states, and geographical locations. We generated and aligned the core genomes of these sequences and produced a maximum likelihood phylogenetic tree (**Figure 7.1**). In agreement with published work (5,9,10), the most prominent feature of the phylogenetic tree is the presence of a distinct and strongly supported branch that contains the conjunctivitis strains (**Figure 7.1**).

A second noteworthy feature of the phylogenetic tree is the tight grouping between conjunctivitis and nasopharyngeal strains in the distinct phyletic group. Specifically, there is a short distance between the conjunctivitis strains B1599 and B1598 and the non-encapsulated nasopharyngeal carriage strain MNZ85; further there is strong support for the branch that groups these strains together (**Figure 7.1**). The mix of nasopharyngeal and conjunctivitis isolates on this distinct phyletic group suggests that diversification of these strains is not a consequence of niche isolation. Finally, in stark contrast to the conjunctivitis strains, the endophthalmitis and keratitis isolates are

grouped with the remaining pneumococcal strains (**Figure 7.1**). This arrangement argues that the distinct phyletic group is not a feature shared by all strains that infect the eye; rather, it is a specific feature of conjunctivitis. In summary, phylogenetic analysis of six pneumococcal strains isolated for three types of eye infection revealed that conjunctivitis isolates are grouped in a distinct phyletic group together with non-encapsulated nasopharyngeal strains, while keratitis and endophthalmitis isolates are grouped with the remainder of strains.

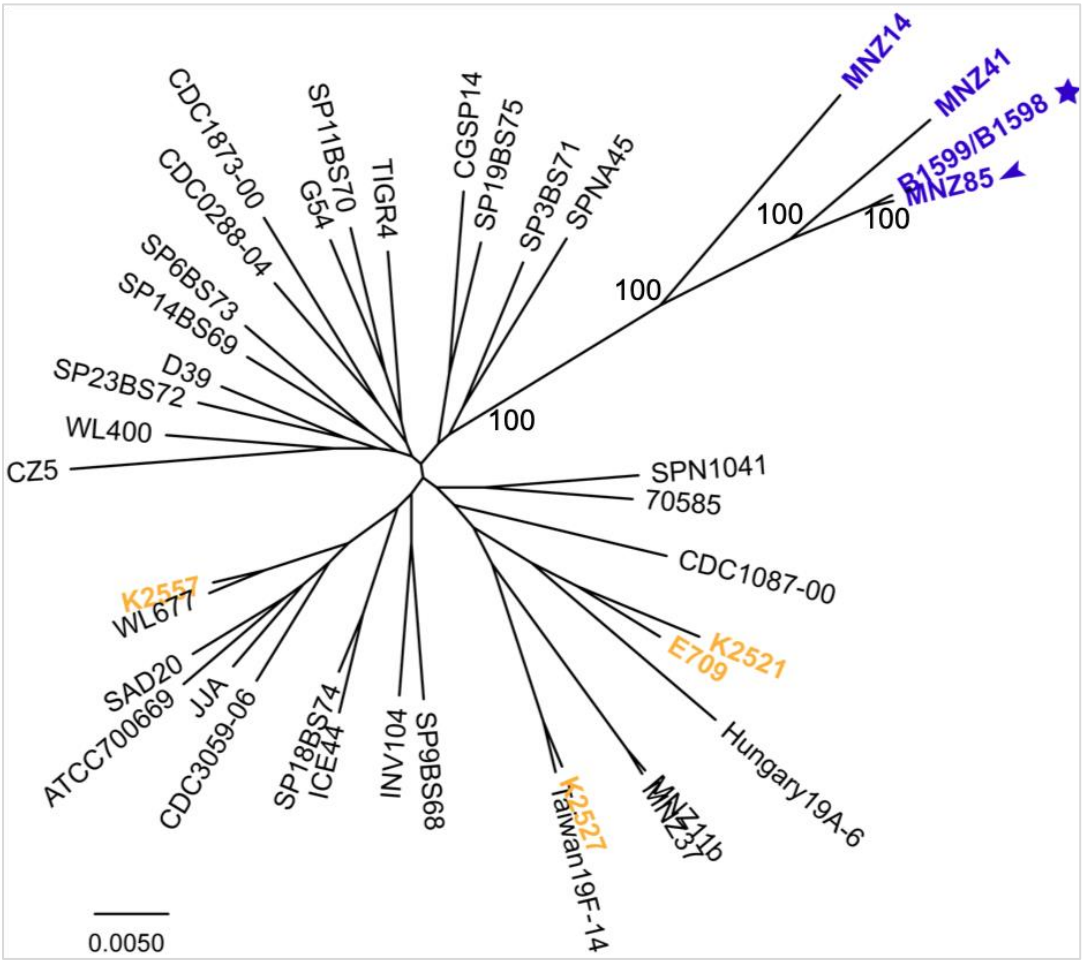



**Figure 7.1 Phylogenetic analysis of pneumococcal strains.**

Maximum likelihood phylogeny of forty pneumococcal strains generated from the core genome. Purple: distinct phyletic group; arrow: nasopharyngeal isolate; star: conjunctivitis isolates.

Yellow: endophthalmitis and keratitis associated strains**.** The scale bar indicates the number of substitutions per site. The numbers on the branch denote bootstrap values. The basal position of the distinct phyletic group presented in the context of other streptococcal species is illustrated in Figure S2.

To identify the genes that distinguish the distinct phyletic group from other pneumococcal strains, we singled out the coding sequences (CDS) that are present in the conjunctivitis isolates (B1598 and B1599) but absent in strains from the other major branch. We identified seventy-seven coding sequences, many of which were grouped together on the genome (**Suppl. Figure 7.1**). The most widespread functional feature of these genes is a predicted localization to the bacterial surface, that is, surface-localization motifs (LPXTG or YSIRK), surface related functions (such as ABC transporter) and/or motifs predicted to interact with the host (β-galactosidase, choline binding, sialidase, and two agglutinin receptors (ARs)) (8,21–25).

The association between this distinct phyletic group and conjunctivitis isolates (prominently those of ST448 and ST344) has been observed in three independent large-scale genomic studies (8–10). To compare our finding to previous work, we utilized a set of 616 genomes isolated in Massachusetts from 2007-2010 (26). This set includes ten strains from the distinct phyletic group identified by Croucher and colleagues, and termed SC12, as well as 606 additional genomes. The datasets are highly consistent: genes captured exclusively in our distinct phyletic group are present in multiple SC12 genomes and are either absent or rare (<2%) in the 606 non-SC12 genomes (**Suppl. Figure 7.1**). The set of uncharacterized CDSs, predicted to be surface exposed, are likely candidates for molecular components of tissue tropism to the ocular epithelium during conjunctivitis.

### 7.3.3 Morphological features that distinguish strains isolated from conjunctivitis relative to other eye-infections.

The core (**Figure 7.1**) and distributed (**Suppl. Figure 7.1**) genes of conjunctivitis strains and of a subset of nasopharyngeal strains differ from that of the majority of pneumococcal strains. This raises the question as to whether and how these strains differ regarding morphology.

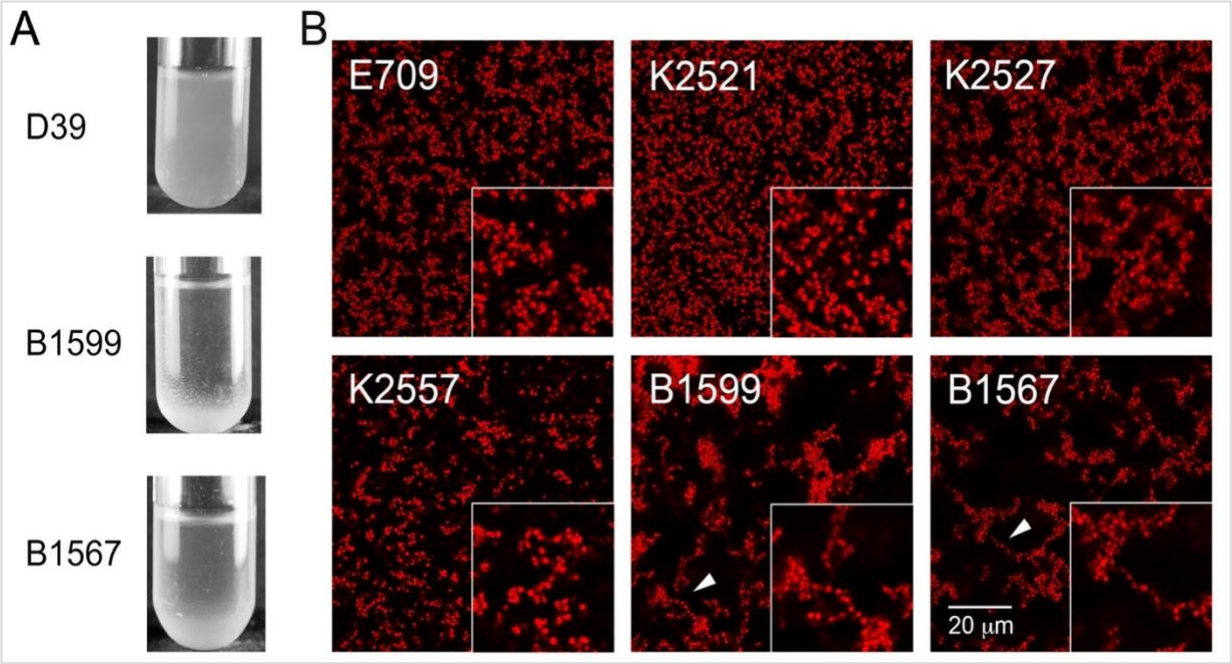



**Figure 7.2 Conjunctivitis-associated strains B1599 and B1567 exhibit aggregates in planktonic culture and form abundant and long chains in biofilms.**
A. Planktonic cultures of strains B1599, B1567, and D39. Conjunctivitis isolates precipitate at the bottom of the test tube, but no precipitate is observed for model strain D39. B. Confocal images of 72h biofilms, fixed and stained with Syto59. Boxes on the bottom right display a magnified view. Conjunctivitis isolates B1599 and B1567 form long chains (white arrows point to examples of long chains); however, these chains are not observed in other eye-associated strains.

To explore this question, we investigated the morphology of conjunctivitis strains in planktonic and biofilm modes of growth. Multiple conjunctivitis outbreaks have been associated with strains from ST488 (represented in Figure 1 as isolate MNZ14); however, our conjunctivitis isolate is from the rare ST2315 (27,28). Thus, for phenotypic characterization we added a conjunctivitis isolate obtained from ST488 from our clinical library (BS1567). First, we analyzed planktonic cultures of two conjunctivitis strains (BS1599 (ST2315) and BS1567 (ST488)) and compared them to a model strain D39. In contrast to the control, both conjunctivitis isolates formed clumps in stationary phase planktonic cultures (**Figure 7.2A**). Second, we investigated biofilm-related phenotypes, given that pneumococci can colonize the epithelium by growing in a biofilm mode of

growth (29–34). We employed confocal microscopy to compared biofilm growth of three-day biofilms of conjunctivitis strains with those of other ocular strains. In contrast to previously characterized pneumococcus biofilms (35) and to the keratitis and endophthalmitis isolates, the majority of cells in strain B1599 were organized into long chains, often with over 20 connected cells (**Figure 7.2B**). Long chains were also formed in strain B1567, but were less abundant than B1599. To our knowledge, this phenotype has not been previously reported in wild-type pneumococcal strains, only pneumococcal mutants (36,37). These observations suggest that the conjunctivitis isolates may display differences in planktonic and biofilm growth that discriminate them from most characterized pneumococcal isolates, as well our isolates from other ocular infections.

### 7.3.4 Functional studies of the AR1 protein.

The Agglutinin Receptor 1 (*ar1*) represents a notable gene among the distributed genes unique to the distinct phyletic group. The predicted *ar1* CDS contains an N-terminal Sec-type signal sequence and a C-terminal LPXTG motif, strongly suggesting AR1 is secreted from the cytosol and attached to the peptidoglycan cell wall by a sortase (21,24) . AR1 contains a glucan-binding protein C (GbpC) domain (CDD: e-value 8.2e-80), followed by three adhesin isopeptide-forming domains (SspB-C2 type) (CDD: e-values 3.12e-69, 4.61e-68 and 3.57e-63, respectively) (38) (**Figure 7.3A**). Proteins containing the GbpC domain are found in several species of oral *Streptococci* (39,40), where they participate in dextran binding and biofilm formation (41). SspB-C2 type domains are present in oral streptococci as components of SspB (42). In *S. gordonii*, SspB is a three-domain adhesion unit that facilitates the formation of mixed species biofilms in the oral cavity (41,42). We found that AR1 is expressed in planktonic cultures of strain B1599, as determined by qRT-PCR (**Figure 7.3B, black bars**). The predicted surface localization and the presence of adhesive domains are consistent with a role for AR1 in host interactions, and its restricted genomic distribution is consistent with a tropism to the ocular epithelium.

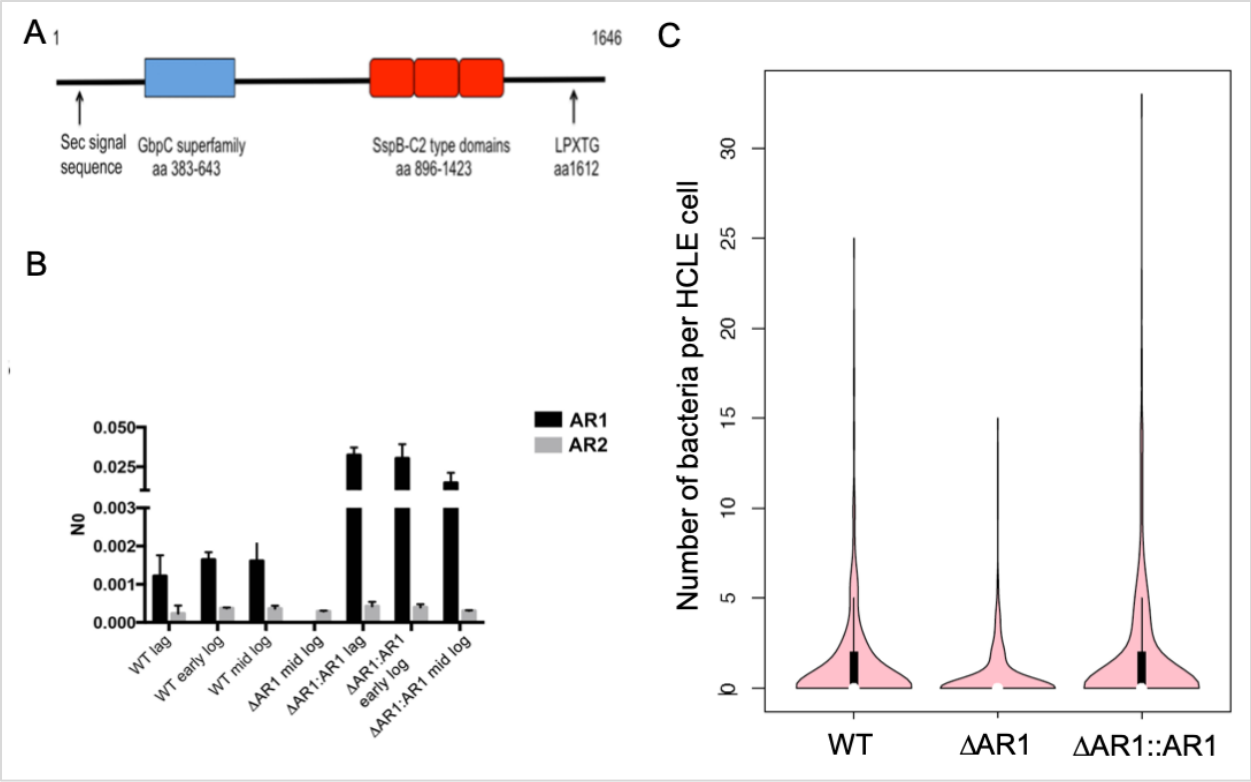



**Figure 7.3 Functional Analyses of the Agglutinin Receptor Protein (AR1).**

**A.** Schematic of AR1 protein illustrating domains implicated in adhesion and biofilm formation. **B.** Gene expression of *ar1* in *S. pneumoniae* B1599 (black bars); for comparison we show the expression of a second predicted agglutinin receptor protein- *ar2* (gray bars)**.** Y-axis displays $N_0$, an arbitrary fluorescence unit representing RNA amount in the respective samples. **(C)** Role of AR1 in attachment to HCLE cells. HCLE cells were exposed to *S. pneumoniae* strains B1599, B1599ΔAR1 or B1599ΔAR1::AR1 for 30 minutes, and HCLE with bacteria attached were enumerated. Experiments were performed in triplicate, and for each strain 460 HCLE cells were analyzed for presence of bacteria and the number of bacteria attached. The results are plotted using a violin plot, generated in the R-statistical package. The violin plot displays the distribution of the data: the pink areas displays the density plot, the thick black bars represent the midspread of the data (interquartile range), the thin black lines display the 95% confidence interval, and the white circles are the median.

To test whether AR1 has a role in adhesion to the ocular epithelium, or to planktonic clumping and long chain phenotype we observed, we generated a deletion mutant (B1599ΔAR1) and an overexpressing strain (B1599ΔAR1::AR1) (**Suppl. Table 7.2**). The levels of AR1 in the overexpression strain vary from 9-26 fold the levels of wild-type strains, depending on the stage of planktonic growth (**Figure 7.3B**). The mutant and wild-type strains displayed the same morphology; thus, *ar1* is not

responsible for the differences characterized in **Figure 7.2** (data not shown). Next, we investigated whether AR1 influences adhesion to human corneal limbal epithelial cells (HCLE). HCLE were selected because they are at the interface between the conjunctiva and the cornea, they are representative of the ocular surface in that they produce many mucins and compounds associated with cornea and conjunctiva (43,44). Independently, each strain was allowed to adhere to HCLE cells, and the number of bacteria attached per cell was enumerated after gentle washing to remove non-adherent bacteria. Tukey-test was used to establish the statistical significance of the differences in adherence to 460 HCLE cells among the three strains (45). The attachment for the mutant was significantly different from that of the wild-type ($p=0.0019$) and the complement ($p=0.000003$) strains. Specifically, the loss of AR1 led to a decrease in the number of HCLE cells with any bacteria attached, as well as a decrease in the number of bacteria attached per HCLE cell (**Figure 7.3C**). Overexpression of AR1 in the deletion strains restored the wild-type phenotype. These data suggest that AR1 plays a role in adhesion to the ocular epithelium.

### 7.3.5 Origin of the pneumococcal AR1 gene

A scan of the genomic region surrounding AR1 using CONJscan-T4SSscan software (46) identified genes encoding components of Type IV secretion systems and relaxases associated with Integrative Conjugative Elements (ICEs), consistent with a foreign origin for *ar1*. The closest relatives of the region encoding AR1 and neighboring ICE components were identified via blastn search, resulting in 26 genomic regions (**Suppl. Table S3**). Of these, five were derived from ICE sequences, further supporting the inference that *ar1* is part of a mobile genetic element. One sequence is from outside the streptococcal genus: an ICE from *Enterococcus faecium*. Two sequences are from within *S. pneumoniae*: an ICE from isolate 403790 and a non-typable pneumococcal strain (NT_110_58) previously localized to a distinct phyletic group (9). All other matches are streptococcal sequences from species outside *S. pneumoniae*: 13 *Streptococcus suis* strains, two *Streptococcus anginosus* strains, six strains from

various pyogenic species, and one each from *Streptococcus thermophilus* and *Streptococcus gallolyticus subsp. macedonicus (Streptococcus macedonicus).*

While this patchy phyletic distribution, covering multiple dispersed groups within the genus, is suggestive of lateral gene transfer, phylogenetic analysis provides stronger evidence for lateral transfer. Explicit comparison of a gene tree with the associated species tree can distinguish between parallel loss and horizontal transfer and infer the specific transfer events that occurred. A Bayesian gene tree (**Figure 7.4A**) was constructed from a codon-aware multiple alignment of *ar1* sequences extracted from the 21 closely related genomic regions remaining after redundant sequences and sequences lacking ORF predictions were removed (**Suppl. Table S3**). Inter-species transfers were inferred by reconciling this gene tree with a previously published streptococcal species tree (**Suppl. Figure S3**) (47) using Notung 2.9 (48).

Reconciliation infers the minimal set of events that explains the topological incongruence between gene and species trees. Reconciliation with Notung 2.9 yielded four possible evolutionary scenarios with minimal event histories (**Figure 7.4B and Suppl. Figure 4A-D**). The four scenarios are largely in agreement, differing only in the events involving pyogenic strains, which are unrelated to our central question. In particular, all four scenarios predict a single horizontal transfer from *Streptococcus suis* to *Streptococcus pneumoniae*. This suggests that pneumococcal *ar1* was acquired from *Streptococcus suis* at the base of the distinct phyletic lineage.

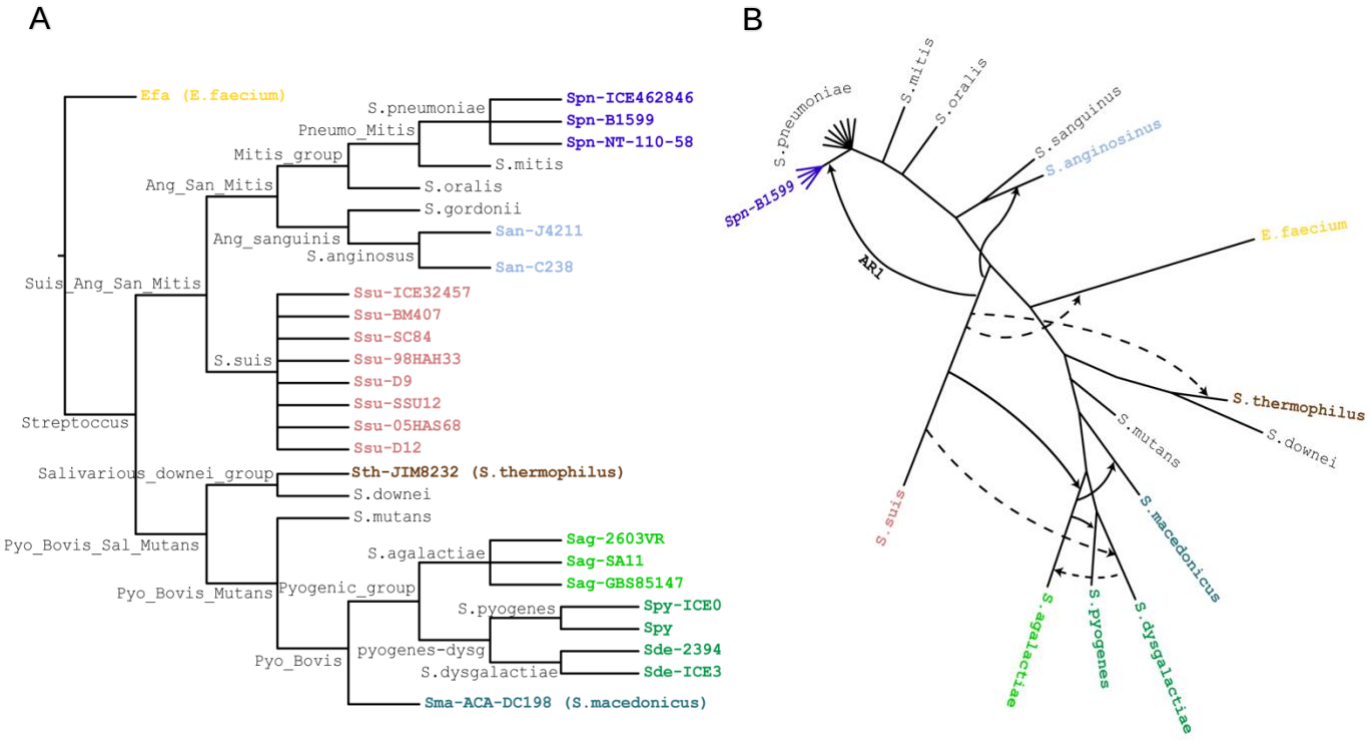



**Figure 7.4 Phylogenetic reconciliation reveals history of transfers in the origin of *ar1.***

**A.** Bayesian phylogeny of *ar1 g*ene sequences constructed with MrBayes v.3.2.6 from codon-aware multiple alignment of nucleic acid sequences. The tree is midpoint rooted. Branches labeled with posterior probabilities representing statistical support. The scale bar represents nucleotide substitutions per site. Colors according to label in (B). **B.** Interspecies *ar1* gene transfers in the context of the Streptococcus phylogeny. Transfers were inferred by reconciling the gene tree (A) with a species tree (Fig S3) based in the streptococcal species tree from Richards *et al.* (generated from a core set of 136 genes sampled from 44 streptococcal species). The evolutionary history contains all species in (**A)** (colored labels), plus additional species (black labels) to provide a representative sample of well-studied taxa with at least one species for each major taxonomic group. The distinct phyletic lineage in pneumococcus is shown in dark blue. Four evolutionary scenarios were inferred with Notung 2.9 (see also Figs S4 and S5), each with 9 transfers, represented as arrows from the donor to the recipient species (solid and dashed lines represent transfers inferred in the upper and lower subtrees in **A,** respectively). All evolutionary scenarios support a horizontal transfer from *S. suis* to the base of the distinct lineage (dark blue).

## 7.4 Discussion

The clustering of conjunctivitis isolates into a distinct phyletic group is the only clear instance of niche specialization in pneumococcus, known to-date. In this study, we compare the genomes and evolutionary histories of pneumococcal strains isolated from various body sites, focusing on multiple types of eye infections. We observe that conjunctivitis isolates cluster with a subset of nasopharyngeal isolates and away from all other strains, including those isolated from patients with keratitis and endophthalmitis. Phenotypically, the conjunctivitis isolates display clumping in planktonic culture and long chains in biofilms grown on an abiotic surface. Further, we demonstrate that the *ar1* CDS, was acquired from *S. suis* by gene transfer, and that its product plays a role in adhesion to the ocular epithelium. Our work combines comparative genomics, phylogenetics, and cell biology to explore the evolution and molecular mechanisms that underlie this unique instance of pneumococcus niche specialization.

The long branch that split up the distinct phyletic group from other pneumococcal strains highlights the extensive differences in their core genomes, and thus provides evidence of strain differentiation. This differentiation does not appear to be the consequence of niche separation, as some members of the distinct phyletic group are nasopharyngeal isolates and likely co-colonize with pneumococcal isolates from the main branch. Another mechanism promoting strain differentiation, in pneumococcus and other bacteria, is restriction-modification systems (49,50). Whereas genomes in the distinct phyletic group do encode a type I restriction-modification system absent in the major branch, it is only encoded in a subset of these distinct strains (**Figure S1**). This limited distribution suggests that this restriction modification system is not driving strain differentiation. Thus, the molecular mechanisms driving this instance of strain differentiation remain a fascinating topic of study.

The strains in the distinct phyletic group share a set of genes not observed in strains outside this group. Do these differences in strain possession translate into differences in transmission routes and/or tissue tropism? Pneumococcal strains from the main branch are transmitted by nasal shedding, and the nasopharynx is the main resevoir of circulating strains (51). The phylogenetic grouping of conjunctivitis and

nasopharyngeal strains is consistent with this model, in that strains from the distinct phyletic group may be transmitted to and from the nasopharynx and then disseminated to the ocular epithelium. Alternatively, strains acquired directly into the eye may have the ability to colonize the nasopharynx. Once in contact with the conjunctiva, strain in the distinct phyletic branch may be able to overcome a specific host barrier and gain access to the ocular epithelium. Phylogenetic positioning of the keratitis and endophthalmitis isolates outside of the distinct phyletic group suggests that this ability is not associated with other eye tissues or the ability to survive immunity in the eye. Instead, the ability to colonize the ocular epithelium could be guided by conjunctive specific features, such as adhesion and/or colonization of the conjunctiva epithelium, or by an intermediate niche such as eyelids, eyelashes or even fingers that people use to rub their eyes. Moreover, these strains may be more resistant to desiccation or more competent at survival in the tear film, ultimately allowing productive infection of the conjunctiva. We postulate that the distinct phyletic group encodes genes that allow these strains to overcome a host barrier associated with transmission and/or colonization of the ocular epithelium.

We captured phenotypic features that differentiated our set of conjunctivitis isolates from pneumococcal strains from outside the distinct phyletic group. We observed long pneumococcal chains within biofilms of wild-type strains from this distinct phyletic group. This extent of chain formation has not been reported for biofilms of wild-type strains grown on abiotic surfaces, and consistently we did not observe these chains in isolates from keratitis or endophthalmitis. Long chains have been previously described in other growth conditions and have been implicated in colonization. Specifically, chaining in the TIGR4 strain (which is not a member of the distinct phyletic group) can be induced by growth in human nasal airway surface fluid (52). Deletion mutants in lytic enzymes (LytA and LytB), and in a putative flippase involved in cell wall maintenance (*tacF)* also lead to chaining. The latter leads to failure to colonize in an intranasal and intraperitoneal mouse model, but the former mutant result in greater attachment to lung cells in culture and to the upper respiratory tract in a mouse model. A second, perhaps related, phenotype associated with conjunctivitis isolates is formation of aggregates in planktonic culture. Valentino and colleagues reported

clumping for conjunctivitis isolates after addition of gp340, a glycoprotein found in tears (10). They hypothesized that orthologs of AR1 and AR2 (SspB-C1 and SspB-C2) may be important for this phenotype. In our strains, we did not observe a difference between wild-type and □*ar1* strains regarding clumping of planktonic cultures or long chains in biofilms. It seems probable that the propensity to form long chains in early biofilms and/or to clump in culture is widespread across isolates from the distinct phyletic group and may play an important role in host-interactions. Finally, our findings suggest that AR1 contributes, but is not required, for adhesion to the ocular epithelium. It seems probable that multiple redundant adhesins influence epithelial binding. For instance, a second agglutinin receptor molecule (AR2) is also expressed in the distinct phyletic group (**Figure S1**). Like AR1, AR2 encodes multiple SspB-C2 domains (CDD: e-values 1.89e-70, 1.8e-69, 2.17e-67 and 6.9e-66, respectively) and a C-terminal LPXTG motif, consistent with a cell-wall attachment adhesin. This study is an initial step in dissecting the phenotypes and molecular components that result in binding to the ocular epithelium.

The *ar1* gene is encoded within an ICE, and reconciliation between the streptococcal species-tree and the *ar1* gene-tree suggest *ar1* was acquired by gene transfer from *S. suis*. Further, multiple *ar1* xenologs are encoded within ICEs, such as MB56Spyo009, ICESde3396, ICEsu32457, and ICESsu32457 (53–56). Similarly, the reconciliation suggests these ICEs were also acquired by gene transfer from *S. suis*. Thus, it is likely that *S. suis* serves as an ICE resevoir dispersing these elements to many species within the streptococcal genus (56). It is tempting to speculate that ICE-associated adhesins may modify strain-tissue interactions, either promoting or suppressing dissemination to specific hosts, body sites, and/or tissues.

The distinct phyletic group in the pneumococcal species tree provides a striking example of strain differentiation and tissue tropism. It generates many open questions on how these strains differ from other pneumococcal strains regarding evolution, transmission routes, morphology, gene expression, and host-interactions. Studies of these strains will provide exciting insight into the evolution and the biology of pneumococcus.

## 7.5 Materials and Methods

**Bacterial strains**: The *Streptococcus pneumoniae* isolates B1567, B1598, B1599, E709, K2521, K2527 and K2557 were obtained from patient ocular infections and stored by the Charles T. Campbell Eye Microbiology Laboratory at the UPMC Eye Center (Table 1). Strains are named according to infection type and order of isolation: E for endophthalmitis, K for keratitis and B for conjunctivitis, and numbers indicates the order in which the isolate was obtained. For instance, isolate E709 is the 709th endophthalmitis-derived strain in the collection.

**Bacterial growth conditions:** Frozen bacterial stocks were streaked onto trypticase soy agar plates containing 5% sheep blood (BD BBL™). All *S. pneumoniae* strains were grown in Columbia Broth (Thermo Scientific) at 37°C with 5% $CO_2$ without shaking. Medium was supplemented with antibiotics at 1ug/ml for tetracycline and 100 ug/ml for spectinomycin.

**Pacific Biosciences Single Molecule Real Time (SMRT) sequencing:** Genomic DNA (10ug) was extracted from strains E709, K2527, K2521, K2557, B1598, and B1599. Following digestion into ~10kb fragments, the DNA was end-repaired, purified with AMPure PB beads, and ligated to SMRTbell hairpin adapters. The SMRTbell libraries were further purified and quantified with a NanoDrop spectrophotometer and an Agilent 2100 Bioanalyzer. Polymerase-bound libraries were loaded onto a PacBio RS for sequencing with two SMRT cells per strain after completing primer and polymerase binding.

**Multi locus sequence typing of strains:** Sequences for the seven MLST alleles with available whole genome sequences were extracted from sequence data and are listed in Table 1. For strain B1567, the ST was acquired by Sanger sequencing of PCR amplimers (18). For typing, the allele sequences were submitted to the *S. pneumoniae* MLST website (pubmlst.org/spneumoniae) (57).

**Gene annotation:** Genomes were submitted to RAST for CDS prediction and annotation (58).

**Generation of *S. pneumoniae* species tree***:* Forty streptococcal strains were selected for phylogenetic analysis (Table S1). These genome correspond to those used

for the first large-scale pneumococcal pangenome studies (2,6), as well as additional genomes from PCV-7 immunized children (19), and non-encapsulated strains (20). Combined these strains reflect a large variety of multilocus sequence types (MLSTs) and serotypes, as well as strains isolated from different disease states and geographic locations. The whole genome sequence (WGS) for all 40 strains were aligned using MAUVE (59) and the core region corresponding to 1,345,780 total sites and 92,728 informative sites was extracted from the MAUVE output files. Alignment of the core region was performed using MAFFT (MODEL) (60) and model selection was performed using MODELTEST (61). The phylogenetic tree was built with PhyML 3.0 ,model GTR+I(0.63) using maximum likelihood and 100 bootstrap replicates (62).

**Gene clustering and selection of genes unique to distinct phylogenetic clade:** The CDSs were organized into gene clusters as previously described (63). Briefly, similar genes were identified by tfasty36 (FASTA v.3.6 package) for six-frame translation homology searches of all predicted proteins against all possible translations (64). The output was parsed such that genes with at least 70% identity over 70% of their sequence were grouped into gene clusters allowing the strains to be analyzed for presence or absence of clusters. We selected genes present in strains B1598, B1599, MNZ14, MNZ85, MNZ41 but absent in the remaining pneumococcal strains.

We utilized a set of 616 genomes isolated in Massachusetts from 2007-2010 (26), to compare our findings to previous work. This set, analyzed by Croucher and colleagues, contains ten strains from a distinct phylogenetic branch termed SC12, as well as 606 additional genomes. To compare the sets, we used blastP and an E-value cutoff of 1e-10, to query the seventy-seven genes unique to our distinct phyletic group against a database of 1,231,516 sequences from the 616 genomes in the Massachusetts from 2007-2010 set. The blast results were parsed to exclude hits with less than 70% identity and/or less than 70% coverage. A python script was used to parse the Blast output into a matrix correlating the query with the output. The data was represented as a heatmap using the ggplot2 package in the R statistical package (65). All positive hits are plotted in Figure S1 (genomes without any hits are not represented in the heat map).

**Phyletic distribution:** Sequences were retrieved using NCBI blastn to search the non-redundant (nr) database, restricted to *Bacillus*, *Lactobacillus* and *Streptococcus* (taxid: 91061). The query corresponded to a 26,160bp region of B1599 (within B1599_contig 202) that comprises *ar1* and adjacent genes. The resulting matches were curated to include only sequences with an E-value of zero and a maximum bit score above 8100; all of these matches entailed high-scoring pairs (HSPs) with at least 74% identity distributed along the full length of the query sequence, suggesting that these are genomic regions that are homologous to the query sequence in its entirety. The final set includes 26 genomic sequences, of which five are ICE sequences (Table S3).

**Gene tree reconstruction:** To analyze the evolutionary origin of the *ar1* gene, we extracted the sequences that encode AR1 and inferred a gene tree. Genomes lacking ORF predictions were discarded. For each of the remaining genomes, the *ar1* nucleotide and amino acid sequences were extracted from the Genbank file using genbank_to_fasta.py, downloaded from the Rocab lab website (http://rocaplab.ocean.washington.edu/tools/genbank_to_fasta/). This resulted in 21 sequences after removing redundant sequences and adding B1599AR1 to the set.

Multiple alignment of the AR1 protein sequences was performed using MAFFT (60), with the “E-INS-I” option via the Jalview dashboard (66). The alignment was converted into a codon-aware nucleic acid alignment using PAL2NAL (Supplementary File 1) (67), which was trimmed manually in Jalview to remove columns with more than 25% gaps. The best phylogenetic model for each of the three codon positions, GTR+G, HKY+G and HKY+G, was selected by the Bayesian Information Criterion using MODELGENERATOR (68) in Topali (69). A gene tree (Figure 4a) was then constructed from the trimmed alignment using MrBayes v3.2.6 (70); model parameters were fit by MrBayes for each codon site. The Markov chain Monte Carlo (MCMC) process was run for 500,000 generations with a sampling frequency of 15 generations and default settings for all other MCMC parameters. The gene tree was then midpoint rooted in FigTree, which was also used in figure generation (http://tree.bio.ed.ac.uk/software/figtree/).

**Gene tree-species tree reconciliation:** To infer the history of evolutionary events during *ar1* evolution, the resulting gene tree was reconciled with a species tree (Supplementary Figure S02), adapted from the tree from Richards *et al*. for a reduced set of taxa. The Richards species tree was constructed from a concatenation of a core set of 136 genes across 44 Streptococcus species (46 strains), representing 8 major groups. All species that encode a putative homolog of *ar1* were retained in the species tree for this study, as were additional species (that do not harbor a putative *ar1* homolog) to provide a broad representation of the streptococcal genus, including at least one species for each of the major taxonomic groups identified by Richards *et al*. (Mitis, Sanguinis, Anginosus, Salivarius, Downei, Mutans, Pyogenic, Bovis). When an *ar1* homolog was predicted in more than one strain per species, all strains were added to the species tree. Strain relationships are unresolved in this tree; if more than two strains were included for a single species, their relationships are represented as a non-binary node (i.e, polytomies).

We utilized Notung 2.9 to reconcile the *ar1* gene tree with this non-binary species tree under a duplication, transfer, and loss model (Supplementary Figure S4A). Notung infers the history of events that minimizes the weighted sum of events when fitting a gene tree to a species tree; in this analysis we used weights of 3.0, 1.5, and 1.6 for transfers, duplications, and losses, respectively. Notung does not infer events between taxa within an unresolved clade. We chose to represent strain relationships as non-binary nodes (i.e, polytomies) in order to focus the analysis to inter-species transfers and not intra-species transfers. A schematic (Figure 4B and Supplementary Figures 4B-D) displaying the predicted history of gene transfers on the species tree was generated in FigTree and Adobe Illustrator.

**Biofilm growth and imaging:** All strains were grown in Columbia broth to $OD_{600}$=0.05 before seeding culture onto MatTek dishes. At 24h and 48h, media was exchanged using 1/5$^{th}$ Columbia broth. At 72 hours, biofilms were washed 2 times with PBS and fixed with 4% PFA for 30 minutes. Fixed biofilms were stained with Syto59 fluorescent dye according to manufacturers instructions (LifeTech). Biofilms were imaged using Zeis 510 Meta Confocor3 Laser Scanning Microscope, and images were processed using ImageJ.

**Bacterial Cell Clumping Assays:** Strains D39, B1599 and B1567 were inoculated into Columbia broth and grown until $OD_{600}$=0.05. Each culture was diluted 10-fold in full-strength Columbia broth and incubated over night. Photographs were taken the following day, after 18-20 hours of growth. Test tubes were photographed with Olympus Pen E-P1 for documentation.

**Construction of deletion mutant and complemented strain:** The AR1 deletion mutant in strain B1599 was generated by replacement of this gene with a spectinomycin resistance cassette. Specifically, we amplified the 1kb upstream and downstream region of AR1 and ligated the flanking sequences to the resistance cassette by sticky end ligation with T4 DNA ligase. We amplified the ligation mixture by PCR to generate the transforming DNA (primers in Table S4). The AR1 complement strain (B1599ΔAR1::AR1) was generated in the ΔAR1 background by reintroducing the AR1 gene into a conserved intragenic region (contig208, position 161,851) previously used for complementation (50).

**Bacterial Transformations:** For all bacterial transformations, about 1µg of transforming DNA was added to the growing culture of a target strain at $OD_{600}$ of 0.05, supplemented with 125µg/mL of CSP2 (sequence: EMRISRIILDFLFLRKK; purchased from GenScript, NJ, USA), and incubated at 37°C. After 4 hours, the treated cultures were plated on Columbia agar containing spectinomycin, 100µg/mL. Resistant colonies were cultured in media, the region of interest was amplified by PCR and the amplimer was submitted for Sanger sequencing (Genewiz, Inc., USA) to verify the sequence of the mutants.

**mRNA isolation and qRT-PCR analysis:** Strains B1599, B1599ΔAR1 and B1599 ΔAR1::AR1 were grown in Columbia broth until reaching $OD_{600}$=0.05, 0.2 and 0.5. At each time point 5 ml of culture was collected and mixed with RNALater. Pelleted cultures were frozen until RNA extraction. RNEasy Plus Mini Kit from Qiagen was used to extract and purify RNA from each sample. Each sample was DNase treated to remove DNA contamination. Expression of AR1 was assayed by qRT-PCR and normalized to GAPDH. Primers for each locus were designed using Roche Universal Probe Library software. Experiment was performed in triplicate and data was analyzed

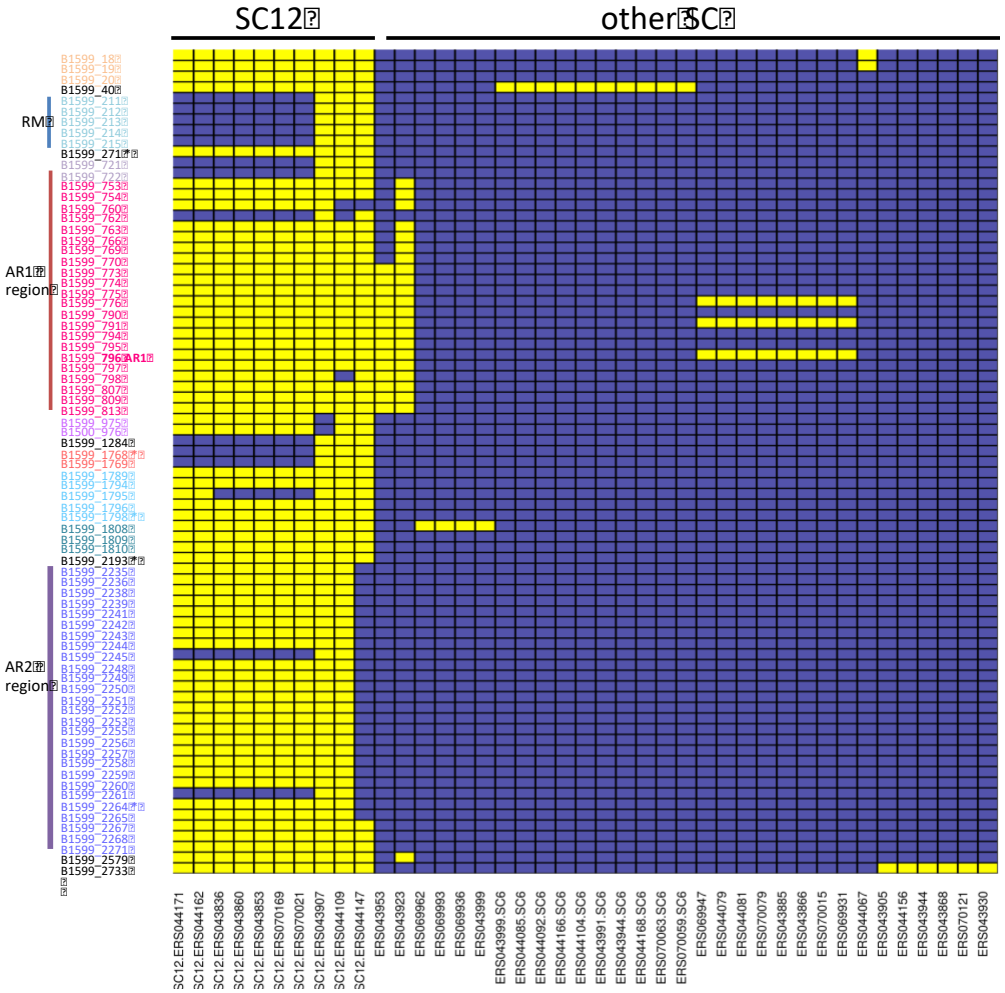
using LinReg PCR software. Statistical analysis was performed using Wilcoxon paired rank test using GraphPad.

**Mammalian tissue culture conditions:** Human corneal limbal epithelial (HCLE) cells were cultured in KSFM (keratinocyte serum-free medium) (Gibco Catalog number 10724-011) containing 25 μg/ml bovine pituitary extract (Gibco # 13028-014), 10.2 ng/ml embryonic growth factor (EGF) (Gibco # 10450-013), 100 μg/ml penicillin, and 100 μg/ml streptomycin (Corning # 30-002-CL).

**Bacterial attachment to HCLEs:** HCLEs were seeded into 12-well MatTek glass bottom dishes (MatTek P12G-1.5-14-F) in antibiotic free KSFM at a density of 1.50 x $10_5$ cells per well and allowed to adhere overnight at 37°C + 5% $CO_2$. *Streptococcus pneumoniae* strain B1599 (wild type), ΔAR1, and 1599ΔAR1::AR1 were streaked onto blood agar and grown overnight at 37°C + 5% $CO_2$. Bacteria were scraped off of the blood agar with an inoculating loop, added to 2.5ml Columbia broth, and grown for 5 hours to an $OD_{600}$ of 0.3. Cultures were pelleted by centrifugation at 14,000 rpm for 2 minutes. The pellets were washed two times in PBS and resuspended in 1 ml columbia broth. Two-hundred microliters of each strain was added to each well of HCLE cells containing 1 ml of antibiotic free KSFM. The plate was incubated at 37°C + 5% $CO_2$ for 30 minutes. After incubation, HCLEs were washed gently two times in PBS and supplemented with fresh KSFM. The cells were imaged on an Olympus Fluoview FV-1000 laser scanner confocal microscope with a 60X objective. Ten fields per treatment group were imaged and the number of bacteria on each HCLE was manually counted using Fluoview image viewing software version 3.1. The experiment was repeated on three separate days with similar results.

**Accession Numbers:** We present the complete genome sequence of six genomes, the GenBank accession numbers are as follows: E709 (JBOR00000000), K2521 (JBOS00000000), K2527 (JBOT00000000), K2557 (JBOU00000000), B1598 (JBOV00000000), and B1599 (JBOW00000000).

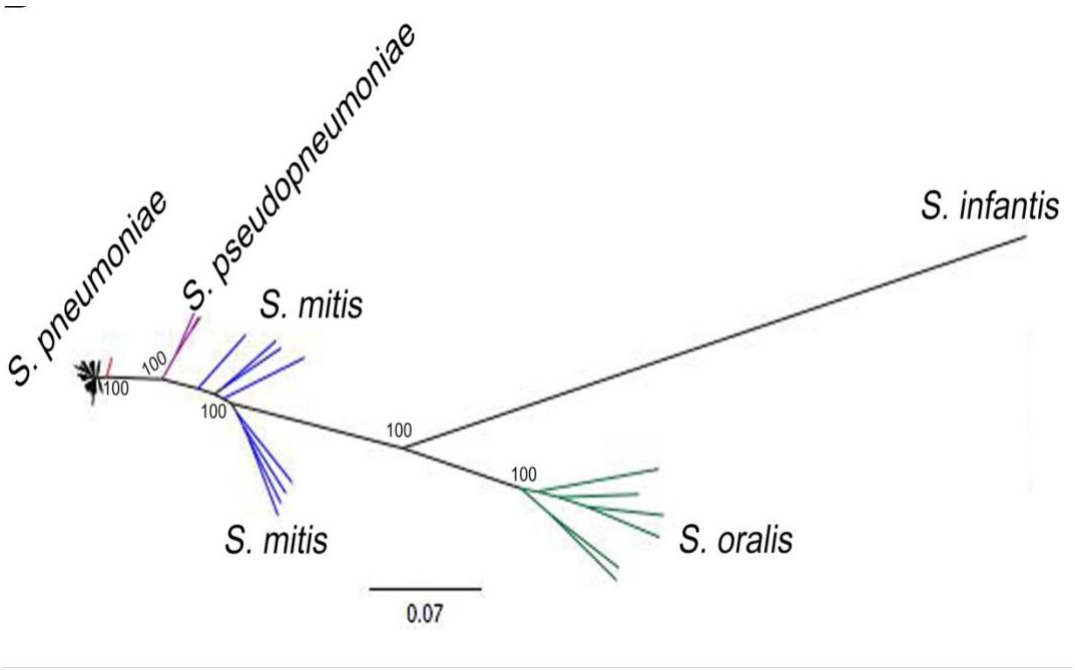
## 7.6 Supplementary Figures




**Supplementary Figure 7.1 Genes enriched in the distinct phylogenetic branch.**

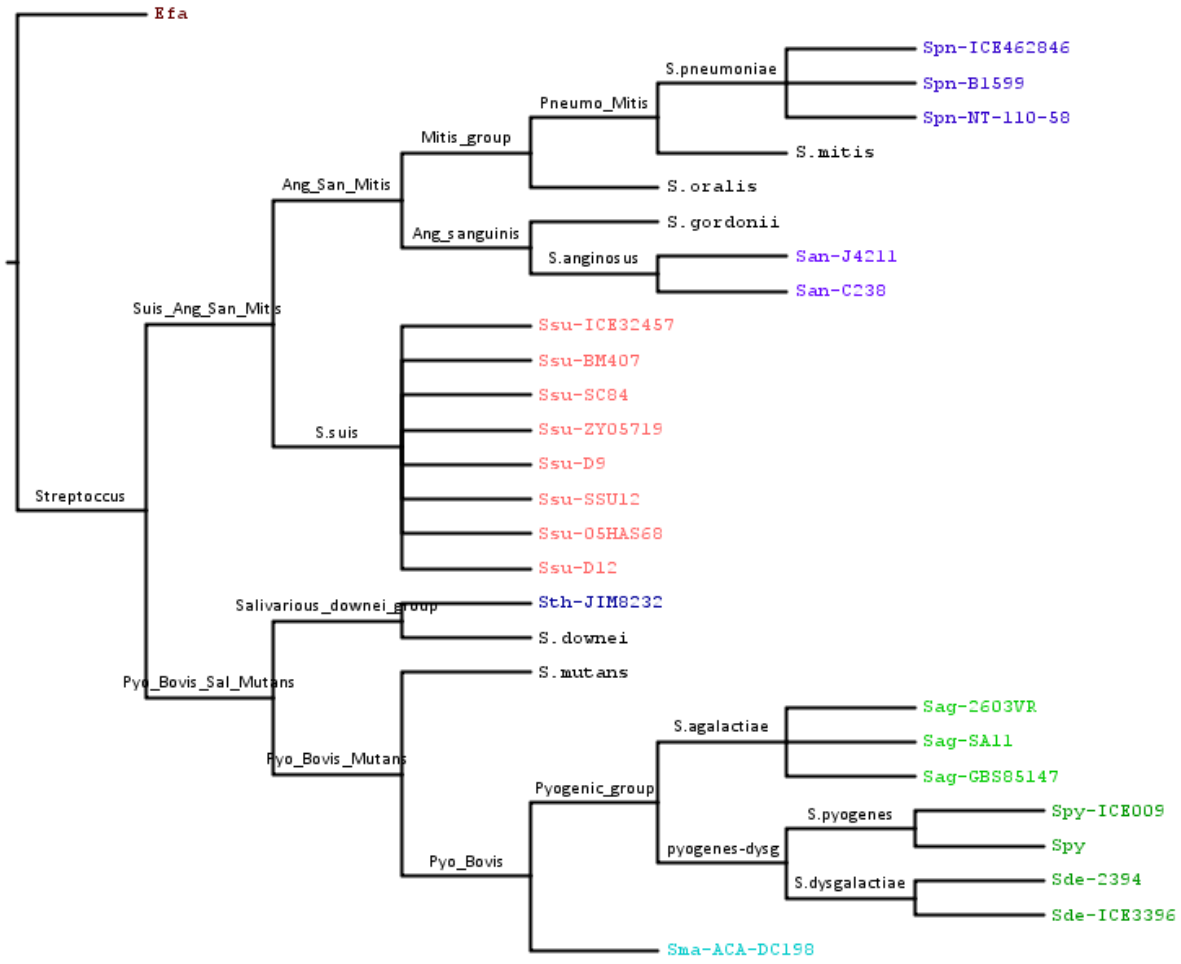
Left column: List of seventy-seven genes present in conjunctivitis isolates (B1598 and B1599) but absent in strains from the other major branch, including those isolates from keratitis and endophthalmitis. The numbers represent the locus number from genome B1599 (GenBank: JBOW00000000). Colors denote genes grouped on the chromosome, and asterisks denote genes predicted to be surface exposed. The heat map summarized the BlastP analysis of the coding sequences this study identified as unique to the distinct phylogenetic branch regarding their distribution in a large set of isolates gathered by Croucher and colleagues (Croucher et al., 2013). The comparative set consists of 616 strains: ten from the SC12 cluster that corresponds to the basal branch, and 606 from fifteen monophyletic sequence clusters and one additional diverse group of less common genotypes (Croucher et al., 2013). Our gene set can be

categorized into two distributions. The first consists of twelve genes, rare in the SC12 group (30% or less of the genomes). The second consists of sixty-five genes present in 90-100% of the SC12 genomes. All these genes are rare outside the SC12 group, and are candidates for biological features that distinguish this phylogenetic group.



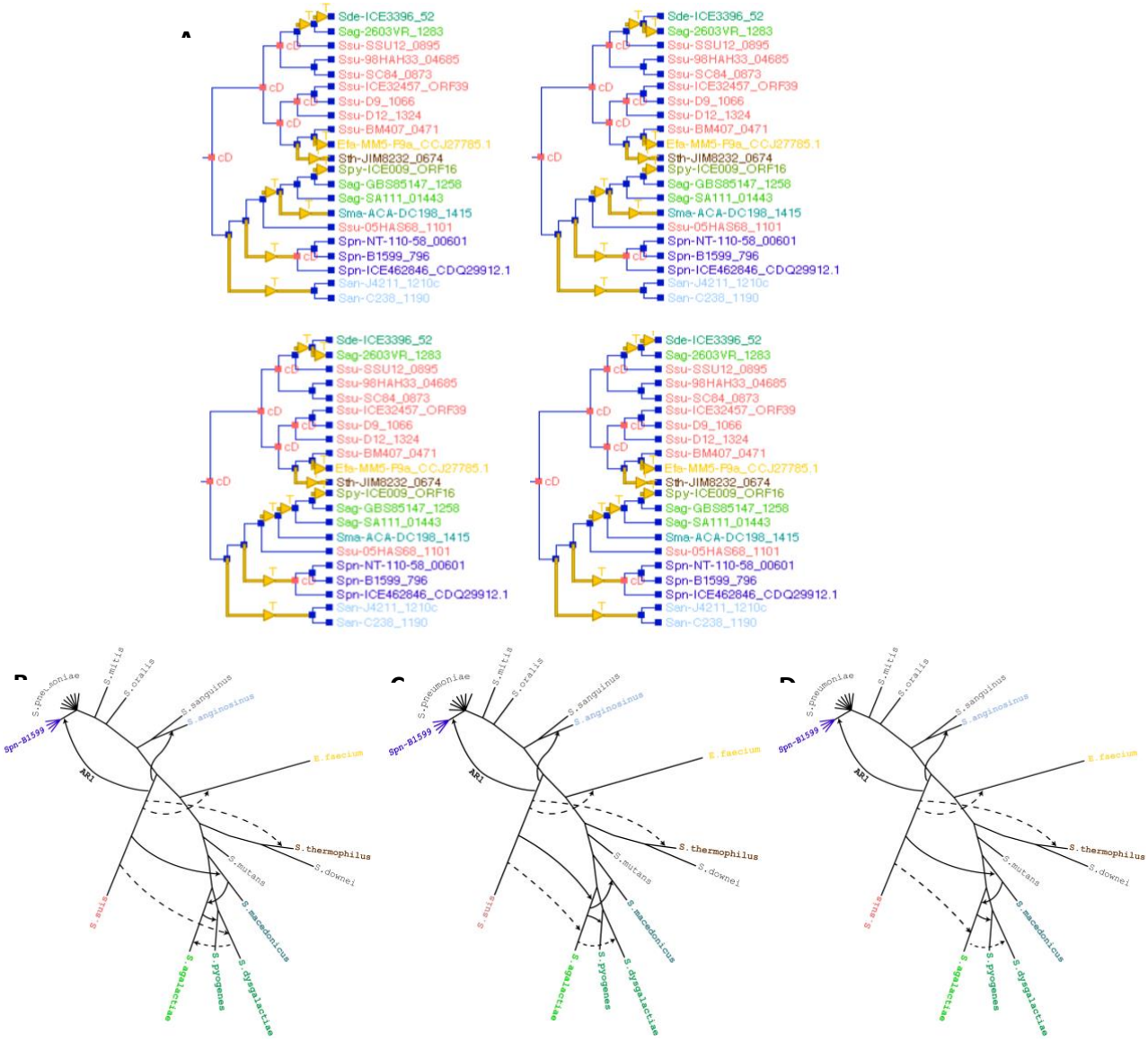

**Supplementary Figure 7.2 Phylogenetic analyses of streptococcal strains, demonstrating the basal position of the distinct phyletic group.**

Maximum likelihood phylogeny of multiple streptococcal species. Colors are used to highlight species. From left to right: (black) majority or pneumococcal strains; (red) distinct phyletic group of pneumococcal isolates; (purple) *S. pseudopneumoniae;* (blue)*: S. mitis*; (black): *S. infantis*; (green): *S. oralis* (includes strains previous annotated as *S. mitis*). The scale bar indicates the number of substitutions per site. The numbers on the branches indicate bootstrap values. Red arrow highlights the basal position of the distinct phyletic branch.




**Supplementary Figure 7.3 Streptococcus species tree used in reconciliation.**

The phylogenetic reconciliation analysis was carried out with a tree representing the phylogenetic relationships of *Enterrococcus faecium* and 13 *Streptoccocus* species. This tree is based on the 44-species *Streptococcus* phylogeny inferred by Richards *et al.* (Richards *et al)* from the concatenated sequences of 136 core genes. All species harboring putative *ar1* xenologs were retained in this reduced tree, as were five additional species in order to maintain a representative sample of well-studied taxa and all major taxonomic groups delineated in Richards *et al*. For those species where an *ar1* homolog was predicted in multiple strains, all strains represented in the gene tree in Figure 4A were added to the species tree. Strain relationships were not resolved, resulting in a terminal polytomy for each species with more than two strains. Strains are labeled with the abbreviations in **Table S3**.




**Supplementary Figure 7.4 Interspecies *ar1* gene transfers in the context of the *Streptococcus* phylogeny.**

The *ar1* gene tree (Fig. 4A) was reconciled with the species tree in Fig S3 using Notung 2.9 to infer the the event history that minimizes the weighted sum of duplications, transfers, and losses (weights: 1.5, 3.0, 1.6). **(A)** Reconciliation results in four minimal evolutionary scenarios, each with 9 inferred transfers. (**B- D**) Evolutionary scenarios shown in the context of the species tree. All trees support the conclusion that *ar1* was first acquired by *S. suis* and later dispersed to other species though multiple gene transfer events. These scenarios differ only in the direction of gene transfers between (i) *S. dysgalactiae* and *S. agalactiae* 2603 V/R and (ii) *S. agalactiae* isolate SA111 and *S. macedonicus* strain ACA-DC 198. The fourth scenario is shown in Figure 4B. All trees agree regarding the transfer of *ar1* from *S. suis* to *S. pneumoniae*.

## 7.7 Supplementary Tables

### Supplementary Table 7.1 Pneumococcal genomes used to generate a strain tree

Bold: eye-associated strains

| Strain | Serotype | MLST | Genome (bp) | Number ORFs | Accession number | Location of isolation |
|---|---|---|---|---|---|---|
| MNZ41 | NT | 6153 | 2130281 | 2185 | ASJW00000000 | South Korea[19] |
| **B1068B1598** | NT | 2315 | 2902565 | 3874 | JBOV00000000 | Pittsburgh, US |
| **B1069B1599** | NT | 2315 | 2422872 | 2700 | JBOW00000000 | Pittsburgh, US |
| MNZ85 | NT | 2315 | 2146742 | 2190 | ASJF00000000 | Brazil[19] |
| MNZ14 | NT | 448 | 2131549 | 2167 | ASJO00000000 | South Korea[19] |
| SPNA45 | 3 | New | 2041833 | 1932 | HE983624.1 (SpnA45) | Newmarket |
| Sp3BS71 | 3 | 180 | 2036967 | 2177 | AAZZ00000000 | Pittsburgh, US |
| Sp11BS70 | 11 | 62 | 2060705 | 2127 | ABAC00000000 | Pittsburgh, US |
| G54 | 19F | 63 | 2078953 | 2115 | NC_011072.1 | Italy |
| 70585 | 5 | 289 | 2184682 | 2202 | NC_012468.1 | Bangladesh |
| SPN1041 | 1 | 217 | 2166490 | 1905 | NZ_CACE01000052 | Ghana |
| INV104 | 1 | 227 | 2142122 | 1941 | FQ312030.1 | Oxford, UK |
| Sp9BS68 | 9 | 1269 | 2117908 | 2241 | ABAB00000000 | Pittsburgh, US |
| **B1066K2557** | 17F | 2355 | 2051681 | 2353 | JBOU00000000 | Pittsburgh, US |
| WL677 | 19A | 276 | 2085829 | | NZ_AUWZ00000000.1 | Portugal |
| CZ5(61) | 19A | New | 2239769 | | in process | Czech Republic |
| ATCC700669 | 23F | 81 | 2221315 | 2135 | FM211187 | Barcelona, Spain[54] |
| CDC305906 | 19A | 199 | 2293277 | 2379 | ABGG000000000 | US, UK |
| JJA | 14 | 66 | 2120234 | 2123 | NC_012466.1 | Brazil |
| Sp19BS75 | 19 | 485 | 2136434 | 2301 | ABAF00000000 | Pittsburgh, US |
| CGSP14 | 14 | 15 | 2209198 | 2206 | NC_010582.1 | Beijing, China |
| SAD20 (63) | 19A | 1656 | 2138581 | | In process | South Africa |
| CDC108700 | 7F | 191 | 2190853 | 2232 | ABFT00000000 | Brazil, Denmark, Finland, Norway, Uruguay, UK, US |
| Hungary19A | 19A | 236 | 2245615 | 2155 | CP000936.1 | Hungary |
| **B1064K2521** | 15 | 3280 | 2184812 | 2451 | JBOS00000000 | Pittsburgh, US |
| **B1063E709** | 23A | 338 | 2155121 | 2362 | JBOR00000000 | Pittsburgh, US |
| Taiwan19F | 19F | 236 | 2112148 | 2044 | NC_012469.1 | Taiwan |
| **B1065K2527** | 19A | 320 | 2180290 | 2392 | JBOT00000000 | Pittsburgh, US |
| MNZ11b | NT | 8966 | 2021497 | 2090 | ASJW00000000 | South Korea[19] |
| MNZ37 | NT | 1106 | 1946911 | 1997 | ASJP00000000 | South Korea[19] |
| WL400 | 6A | 2191 | 2149529 | | NZ_AVFA01000087.1 | Portugal |
| ICE44 | 6B | 71 | 2213083 | 2378 | NZ_AUYF00000000.1 | Iceland |
| Sp18BS74 | 6 | new | 2105593 | 2211 | ABAE00000000 | Pittsburgh, US |
| CDC1873-00 | 6A | 376 | 2265195 | 2402 | ABFS00000000 | USA |
| TIGR4 | 4 | 205 | 2160842 | 2125 | NZ_AKBW01000001.1 | Norway |
| Sp6BS73 | 6 | 460 | 2162916 | 2325 | ABAA00000000 | Pittsburgh, US |
| CDC0288-04 | 12F | 220 | 2051140 | 2150 | ABFG00000000 | USA, UK |
| D39 | 2 | 595 | 2046115 | 2043 | NC_008533.1 | Laboratory |
| Sp23BS72 | 23 | 37 | 2103479 | 2200 | ABAG00000000 | Pittsburgh, US |
| Sp14BS69 | 14 | 124 | 2148093 | 2624 | ABAE00000000 | Pittsburgh, US |

## Supplementary Table 7.2 Strains used in this study

| Strain name | Description | Origin |
|---|---|---|
| B1599 | Wild type conjunctivitis strain | UPMC |
| B1599 ΔAR1 | AR1 deletion mutant, $Spec^R$ | This study |
| B1599 ΔAR1AR2 | AR1 and AR2 deletion mutant, $Spec^R$ $Tet^R$ | This study |
| B1599 ΔAR1::AR1 | AR1 overexpressing strain, $Spec^R$ $Tet^R$ | This study |

## Supplementary Table 7.3 List of *ar1* homologues used to generate a gene-tree-species-tree reconciliation

| LOCUS | Abbreviation | Definition |
|---|---|---|
| LT545678.1 | Sag-SA111 | Streptococcus agalactiae isolate SA111 genome assembly, chromosome |
| CP010319.1 | Sag-GBS85147 | Streptococcus agalactiae strain GBS85147, complete genome |
| AE009948.1 | Sag-2603VR | Streptococcus agalactiae 2603V/R, complete genome |
| CP012805.1 | San-J4211 | Streptococcus anginosus strain J4211, complete genome |
| CP003861.1 | San-C238 | Streptococcus anginosus C238, complete genome |
| EU142041.1 | Sde-ICE3396 | Streptococcus dysgalactiae subsp. equisimilis strain NS3396 integrative conjugative element ICESde3396, complete sequence |
| CP002215.1 | | Streptococcus dysgalactiae subsp. equisimilis ATCC 12394, complete genome |
| HE963029.1 | Efa-MM5-F9a_CCJ2 | Enterococcus faecium transfer chromosomal element, strain MM5-F9a |
| HE613569.1 | Sma-ACA-DC198 | Streptococcus macedonicus ACA-DC 198 main chromosome complete genome |
| CP007593.1 | Spn-NT-110-58 | Streptococcus pneumoniae strain NT_110_58, complete genome |
| LK020693.1 | Spn-ICE462846_CDQ2 | Streptococcus pneumoniae putatative integrative and conjugative element sequence, isolate 403790 |
| KU056701.1 | Spy-ICE009_ORF | Streptococcus pyogenes strain MB56Spyo009 transposon ICESpy009 complete sequence |
| CP002007.2 | Ssu-05HAS68 | Streptococcus suis 05HAS68, complete genome |
| LT671674.1 | | Streptococcus suis strain LS9N genome assembly, chromosome: I |
| CP008921.1 | | Streptococcus suis 6407, complete genome |
| CP002641.1 | Ssu-D9 | Streptococcus suis D9, complete genome |
| FR823304.2 | Ssu-ICE32457_ORF | Streptococcus suis integrative conjugative element ICESsu32457, strain 32457 |
| CP018908.1 | | Streptococcus suis strain SS2-1, complete genome |
| CP007497.1 | Ssu-ZY05719 | Streptococcus suis strain ZY05719, complete genome |
| FM252031.1 | Ssu-SC84 | Streptococcus suis SC84 complete genome, strain SC84 |
| CP002644.1 | Ssu-SSU12 | Streptococcus suis D12, complete genome |
| CP000407.1 | | Streptococcus suis 05ZYH33, complete genome |
| FM252032.1 | Ssu-BM407 | Streptococcus suis BM407 complet genome, strain BM407 |
| CP000408.1 | Ssu-98HAH33 | Streptococcus suis 98HAH33, complete genome |
| CP002640.1 | Ssu-D12 | Streptococcus suis SS12, complete genome |
| FR875178.1 | Sth-JIM8232 | Streptococcus thermophilus JIM 8232 complete genome |

## Supplementary Table 7.4 Primers used in this study

| Primer name | Primer Sequence |
|---|---|
| SpecF XhoI | 5'-ATATATCTCGAGGGATCCCCCGTTTGATTTTTAATGGTAATG-3' |
| SpecR TERM NheI | 5'-ATATATGCTAGCGCGAAAAAACCCCGCCGAAGCGGGGTTTTTTGCGAATTGACGCGGAATGGATCC-3' |
| Tet F BamHI | 5'-ATATATGGATCCAAAATTTGTTTGAT-3' |
| Tet R NheI | 5'-ATATATGCTAGCTCCCAAAGTTGATCCCTTAACG-3' |
| Agglut R1 F1F 2kb | 5'-GCAGGAAATTCACCATTATTGA-3' |
| Agglut R1 F1R Xho1 2kb | 5'-ATATATCTCGAGTGACATTGTTTGGTC-3' |
| Agglut R1 F2F Nhe1/2kb | 5'-ATATATGCTAGCCGCAAAATGGAGGACAATTAAATG-3' |
| Agglut R1 F2R 2kb | 5'-CAATATCGAGGTATTCATGTAAGC-3' |
| R1_comp_F1_fwd | 5'-CGCTTGCAAAATTCGTTG-3' |
| R1_comp_F1_rev | 5'-AAACAAATTTTGGGCCTCCAATTTCTTGATAGTTTATAAAAAAC-3' |
| Tet_(R1_comp)_fwd | 5'-ATCAAGAAATTGGAGGCCCAAAATTTGTTTGATTTG-3' |
| Tet_(R1_comp)_rev | 5'-AGCAACAAAGAAAGGCCCAAAGTTGATCCCTTAAC-3' |
| R1_fwd | 5'-GGGATCAACTTTGGGCCTTTCTTTGTTGCTTTTGTAC-3' |
| R1_rev | 5'-TTCTATCTCAAACCCTTAATTGTCCTCCATTTTGC-3' |
| R1_comp_F2_fwd | 5'-ATGGAGGACAATTAAGGGTTTGAGATAGAAGTTTTATAAAAAAG-3' |
| R1_comp_F2_rev | 5'-CCCTTAATTGAGAAGAAATCTTTC-3' |
| R1 locus left | 5'-ATG GAT TTA GGC AAC ACG ACT T-3' |
| R1 locus right | 5'-TTC AGC GGT ATC CAA TTG CT-3' |
| R2 locus left | 5'-CAT GGT TTC GGA TTT AGG AGA-3' |
| R2 locus right | 5'-TCT GTC ACA GGA ACA CCA TTT C-3' |

## 7.8 References


1. Bessen DE, Kumar N, Hall GS, Riley DR, Luo F, Lizano S, et al. Whole-Genome Association Study on Tissue Tropism Phenotypes in Group A Streptococcus ▿. J Bacteriol. 2011 Dec;193(23):6651–63.
2. Donati C, Hiller NL, Tettelin H, Muzzi A, Croucher NJ, Angiuoli SV, et al. Structure and dynamics of the pan-genome of Streptococcus pneumoniae and closely related species. Genome Biol. 2010;11(10):R107.
3. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, et al. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome." PNAS. 2005 Sep 27;102(39):13950–5.
4. Kulohoma BW, Cornick JE, Chaguza C, Yalcin F, Harris SR, Gray KJ, et al. Comparative Genomic Analysis of Meningitis- and Bacteremia-Causing Pneumococci Identifies a Common Core Genome. Infect Immun. 2015 Oct;83(10):4165–73.
5. Croucher NJ, Finkelstein JA, Pelton SI, Mitchell PK, Lee GM, Parkhill J, et al. Population genomics of post-vaccine changes in pneumococcal epidemiology. Nat Genet. 2013 Jun;45(6):656–63.
6. Hiller NL, Janto B, Hogg JS, Boissy R, Yu S, Powell E, et al. Comparative genomic analyses of seventeen Streptococcus pneumoniae strains: insights into the pneumococcal supragenome. J Bacteriol. 2007 Nov;189(22):8186–95.
7. Kadioglu A, Weiser JN, Paton JC, Andrew PW. The role of Streptococcus pneumoniae virulence factors in host respiratory colonization and disease. Nat Rev Micro. 2008 Apr;6(4):288–301.
8. Croucher NJ, Coupland PG, Stevenson AE, Callendrello A, Bentley SD, Hanage WP. Diversification of bacterial genome content through distinct mechanisms over different timescales. Nat Commun [Internet]. 2014 Nov 19 [cited 2015 Feb 1];5. Available from: http://www.nature.com/ncomms/2014/141119/ncomms6471/abs/ncomms6471.html
9. Hilty M, Wüthrich D, Salter SJ, Engel H, Campbell S, Sá-Leão R, et al. Global phylogenomic analysis of nonencapsulated Streptococcus pneumoniae reveals a deep-branching classic lineage that is distinct from multiple sporadic lineages. Genome Biol Evol. 2014 Dec 4;6(12):3281–94.
10. Valentino MD, McGuire AM, Rosch JW, Bispo PJM, Burnham C, Sanfilippo CM, et al. Unencapsulated Streptococcus pneumoniae from conjunctivitis encode variant traits and belong to a distinct phylogenetic cluster. Nat Commun. 2014;5:5411.
11. Keller LE, Robinson DA, McDaniel LS. Nonencapsulated Streptococcus pneumoniae: Emergence and Pathogenesis. mBio. 2016 May 4;7(2):e01792-15.
12. Murrah KA, Pang B, Richardson S, Perez A, Reimche J, King L, et al. Nonencapsulated Streptococcus pneumoniae causes otitis media during single-species infection and during polymicrobial infection with nontypeable Haemophilus influenzae. Pathog Dis [Internet]. 2015 Jul [cited 2017 Apr 15];73(5). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4849347/
13. Chewapreecha C, Harris SR, Croucher NJ, Turner C, Marttinen P, Cheng L, et al. Dense genomic sampling identifies highways of pneumococcal recombination. Nat Genet. 2014 Mar;46(3):305–9.

14. Hiller NL, Ahmed A, Powell E, Martin DP, Eutsey R, Earl J, et al. Generation of genic diversity among Streptococcus pneumoniae strains via horizontal gene transfer during a chronic polyclonal pediatric infection. PLoS Pathog. 2010;6(9):e1001108.

15. Marquart ME, O'Callaghan RJ. Infectious Keratitis: Secreted Bacterial Proteins That Mediate Corneal Damage. Journal of Ophthalmology. 2013 Jan 8;2013:e369094.

16. Miller JJ, Scott IU, Flynn HW, Smiddy WE, Corey RP, Miller D. Endophthalmitis caused by Streptococcus pneumoniae. Am J Ophthalmol. 2004 Aug;138(2):231–6.

17. Parmar P, Salman A, Kalavathy CM, Jesudasan CAN, Thomas PA. Pneumococcal keratitis: a clinical profile. Clin Experiment Ophthalmol. 2003 Feb;31(1):44–7.

18. Enright MC, Spratt BG. A multilocus sequence typing scheme for Streptococcus pneumoniae: identification of clones associated with serious invasive disease. Microbiology (Reading, Engl). 1998 Nov;144 ( Pt 11):3049–60.

19. Frazão N, Hiller NL, Powell E, Earl J, Ahmed A, Sá-Leão R, et al. Virulence potential and genome-wide characterization of drug resistant Streptococcus pneumoniae clones selected in vivo by the 7-valent pneumococcal conjugate vaccine. PloS One. 2013;8(9):e74867.

20. Keller LE, Thomas JC, Luo X, Nahm MH, McDaniel LS, Robinson DA. Draft Genome Sequences of Five Multilocus Sequence Types of Nonencapsulated Streptococcus pneumoniae. Genome Announc. 2013 Jul 25;1(4).

21. Bae T, Schneewind O. The YSIRK-G/S Motif of Staphylococcal Protein A and Its Role in Efficiency of Signal Peptide Processing. J Bacteriol. 2003 May 1;185(9):2910–9.

22. Bergmann S, Hammerschmidt S. Versatility of pneumococcal surface proteins. Microbiology (Reading, Engl). 2006 Feb;152(Pt 2):295–303.

23. Jeong JK, Kwon O, Lee YM, Oh D-B, Lee JM, Kim S, et al. Characterization of the Streptococcus pneumoniae BgaC Protein as a Novel Surface β-Galactosidase with Specific Hydrolysis Activity for the Galβ1-3GlcNAc Moiety of Oligosaccharides. J Bacteriol. 2009 May 1;191(9):3011–23.

24. Schneewind O, Missiakas DM. Protein secretion and surface display in Gram-positive bacteria. Philos Trans R Soc Lond B Biol Sci. 2012 Apr 19;367(1592):1123–39.

25. Zähner D, Hakenbeck R. The Streptococcus pneumoniae Beta-Galactosidase Is a Surface Protein. J Bacteriol. 2000 Oct;182(20):5919–21.

26. Croucher NJ, Finkelstein JA, Pelton SI, Parkhill J, Bentley SD, Lipsitch M, et al. Population genomic datasets describing the post-vaccine evolutionary epidemiology of Streptococcus pneumoniae. Sci Data. 2015;2:150058.

27. Porat N, Greenberg D, Givon-Lavi N, Shuval DS, Trefler R, Segev O, et al. The important role of nontypable Streptococcus pneumoniae international clones in acute conjunctivitis. J Infect Dis. 2006 Sep 1;194(5):689–96.

28. Zegans ME, Sanchez PA, Likosky DS, Allar RT, Martin M, Schwartzman JD, et al. Clinical features, outcomes, and costs of a conjunctivitis outbreak caused by the ST448 strain of Streptococcus pneumoniae. Cornea. 2009 Jun;28(5):503–9.

29. Blanchette-Cain K, Hinojosa CA, Akula Suresh Babu R, Lizcano A, Gonzalez-Juarbe N, Munoz-Almagro C, et al. Streptococcus pneumoniae biofilm formation is strain dependent, multifactorial, and associated with reduced invasiveness and immunoreactivity during colonization. MBio. 2013;4(5):e00745-00713.

30. Hall-Stoodley L, Hu FZ, Gieseke A, Nistico L, Nguyen D, Hayes J, et al. Direct Detection of Bacterial Biofilms on the Middle-Ear Mucosa of Children With Chronic Otitis Media. JAMA. 2006 Jul 12;296(2):202–11.

31. Hoa M, Syamal M, Sachdeva L, Berk R, Coticchia J. Demonstration of nasopharyngeal and middle ear mucosal biofilms in an animal model of acute otitis media. Ann Otol Rhinol Laryngol. 2009 Apr;118(4):292–8.

32. Marks LR, Parameswaran GI, Hakansson AP. Pneumococcal interactions with epithelial cells are crucial for optimal biofilm formation and colonization in vitro and in vivo. Infect Immun. 2012 Aug;80(8):2744–60.

33. Post JC, Hiller NL, Nistico L, Stoodley P, Ehrlich GD. The role of biofilms in otolaryngologic infections: update 2007. Curr Opin Otolaryngol Head Neck Surg. 2007 Oct;15(5):347–51.

34. Sanderson AR, Leid JG, Hunsaker D. Bacterial biofilms on the sinus mucosa of human subjects with chronic rhinosinusitis. Laryngoscope. 2006 Jul;116(7):1121–6.

35. Hall-Stoodley L, Nistico L, Sambanthamoorthy K, Dice B, Nguyen D, Mershon WJ, et al. Characterization of biofilm matrix, degradation by DNase treatment and evidence of capsule downregulation in Streptococcus pneumoniae clinical isolates. BMC Microbiology. 2008 Oct 8;8(1):173.

36. Berg KH, Stamsås GA, Straume D, Håvarstein LS. Effects of low PBP2b levels on cell morphology and peptidoglycan composition in Streptococcus pneumoniae R6. J Bacteriol. 2013 Oct;195(19):4342–54.

37. Damjanovic M, Kharat AS, Eberhardt A, Tomasz A, Vollmer W. The essential tacF gene is responsible for the choline-dependent growth phenotype of Streptococcus pneumoniae. J Bacteriol. 2007 Oct;189(19):7105–11.

38. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. Nucleic Acids Res. 2015 Jan;43(Database issue):D222-226.

39. Biswas I, Drake L, Biswas S. Regulation of gbpC Expression in Streptococcus mutans. J Bacteriol. 2007 Sep;189(18):6521–31.

40. Kagami A, Okamoto-Shibayama K, Yamamoto Y, Sato Y, Kizaki H. One of two gbpC gene homologues is involved in dextran-dependent aggregation of Streptococcus sobrinus. Oral Microbiol Immunol. 2007 Aug;22(4):240–7.

41. Forsgren N, Lamont RJ, Persson K. Two intramolecular isopeptide bonds are identified in the crystal structure of the Streptococcus gordonii SspB C-terminal domain. J Mol Biol. 2010 Apr 2;397(3):740–51.

42. Demuth DR, Irvine DC, Costerton JW, Cook GS, Lamont RJ. Discrete protein determinant directs the species-specific adherence of Porphyromonas gingivalis to oral streptococci. Infect Immun. 2001 Sep;69(9):5736–41.

43. Gipson IK, Spurr-Michaud S, Argüeso P, Tisdale A, Ng TF, Russo CL. Mucin gene expression in immortalized human corneal-limbal and conjunctival epithelial cell lines. Invest Ophthalmol Vis Sci. 2003 Jun;44(6):2496–506.

44. Williamson YM, Gowrisankar R, Longo DL, Facklam R, Gipson IK, Ades EP, et al. Adherence of nontypeable Streptococcus pneumoniae to human conjunctival epithelial cells. Microbial Pathogenesis. 2008;3(44):175–85.

45. Tukey JW. Comparing individual means in the analysis of variance. Biometrics. 1949 Jun;5(2):99–114.

46. Guglielmini J, Néron B, Abby SS, Garcillán-Barcia MP, Cruz L, De F, et al. Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. Nucleic Acids Res. 2014 May 14;42(9):5715–27.

47. Richards VP, Palmer SR, Pavinski Bitar PD, Qin X, Weinstock GM, Highlander SK, et al. Phylogenomics and the dynamic genome evolution of the genus Streptococcus. Genome Biol Evol. 2014 Apr;6(4):741–53.

48. Stolzer M, Lai H, Xu M, Sathaye D, Vernot B, Durand D. Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. Bioinformatics. 2012 Sep 15;28(18):i409–15.

49. Budroni S, Siena E, Dunning Hotopp JC, Seib KL, Serruto D, Nofroni C, et al. Neisseria meningitidis is structured in clades associated with restriction modification systems that modulate homologous recombination. Proc Natl Acad Sci USA. 2011 Mar 15;108(11):4494–9.

50. Eutsey RA, Powell E, Dordel J, Salter SJ, Clark TA, Korlach J, et al. Genetic Stabilization of the Drug-Resistant PMEN1 Pneumococcus Lineage by Its Distinctive DpnIII Restriction-Modification System. mBio. 2015 Jul 1;6(3):e00173-15.

51. Zangari T, Wang Y, Weiser JN. Streptococcus pneumoniae Transmission Is Blocked by Type-Specific Immunity in an Infant Mouse Model. mBio. 2017 May 3;8(2):e00188-17.

52. Rodriguez JL, Dalia AB, Weiser JN. Increased Chain Length Promotes Pneumococcal Adherence and Colonization. Infect Immun. 2012 Oct;80(10):3454–9.

53. Bjørkeng EK, Hjerde E, Pedersen T, Sundsfjord A, Hegstad K. ICESluvan, a 94-kilobase mosaic integrative conjugative element conferring interspecies transfer of VanB-type glycopeptide resistance, a novel bacitracin resistance locus, and a toxin-antitoxin stabilization system. J Bacteriol. 2013 Dec;195(23):5381–90.

54. Davies D. Understanding biofilm resistance to antibacterial agents. Nat Rev Drug Discov. 2003 Feb;2(2):114–22.

55. Huang J, Liang Y, Guo D, Shang K, Ge L, Kashif J, et al. Comparative Genomic Analysis of the ICESa2603 Family ICEs and Spread of erm(B)- and tet(O)-Carrying Transferable 89K-Subtype ICEs in Swine and Bovine Isolates in China. Front Microbiol [Internet]. 2016 Feb 2 [cited 2017 Apr 23];7. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4735348/

56. Palmieri C, Magi G, Mingoia M, Bagnarelli P, Ripa S, Varaldo PE, et al. Characterization of a Streptococcus suis tet(O/W/32/O)-carrying element transferable to major streptococcal pathogens. Antimicrob Agents Chemother. 2012 Sep;56(9):4697–702.

57. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. 2010;11:595.

58. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucl Acids Res. 2014 Jan 1;42(D1):D206–14.

59. Darling AE, Mau B, Perna NT. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. PLoS ONE. 2010 Jun 25;5(6):e11147.

60. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002 Jul 15;30(14):3059–66.

61. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. Bioinformatics. 1998;14(9):817–8.

62. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010 May;59(3):307–21.

63. Hogg JS, Hu FZ, Janto B, Boissy R, Hayes J, Keefe R, et al. Characterization and modeling of the Haemophilus influenzae core and supragenomes based on the complete genomic sequences of Rd and 12 clinical nontypeable strains. Genome Biol. 2007;8(6):R103.

64. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct 5;215(3):403–10.

65. Anonymous. R: a language and environment for statistical computing [Internet]. GBIF.ORG. 2015 [cited 2015 Jun 1]. Available from: http://www.gbif.org/resource/81287

66. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009 May 1;25(9):1189–91.

67. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 2006 Jul 1;34(Web Server issue):W609-612.

68. Keane TM, Creevey CJ, Pentony MM, Naughton TJ, Mclnerney JO. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evolutionary Biology. 2006;6:29.

69. Milne I, Lindner D, Bayer M, Husmeier D, McGuire G, Marshall DF, et al. TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. Bioinformatics. 2009 Jan 1;25(1):126–7.

70. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Syst Biol. 2012 May 1;61(3):539–42.