Investigations into the Role of Inter and Intra Strain Quorum Sensing in Pneumococcal Pathogenesis

A dissertation

submitted to the

Department of Biological Sciences

Carnegie Mellon University

towards partial fulfillment of the requirements for

the degree of

Doctor of Philosophy

by

Anagha Kadam

September 2017

Abstract

Streptococcus pneumoniae is the leading cause of pediatric morbidity and mortality worldwide. The rich genomic diversity among strains of pneumococci contributes to varied phenotypes exhibited by strains such as propensity to cause disease, colonization capabilities, and responses to antibiotics and vaccines. In this thesis, I explore the role of genes acquired by gene transfer on the ability of pneumococcus to cause disease in single and multi-strain infections. I describe my work on the identification and molecular characterization of a novel quorum sensing system comprising a transcription regulator and peptide (TprA2/PhrA2). This system is unique to the PMEN1 lineage, a group of pandemic and multi-drug resistant pneumococcal strains. I demonstrate the detailed mechanism of regulation of this peptide-regulator pair and its ability to control a downstream lanthionine-containing peptide. Through *in vivo* studies I observe that the TprA2 regulator acts to promote commensalism over tissue dissemination. In addition, I show that the downstream lanthionine-containing peptide is a novel virulence determinant. My findings extend beyond this specific cell-cell communication system, since PhrA2 unidirectionally influences gene expression of a second quorum sensing system, TprA/PhrA, widespread across the pneumococcal species. I participated in studies that conclusively demonstrate the role of TprA/PhrA as a potent virulence determinant, and I show that the PMEN1-unique peptide PhrA2 can control gene expression of TprA/PhrA in a non-PMEN1 strain. These data open the door to the potential role for PMEN1 strains in modifying the virulence potential of other strains in multi-strain infections. In summary, my findings suggest that a

horizontally acquired PMEN1-unique genomic element has the ability to modify regulatory circuits and fine tune the net regulation of multiple downstream effectors via intra- and inter-strain communication in single and multi strain infections.

To my grandfather,
who hailed from humble beginnings in rural India but was the greatest champion
of scientists and research I have ever known.
He journeyed to another world earlier this year, but I know he is giving me a pat on the back
right about now. Thank you, Aazoba!

Acknowledgements

In my journey through graduate school, a number of very kind people have helped me along the way. First and foremost, I would like to thank my advisor, Dr. Luisa Hiller, for her patience, scientific guidance and encouragement that has shaped me and this work to the present form. Her optimistic outlook and unwavering support at every step has helped me emerge stronger after every setback I faced. I shall forever be grateful for the special relationship I have with her.

I would like to thank all members of my thesis committee for their invaluable time and feedback. I became a better thinker after every committee meeting because of them. Dr. Aaron Mitchell's insightful advice and support from his lab has been an integral part of this work. Dr. Dannie Durand provided phylogenetic expertise and lent the unique perspective of a computational biologist that greatly helped this work. Dr. Benjamin Janto was immensely supportive during my early days as a novice graduate student and I have continued to learn from him. Special thanks to Dr. Hasan Yesilkaya for our collaborative work and for graciously becoming a part of my committee. His expertise in the field of pneumococcus has helped raise the quality of this work.

I would also like to thank Mr. Glen de Vries for his support through the Presidential Fellowship sponsored by him. His kind words and appreciation for my work have been truly rewarding. My sincere thanks to all members of the lab at the Allegheny General Hospital, where I spent the first two years of grad school. Special thanks to Mark Longwell, with whom I have performed many *in vivo* studies and Evan Powell for training me when I was a rookie.

My day-to-day lab experience was made a lot more cheery because, as The Beatles put it, I got by with a little help from my friends. Rory Eutsey and I were a great team and worked together on a number of projects. Over the last five years, he has taught me good science, listened to my woes, laughed with me, challenged me and catalyzed my growth as a scientist and a person. Rory's signature response to my breakthroughs-"What could it all mean?!"- and to my failures-"There, there"-shall continue to echo in my ears for the times to come. Surya Dev Aggarwal is another of Hiller lab's finest who deserves a special mention in my journey. He has been my true friend and the voice of reason that kept me sane and cheered me on without

question or judgement. I also want to thank all past and present members of the Hiller lab and my friends Phil, Brendan and Teresa for their camaraderie and the good times.

Words can never express the gratitude I feel towards my dear family for always letting me fly far and free towards my goals. Their unconditional love and support at every step kept me going. Even though we are continents apart, there was never a day I did not think of them. Receiving letters written by my sister in my mailbox felt like a slice of sunshine and I always looked forward to them. Everything I am today, I owe to the efforts of my dear mother and the never-ending encouragement of my father.

In the summer of 2015, I had the good fortune of meeting John Hubczak, for whom I am truly grateful. Thank you for always believing in me, making me laugh and putting up with my moods during the writing of this thesis. I am also grateful to the Hubczak family for their warmth and kindness towards me.

Last, but by no means least, I would like to thank a very special bacterium *Streptococcus pneumoniae* and the elegant mysteries it harbors that have been the source of my hardest struggles and the greatest joys for the past five years.

-

Anagha Kadam

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Streptococcus pneumoniae (pneumococcus) is one of the most important community acquired pathogens worldwide. Pneumococcal infections can be either asymptomatic or symptomatic. During asymptomatic infections pneumococcus colonizes the nasopharynx. The pneumococcal carriage rates are estimated to be 20–50% in children and 5–20% in adults in higher resourced countries while even higher rates are seen in resource poor settings where up to 90% of children and over half of adults are colonized [1–4]. Approximately 40% of carriage cases have multiple strains [5,6]. Pneumococcus can also disseminate to tissues, frequently the middle ears and lungs, and relatively rarely the eyes, heart, and brain causing severe disease [7–10]. Worldwide, an estimated 850,000 children under the age of five succumb to pneumococcal disease [2]. The molecular mechanisms that determine whether pneumococcus manifests disease or asymptomatic carriage are not well understood.

To date, over 20,000 genomes of pneumococcus have been sequenced revealing a diverse pangenome with numerous pathways and regulatory networks of unknown function. My goal is to study the mechanisms by which pneumococcus causes disease in the context of an important clinical lineage (termed PMEN1) and in the context of multi-strains infections. This goal is achieved in this thesis through the discovery and characterization of a novel signaling system in the PMEN1 lineage of *Streptococcus pneumoniae,* and the demonstration of its role in disease as well as its impact on multi-strain infections via promiscuous signaling

Pneumococcal disease and public health

Pneumococcus causes a wide variety of illnesses of the respiratory tract and systemic complications. It is one of the top causative agents of pneumonia, ear infections, sinus infections and meningitis. The Center for Disease Control (CDC) classifies pneumococcus at a threat level of "serious", underlining it as a bacterium that requires prompt and sustained action. Infected persons experience a sudden onset of illness with fever and malaise; these symptoms are common to all forms of pneumococcal disease. In pneumococcal pneumonia, symptoms include cough, chest pain, and sputum that is purulent or blood-tinged. Symptoms may present less abruptly in the elderly, sputum may be absent, and additional symptoms may include shortness of breath or altered metal status. In pneumococcal meningitis, symptoms are a stiff neck, headache, lethargy, and seizures. In otitis media and sinusitis, sinus and ear pain are typical symptoms [11].

Infection transmits from person-to-person via respiratory secretions such as saliva or mucus or via droplets during coughing. Risk groups for pneumococcal disease are children below the age of 5 and senior adults above 65 years of age. Susceptibility to infection is increased in individuals who are in group child care, recipients of cochlear implants, or display chronic cardiac, or lung conditions or an immunocompromised state. According to CDC, pneumococcal infections in the United States lead to 160,000 pediatric visits every year. In adults, this number is 600,000. Worldwide, pneumococcal infections cause about 14.5 million cases and are responsible for one-third of the annual

pediatric deaths with highest occurrence in children less than 2 years of age [11]. The burden of pneumococcal disease is further worsened by the emergence of drug resistant strains. In 2013, 34,000 cases of invasive pneumococcal disease were reported in the US, out of which about 30% of the cases had strains that were resistant to at least one antibiotic. Finally, the economic cost of pneumococcal disease is estimated at $96 billion [12].

Presently, there are two types of circulating vaccines for pneumococcus in the US. Additional pneumococcal conjugate vaccine formulations are anticipated to be publically available in the near future, such as the glycoconjugate vaccine [13]. The pneumococcal polysaccharide vaccine (PPV, Pneumovax by Merck) was first introduced in the 1970's and modified in 1983. It contains capsular polysaccharides from 23 serotypes. Today, PPSV covers serotypes that cause about 60% of infections in adults and is used in adults and children above two years of age [14]. Polysaccharide vaccine PPSV are poorly immunogenic to infants and toddler below the age of two [15], the group with the highest incidence of invasive disease [16]. The second type of pneumococcal vaccine is pneumococcal conjugate vaccine (PCV, Prevnar by Pfizer). PCVs contain polysaccharides from target serotypes covalently linked to a nontoxic immunogen carrier. For this reason, PCVs have been very effective in infants and are a part of regular immunization plan for that age group [17–19]. PCV7 was first introduced in 2007, and in 2010 it was replaced by PCV 13. PCV vaccines induce high levels of mucosal antibody production than the PPV23 vaccine, this in turn reduces nasopharyngeal carriage of pneumococcus and triggers greater immune memory in

infants, children, the elderly, and the immunosuppressed [20,21]. PCVs have been very effective in reducing pneumococcal mortality in the US and the rest of the world. In a trial to evaluate the efficacy of PCV7 vaccine pre- and post- vaccine carriage rates demonstrated strong reduction in the carriage of serotypes included in the vaccine [16] [20]. However, the overall carriage remained unchanged due to colonization by strains with capsular types not included in the PCV7 [22–24]. For example, serotypes 11A and 19A (non included in PCV7), were the most frequently isolated post-PCV7; they are now included in PCV13. The largest trial to assess the efficacy of PCV13 in adults was the CAPiTA trial. CAPiTA trial compared PCV13 to placebo in nearly 85,000 immunocompetent adults ≥65 years of age in the Netherlands. Participants were enrolled between 2008 and 2010, had not received a pneumococcal vaccine previously and had no prior history of pneumococcal disease. The trial demonstrated 46 percent efficacy of PCV13 against vaccine-type pneumococcal pneumonia and 75 percent efficacy (95%, confidence interval 41 to 91 percent) against vaccine-type invasive pneumococcal disease [25]. In summary, while the vaccine is effective in reducing invasive disease, the reduction in nasopharyngeal carriage is limited [26].

In conclusion, the human nasopharynx is the only naturally existing reservoir of *S. pneumoniae* and the carriage rates in healthy population remain high in post-vaccine era. Nearly all children harbor at least one strain as early as the first year of life. Thus, the nasopharynx of both vaccinated and unvaccinated children is a niche for pneumococcal growth and adaptation and pneumococcal disease continues to pose a challenge to human health worldwide.

Classification of pneumococcus

Scientific classification:

Domain: Bacteria

Phylum: Firmicutes

Class: Bacilli

Order: Lactobacillales

Family: Streptococcaceae

Genus: Streptococcus

Species: *S. pneumoniae*

S. pneumoniae was discovered in 1881 by Louis Pasteur. In 1884, it became the first pathogenic bacterium to be observed by Christian Gram during the development of his staining technique. *S. pneumoniae* is Gram-positive, bile-soluble, and aerotolerant and alpha-haemolytic. Viewed microscopically, cells are lancet-shaped, ranging in diameters between 0.5 and 1.25 micrometers. Most often, they are seen in a diplococcal arrangement, but can also be found singly or in short chains. Pneumococci produce autolysin, the enzyme that causes self-destruction of cells. On an agar plate, the colonies are seen with a plateaued elevation initially but these later collapse in the center, signaling the autolytic event and yielding concave shaped colonies.

The pneumococcal capsule is commonly used for classification as it the primary virulent factor of this bacterium, and has been widely studied. In accordance, loss of the capsule leads to dramatic decrease in virulence [27]. There are over 100 capsular types, and the human immune system produces protective antibodies that are specific to serotypes or serogroups. This diversity prompted the famous immunologist Charles

Janeway, Jr., to state "from the point of view of the adaptive immune system, each serotype of *S. pneumoniae* represents a distinct organism". The capsule aids in evading the host immune system and its antigenicity is the basis for classification of strains into serotypes [28]. In the early 20th century, serotype-specific antisera were widely used for treating patients leading to the discovery of a large number of serotypes as antiserum therapy developed in Denmark and America. In the early 1900's Neufeld described the Quellung reaction, which later became the gold standard method in serotype classification of pneumococcal strains. This method of serotype identification requires the use of antisera pools that specifically target the capsule of the bacterium leading to a swelling appearance under the microscope. The composition of the polysaccharide units in the capsule defines the serotype of the strain. To date 101 serotypes are known and more are on the way as new isolates gain distribution or continue to evolve due to capsular switches induced by the fitness pressure from host immunity and vaccination. The capsules belong to 45 serogroups [29,30]. A serogroup was defined to include serotypes that share many serologic properties (i.e., cross-reactive antibodies). Serotype classification provides the basis for vaccine development, epidemiologic surveillance and phenotypes in the clinical spectrum where invasive diseases are found to be associated with specific serotypes reviewed in [28].

With advancements in sequencing technologies, pneumococci are commonly classified based on their genomic backgrounds. PCR has allowed rapid, less laborious and cost-effective identification of serotypes and strain backgrounds. Numerous assays have been described in which PCRs have been developed and multiplexed, in different

combinations, to identify variable numbers of commonly occurring serotypes and a present-day panel of housekeeping genes. The most acceptable classification today is the multi locus sequence type (MLST), which was first developed by Enright and Spratt in 1998. The authors developed a pneumococcal MLST scheme and database by sequencing approximately 450 bp fragments of seven housekeeping loci from 295 isolates [31]. The combination of alleles at the seven loci provided an allelic profile that is used for identification and evolutionary characterization of widely distributed and emerging strains. MLST data is electronically portable, available on the MLST website (PubMLST) and widely used (https://pubmlst.org/spneumoniae/).

The Pneumococcal Molecular Epidemiology Network has classified strains that are widespread and drug resistant into PMEN clones. This classification is based on MLST, pulsed field gel electrophoresis (PFGE), and penicillin binding protein (PBP) profiles [32]. There are currently 43 PMEN clones.

The PMEN1 lineage

The Pneumococcal Molecular Epidemiology Network has grouped strains of multi locus sequencing type (MLST) 81 (also known as Spain23F-1 and SPN23F) into the PMEN1 lineage [32]. Over the past 30 years, PMEN1 has distinguished itself by its wide-distribution worldwide, its multi-drug resistant profile, and emergence of vaccine-escape strains. Historically, the PMEN1 were responsible for the Spanish epidemic of the 1980s and have since spread to North and South America, Europe, Asia, Africa, and Australia [7,32]. Most strains in the PMEN1 lineage are resistant to

penicillin, chloramphenicol, and tetracyclin, and many isolates have additional resistances to fluoroquinilones and macrolides[33,34]. PMEN1 strains are predominantly of serotype 23F, but there are also capsular switches to other serotypes thus generating vaccine-escape isolates [35].

PMEN1 (ST81) lineage is postulated to have evolved from an ancestor in 1967 in Australia, and by the end of 1990s it represented an estimated 40% of penicillin resistant strains in US [36,37]. These strains display very high rates of carriage [7,37–39]. PMEN1 also displays very high rates of disease [7,38,40]. Is the prevalence of PMEN1 in invasive disease a function of its carriage rates or does it reflect a propensity to cause disease? Multiple studies have shown that sequence types vary regarding their propensity to cause disease and Sjostrom *et al*. show that PMEN1 displays a low propensity to cause invasive disease [41]. Thus, high rates of PMEN1 invasive disease in the population likely reflect high carriage rates, and not heightened virulence potential.

The PMEN1 lineage is not only itself widespread; it has also impacted the genome content of the pneumococcal population by virtue of a high frequency of DNA donation to other pneumococcal strains [36]. This is particularly clear for the cell wall transpeptidases, the penicillin binding proteins (PBPs), where the PMEN1 alleles confer resistance to beta-lactams. The spread of PMEN1 and PMEN1 genes have earned this lineage the distinction of a paradigm of genetic success [36].

Genomic diversity and plasticity

There is extensive genomic diversity among strains of pneumococci. Approximately 50% of the pangenome is unevenly distributed amongst isolates, while the other half is shared across all strains (core set) [42,43]. In addition to differences in gene possession, there is also extensive allelic variation. Further, some loci are notorious for their extensive allelic diversity. In fact, thirteen large loci, termed regions of diversity by Tettelin *et al*., account for greater than half the allelic diversity observed between isolates [44]. This extensive genomic and allelic diversity across strains poses a challenge to cataloging mechanisms of pathogenesis in pneumococcus.

Pneumococcus is endowed with the ability of natural transformation where DNA from outside the cells, released by dead cells of either the same or different bacterial species, is readily taken up by cells where it can recombine into the recipient genomes leading to horizontal gene transfer. The genomic plasticity of pneumococci is so extensive that it blurs the phylogenetic signal of the species [42]. Moreover, recombination events that lead to switches of an estimated 10% of the pneumococcal genome can be detected in a single chronic infection [45]. Genomic regions often display features that are consistent with horizontal transfer: atypical GC content, insertion sequences or remnants of mobile genetic elements, and phage components [46,47]. Thus, horizontal gene transfer from transformation, conjugation, or transduction events are commonplace in pneumococcal genomes.

Differences in alleles and in gene possession may contribute to the diverse phenotypes of different pneumococcal lineages as observed in terms of virulence

potential, colonization capabilities, and response to vaccines and clinical interventions. Acquisition of new gene content via horizontal gene transfer events may lead to the acquisition of new functions as well as re-wiring of existing regulatory networks in the recipient strain [47,48]. In this manner, characterization of strain-specific genes can provide molecular clues to explain virulence and colonization differences across strains. This thesis focuses on a set of genes present only in strains from the PMEN1 lineage and their role in carriage and disease.

Biofilms and cell-cell communication

Pneumococcus forms biofilms on mucosal surfaces, such as the nasopharynx during asymptomatic colonization, the middle ear during otitis media, and the sinus during rhinosinusitis [49–52]. Biofilm are surface-associated microbial communities surrounded by an extra-cellular matrix [53]. Pneumococci in a biofilm mode of growth display increased rates of gene recombination relative to a planktonic mode of growth, and are recalcitrant to antimicrobial treatments [54,55]. Furthermore, pneumococci in a biofilm mode of growth display coordinated and cooperative behaviors that depend on cell-cell signaling. These signaling pathways are at the center of this thesis.

Transcriptional factors are the gatekeepers that bring about changes in the gene expression landscape of a cell. They usually fit into the regulatory circuits downstream of sensory machinery. In bacteria, the best-characterized sensory-response machinery are the two-component systems defined by a surface-located histidine kinase and a

cytosolic response regulator. Pneumococcal genomes encode thirteen two-component systems [56,57], and they have been implicated in virulence, response to host ammunition, competence, biofilm formation, bacteriocidal activity, and cell wall synthesis [58–69].

In Gram positive bacteria, mainly bacilli, streptococci and enterococci, there is also a second architecture for sensory-response machinery. These systems consists of a transcriptional factor and a regulatory peptide that is exported, processed, and internalized into the cytosol where it modulates the activity of the cognate transcriptional factor. These regulator-peptide systems belong to the RRNPP family (Rgg, Rap, Npr, PlcR, PrgX) [70,71]. Different RRNPP members and their cognate peptides have been characterized across Firmicutes and have been shown to regulate virulence and commensalism, demonstrating their important in the physiology of pathogens in the context of host interactions. For example, PrgX/TraA from *Enterococcus faecalis* controls conjugation *in vivo* in response to human plasma (REF). PlcR/PapR control the production of extracellular toxins and modulate virulence in Bacilli [72,73]. TprA/PhrA plays a role in utilization of host sugars in pneumococci [74]. NprR/NprX controls necrotrophism in Bacilli in an insect larval model [75]. Rgg/SHP in the Streptococcus family regulates commensalism, virulence, oxidative response, and biofilm formation [76–79].

In both two component systems and RRNPP systems, the signal is mediated via a secreted peptide. In this way, peptides are prime candidates for transmission of

environmental signals, anti-microbial activity, and host-toxicity. Their small size and activation via processing make peptides ideal responders to specific conditions, samplers of the extracellular milieu and whistle-blowers to the bacterial cells. Further, secreted peptides may also function in quorum sensing, as proxy for the number of producer cells, to bring about a larger impact on the microbial community by instigating group behavior in bacterial cells that sense and respond to the peptide [80]. In conclusion, peptide-mediated signal transduction systems allow for both intra and inter strain communication, and are poised to play important roles in single strains and multi-strain infections. In Chapter 2, I introduce a novel member of the RRNPP superfamily, and in chapter 3, I describe a novel function for the recently identified TprA/PhrA members of the RRNPP superfamily.

Multi-strain infections

An estimated forty percent of carriage are comprised of multiple strains [5]. However in the majority of the treatment landscape, infections are treated as a single, uniform entity. The lack of understanding of the molecular mechanisms and cross-strain interaction in multi-strain infections represent a major gap in current knowledge [81]. This is not unique to pneumococcus, in a report by Balmer and Tanner from 2001, about 51 of human pathogens and 21 non-human ones have been reported to exist as multi-strain infections [82]. It is probable, that multi-strain infections do not respond to treatment and management in the same manner as their single infection counterparts. Presence of multiple strains can cause inter-strain interactions of cooperative and

competitive natures, and may modify the host immune response, as well as affect the response to antimicrobial treatments, vaccinations, and outcome from a pre-existing disease [83].

To date, the majority of pneumococcal research has been centered on understanding the behavior and mechanism of single strain infections. This large body of work has greatly contributed to the knowledge of pneumococcal physiology and virulence, and informed vaccination strategies and antibiotic use. However, there is scarcity in the knowledge of multi-strain infections. In Chapter 4, I present a novel example of cross-talk between signal transduction systems across strains, and explore their relevance in the context of cell-cell communication and multi-strain infections involving strains from the PMEN1 lineage.

Thesis summary and hypotheses

The high rates of PMEN1 invasive disease in the population appear to reflect high carriage rates (section 1.3). Thus, I tested the hypothesis that acquisition of a unique signaling system has endowed the PMEN1 lineage with attributes that promote commensalism. To this end, I introduce a novel regulator-peptide signaling system, TprA2/PhrA2, encoded by a clinically relevant lineage, the PMEN1. I provide evidence that this system activates a virulence factor, but yet controls its expression by carefully balancing commensal and pathogenic outcomes (Chapter 2).

Studies of TprA2/PhrA2 revealed that this system cross-reacts with a second regulator-peptide signaling system, TprA/PhrA, widespread across pneumococcal strains (Chapter 4). Further, data from collaborative work revealed that TprA/PhrA is a robust virulence factor (Chapter 3). Together, these finding suggest that TprA2/PhrA2 influences gene expression both intra and inter strain. Thus, I end my thesis with the new hypothesis that PMEN1 strains control the expression of virulence determinants in mixed populations.

Chapter 2. Characterization of the PMEN1-unique TprA2/PhrA2 quorum sensing system

This chapter describes the discovery and characterization of a novel *S. pneumoniae* signaling system. In this study, a PMEN1-unique genomic locus encoding for a regulator-peptide quorum sensing system was identified using comparative genomics, the regulation and function of this signaling system deduced by molecular biology, and the role in infection demonstrated using animal models of pneumococcal disease. This work corresponds to pages 1-9 of Kadam *et al*., 2017

Citation: Kadam A., Eutsey R.A., Rosch J., Miao X., Longwell M., Xu W., Woolford C.A., Hillman T., Yesilkaya H., Mitchell A.P., Hiller N.L. Promiscuous Signaling by a Regulatory System Unique to the Pandemic PMEN1 Pneumococcal Lineage. ***PLoS Pathog.*** 2017

Disclaimer: I led and performed all aspects of experiments and data analyses of this work with contributions from Xinyu Miao for phylogenetics, Rory Eutsey for molecular biology, and Dr. Carol Woolford, Dr. Wenjie Xu and Dr. Aaron Mitchell for transcriptional studies using NanoString technology. I performed experiments in the chinchilla model of middle ear disease with support from Mark Longwell, and otoscopic disease assessments were performed by Dr. Todd Hillman. The experiments in the mouse model of pneumonia were carried out by Dr. Jason Rosch.

Introduction

Streptococcus pneumoniae (pneumococcus) is one of the most important community acquired human pathogens, and is responsible for an estimated 850,000 deaths annually in children under the age of 5 [2]. Pneumococcus colonizes the nasopharynx of young children at very high rates, and is asymptomatic in most cases [7,38]. However, it can also disseminate from the nasopharynx into tissues leading to diseases such as otitis media, pneumonia, bacteremia, meningitis, and inflammation of the heart [8–10]. The pneumococcal molecules responsible for this transition from a commensal to a pathogen are not well understood. Here we characterize a novel quorum sensing (QS) system (TprA2/PhrA2) that limits pneumococcal disease, without affecting nasopharyngeal colonization.

At the genomic level, there is extensive diversity among pneumococccal lineages. These genomic variations contribute to the differences in colonization and virulence potential [84]. Only half of the pangenome is shared across all strains (core set), while the other half is unevenly distributed amongst isolates [43,85]. The Pneumococcal Molecular Epidemiology Network (PMEN) has grouped strains of multi locus sequencing type (MLST) 81 into the PMEN1 lineage (also known as Spain23F-1 and SPN23F) [32]. Over the past 30 years, PMEN1 has distinguished itself by its worldwide distribution, multi-drug resistant profile, and emergence of vaccine-escape strains.

Historically, the PMEN1 lineage was responsible for the Spanish epidemic of the 1980s and has since spread to North and South America, Europe, Asia, Africa, and

Australia [7,32]. Most PMEN1 isolates are resistant to penicillin, chloramphenicol, and tetracycline, and many isolates have additional resistances to fluoroquinilones and macrolides [33,34]. PMEN1 isolates are predominantly of serotype 23F, but there are also capsular switches to other serotypes, some of which represent vaccine-escape isolates [35]. Further, the PMEN1 lineage has impacted the genome content of the pneumococcal population by virtue of its high frequency of DNA donation, including genes for drug-resistance, to other pneumococcal lineages [36]. The PMEN1 genome encodes an integrative conjugative element (ICESp23FST81) [35,47,86]. As described by Croucher and colleagues upon sequencing of the first PMEN1 genome, this ICE encodes drug resistance determinants, a complete lanthionine-peptide gene cluster and a regulator-peptide pair, which in this study we have identified as the TprA2/PhrA2 QS system.

Quorum sensing systems serve as a critical, decision-making process in the response of bacteria to the environment, and their ability to colonize and/or disseminate to tissues. The best characterized kind of QS machinery is the two component system, where the signal is sensed by a surface-localized histidine kinase and transferred to a cytosolic response regulator [87]. Streptococci, enterococci and bacilli have been shown to encode a second kind of QS characterized by the emerging RRNPP (Rgg/Rap/NprR/PlcR/PrgX) superfamily of transcriptional regulators and their cognate peptides [71]. In these systems, the secreted peptide is exported from the producer cell, processed, and imported into the cytosol of producing or neighboring

cells, where it interacts with the RRNPP regulator . RRNPP-peptide systems have been shown to regulate virulence, biofilm formation, and the production of bacteriocins [88–90].

In pneumococcus, the majority of characterized peptides signal via two component systems [87]. These peptides regulate competence and class II bacteriocin production [91,92]. The first RRNPP-peptide pair was recently characterized in the pneumococcus strain D39 [74]. It is composed of the TprA regulator and its cognate peptide PhrA. PhrA alleviates gene inhibition leading to the expression of physiologically important genes [74]. PhrA levels are repressed by glucose and activated by galactose, consistent with activity in the upper respiratory track where galactose is a major source of energy [93].

In this study we characterize the TprA2/PhrA2 QS system, a novel pneumococcal RRNPP-peptide pair, highly expressed in middle ear effusions. TprA2/PhrA2 is present almost exclusively in PMEN1 isolates where it restrains dissemination. Unlike other lineages, the PMEN1 strains encode both the TprA/PhrA and the TprA2/PhrA2 signaling systems. Extracellular PhrA2 leads to induction of TprA in PMEN1 cells as well as in D39 cells. Thus, horizontal acquisition of TprA2/PhrA2 has provided the PMEN1 lineage with a QS system and associated regulon, as well as the molecular machinery to regulate a widespread cell-cell communication system and in doing so, influence not only its own gene expression but also that of other strains.

Results

The genes encoding the TprA2/PhrA2 system are enriched in the PMEN1 lineage

Genes enriched in the PMEN1 strains may provide this lineage with exclusive phenotypic properties, explaining its prevalent occurrence and rapid spread. We performed a comparative genomic screen to search for genes that are present in the majority of the PMEN1 isolates, but absent in other pneumococcal lineages. The analysis was performed on 60 pneumococcal genomes, selected to capture the diversity in the pneumococcal population (S1 Table, labeled "To establish PMEN1 enrichment"). We employed RAST [94] to annotate the whole genome sequences (WGS) into 125,612 coding sequences (CDSs), and organized these into 3,571 clusters of homologous sequences as previously described [95]. The screen identified a genomic region present only in the PMEN1 strains. This region encodes a transcriptional regulator (*tprA2*) on the opposite strand of a small peptide (*phrA2*) and three ABC transporters. Immediately downstream are three genes *lcpA*, *lcpM*, and *lcpT*. LcpA encodes a putative 71aa peptide with the full size weight of 7.5kDa, which we predict is a lanthionine containing peptide. Lanthionine and methyllanthionine are usually formed by the dehydration of threonines or serines, and subsequent cyclization to cysteine (*lcpA* encodes for serine, threonines, and cysteines) [96]. Cyclization is performed by lanthipeptide synthetases, of which there are four known classes [97]. The *lcpM* gene downstream of LcpA is consistent with class II synthetases (CDD score: LanM-like e-value 0e+00 [98]) . Finally,

the *lcpT* encodes a transporter with a C39 peptidase domain, which we predict is involved in LcpA cleavage and export (**Fig 2.1, Table 2.2**).

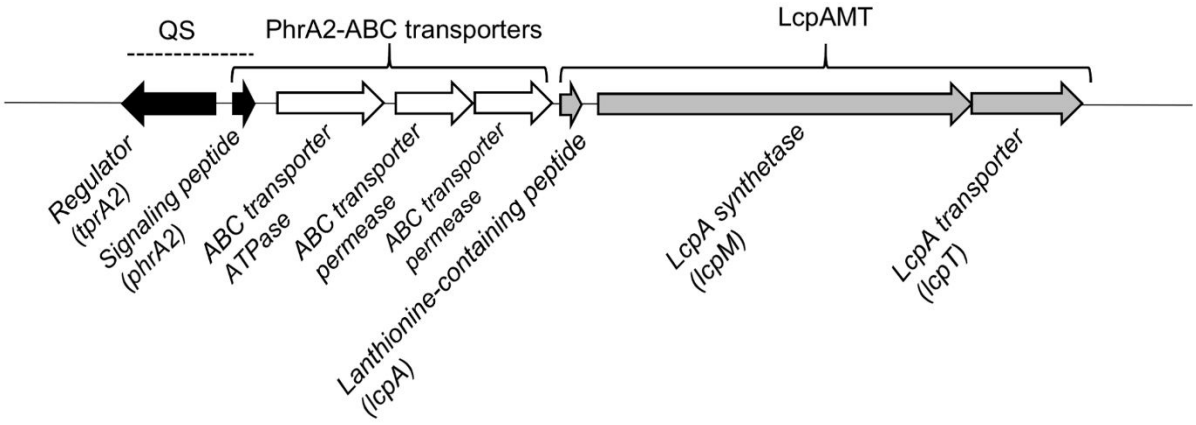


Fig 2.1. Genomic organization of TprA2/PhrA2 region in PMEN1 strains.

We performed a detailed assessment on the phylogenetic distribution of the QS-Lcp genes in the pneumococcus species and the Streptococcus genus. First, for the assessment of the distribution of TprA2 in the PMEN1 lineage, we searched for this gene in 215 PMEN1 isolates. To this end we used either polymerase chain reaction (PCR) or genomic data assembled by Croucher and colleagues [35]. The *tprA2* gene was present in 212 isolates. It was either disrupted or deleted in the genomes of strains 111 (ERS004810), 11933 (ERS005313) and HKP38 (ERS004775) (genome data was confirmed by PCR). Next, we broadened our search into the non-redundant database, which revealed that *tprA2* was present in only one strain outside the PMEN1 lineage (GA13494) [86] (**Fig 2.2**). Finally, we expanded our search for *tprA2* in related streptococcal species, specifically *S. pseudopneumoniae*, *S. mitis*, *S. oralis*, and *S. infantis*

(**Table S1** labeled "Distribution with Streptococcus sp"). We found one occurrence in *S. mitis* and one in *S. infantis*, but these species did not encode the downstream *lcpAMT* locus (**Fig 2.2**). These phylogenetic analyses demonstrate that the QS system and *lcpAMT* are present in >98% of the PMEN1 isolates and are rare outside this lineage. This distribution suggests these genes were acquired via horizontal gene transfer by a PMEN1 ancestral strain.

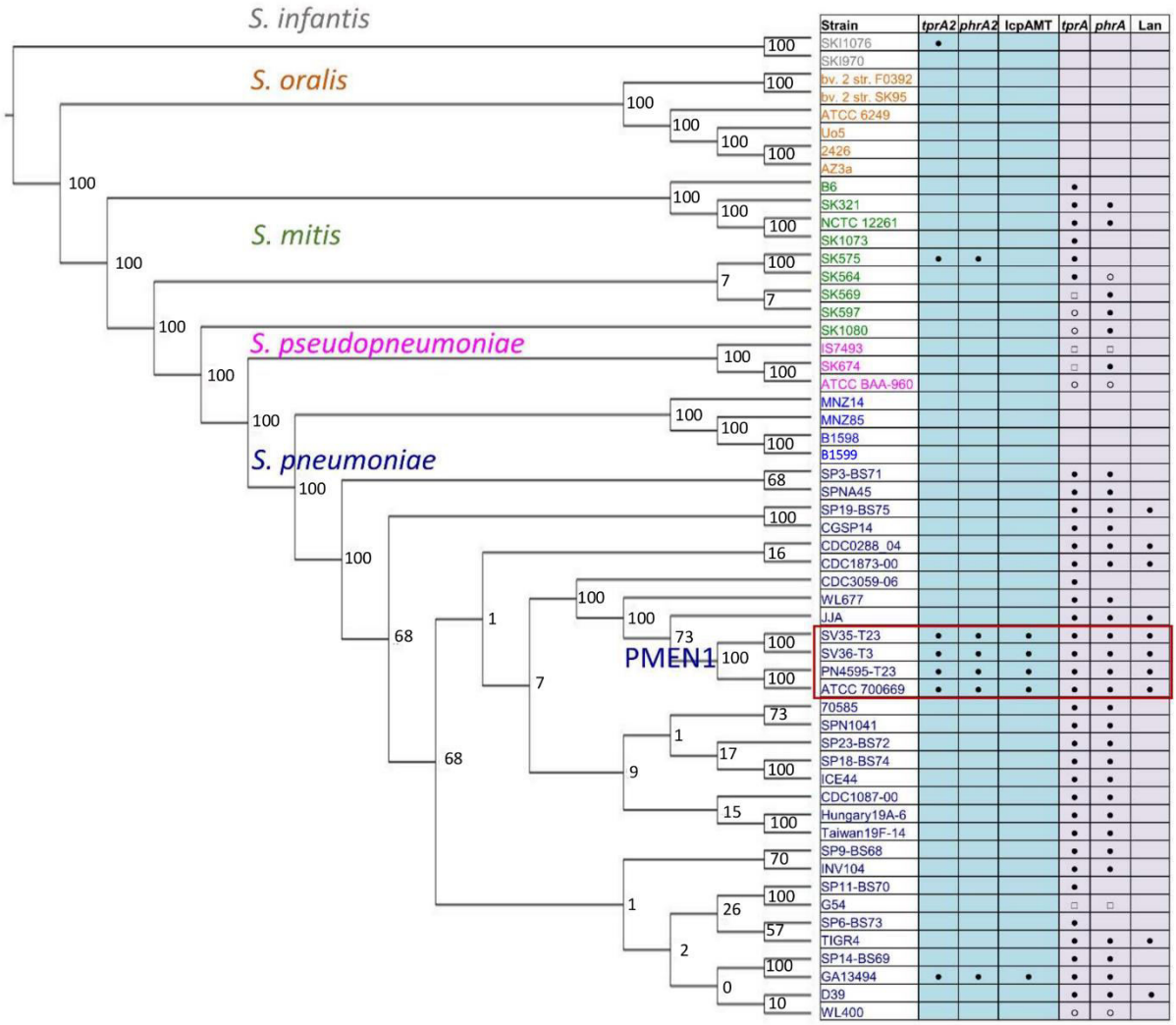


Fig 2.2 Intra- and inter-species distribution of the TprA2/PhrA2 and TprA/PhrA genomic regions. Phylogenetic analyses displaying bootstrap values on the branches.

Left side: Maximum likelihood tree of streptococcal genomes generated from the core genome. Right side: Gene distribution, where blue columns display the distribution of *tprA2*, *phrA2*, and associated *lcpAMT*, and purple columns display the distribution of *tprA*, *phrA*, and downstream lantibiotic genes (seven consecutive genes, including predicted *lanA* and *lanM* labeled as Lan). Presence of the gene is marked with the following symbols: '●' gene present in one copy; '○' low coverage of region; '□'multiple copies of the gene. Red box indicates isolates from the PMEN1 lineage.

QS-Lcp genes are induced and highly expressed *in vivo*

To determine whether QS-Lcp genes are active during infection, we measured their gene expression during middle ear infection. We utilized the nCounter NanoString technology since this allows for an automated, highly sensitive enumeration of pathogen's mRNA transcripts in the infected host tissue. Our probes capture *tprA2*, *lcpA*, *lcpM*, and *lcpT*. Further, since we were unable to design a probe for the short coding sequence of *phrA2*, we used *ABCATPase* as a proxy since it is present on the same transcript (**Fig 2.3**). For normalization we used probes to *gyrB* and *metG*, and normalized to the geometric mean of these housekeeping genes. The PMEN1 strain PN4595-T23 [99] was inoculated transbullarly into the chinchilla. We isolated RNA from effusions of the chinchilla middle ears at 48h post-transbullar inoculation. All five genes were expressed in middle ear effusions (**Fig 2.4**). The average counts for *ABCATPase* and *lcpA* were comparable to those of *psaA* (56,036 counts), which has been shown to be highly expressed *in vivo* [100], consistent with high levels of QS-Lcp *in vivo*.

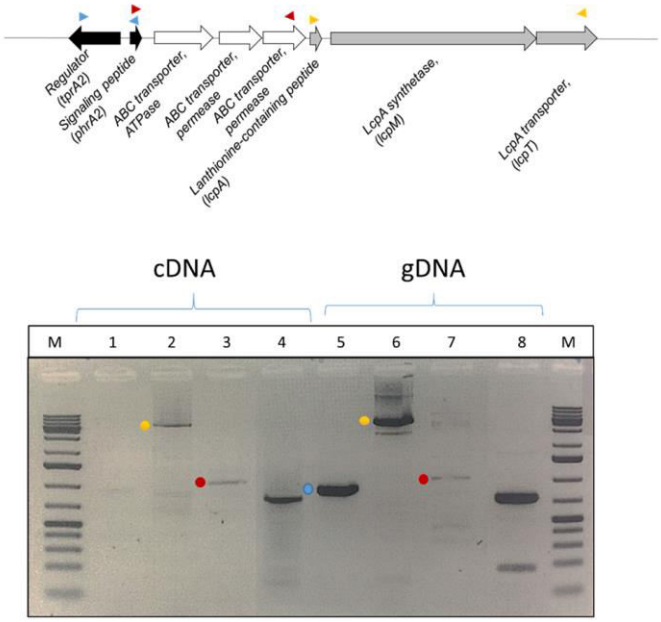


Fig 2.3. **PCR performed on cDNA and genomic DNA to demonstrate transcriptional units.** Lanes 1-4 are PCRs on cDNA template, lanes 5-8 on gDNA template. Primers used are as follows: lanes 1 and 5, *tprA2* fwd and *phrA2* rev; lanes 2 and 6, *lanA* fwd and *lanT* rev; lanes 3 and 7, *phrA2* fwd and *ABCATpase* rev; lanes 4 and 8 *gapdh.* Colored arrow heads in the genomic locus schematic indicate the primer binding sites corresponding to the bands on the gel (marked with the equivalent color). 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.

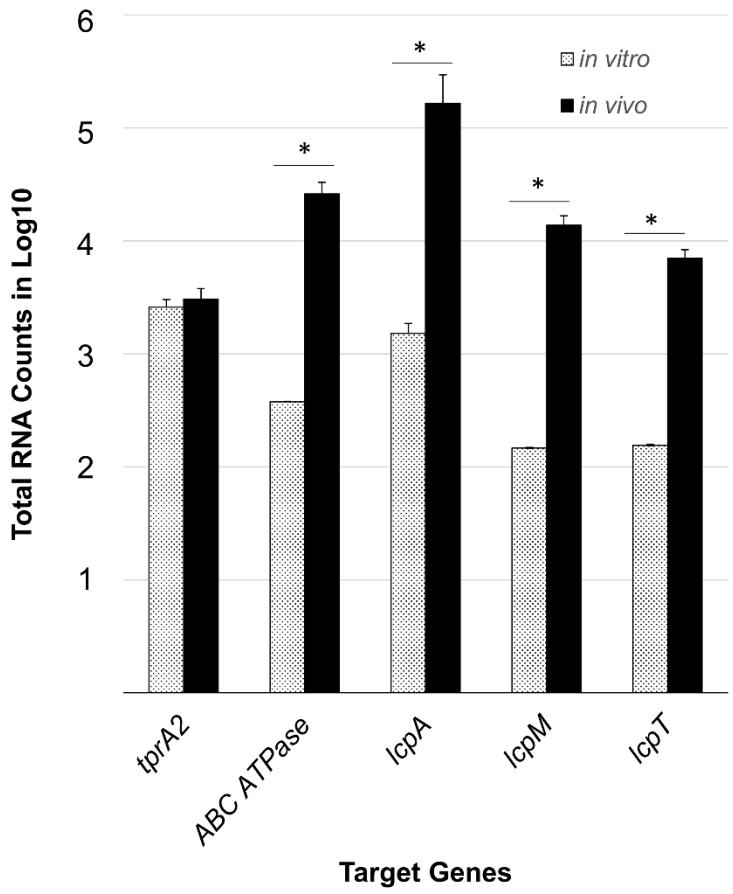


Fig 2.4. Gene expression levels of the TprA2/PhrA2 system and associated lcpAMT locus in chinchilla middle ear effusions and planktonic cultures. nCounter NanoString technology was used to quantify mRNA transcripts from planktonic cultures (dotted bars, n=2) and chinchilla middle ear effusions (black bars, n=3). Data was normalized to the geometric mean of the expression of *gyrB* and *metG* using nSolver software. The X-axis denotes the test genes assayed for gene expression. The Y-axis displays the log10 of the total number of transcripts for each gene averaged over biological replicates. Error bars represent the standard deviation. '*' Significantly higher in vivo expression (P-value < 0.05), as determined by Student's t-test.

To evaluate whether these genes were induced in the middle ear relative to growth in rich media, we calculated the ratio of the average number of transcripts between middle ear effusions and *in vitro* planktonic cultures. The gene expression levels of *ABCATPase*, *lcpA*, *lcpM* and *lcpT* were 69, 108, 93 and 45-fold higher *in vivo* relative to planktonic cultures, respectively. From these *in vitro* and *in vivo* measurements we infer that the QS-Lcp system is both induced and highly expressed during infection.

The expression of phrA2 is regulated in a density-dependent manner

The expression of sensory peptides can be cell-density dependent (reviewed in detail in [80]. Using quantitative real time PCR (qRT-PCR) we found that *phrA2* is regulated in a density-dependent manner. Expression of *phrA2* increases at higher cell density, as observed by measuring gene expression at lag, early-log, mid-log and stationary phase (**Fig 2.5, solid bars**). Further, when a lag phase culture was left to grow for one hour, the levels of *phrA2* expression increased 3 fold. When the same culture was exposed to cell-free supernatant from a wild-type high-density culture, the levels of *phrA2* expression increased 8 fold. Yet, when it was exposed to the cell-free supernatant from a Δ*phrA2-ABC* high-density culture, the levels of *phrA2* did not increase (**Fig 2.5, striped bars**). Thus, the wild-type cells but not the Δ*phrA2-ABC* mutant, secrete a molecule that induces expression of *phrA2* in the population. These data are consistent with secretion and autoinduction of PhrA2.

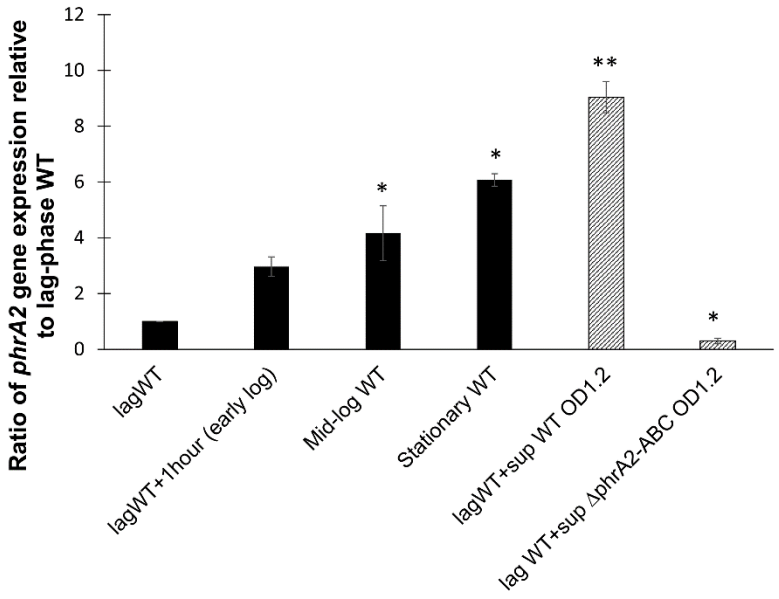


Fig 2.5. Density-dependent gene expression and extracellular secretion of PhrA2 during planktonic growth. qRT-PCR measurements of *phrA2* gene expression in PN4595-T23. The Y-axis displays expression levels as a ratio to expression in lag phase culture. The X-axis denotes culture conditions. Black bars displays density-dependent gene expression at lag phase (OD_{600}0.05), early-log phase (OD_{600}0.2), mid-log phase (OD_{600}0.6), and stationary phase (OD_{600}1.0). Striped bars display treatment by cell-free supernatants. The lag phase culture was divided into three tubes and grown for 1h in one of three ways in: original supernatant (lagWT+1hour), cell-free supernatant from a high density wild type culture (OD_{600}1.2), or cell-free supernatant from a high density Δ*phrA2-ABC* culture (OD_{600}1.2). 16SrRNA was used as normalization control. Error bars represent standard deviations from biological duplicate experiments. '**' P-value<0.01 and '*', P-value<0.05 as determined by Student's t-test.

TprA2 is a negative regulator of *phrA2-ABC* and *lcpAMT*

To identify the TprA2 regulon, we compared the gene expression levels of the wild-type (WT) PMEN1 strain PN4595-T23 and the isogenic *tprA2* deletion mutant (Δ*tprA2*), utilizing a pneumococcal gene array (**S2 Table** in Appendix I) [101]. The expression of the phrA2-ABC and lcpAMT genes were >30-fold higher in Δ*tprA2* relative to the WT strain. These results were verified, using independent biological replicates, by both qRT-PCR and NanoString technology (**Table 2.1**). These findings suggest that TprA2 is a negative regulator of these neighboring genes.

To confirm the role of TprA2, we generated a complemented strain (Δ*tprA2::tprA2*) where *tprA2* was inserted into the Δ*tprA2* strain at a distant chromosomal location, under the influence of the constitutive erythromycin-resistance gene promoter (*ermB*). We measured gene expression of *tprA2*, *phrA2*, ABC transporter *ATPase*, and *lcpA* in the WT, Δ*tprA2* and Δ*tprA2::tprA2* strains (**Fig 2.6**). The *tprA2* gene was expressed in the Δ*tprA2::tprA2* strain, and its expression level was higher than in the WT. Further, low levels of *phrA2*, ABC transporter *ATPase*, and *lcpA* were re-established in the complement strain. These findings strongly support our conclusion that the gene product of *tprA2* is a negative regulator of *phrA2* and *lcpAMT*.

Table 2.1: Gene expression in Δ*tprA2* compared to the PN4595-T23 (WT) in planktonic cultures represented as ratio to the WT level (n=3).

Gene ID	Target Gene	qRT-PCR		Microarray		NanoString	
		Δ*tprA2*/WT	*P*-value	Δ*tprA2*/WT	*P*-value	Δ*tprA2*/WT	*P*-value
CGSSp4595_1262	*tprA2*	0		-19	<2.2E-16	NA	2.18E-02
CGSSp4595_1261	*phrA2*	70.7	5.54E-04	+32.6	<2.2E-16	NA	NA
CGSSp4595_1260	*ABC transporter (ATPase)*	85.6	5.52E-03	+34.7	<2.2E-16	+63.8	1.06E-02
Not annotated	*lcpA*	45.2	9.26E-04	+62.1	<2.2E-16	+71	2.25E-02
CGSSp4595_1257	*lcpM*	37.9	5.32E-03	+45.4	<2.2E-16	+57.8	1.65E-02
CGSSp4595_1256	*lcpT*	37.2	4.11E-03	+40.4	<2.2E-16	+47.3	2.12E-02

Table 2.2: Strains used in this study

Strain	**Description**	**Source**
PN4595-T23	PMEN1 strain isolated in Lisbon, Portugal in 1996	Drs. A.Tomasz and H. deLencastre
Δ*tprA2*	PN4595-T23; CGSSp4595_1262:Spec^R	This study
Δ*tprA2::tprA2*	*tprA2* gene inserted in the intergenic region of equivalent genes spr_0515 and spr_0516 with 90 bp of native promoter downstream of constitutively expressed erythromycin resistant gene (*ermB*)	This study
Δ*phrA2-ABC*	PN4595-T23; CGSSp4595_1258-CGSSp4595_1261:Spec^R	This study
OElcpAMT3x	Δ*phrA2-ABC/OElcpAMT*; Overexpressor of *lcpAMT* in Δ*phrA2-ABC* background. *lcpA* levels *in vivo* are 3.6 times higher than the WT	This study
OElcpAMT33x	Δ*phrA2-ABC/OElcpAMT::OEphrA2-ABC*; Overexpressor of *phrA2-ABC* in Δ*phrA2-ABC/OElcpAMT* background such that *phrA2-ABC* genes are inserted in *bga* region under theconsititutive promoter of the kanamycin cassette. Levels of *lcpA in vivo* are 33.3x higher than the WT	This study
SV36	PMEN1 strain isolated in New York in 1996	Drs. A.Tomasz and H. deLencastre
SV36Δ*tprA2*	SV36; CGSSp4595_1262:Spec^R	This study
SV36Δ*phrA2-ABC*	*phrA2* and downstream ABC transporters (CGSp4595_1261, CGSSpSV36_1147, CGSSpSV36_1146, CGSSpSV36_1145) replaced with spectinomycin resistant gene in SV36	This study
D39Δ*phrA*	D39; SPD_1746:Spec^R	Dr. H. Yesilkaya

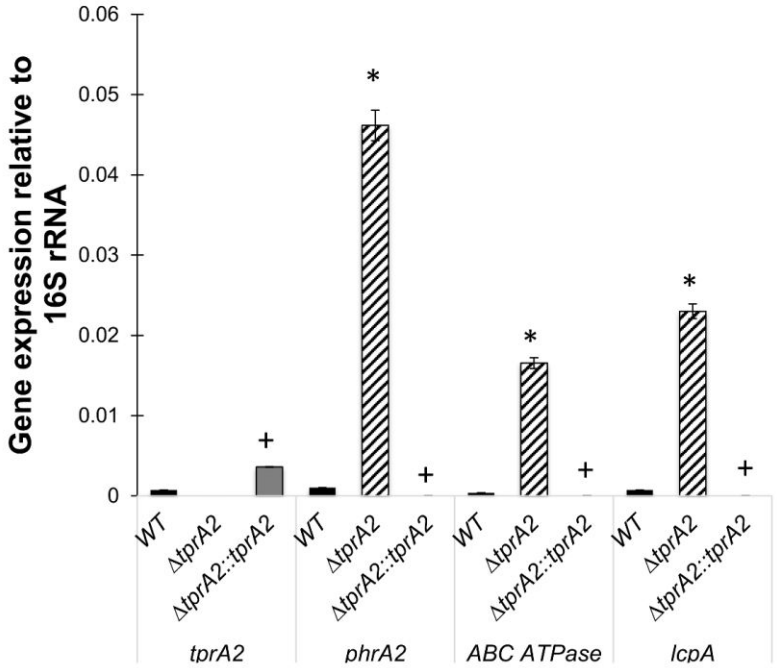


Fig 2.6. Analysis of the TprA2 regulon by comparison of gene expression levels among WT, Δ*tprA2*, and Δ*tprA2::tprA2* strains. qRT-PCR measurements for genes *tprA2*, *phrA2*, *ABCATPase* and *lcpA*. X-axis represents genes that were tested for expression in strains WT, Δ*tprA2* and Δ*tprA2::tprA2*. Y-axis denotes starting concentration of mRNA in arbitrary fluorescence units as calculated from LinRegPCR. Data was normalized to the expression of 16S rRNA. Error bars represent standard deviation for biological replicates (n=3).'*' significantly different expression relative to WT (P-value < 0.005), '+' significantly different expression relative to ΔtprA2 (P-value < 0.005).

PhrA2 modulates the TprA2 regulon

The TprA2 regulator displays sequence similarity to the Bacillus sp. transcription factor, PlcR and to the pneumococcal TprA, which are regulated by extracellular forms of the C-terminal heptapeptides from their cognate peptides [72,74]. Given that TprA2 is part of the PlcR family, we hypothesized that the C-terminal heptapeptide of PhrA2 would encompass a functional peptide capable of influencing TprA2 activity. Thus, we utilized synthetic peptides corresponding to the seven terminal residues of PhrA2 (sequence: VDLGLAD) and a scrambled control (sequence: DAGVLDL). Addition of the PhrA2 peptide, but not the scrambled peptide to planktonic culture led to a significant increase in expression levels of *tprA2*, *phrA2*, ABC transporter *ATPase* and adjacent *lcpAMT* genes (**Fig 2.7**). The PhrA2 peptide up-regulates its own production demonstrating autoinduction of this density-dependent system. We also observed an increase in the levels of *tprA2* suggesting that TprA2 serves as a negative regulator of its own expression.

The induction of gene expression by the synthetic peptide explains the observation that supernatant from a high-density WT culture, but not a Δ*phrA2-ABC*, can induce gene expression (**Fig 2.5**). Further, cell-free supernatant from a PhrA2 overexpressing strain increases levels of *phrA2* and *lcpA* by over 5 fold when compared to media alone (**S1 Fig** in Appendix I). These findings strongly support a model in which the *phrA2* gene product is exported.

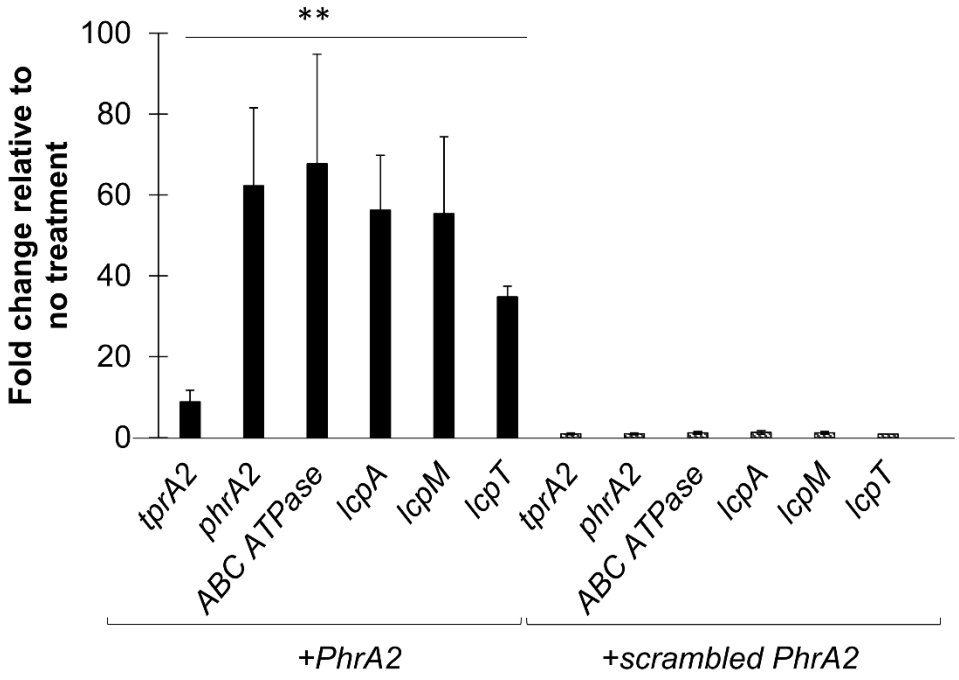


Fig 2.7. Gene expression measured by qRT-PCR of QS-Lcp genes in WT strain PN4595-T23 upon treatments. Data was normalized to 16S rRNA expression. Y-axis displays fold change in gene expression upon exposure to a peptide treatment relative to untreated control. Error bars represent standard deviations for biological replicates (n=3). On the left, dark bars display expression from cells exposed to the PhrA2 C-terminal heptapeptide (VDLGLAD); on the right side, stripped bars display expression from cells exposed to the scrambled control peptide (DAGVLDL). "**" Statistically significant difference in gene expression after PhrA2 treatment compared to scrambled peptide (P-value<0.01).

TprA2 regulon in the middle ear

We investigated the regulation of the TprA2/PhrA2 system *in vivo* to verify whether our *in vitro* finding were relevant to the *in vivo* environment. We analyzed WT,

Δ*tprA2*, and Δ*tprA2::tprA2*. Three chinchillas were independently inoculated with each strain, middle ear effusions were extracted 48 hours post-inoculation, and bacterial mRNA for *tprA2*, *ABCATPase*, *lcpA* and *lcpM* was quantified using NanoString technology. As observed *in vitro*, deletion of *tprA2* led to increase expression of *ABCATPase* (on the same transcript as *phrA2*) and *lcpM* (**Fig 2.8**). LcpA values were also higher in this mutant, but display elevated inter-animal variability such that the change was not statistically significant. The modest fold increase is consistent with our observation that the TprA2-regulon in the WT is highly expressed *in vivo*, such that complete removal of the negative regulator has a moderate effect. In contrast, overexpression of *tprA2* in the complement strain led to a decrease in the levels of *ABCATPase* and *lcpA*. Together, these findings suggest TprA2 is negative regulator of its neighboring genes *in vivo*.

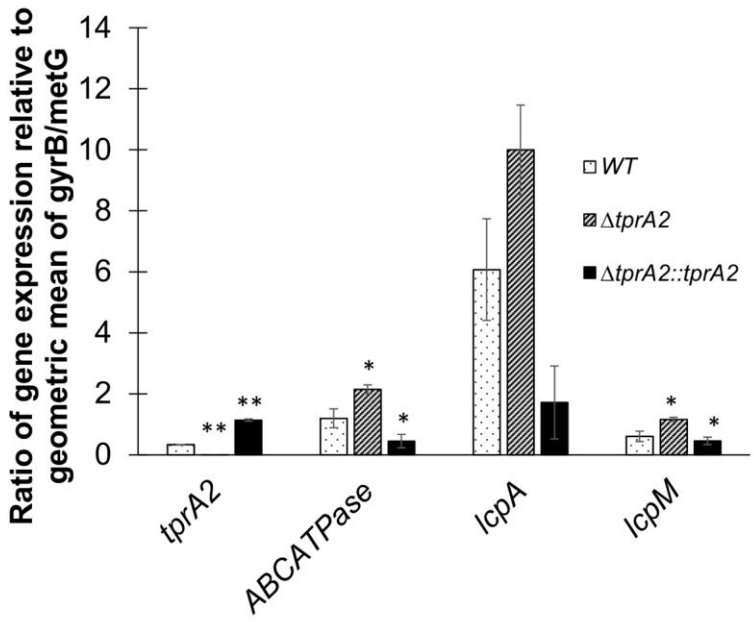


Fig 2.8. Gene expression of TprA2 regulon in the middle ear. Bars represent gene expression as measured by nCounter platform by NanoString technology on RNA extracted from middle ear effusions of chinchillas cohorts (n=3) infected with three different strains: WT (dotted bars), Δ*tprA2* (striped bars), and Δ*tprA2::tprA2* (black bars) individually. The data is represented as ratios relative to the geometric mean of housekeeping genes *gyrB* and *metG* (Y-axis). Target genes are indicated on the X-axis. Error bars represent standard deviations. Statistical significance was determined by Student's t-test and was calculated with reference to WT in each set of test gene; '*', P-value=<0.05; '**', P-value<0.01.

TprA2 promotes commensalism over tissue dissemination

To assess the *in vivo* role of the QS-Lcp region we made use of two pneumococcal infection models. To study colonization of the nasopharynx and spread to the lungs we utilized a murine model where animals are inoculated intranasally and disease progresses causing pneumonia or sepsis or both [102,103]. To study middle ear disease we utilized the chinchilla otitis media model.

The murine model revealed that TprA2 protects against lung disease. We did not observe infection in mice inoculated with PN4595-T23 strains, thus we generated the parallel mutants in another naturally occurring PMEN1 strain with a type 3 capsule (SV36). Cohorts of ten BALB/c mice were infected with SV36, SV36Δ*tprA2* or SV36Δ*phrA2-ABC* and observed over 4 days. The bacterial titers in the nasal lavages were similar for all three strains when tested at 48 hours post-inoculation (**Fig 2.9B**). Notably, SV36Δ*tprA2* displayed a statistically significant increase in mortality (**Fig 2.9A**).

TprA2 is a negative regulator of *lcpAMT* (**Fig 2.6**). To test whether overexpression of *lcpAMT* in the SV36Δ*tprA2* was associated with the increase virulence of this strain, we tested a double mutant with deletions in *tprA2* and *lcpAMT* and observed that it restored the wild-type phenotype. These results strongly suggest that LcpA is a virulence determinant, and that TprA2 can modulate virulence by controlling levels of *lcpAMT*.

Finally, to study middle ear disease, bacteria were inoculated directly into the middle ear of chinchillas. The overall mortality was the same for all three strains, perhaps reflecting differences in peripheral disease progression from the chinchilla middle ear versus the murine nasopharynx (**Fig 2.9C**). Further, we observed a trend toward increased middle ear disease in the Δ*tprA2* (**Fig 2.9D**), and the Δ*tprA2* displayed the highest lung dissemination (**S3 Table** in Appendix I), consistent with our finding that *lcpAMT* plays a role in virulence. In conclusion, our findings suggest that TprA2 controls *lcpA* expression and in doing so can promote commensalism over dissemination.

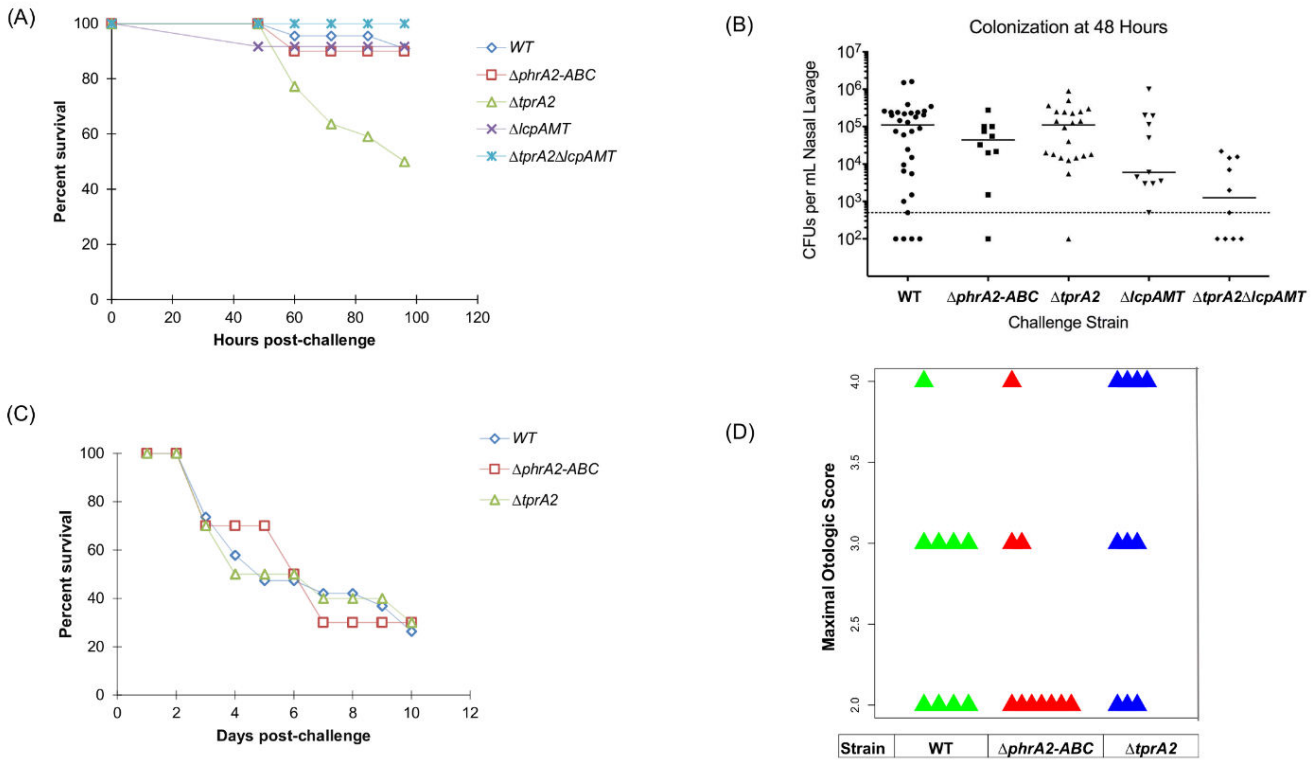


Fig 2.9. *In vivo* effects of TprA2/PhrA2 system. (A,B) Analysis of PMEN1 strain SV36 WT and isogenic mutants Δ*tprA2* ; Δ*phrA2-ABC* ; Δ*lcpAMT*; and Δ*tprA2*/Δ*lcpAMT* in the murine model with intranasal inoculations. (A) Percentage survival of mice after intranasal inoculation. Cohorts of at least ten mice were assessed for the duration of four days. Statistical significance relative to WT was calculated using Mann-Whitney U test; '*', P-value<0.05. (B) Bacterial counts from nasal lavages of mice 48h post-inoculation. (C,D) Analysis of PMEN1 strain (4595-T23) WT and isogenic mutants Δ*tprA2* and Δ*phrA2-ABC* in the chinchilla model of otitis media. (C) Percentage survival of chinchillas after transbullar inoculation. Cohorts of at least ten chinchillas were assessed for the duration of ten days. (D) Scatter plots illustrate the maximal otologic score for animals infected with WT (green), Δ*phrA2-ABC* (red) or Δ*tprA2* (blue). Each triangle represents one animal. Otologic disease ranged from no disease to a ruptured tympanic membrane, where a score of '1' is given for animals with mild or no disease, '2' with moderate disease, '3' with frank purulence, and "4" with tympanic membrane rupture.

Discussion

Our findings demonstrate that acquisition of the TprA2/PhrA2 QS system by horizontal gene transfer into the PMEN1 lineage has endowed these strains with a virulence determinant and a mechanism to regulate its expression and thereby control disease. PMEN1 (ST81) lineage is postulated to have evolved from an ancestor in 1967, and by the end of 1990s it represented an estimated 40% of penicillin resistant strains in US [36,37] . These strains display very high rates of carriage [7,37–39]. PMEN1 also displays very high rates of disease [7,38,40]. Is the prevalence of PMEN1 in invasive disease a function of its carriage rates or does it reflect a propensity to cause disease? Multiple studies have shown that sequence types vary regarding their propensity to cause disease [41,104–106] and Sjostrom *et al*. show that PMEN1 displays a low propensity to cause invasive disease [41]. Thus, high rates of PMEN1 invasive disease in the population likely reflect high carriage rates, and not heightened virulence potential. In this context, it is possible that acquisition of the TprA2/PhrA2 by PMEN1 strains contributes to its low proclivity to cause invasive disease.

Model for the regulation of TprA2/PhrA2 system

Results from the characterization of the TprA2/PhrA2 system led to propose a model for the operation of this system (**Fig. 2.10**). TprA2 represses gene expression by binding to DNA in the promoter region of its target genes, including *phrA2* and *lcpA*. Baseline levels of *phrA2* may be produced and its product secreted via the Sec pathway. When propagation of cells leads to increased density, secreted PhrA2 reaches a

threshold quantity, is internalized into producing and neighboring cells, and binds to TprA2. Our best candidate for the import of PhrA2 into the cell is Opp oligopeptide permease system that is known to internalize small peptides into bacterial cells [74,107]. We propose that binding of TprA2 to PhrA2 leads to a conformational change such that TprA2 undoes its hold on the DNA, similar to the PrgX system of Entercoccus [108], thereby causing release of TprA2 from the DNA. As the promoter becomes free, RNA polymerase may bind to promoter region leading to transcription of target genes. Notably, TprA2 inhibits the expression of *phrA2* and PhrA2 activates the activity of its repressor TprA2, indicative of a complex regulatory circuit.

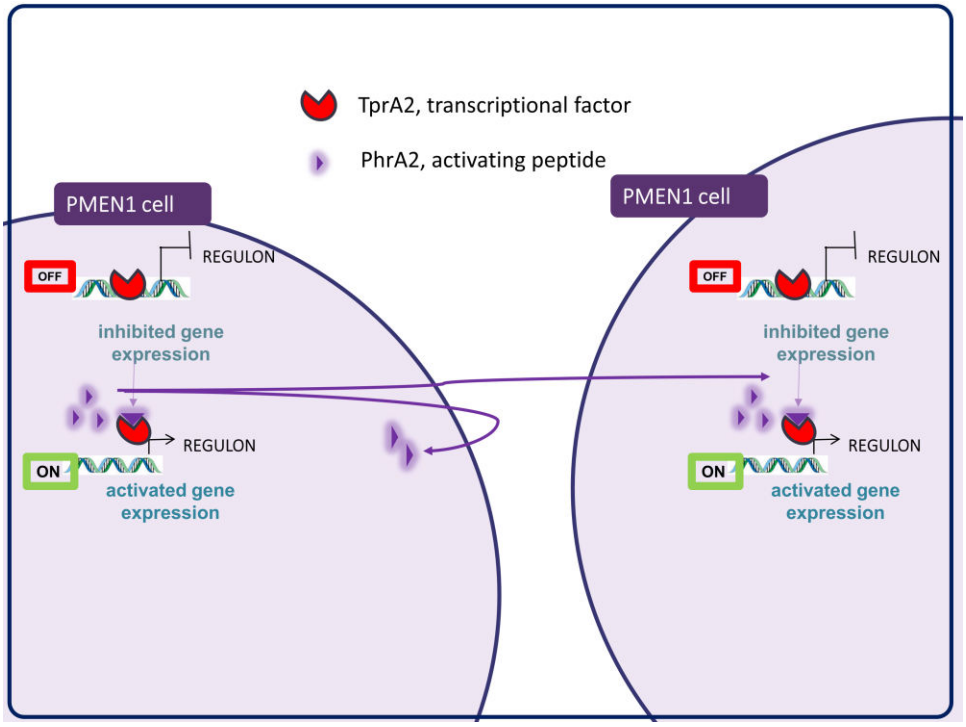


Fig 2.10. Model of the intra-strain action of the TprA2/PhrA2 system

Chapter 3. Distribution and Virulence of TprA/PhrA system

In addition to TprA2/PhrA2, pneumococcus encodes for a second signaling system, TprA/PhrA. Like TprA2/PhrA2, TprA/PhrA is composed of a regulator and its cognate peptide and controls a predicted lantibiotic biosynthesis locus. However, in contrast to TprA2/PhrA2, the TprA/PhrA is widespread across pneumococcal strains.

At the onset of my graduate work, I initiated the characterization of both TprA/PhrA and TprA2/PhrA2 systems in the PMEN1 strains. Hoover and colleagues reported the discovery and regulatory mechanism of the TprA/PhrA system in strain D39 [74]. Following this publication, our work in collaboration with Dr. Hasan Yesilkaya extended the understanding of TprA/PhrA to reveal the fine regulation of TprA in strain D39, and importantly to establish that TprA/PhrA plays a crucial role in virulence. Additionally, in the same manuscript, Dr. Yesilkaya demonstrated that synthetic, neutralizing peptides inhibit PhrA-mediated quorum sensing in pneumococcus; thus providing proof of concept that anti-QS peptides have therapeutic value in Gram positive bacteria. This chapter outlines my contribution to the study of the the pneumococcal TprA/PhrA system.

Citation: Motib A.S., Guerreiro A., Al-Bayati F.A.Y., Piletska E.V., Manzoor I., Shafeeq S., **Kadam A**., Kuipers O.P, Hiller N.L., Piletsky S.A., Andrew P.W., Yesilkaya H. Characterisation and modulation of TprA/PhrA Gram positive quorum sensing system by custom made plastic aptamers. Under consideration at Advanced Materials.

Disclaimer: The study of TprA/PhrA was initiated by Dr. Hasan Yesilkaya and most molecular characterization was carried out by his lab. In this chapter, I present my contribution to this work. First, virulence studies in the chinchilla model of pneumococcal disease that demonstrate a dramatic role for TprA in virulence. Second, phylogenetic studies that suggest TprA has undergone intra-species horizontal gene transfer events.

Abstract of the manuscript

There is an urgent need to develop new antibiotics that are effective, less prone to microbial resistance, and cheap to produce. Quorum-sensing (QS) mechanisms are pivotal for microbial adaptation to host environments, and often required for pathogenesis without affecting bacterial vitality. Hence targeting QS diminish the fitness cost of inhibition, and the emergence and spread of antibiotic resistance. We characterized the TprA/PhrA QS system in the important human pathogen *Streptococcus pneumoniae* with a view to target its operation using novel soluble linear molecularly imprinted polymers (LMIP). We found that TprA/PhrA system is commonly found in pneumococcal strains, and is required for mucin, galactose and mannose utilization. On galactose, TprA is an activator of the virulence determinant neuraminidase (*nanA)*, and TprA controls the expression of nine different operons. TprA expression is modulated by a complex regulatory network, where the master regulators CcpA and GlnR are involved in a sugar dependent manner. Mutants in the

TprA/PhrA system are highly attenuated in the mouse model of pneumonia and septicemia, as well as in the chinchilla model of otitis media, indicating that the TprA/PhrA system is a major virulence determinant and a highly relevant anti-infective target. To interfere with the operation of TprA/PhrA, we used, for the first time, highly homogenous soluble LMIP specific to the PhrA peptide. LMIP decreased PhrA-induction in a dose-dependent and sequence-specific manner, and possessed no visible toxicity in the murine model. Our findings provide proof of principle that LMIP can be used to block Gram-positive quorum-sensing peptides, setting the stage for studies on a novel class of drugs to target Gram positive pathogens.

Virulence studies of TprA/PhrA system

The involvement of the TprA/PhrA system in host derived sugar metabolism and neuraminidase activity led us to hypothesis that this system also plays a role in virulence. To test this hypothesis, we used (i.) a mouse model of pneumonia that develops after intranasal infection, (ii.) systemic infection model initiated by intravenous injection, and (iii.) the chinchilla otitis media model. My contribution was the work in the chinchilla otitis media model.

To test the role of TprA/PhrA in the middle ear, composed of a complex sugar environment [109,110], we employed the chinchilla otitis media model [111,112]. To this end, wild-type D39 and *ΔtprA* strains were transbullarly inoculated (100 CFUs/ear) in cohorts of 10 animals each. Animals were monitored daily for a duration of 10 days

during which time they either succumbed to disease or were euthanized upon reaching extreme morbidity. When compared to wild type, the *ΔtprA* mutant caused significantly less mortality ($p<0.001$). Only one out of ten animals inoculated with wild type survived, while nine out of ten animals inoculated with the mutant survived. Further, virulence was partially restored in animals inoculated with *tprA* complement strain *ΔtprAcom,* where six animals survived (**Fig.3.1**).

Thus, these combined animal studies suggest that the TprA/PhrA system plays an important role in the growth of pneumococcus in the nasopharynx, the middle ear, the lungs, and the blood.

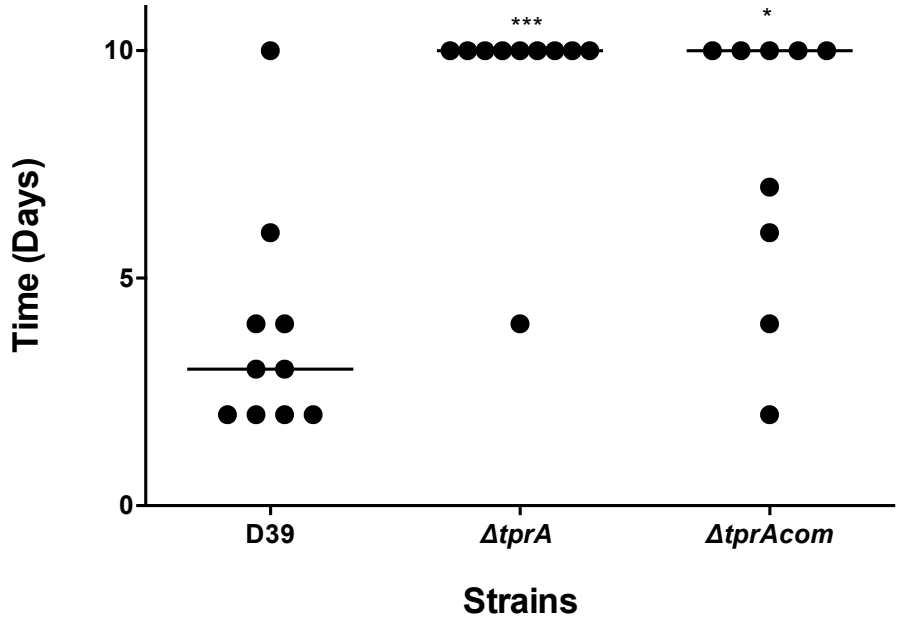


Fig. 3.1: Analysis of pneumococcal strains in chinchilla otitis media model. Chinchillas were injected through ear canal either with D39 (40 CFU/ear), ΔtprA (85 CFU/ear), or ΔtprAcom (24 CFU/ear). Each dot represents survival time of individual mouse. Horizontal line represents median survival time. *p<0.05, and ***P<0.001.

Phylogenetic analysis of the TprA/PhrA system

The majority of pneumococcal strains code for the TprA/PhrA signal transduction system [74] (**Fig 2.2)**. Pneumococcal strains undergo extensive gene transfer, and these transfer events are likely associated with the selective pressure on the variable alleles and/or possession of the gene. Horizontal gene transfer events can be inferred by comparing the topology of a gene tree with that of a species tree using reconciliation [113,114]. Horizontal gene transfer, gene duplications and gene losses produce gene trees with topologies that differ from that of the species tree.

To generate a species tree we utilized fifty-five *Streptococcal* genomes: thirty-five from *S. pneumoniae*, three from *S. pseudopneumoniae,* nine from *S. mitis,* sic from *S. oralis*, and two from *S. infantis* (**Table S1** in Appendix I). PhyML was used to build a maximum likelihood tree from the core genome of these fifty-five isolates (352,371 informative sites out of 995,531 sites). Predicted TprA gene sequences were extracted from these genomes using BLASTn. RaxML was used to generate the TprA gene tree. The tree topologies were re-arranged using Notung-2.8.1.2-beta with 100 bootstrap replicates.

Gene-species tree reconciliation by Notung generated sixteen solutions that represented the least number of changes to reach the solution, and a representative output is displayed in **Fig. 3.2**. In all solutions, TprA was transferred from *S. pseudopneumoniae* into a subset of *S. pneumoniae* isolates that includes D39. Furthermore, all solutions displayed transfer events between the subset of *S. pneumoniae* isolates

consisting of TIGR4, D39, and PMEN1 isolates, but the directionality within this pneumococcal set could not be established. Thus it is likely that *tprA* has incurred both inter- and intra-species transfer events, suggesting the variable alleles maybe under selective pressure.

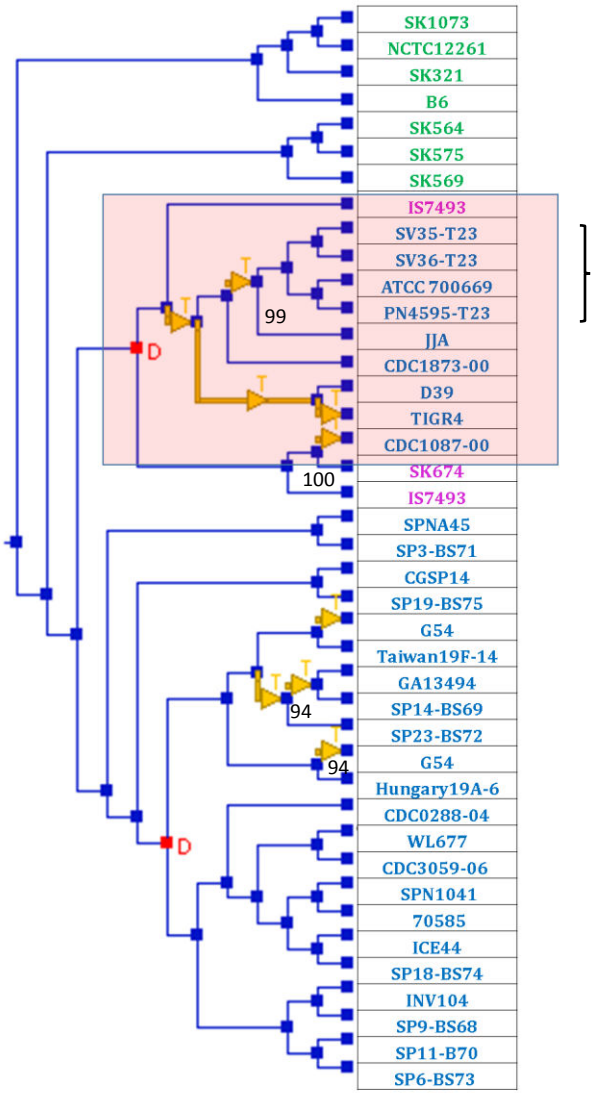


Fig. 3.2: Evidence of intra and inter-species horizontal gene transfer of TprA. The phylogenetic tree is a reconciliation between a Streptococcal species tree and a TprA gene tree. Bootstrap values above 90 are indicated on the tree. Yellow arrows: display predicted gene transfer events; red "D": displays predicted gene duplication events; curly bracket displays PMEN1 strains; red box highlights the transfer event from *S.*

pseudopneumoniae into a subset of *S. pneumoniae* isolates. Species are labeled by color, where green: *S. mitis*, pink: *S. pseudopneumoniae*, and blue: *S. pneumoniae*.

Discussion

This work in the chinchilla model of pneumococcal disease demonstrates that TprA/PhrA is a pneumococcal virulence determinant. When combined with parallel experiments in three different mouse models of pneumococcal disease, our data unambiguously demonstrates that TprA/PhrA is pivotal to disease. In addition, phylogenetic evidence is consistent with extensive horizontal gene transfer of TprA/PhrA across strains, likely due to selective pressure on allelic forms. The next chapter connects TprA2/PhrA2 to the TprA/PhrA system.

Chapter 4. The PMEN1 TprA2/PhrA2 regulates TprA signaling

As described in chapters 2 and 3, TprA2/PhrA2 and TprA/PhrA are regulator-peptide signaling transduction systems that modulate the ability of pneumococcus to cause disease. In this chapter, I describe unilateral signaling from TprA2/PhrA2 to the TprA/PhrA system, thus demonstrating promiscuous signaling by the TprA2/PhrA2 PMEN1-unique genes. Moreover, I show that this cross-talk is not limited to PMEN1 strains, but functions across lineages. This work opens the door to studies that explore the consequences of inter-strain signaling in the outcome of multi-strain infections that include PMEN1 strains. The work presented below corresponds to Kadam *et al.,* 2017 Pages 10-15.

Citation: Kadam A., Eutsey R.A., Rosch J., Miao X., Longwell M., Xu W., Woolford C.A., Hillman T., Yesilkaya H., Mitchell A.P., Hiller N.L. Promiscuous Signaling by a Regulatory System Unique to the Pandemic PMEN1 Pneumococcal Lineage. ***PLoS Pathog.*** **2017**

Disclaimer: I led and performed all aspects of experiments and data analyses of this work with contributions from Rory Eutsey for molecular biology, and Dr. Carol Woolford and Dr. Aaron Mitchell for transcriptional studies using NanoString technology.

PMEN1 Codes for Two Related Regulator/Peptide Systems

TprA2 shares moderate homology to TprA, another streptococcal transcription factor that belongs to the recently characterized TprA/PhrA system, where TprA inhibits expression of PhrA and downstream lantibiotic genes [24] . Unlike *tprA2*, which occurs rarely outside the PMEN1 lineage, *tprA* has a wide distribution in pneumococci. Using a set of highly curated WGSs, with representatives of the major lineages *of S. pneumoniae*, we found that *tprA* was present in over 90% of the isolates in our set (**Fig 2.2**, all tprA genes displayed >= 86% similarity). The prominent exception is a set of strains in a basal pneumococcal branch associated with unencapsulated strains and conjunctivitis infections [115,116] (**Fig 2.2**). Hoover and colleagues first characterized the TprA/PhrA system, and also reported a wide distribution (approximately 60%) in pneumococcal strains [74].

PMEN1 strains are notable in that they code for both the TprA2/PhrA2 and TprA/PhrA QS systems. In the PMEN1 strain PN4595-T23, the TprA and TprA2 protein sequences share approximately 60% identity. We searched the genomes of 55 streptococcal strains, identified 48 sequences to construct a phylogenetic tree of these regulators using maximum likelihood, and found that the *tprA2* and *tprA* homologues are separated into two distinct branches (**Fig 4.1A**). Their cognate peptides in PMEN1, PhrA2 and PhrA share only 28% identity over the full length, but display very high similarity at their C-termini. To analyze the extent of conservation of the C-terminal residues, we generated a consensus logo from the six PhrA2 sequences and the thirty-

six PhrA sequences. The C-terminal residues are either identical or share similar charge in 6/7 residues; but can be distinguished by position -3 that codes for a conserved leucine in PhrA2 and a lysine in PhrA (**Fig 4.1A, 4.1B**). The sequence separation between the QS components suggests that the *tprA2/phrA2* genes did not originate from a recent duplication within PMEN1, and is consistent with acquisition of TprA2/PhrA2 by horizontal gene transfer.

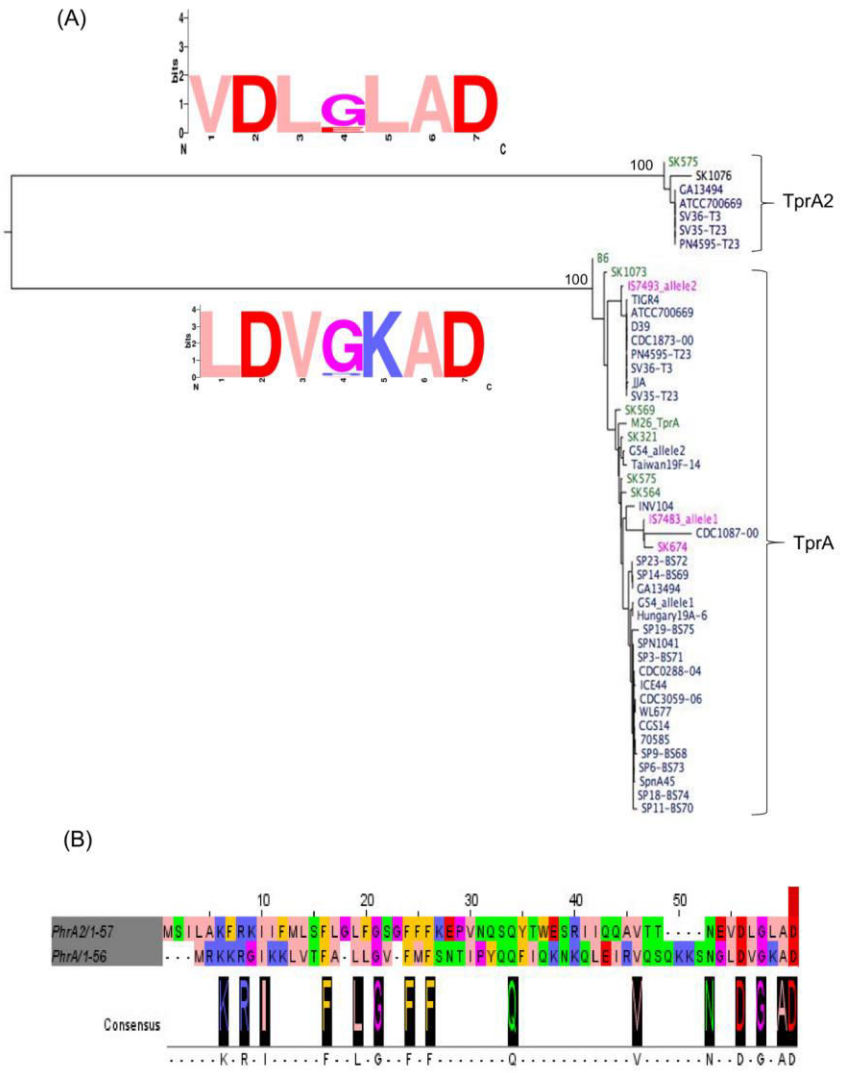


Fig 4.1.Phylogenetic analysis of separation between TprA2 and TprA systems. (A) Gene tree generated from the coding sequences for *tprA* and *tprA2* using maximum likelihood. Each branch displays a sequence logo, derived from the predicted C-terminal heptapeptide of PhrA2 (top) and PhrA (bottom). In the logo, amino acids are represented in one letter abbreviation where their height within the stack represents its relative frequency at a given position, in zappo color-coding scheme: blue/positive; red/negative; salmon/hydrophobic; orange/aromatics; purple/glycine or proline; green/hydrophilic. (B) Alignment of predicted coding sequence of PhrA and PhrA2 in PMEN1 strain PN4595-T23. Representation showing alignment (top) and consensus (bottom). Seven amino acids of the C-termini are highlighted in the red box indicating the sequence of synthetic peptides used in this study.

Interaction of TprA2/PhrA2 QS system with the TprA/PhrA QS system

The co-occurrence of both QS systems in the PMEN1 strains led us to investigate whether PhrA2 and PhrA peptides can exert regulatory effects on their non-cognate QS systems, TprA/PhrA and TprA2/PhrA2 respectively. To test this, we measured how the addition of synthetic peptides to the extracellular milieu affects gene expression of the non-cognate regulon. Addition of synthetic PhrA2 (VDLGLAD), but not the scrambled peptide, induced gene expression of the TprA regulon (*tprA*, *phrA*, and the TprA-associated *lanA*, *lanM*, and *lanT*) at levels similar to those induced by cognate PhrA (LDVGKAD) itself (**Fig 4.2A**). In contrast, neither the addition of synthetic PhrA nor the addition of the scrambled peptide had any effect on expression of the *tprA2*, *phrA2*, or *lcpA* genes in the TprA2/PhrA2 regulon (**Fig 4.2B**). These findings suggest that PhrA2 regulates gene expression of the TprA regulon, and PhrA has no effect on the TprA2 regulon.

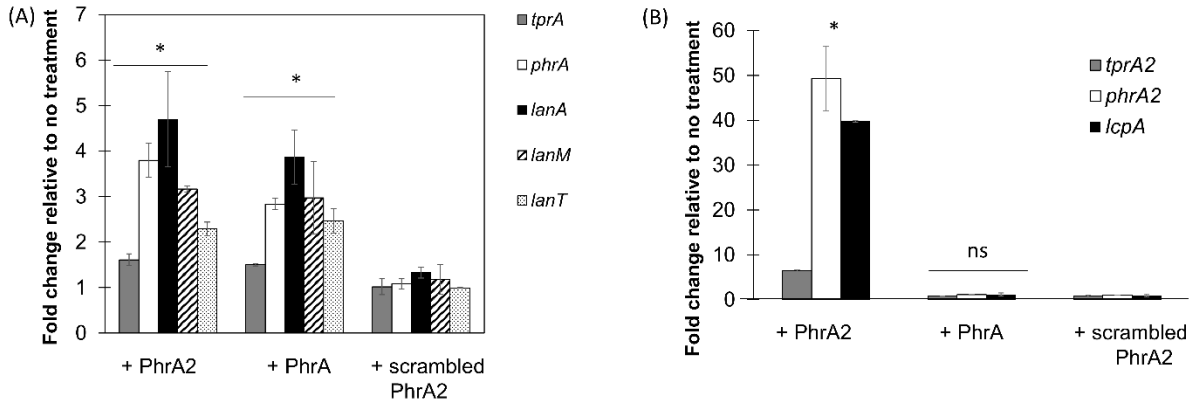


Fig 4.2. PhrA2 influences the gene expression levels of the TprA/PhrA system. qRT-PCR measurements of gene expression for target genes performed in strain PN4595-T23. Data was normalized to levels of 16S rRNA. The X-axis denotes test genes of TprA/PhrA system and treatment conditions. The Y-axis reflects the fold change in the treatment group relative to the no treatment control. Treatments correspond to: (i) PhrA2 C-terminal heptapeptide (VDLGLAD); (ii) PhrA C-terminal heptapeptide (LDVGKAD); or (iii) scrambled peptide (DAGVLDL). Error bars represent standard deviations for biological replicates (n=3). (A) Target genes correspond to *tprA* regulator (gray bar), its cognate *phrA* peptide (white bar), and lantibiotic genes in the TprA regulon (*lanA*/dark bar; *lanM* stripped bar; and *lanT*/dotted bar). * Statistically significant difference in gene expression compared to scrambled peptide (P-value<0.05). (B) Target genes correspond to *tprA2* regulator (gray bar), its cognate *phrA2* peptide (white bar), and *lcpA* /dark bar. * Statistically significant difference in gene expression compared to scrambled peptide (P-value<0.01, ns= not significant).

PhrA2 regulates the TprA/PhrA system in non-PMEN1 strains

The unidirectional influence of PhrA2 gene expression upon TprA/PhrA led us to investigate whether the PMEN1 peptide could influence gene expression in non-PMEN1 cells. We used strain D39 as a representative of the non-PMEN1 strains since TprA/PhrA system has been previously described in D39. Hoover *et al*. have demonstrated that *phrA* is under catabolite repression.

The gene encoding *phrA* is expressed in galactose and repressed in glucose, and the *phrA* promoter region contains a *cre* (catabolite response element) site for CcpA catabolite repression [74,117]. In contrast, we have not identified a *cre* site in the *phrA2* promoter region. Therefore, to maximally discern the input through PhrA2 in our experiment, we used a D39-derived strain with a deletion of *phrA* and grew it in chemically-defined medium with galactose as the sole sugar.

We found that exogenous PhrA2 interacts with the TprA regulon in non-PMEN1 strains. Specifically, D39Δ*phrA* cultures were exposed to treatments with synthetic PhrA2, PhrA, and scrambled peptides for an hour and gene expression of *tprA* and *lanA* was measured relative to no treatment. Treatment with PhrA2 significantly induced expression of *tprA* and *lanA* by 11-fold and 2-fold, respectively (**Fig 4.3**). Treatment with scrambled peptide showed no induction of gene expression in D39Δ*phrA*. The extent of LanA induction by PhrA is lower in the D39ΔphrA strain than in experiments with the WT strain (**Fig. 4.2A**), we presume this difference is due to the absence of *phrA*-autoinduction in the mutant strain. These findings suggests that PhrA2 can be

internalized by strains outside the PMEN1 lineage and induce changes in their gene expression.

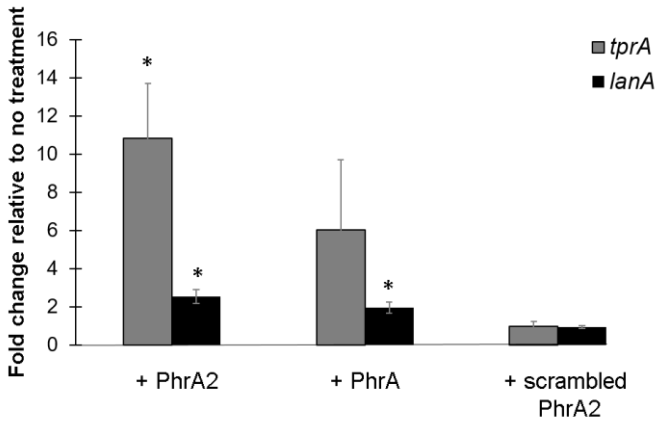


Fig 4.3. PhrA2 influences the gene expression levels of the TprA/PhrA system in non-PMEN1 strain D39. qRT-PCR measurements of gene expression for target genes performed in strain D39Δ*phrA* upon treatments indicated on the X-axis. Data was normalized to levels of 16S rRNA. The Y-axis reflects the fold change in the treatment group relative to the no treatment control. Treatments correspond to: (i) PhrA2 C-terminal heptapeptide (VDLGLAD); (ii) PhrA C-terminal heptapeptide (LDVGKAD); or (iii) scrambled peptide (DAGVLDL). Target genes correspond to *tprA* regulator (gray bar), and its associated *lanA* gene (black bar). Error bars represent standard deviations for biological replicates (n=3), * Statistically significant difference in gene expression compared to scrambled peptide (P-value<0.01).

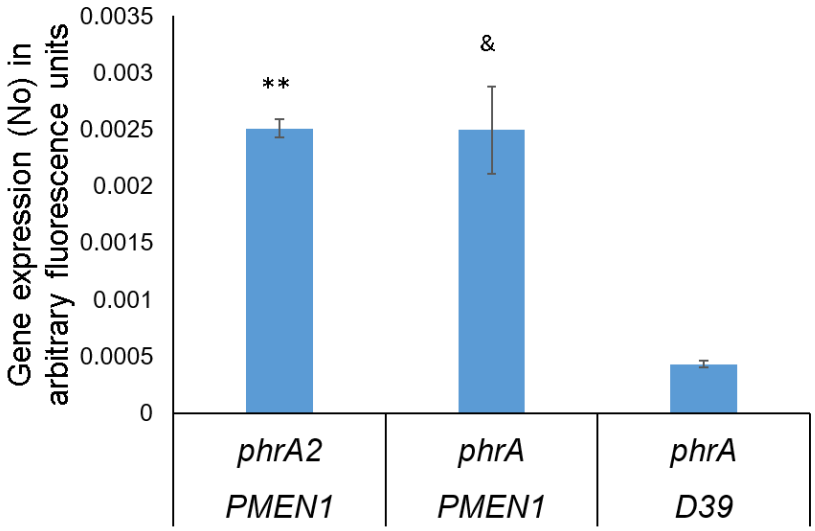


Fig 4.4. *In vitro* condition where expression of PMEN1-*phrA*2 and PMEN1-*phrA* is higher than that of D39-*phrA*. qRT-PCR measurement of cultures of PMEN1 and D39 grown independently in rich media (Columbia broth) to mid-log phase (n=2).Statistical tests for gene expression: '**' *P*-value=0.006 and '&' *P*-value=0.057.

Discussion

TprA2/PhrA2 may provide PMEN1 strains with the means to manipulate gene expression in neighboring strains from other lineages in multi-strain infections. We show that synthetic C-terminal PhrA2 can stimulate expression of the TprA/PhrA system as well as its associated lantibiotic biosynthesis cluster in distantly related strain D39 (**Fig 4.2, 4.3**). We have observed that the expression of PMEN1-*phrA2* is six fold that of D39-*phrA* in rich media, thus exemplifying a condition where PMEN1-*phrA2* expression is high when D39-phrA is low (**Fig 4.4**). We are currently investigating this interaction in physiologically relevant conditions. The activation of *phrA* in response to galactose has led to the conclusion that TprA/PhrA may promote colonization in the nasopharynx where free sugars are rare and pneumococci survive by breaking down host mucins to free complex sugars, most prominently galactose [74]. However, experiments with TprA/PhrA in the murine model demonstrate that this system is a virulence determinant in multiple models of pneumococcal disease (personal communication, Motib and Yesilkaya, Chapter 3), in this manner, PhrA2 may trigger a virulence regulon in neighboring strains. We propose that PhrA2 signaling across systems is physiologically relevant in multi-strain infections.

We conclude that PhrA2 peptide is secreted by PMEN1-cells, since cell-free culture supernatants reiterate the function of extracellular addition of synthetic PhrA2. We predict that export occurs via the Sec secretion system, consistent with other peptides from the PlcR family of regulator-peptide pairs [77,118,119]. Import must

occur via a relatively widespread transporter, given that PhrA2 can influence D39 gene expression. Further, the high sequence similarity between the functional C-termini of PhrA and PhrA2 suggests common import machinery. The oligopeptide permease *amiACDEF* has been shown to be required for import of processed PhrA, and its homologues are required for import of PlcR-associated peptides in other species [77,118,119]. Thus, *amiACDEF* is a high value candidate for a PhrA2 importer.

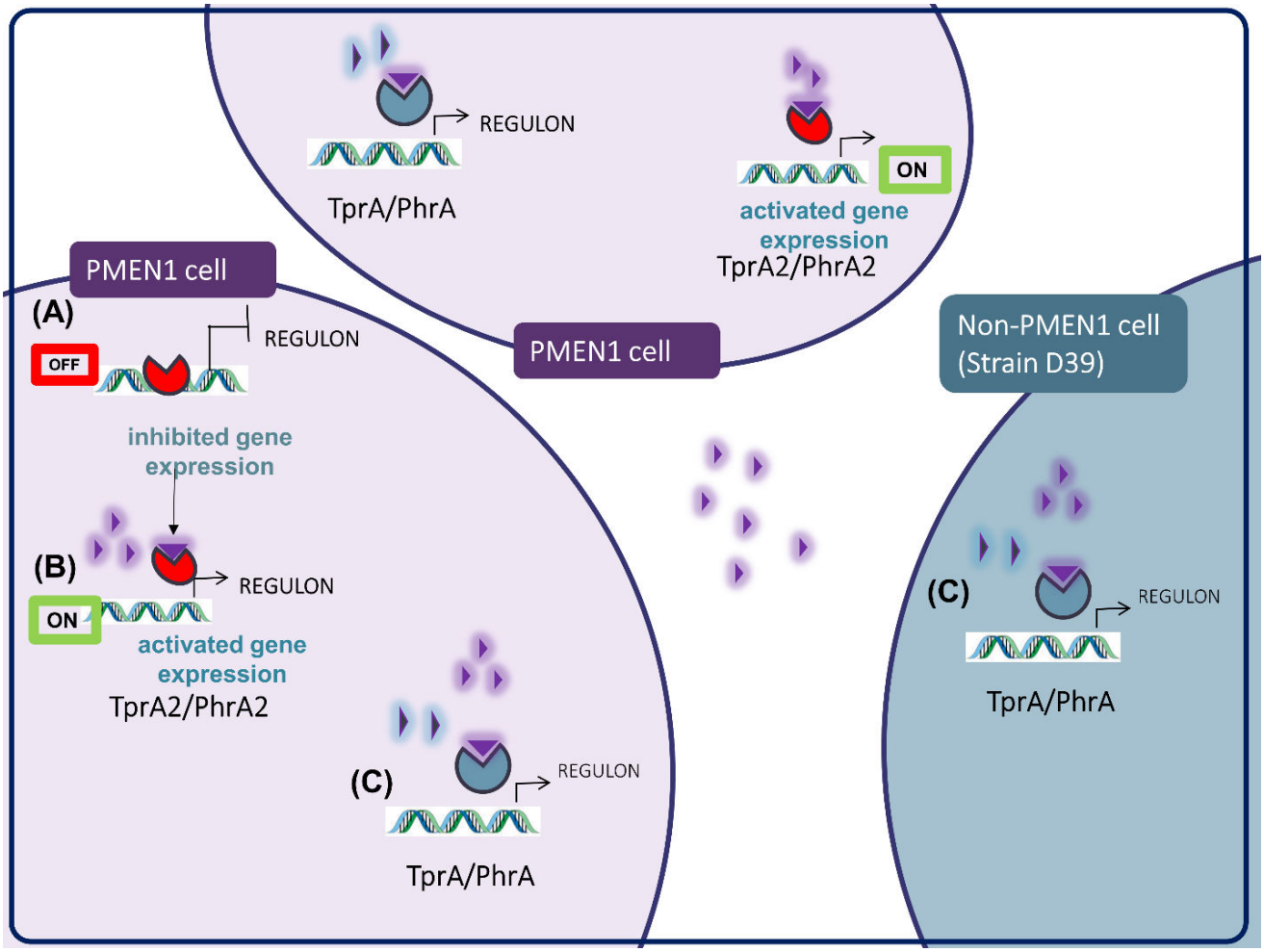


Fig 4.5. Model for regulation of gene expression by TprA2-PhrA2 in intra- and inter-strain infections. (A) In the OFF state, TprA2 inhibits gene expression. (B) In the ON state, PhrA2 releases TprA2-mediated gene inhibition. This effect of PhrA2 is observed from synthetic peptide added to the extracellular milieu and cell-free supernatant, suggesting that PhrA2 is exported, activated and re-imported before it modulates TprA2 activity, in both the producer PMEN1 cells and surrounding PMEN1 population. (C) PhrA2 secreted by PMEN1 cells activates gene expression of tprA and associated lanA, in both PMEN1 and non-PMEN1 cells. Red circular shape/TprA2, purple triangle/PhrA2, blue circular shape/TprA; blue triangle/PhrA.

Chapter 5. Functional characterization of LcpA, a PMEN1-unique lanthionine peptide

Lanthionine-containing peptides of Gram-positive bacteria

Lanthionine-containing peptides or lantipeptides are peptides produced by a number of Gram positive bacteria. These are heavily post-translationally modified and include unusual cross-linked amino acids with a thioether linkage such as lanthionine, 3-methyllanthionine, dehydroalanine, or dehydrobutyrine [120]. They are ribosomally synthesized and the genes involved in their biosynthesis are usually clustered together on the genome. A typical lantipeptide biosynthesis cluster includes structural genes for a precursor peptide, synthetases that catalyze the maturation of the precursor, transporters to export the lantipeptide and immunity proteins to shield the producer cell [121,122]. The precursor peptide includes an N-terminal leader sequence, believed to harbor a signal for export and to maintain the peptide inactive until export [122,123]. Synthetases process the precursor peptide by dehydration of serine or threonine, followed by cyclization to produce the lanthionine rings. LanM type synthetases are bifunctional and perform both dehydration and cyclization reactions (designated as Group II lantibiotics) [123,124]. Once the leader peptide is cleaved by the transporter, the mature peptide is secreted out of the cell or, in some cases, remains attached to the cell wall [122,125].

Most lanthionine-containing peptides have been characterized for the antimicrobial activity and termed lantibiotics. Lantibiotics have been isolated from

various lactic acid bacteria. Nisin from *Lactococcus lactis* was the first to be discovered in 1928 [126] . Other examples include, lacticin 481, mersacidin from Bacillus sp., subtilisin from *Bacillus subtilis* and lichenicidin from *Bacillus licheniformis* [127–129]. Many lantibotics are inhibitory via targeting of the lipidII component of the peptidoglycan cell wall [96,130–132]. In contrast to most antibiotics, there are no reports of acquired resistance to lantibiotics Thus, lantibiotics are promising candidates for future generations of antibacterial drugs.

Lanthionine containing peptide also have additional functions. For example, SalY from *Streptococcus pyogenes* is required for full virulence and survival inside macrophages [133]. Nisin, subtilin, and sublancin, play a role in prevention of spore outgrowth in Bacillus and Clostridium species [134–136]. SapT is involved in aerial hyphae formation in *Streptomyces* [137]. Futhermore, some lantipeptides are signaling molecules. Nisin acts on the TCS NisRK and *Streptococcus mutans Smb* binds to LsrS [138]. In summary, lantibiotics are important players in microbiological warfare and signaling, and have both medical and commercial applications.

Regulation of LcpA in the PMEN1 strains

As described in Chapter 1, PMEN1 strains code for a lantipeptide that is uniquely present in these strains. TprA2 is the negative regulator of *lcpA*. In the presence of the activating peptide PhrA2, TprA2 inhibition is released leading to induction of *lcpA*. Using a mouse model of pneumococcal disease, we show that *lcpA* is

a virulence factor. In this way, LcpA is an exciting example of a lantipeptide playing an important role in pathogenesis.

Investigation into the function of LcpA

To identify an *in vitro* condition where *lcpA* expression is high, we performed a microarray screen comparing the gene expression profile of the wildtype strain PN4595-T23 in two media condition- rich media (Columbia broth) and chemically defined medium that containing 55mM of glucose as the sole sugar (CDM-Glucose). We found that the expression of *lcpA* was 62 fold higher in CDM glucose when compared to that in rich media (**Table 5.1**). Thus, CDM-glucose media represents a suitable condition for *lcpA* studies.

Table 5.1. Comparison of gene expression for *lcpA* and regulatory genes in Columbia broth and CDM supplemented with glucose.

Gene	**Ratio of gene expression CDM-Glucose/Columbia**	***P*-value**
lcpA	61.96	2.1E-04
ABC permease	56.18	3.46E-06
lcpM	32.23	4.28E-05
ABC ATPase	31.68	5.55E-07
lcpT	30.93	1.7E-04
phrA2	30.39	8.81E-08
ABC permease	21.76	3.63E-08
tprA2	7.98	7.23E-05

To investigate the function of *lcpA*, we performed RNA-Seq comparing the expression profiles of wildtype PN4595-T23 and isogenic Δ*lcpA,* under CDM-glucose growth conditions (n=1). The screen revealed putative differentially expressed genes indicating a regulatory role of *lcpA* (**Table 5.2**).

The differentially expressed genes are organized into a few genomic regions. Genes predicted to be upregulated in the presence of LcpA include: (1) the genes upstream of *lcpA*, including *tprA2* and *the phrA2-ABC operon*, consistent with a positive feedback between *lcpA* and tprA2/phrA2 levels. (2) an operon of six genes annotated to participate in nuleic acid production. (3) an operon with the components of a phosphate transport system. Gene predicted to be downregulated in the presence of *lcpA* include a locus with a hypothetical and a SdrI protein, containing serine-aspartate repeats suggesting it is exposed on the cell surface.

Three of these genes, SdrI and a representative of the phosphate transport operon and the purine biosynthesis operon were selected for validation. To this end, we measured their expression levels in wildtype and Δ*lcpA* using qRT-PCR and the result reflect those observed in the RNA Seq (highlighted in **Table 5.2**). We conclude that the level of LcpA is linked to expression of multiple pneumococcal loci.

Table 5.2 Differentially regulated gene in RNA-Seq of PN4595-T23 vs Δ*lcpA*. Gray and white highlight gene in the same genomic region.

RAST peg number	Annotation	Fold change lcpAKO/WT
fig|1313.153.peg.1272	putative ABC transporter, permease protein	0.06
fig|1313.153.peg.1273	hypothetical protein	0.06
fig|1313.153.peg.1274	ABC transporter, ATP-binding protein	0.06
fig|1313.153.peg.1275	Signaling peptide, PhrA2	0.06
fig|1313.153.peg.1276	Transcriptional regulator, TprA2	0.17
fig|1313.153.peg.1287	TrsE-like protein	0.13
fig|1313.153.peg.1288	FIG01116986: hypothetical protein	0.15
fig|1313.153.peg.1289	Tn5252, Orf23	0.20
fig|1313.153.peg.1290	FIG01114970: hypothetical protein	0.15
fig|1313.153.peg.1292	FIG01114872: hypothetical protein	0.23
fig|1313.153.peg.1296	FIG00627241: hypothetical protein	0.18
fig|1313.153.peg.136	Asparagine synthetase [glutamine-hydrolyzing] (EC 6.3.5.4)	2.05
fig|1313.153.peg.1691	FIG01114788: hypothetical protein	2.30
fig|1313.153.peg.1700	Conserved domain protein	3.83
fig|1313.153.peg.1701	ABC transporter, ATP-binding protein	4.19
fig|1313.153.peg.1702	FIG01114494: hypothetical protein	4.74
fig|1313.153.peg.1703	FIG01114528: hypothetical protein	4.64
fig|1313.153.peg.1704	FIG01115614: hypothetical protein	5.55
fig|1313.153.peg.1722	FIG01113976: hypothetical protein	0.14
fig|1313.153.peg.1781	hypothetical protein	2.50
fig|1313.153.peg.1782	**Surface protein SdrI**	**2.73**
fig|1313.153.peg.18	Cell wall-associated murein hydrolase LytA	0.01
fig|1313.153.peg.1864	Xanthine phosphoribosyltransferase (EC 2.4.2.22)	0.13
fig|1313.153.peg.1865	Xanthine permease	0.11
fig|1313.153.peg.2053	L-xylulose 5-phosphate 3-epimerase (EC 5.1.3.-)	2.10
fig|1313.153.peg.2105	Phosphate ABC transporter, periplasmic phosphate-binding protein PstS (TC 3.A.1.7.1)	0.20
fig|1313.153.peg.2106	Phosphate transport system permease protein PstC (TC 3.A.1.7.1)	0.19
fig|1313.153.peg.2107	**Phosphate transport system permease protein PstA (TC 3.A.1.7.1)**	**0.19**
fig|1313.153.peg.2108	Phosphate transport ATP-binding protein PstB (TC 3.A.1.7.1)	0.17
fig|1313.153.peg.2109	Phosphate transport system regulatory protein PhoU	0.18

fig|1313.153.peg.42	Phosphoribosylaminoimidazole-succinocarboxamide synthase (EC 6.3.2.6)	0.18
fig|1313.153.peg.43	Phosphoribosylformylglycinamidine synthase, synthetase subunit (EC 6.3.5.3) / Phosphoribosylformylglycinamidine synthase, glutamine amidotransferase subunit (EC 6.3.5.3)	0.18
fig|1313.153.peg.44	Amidophosphoribosyltransferase (EC 2.4.2.14)	0.19
fig|1313.153.peg.45	**Phosphoribosylformylglycinamidine cyclo-ligase (EC 6.3.3.1)**	0.19
fig|1313.153.peg.46	Phosphoribosylglycinamide formyltransferase (EC 2.1.2.2)	0.20
fig|1313.153.peg.47	Teicoplanin resistance protein	0.21
fig|1313.153.peg.664	6-phospho-beta-glucosidase (EC 3.2.1.86)	0.22
fig|1313.153.peg.91	Putative uncharacterized protein spr0086	2.01
fig|1313.153.peg.92	FIG015389: hypothetical membrane associated protein	2.10
fig|1313.153.peg.93	FIG04612: Integral membrane protein (putative)	2.10
fig|1313.153.peg.94	FIG014387: Transcriptional regulator, PadR family	2.15

Future directions

The differentially regulated loci contained genes belonging to surface-exposed proteins. This finding led us to hypothesize that LcpA may function as a signaling molecule by binding to a receptor on the pneumococcal surface. Validation of the RNA-Seq results by performing qRT-PCR will shed further light on the regulatory role of *lcpA*. Through RNA-Seq data, LcpA appears to regulate TprA2/PhrA2 system (*tprA2*, *phrA2* and ABC transporters) suggesting an additional circuit in the regulation of this quorum sensing system. Some other differentially regulated genes play a role in nucleotide synthesis. One of the highest fold changes was observed in cell-wall murein hydrolase, lytA, which is responsible for autolysis in pneumococcal cells. *SdrI* is a notable gene in the list of genes differentially regulated in the absence of *lcpA*. It belongs

to the family of serine-aspartate repeat proteins that have been shown to play a role during host response. While not characterized in pneumococcus, in the Staphylococcus genus, different members of Sdr protein family play a role in bacterial survival in blood, adhesion to host molecules such as collagen and host clumping factor [139–141].

Our data shows that LcpA levels are linked to increased survival of bacteria in the lung and increased death. One of my hypotheses is that changes in SdrI levels may play a role in virulence. Specifically, *lcpA* expression leads to decreased in Sdrl. This could in turn lead to a release of bacteria from the nasopharynx and increase dissemination into the lungs. Alternatively or in addition, LcpA may function as a lantibiotic and exhibit bacteriocidal activity. In conclusion, our findings suggest that LcpA is a virulence determinant associated with changes in gene expression and sets the stages for exciting new hypotheses regarding its molecular function.

Chapter 6. Development of technology

In the course of my graduate studies I have been involved in the improvement of techniques to assay genomic and transcriptomic variability. In this chapter, I describe the first adaptation of NanoString technology for screening pneumococcal gene expression from animal tissue, the development of a *Streptococcal* pangenome microarray, as well as provide a description of all the techniques used in this work.

The challenges of pneumococcal gene expression studies

I have contributed to the development of two transcriptomic techniques motivated by the goal to overcome two challenges: (1) measurement of gene expression when the genes of interest corresponds to a low percentage of total transcripts, and (2) measurement of bacterial gene expression in the context of high genomic variability.

NanoString technology was adapted for highly sensitive gene expression profiling *in vivo*. Measurement of bacterial gene expression *in vivo* has been a challenge since the pathogen's mRNA amounts to a very small fraction of the host's RNA that is extracted during isolation. An alternative technology is RNAseq, where only the most highly expressed molecules can be interrogated; RNAseq of pneumococcus in the context of the host has been reported in cell line infections but not in animal models with active infection [142,143]. Our NanoString experiment on pneumococcus displayed a range of probe signals that spaned five orders of magnitude (0 to 200,000 counts). Further, the signal to noise ratio was between 10 and 10,000, when using effusions from

animals that cleared infection (CFU<1/ml after 48h) as a control. In addition, unlike I contrast to qRT-PCR and RNA-Seq, the NanoString does not require preparation of cDNA that involve multiple steps of enzymatic processing and lead to loss of sample and quality. Thus, we find that the NanoString is a sensitive technique to measure pneumococcal gene expression *in vivo*.

The pangenome array (SpSGH) was designed to assess the genomic/transcriptomic content of pneumococcal strains without prior knowledge of their genomic sequence. As described in Chapter 1 section 1.4, the genomic variability among pneumococcal strains is very high. Previous arrays are designed to one or a best a few model strains, thus they are likely to miss 10-30% of genes when other pneumococcal strains are used, due to differences in gene content and allelic variability. The pangenome array was designed to overcome this challenge. It captures most genes in the species, as the probes were designed to the majority of genes in the pangenome. It captures allelic variation, as probes were designed to the conserved regions of each gene, and when necessary probes were designed to multiple alleles.

6.1 NanoString technology for *in vivo* gene expression

nCounter Analysis System from NanoString technology provides a highly sensitive platform to measure gene expression of a pathogen during host infection [144]. The fully-automated, barcode technology directly detects mRNA transcripts, thereby

eliminating the amplification and enzymatic steps of DNase treatment and cDNA synthesis. The probes are custom designed by NanoString Technologies.

During probe design, five genes were added to the probe set as putative normalization controls: *gapdh*, *lytA, gyrB, metG* and *rpoB.* These were selected based on previous reports of pneumococcal gene expression *in vitro* [117,145,146]. Our goal was to identify genes with expression levels that remain within a stable range under both *in vivo* and *in vitro* conditions. In addition, together the controls must display a wide range of *in vivo* expression counts. After comparing *in vitro* and *in vivo* expression in multiple conditions, we selected *gyrB* and *metG*. Normalization with the geometric mean of these two transcripts provided consistent results in both *in vitro* and *in vivo* samples, and for molecules with both high and low counts relative to the average number of counts. Importantly, for probe sets with a high number of probes normalization with *gyrB, metG* resembles normalization using total counts. We discarded *gapdh* and *lytA*, because we find that they are high variable during *in vivo* conditions and therefore are not suitable for normalization across *in vitro* and *in vivo* samples.

For NanoString experiment, 5µL of total extracted RNA sample was isolated directly from processing of middle ear effusions or mouse lung tissue with the RNeasy Mini Kit and hybridized onto the nCounter chip following manufacture's instruction. No step for enrichment of microbial RNA was necessary. RNA concentration ranged from 80-200 ng/µl for *in vivo* samples, and 50ng total nucleic acid for planktonic samples. Manufacturer's software, nSolver, was used for quality assessment of the raw

data and normalization. The data was normalized across samples against the geometric mean of the housekeeping genes, *gyrB* and *metG* [117,146]. Finally, the *in vitro* and *in vivo* levels were compared using Student's t-test in the GraphPad Prism 6 tool.

This technology was applied in Chapter 2 to demonstrate high levels of *phrA2*, *phrA*, and *lcpA in vivo*. Subsequently, it was used in the laboratory to screen for peptides highly expressed *in vivo*, serving as a foundation for multiple ongoing projects. Here I highlight one:

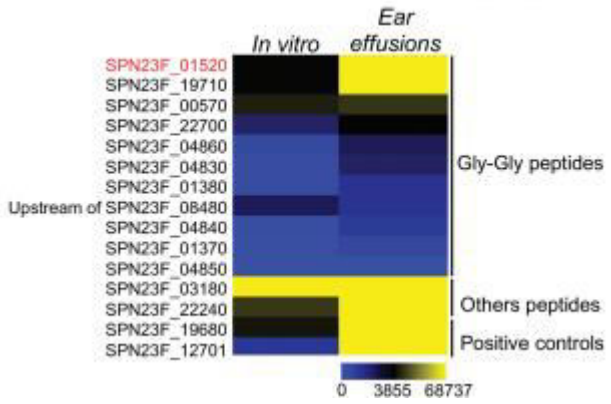


Fig 6.1 Analysis of peptide expression during host infection. NanoString RNA count averages for the set of predicted secreted peptides in PMEN1 strain PN4595-T23 in mid-log planktonic cultures (left column) and effusions extracted from the chinchilla middle ear 48 h post-inoculation (right column). As positive controls, we utilized the quorum sensing peptide encoding gene *phrA*, and the predicted lanthiopeptides *lanA* and *lcpA* genes, which are highly expressed during *in vivo* expression. 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 gene on the genome. ID's are represented in the PMEN1 strain ATCC 700669 (GenBank FM211187).

Citation: Cuevas R.A, Eutsey R.A., **Kadam A**,. West-Roberts J, Woolford C.A., Mitchell A.P., Mason K., Hiller N.L. A novel streptococcal cell-cell communication peptide promotes pneumococcal virulence and biofilm formation. ***Mol. Microbiol.*** **2017**

6.2 Design and Implementation of the *Streptococcus pneumoniae* Supragenome Hybridization Array for Profiling of Genetic Content and Gene Expression.

Citation **Kadam A.**, Janto B., Eutsey R.A., Earl J.P., Powell E., Dahlgren M.E., Hu F.Z., Ehrlich G.D., Hiller N.L., *Streptococcus pneumoniae* Supragenome Hybridization Arrays for Profiling of Genetic Content and Gene Expression. ***Curr Protoc Microbiol.*** **2015**

Disclaimer: The compilation of the pneumococcal pangenome, its analyses, and the probe design were performed by Dr. Luisa Hiller, Dr. Benjamin Janto, Dr. Garth Ehrlich, Dr. Fen Hu, Margaret Dahlgren, Evan Powell, and Josh Earl. Together with Dr. Benjamin Janto and Rory Eutsey, I validated the array with multiple strains, and extended its use from genomic analysis to transcriptomic analysis.

Abstract of the manuscript

There is extensive genomic diversity among *Streptococcus pneumoniae* isolates. Approximately half of the comprehensive set of genes in the species (the supragenome or pangenome) is present in all the isolates (core set), and the remaining is unevenly distributed among strains (distributed set). The *Streptococcus pneumoniae* Supragenome Hybridization (SpSGH) array provides coverage for an extensive set of genes and polymorphisms encountered within this species, capturing this genomic diversity. Further, the capture is quantitative. In this manner, the SpSGH array allows for both genomic and transcriptomic analyses of diverse *S. pneumoniae* isolates on a single platform. In this unit, we present the SpSGH array, and describe in detail its design and implementation for both genomic and transcriptomic analyses. The methodology can be

applied to construction and modification of SpSGH array platforms, as well as applied to other bacterial species as long as multiple whole genome sequences are available that collectively capture the vast majority of the species supragenome.

Introduction

In many bacterial species, isolates differ from one another by extensive genomic variability [43,45,85,147–154]. This variability is observed as single nucleotide polymorphisms (allelic differences), as well as extensive differences in gene possession where a percentage of the genes are shared across all strains (core set), and the remainder are unevenly distributed across isolates (distributed/accessory/variable set). The comprehensive set of all the genes across all the strains of the species is referred to as the pangenome or supragenome [149,151,155]. This variability in gene possession is an important factor in determining the broad array of phenotypes displayed by various isolates with respect to disease, as well as drug and vaccine resistance [48,84,156–159].

We have developed supragenome hybridization (SGH) arrays to study the differences in gene content and gene expression for *Haemophilus influenzae* [160,161] and *Streptococcus pneumoniae* (pneumococcus), two opportunistic pathogens that colonize the human nasopharynx. For both of these species, ~50% of the supragenome is core, and strain pairs often differ by ~20% of their genomic content [43,85,148,162]. Thus, in these species, when an array is designed to a reference strain, the analysis of a distantly related isolate will be hampered due to the loss of information associated with

the lack of probes that can capture highly variable alleles or genes that were absent in the reference strain. The SGH array is designed to capture the diversity of the species by providing coverage of multiple alleles and most genes in the supragenome; and in doing so, allow for the analysis of diverse isolates on the same platform.

The *H. influenzae* SGH array (HiSGH) was designed based on a supragenome (pangenome) analysis of 24 clinical *H. influenzae* strains. The array contains 31,307 probes that collectively cover 2,890 *H. influenzae* genes, corresponding to greater than 85% of all non-rare genes (that is, those present in 10% or more of isolates). This array has been used to investigate the gene content of a library of isolates [160]; as well as to measure transcriptomic differences between a wild-type (WT) strain and an associated deletion mutant [161]. For genome content studies, the HiSGH array accuracy was shown to be ~ 98% by comparing whole genome sequence (WGS) of eight strains with their hybridization data obtained using the HiSGH array. Once tested, the array was used to investigate the gene content of 193 geographically and clinically diverse *H. influenzae* clinical strains [160]. In transcriptomic studies, the HiSGH aray was used to compare transcripts levels between a WT strain and the cognate AI-2 sensing mutant. The strains were grown in multiple conditions using different media and sampling time points. Additionally, technical and biological replicates were analyzed for reproducibility. The results were highly reproducible, and the differentially expressed genes were confirmed by quantitative reverse transcription polymerase chain reaction (qRTPCR) [161].

In this unit, we describe the development and testing of an SGH array, and use the SpSGH array as an example. Each section has a general description, followed by details from the SpSGH array. The unit is divided into three major sections: I) probe design; II) analysis of genomic content; and III) analysis of transcriptomic content. These are organized into the following basic protocols: 1) probe design for the SpSGH Array; 2) SpSGH Array to determine gene content; 3) Data Analysis of SpSGH Array to determine gene content; 4) SpSGH Array for gene expression profiling; and 5) Data Analysis of SpSGH Array for gene expression profiling. Development of an SGH array can assist in understanding the finer genomic and transcriptomic differences contributing to diverse phenotypes with respect to disease and carriage of historical, present day as well as emerging pneumococcal strains. In addition, it provides a holistic view to elucidate gene regulatory networks differentially regulated in selected *in vitro* conditions.

BASIC PROTOCOL 1: Probe Design for SpSGH Array

The goal of probe design is to generate DNA probes that recognize most genes in the supragenome, but which do not cross hybridize with paralogous genes. Probe design is a multistep process that requires: 1) preparation of sequence input by selection and annotation of WGS; 2) organization of genes into homologous clusters and then into allelic groups; and finally 3) selection of sequences for probe design and manufacturing. A schematic of SGH probe design is provided in **Fig 6.2.1**.

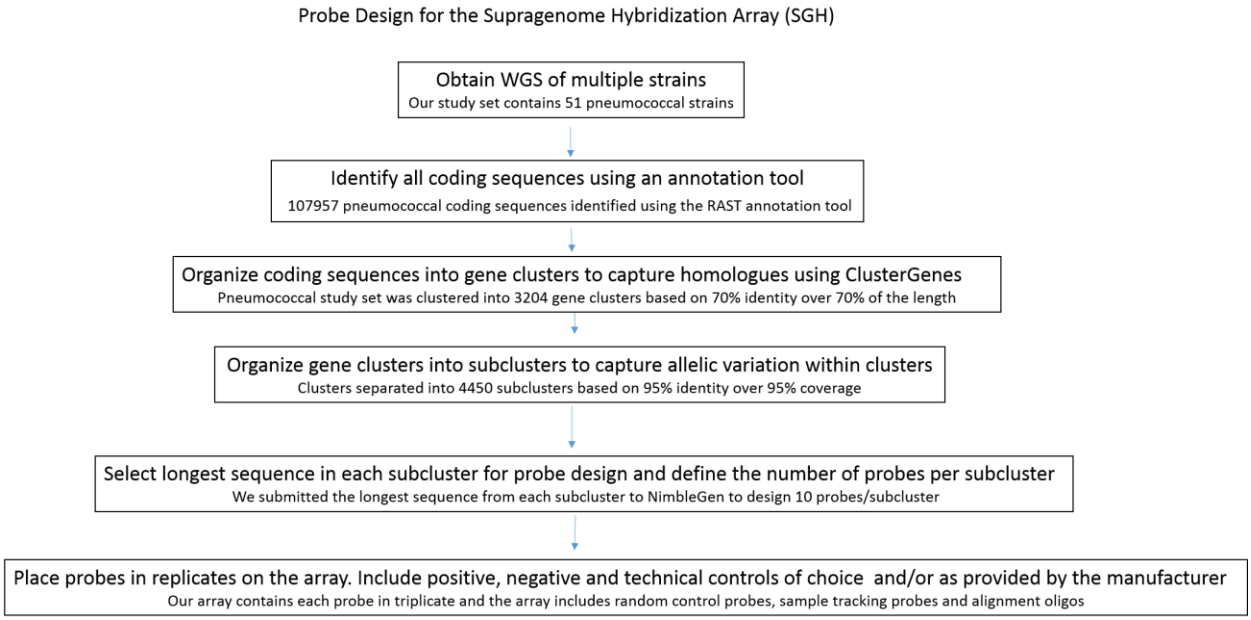


Fig 6.2.1. Schematic of probe design for the SGH array. Specific information on the SpSGH array is indicated in smaller fonts.

Materials

WGS of multiple strains that capture the diversity of the set of interest

Rapid Annotations with Subsystems Technology (RAST) (http://rast.nmpdr.org/)

FASTA36 from the FASTA package

(http://faculty.virginia.edu/wrpearson/fasta/fasta36/.)

Scripts for gene clustering (available from the authors on request, or at

https://github.com/jpearl01/)

Array manufacturer's probe design tool

1) **Prepare Input: Select Strains and Annotate to Obtain CDSs**

The coverage potential of the final probe set will depend on how well the input sequences capture the distributed gene content within the species. The goal is to capture as many genes as possible, by selecting not only a large number of strains with available high quality WGS, but also highly variable strains with respect to gene content, geographic isolation and clinical phenotypes. Boissy and colleagues describe in detail a model to predict the coverage of a supragenome/pangenome given a subset of strains [163]. For *H. influenzae*, *S. pneumoniae* and *Staphylococcus aureus* it was found that less than 50 strains cover >95% of the non-rare ($v < 0.1$) genes.

After selecting the strains, annotate the WGSs to identify the CDSs. We recommend that all strains be annotated in parallel given that gene annotations vary significantly depending on the tool selected and the version of the algorithm at the time of submission. We used Rapid Annotations with Subsystems Technology (RAST), a fully-automated web service for annotating bacterial genomes, where the annotated genomes are made available in a GenBank format [94]. RAST is available at http://rast.nmpdr.org/.

2) Organize CDSs into Allelic Groups: Compare all CDSs to Each Other using FASTA36 and Parse Data into Clusters and Subclusters

Organize the gene sequences into clusters of related sequences, so that cluster-specific probes can be designed. The clustering requires the following steps: A) prepare the input; B) compare all sequences using FASTA36, C) parse the sequence comparison

into clusters of homologous genes with (presumed) shared function, D) parse each cluster into subclusters to ensure probes will recognize all alleles, and E) submit selected sequences for probe selection and manufacturing.

A. Prepare the Input for FASTA36 Comparisons: Organize the GenBank files from the RAST output into three files: i) a multi-fasta with all CDS as amino acid sequences, ii) a multi-fasta with CDS as nucleotide sequences, and iii) a multi-fasta with all the contigs. An in-house program for these functions is available from authors by request or can be downloaded from:

https://github.com/jpearl01/prepare_supragenome_project

B. Compare all CDSs and Contigs using FASTA36: Use the programs within the FASTA Package (FASTA36) to compare all the sequences [164]. Tfasty36 is used to compare the protein CDSs to the DNA CDSs database, calculating similarities with frameshifts to the forward and reverse orientations. Fasta36 is used to compare the DNA CDSs to the contigs, and capture any genes that may have been missed in the annotation process. The programs can be downloaded from http://faculty.virginia.edu/wrpearson/fasta/fasta36/.

Run the fasta36 programs using the following parameters:

Fasta36: fasta36 –E 1 –m 9 –n –Q –d 0 input ii (multi-fasta of all CDS as nucleic acids) input iii (multi-fasta of all contig sequences) > output name

Tfasty36: tfasty36 –E 1 –m 9 –p –Q –d 0 input i (multi-fasta of all CDS as amino acids) input ii (multi-fasta of all CDS as nucleic acids) > output name

C. Group the CDSs into Gene Clusters to Capture Similar Sequences. To parse the gene comparison into clusters we recommend our in-house Perl script – termed ClusterGenes - developed by Justin Hogg and originally presented by Hogg and colleagues [150]. A cluster is defined as a group of genes that share at least 70% identity, over 70% of their length, with one or more of the other genes in the group, and where at least one sequence in the cluster is equal to or longer than 120 amino acids. This script also organizes the cluster as either core or distributed. ClusterGenes is available from authors by request or can be downloaded from https://github.com/jpearl01/

D. Group the Gene Clusters into Subclusters to Capture Allelic Differences. Many of these clusters contain multiple allelic variants, such that if probes are designed to only one representative sequence from each cluster, they may not hybridize to all the alleles. To ensure that probes are designed that will collectively hybridize to all known alleles, each cluster should be further split into subclusters. Within a subcluster, all sequences are 95% identical over 95% of the length of the shorter sequence. For the subclustering step, apply the sequence comparison and parsing to each individual cluster using the same ClusterGenes script.

E. Submit the Longest Sequence in Each Subcluster for Probe Design and Manufacturing. The selected company that manufactures the probes will apply their in-house algorithms to design probes when given a user-defined set of sequences. Our probes were designed by Roche NimbleGen, and currently could be designed by the Agilent platform.

The company algorithms will ensure the design of probes of the desired length (60-200 bp), while avoiding homopolymers and low complexity regions. To this end, submit a multi-fasta file with the longest sequence in each subcluster for probe design and manufacturing. Regarding the number of probes per sequence, we suggest generating the maximum number that would fit one array while allowing each probe to be placed in triplicate. For the NimbleGen pneumococcal array, this meant that we could include up to 10 probes for each subcluster.

Control probes should also be included; two such sets are included in the SpSGH array. First, a set of 1000 random control probes (generated by NimbleGen) with the same length and GC characteristics as the experimental probes on the array and these can be used to estimate non-specific hybridization for background correction. Second, a set of alignment and tracking probes that serve for accurate positioning of the probe grid during image analysis, detection of erroneous mixing of samples, and gauging the uniformity of hybridization over the probe covered area of the array.

Probe Design for the SpSGH Array

For design of the SpSGH array, 51 strains were selected (**Table 6.2.1**). The strains include multiple representatives of the major pathogenic lineages, and multiple serotypes and multi locus sequencing types (MLST). Furthermore, the chosen strains were isolated from subjects on multiple continents, and included representatives associated with nasopharyngeal carriage as well as disease. Together, the 51 genomes code for 107,957 CDSs. These were compared and organized into 3,204 pneumococcal

gene clusters of which 1,597 are core and 1,607 are distributed. All clusters were further subdivided into subclusters and the longest sequences from each subcluster were used to design probes by the manufacturing company. Some subclusters were eliminated because no suitable probes could be designed and/or only suitable probes were predicted to cross react with multiple subclusters. 40,988 experimental probes were designed to 3,027 clusters subdivided into 4,450 subclusters (9.2 probes/subcluster), of which 2,344 are core and 2,106 are distributed. The final probes for the SpSGH are provided in **Table S4** Appendix I.

Table 1: List of strains used for design of the SpSGH

Pneumococcal Strain	Serotype	MLST	Genome (bp)	#ORFs	Location of isolation	Carriage / disease	Status	Source	Accession Number	Technology	Coverage
SP3	3	180	2033581	2177	Pittsburgh, US	disease	draft	Ref. [4]	AAZZ00000000	454 GS20	18x
SP6	6	460	2162916	2325	Pittsburgh, US	disease	draft	Ref. [4]	ABAA00000000	454 GS20	19x
SP9	9	1269	2117908	2241	Pittsburgh, US	disease	draft	Ref. [4]	ABAB00000000	454 GS20	20x
SP11	11	62	2060705	2127	Pittsburgh, US	disease	draft	Ref. [4]	ABAC00000000	454 GS20	17x
SP14	14	124	2148093	2624	Pittsburgh, US	disease	draft	Ref. [4]	ABAD00000000	454 GS20	16x
SP18	6	new	2105593	2211	Pittsburgh, US	disease	draft	Ref. [4]	ABAE00000000	454 GS20	19x
SP19	19	485	2136434	2301	Pittsburgh, US	disease	draft	Ref. [4]	ABAF00000000	454 GS20	16x
SP23	23	37	2103479	2200	Pittsburgh, US	disease	draft	Ref. [4]	ABAG00000000	454 GS20	15x
INV104B	1	227	2142122	1941	Oxford, UK	disease	complete	Ref. [4]	FQ312030	Sanger	9x
OXC141	3	180	2036967	1973	Oxford, UK	carriage	complete	Ref. [4]	FQ312027	Sanger	9x
INV200	14	9	2093318	2045	Oxford, UK	disease	complete	Ref. [4]	FQ312029	Sanger / Illumina	12x/85x
SpnATCC700669	23F	81	2221315	2132	Spain	carriage	complete	Ref. [19]	FM211187	Sanger	8x
Sp03_4156	3	180	2058353	1954	The Netherlands	carriage	draft	Donati et al	FQ312045	Sanger / 454 FLX	5x/25x
Sp03_4183	3	180	1993183	1933	The Netherlands	carriage	draft	Donati et al	FQ312043	Sanger / 454 FLX	6x/16x
Sp07_2838	3	180	1990038	1901	Bolivia	carriage	draft	Donati et al	CACI01000000	Sanger / 454 FLX	5x/23x
Sp99_4038	3	180	2010908	1952	Glasgow Reference Lab	disease	draft	Donati et al	FQ312041	Sanger / 454 FLX	7x/26x
Sp99_4039	3	180	2010104	1954	Glasgow Reference Lab	disease	draft	Donati et al	FQ312044	Sanger / 454 FLX	4x/32x
Sp02_1198	3	180	1989367	1938	Glasgow Reference Lab	disease	draft	Donati et al	CACH01000000	Sanger / 454 FLX	5x/28x
A45	3	New	2041833	1932	Newmarket	disease	draft	Donati et al	CACG01000000	Sanger / 454 FLX	6x/15x
P1041	1	217	2166490	1905	Ghana	disease	draft	Donati et al	CACE01000000	Sanger / 454 FLX / Illumina	5x/13x/96x
Sp03_2672	1	306	2144331	1904	Glasgow Reference Lab	disease	draft	Donati et al	FQ312039	Sanger / 454 FLX	5x/25x
Sp03_3038	1	306	2164519	1936	Glasgow Reference Lab	disease	draft	Donati et al	FQ312042	Sanger / 454 FLX	5x/20x
Sp06_1370	1	306	2012346	1874	Glasgow Reference Lab	disease	draft	Donati et al	CACJ01000000	Sanger / 454 FLX	5x/19x
NCTC7465	1	615	2100988	1845	Type strain, Rockefeller USA, 1948	disease	draft	Donati et al	CACF01000000	Sanger / 454 FLX / Illumina	5x/11x/86x
P1031	1	303	2111882	2073	Ghana	disease	complete	Donati et al	CP000920	Sanger / 454 FLX	Finished*
D39	2	595	2046115	1914	US	disease	complete	Ref. [17]	CP000410	Sanger	Finished*
TIGR4	4	205	2160842	2125	Norway	disease	complete	Ref. [15]	AE005672	Sanger	Finished*
70585	5	289	2184682	2202	Bangladesh	disease	complete	Donati et al	CP000918	Sanger / 454 FLX	Finished*
JJA	14	66	2120234	2123	Brazil	disease	complete	Donati et al	CP000919	Sanger / 454 FLX	Finished*
MLV-016	11A	62	2247118	2159	USA, Europe	carriage	draft	Donati et al	ABGH00000000	Sanger / 454 FLX	9x/28x
CDC0288-04	12F	220	2051140	2105	USA, UK	disease	draft	Donati et al	ABGF00000000	Sanger / 454 FLX	10x/31x
CDC3059-06	19A	199	2293277	2379	Iceland, UK, USA, others	disease	draft	Donati et al	ABGG00000000	Sanger / 454 FLX	9x/26x
Hungary19A-6	19A	268	2245615	2155	Hungary	disease	complete	Donati et al	CP000936	Sanger / 454 FLX	Finished*
Taiwan19F-14	19F	236	2112148	2044	Taiwan	disease	complete	Donati et al	CP000921	Sanger / 454 FLX	Finished*
CDC1873-00	6A	376	2265195	2402	USA	disease	draft	Donati et al	ABFS00000000	Sanger / 454 FLX	9x/20x
670-6B	6B	90	2240045	2384	Spain	disease	complete	Ref. [4]	CP002176	Sanger	Finished*
CDC1087-00	7F	191	2190853	2232	Bra, Den, Fin, Neth, Nor, UK, Uru, USA	disease	draft	Donati et al	ABFT00000000	Sanger / 454 FLX	9x/14x
SP195	9V	156	2198294	2287	Worldwide	disease	draft	Donati et al	ABGE00000000	Sanger / 454 FLX	10x/29x
G54	19F	63	2078953	2115	Italy	disease	complete	Ref. [4]	CP001015	Sanger	Finished*
R6	2	595	2038615	2043	Laboratory	laboratory	complete	Ref. [16]	AE007317	Sanger	Finished*
ST13v1-CGSSp14BS292	14	13	2100368	2290	Children's Hospital of Pittsburgh	disease	draft	Hiller 2010	ABWQ00000000.1	454 FLX	21.5x
ST13v12-CGSSpBS293	NT	13	2065452	2242	Children's Hospital of Pittsburgh	disease	draft	Hiller 2010	ABWU00000000.1	454 FLX	23x
ST13v6-CGSSpBS457	NT	13	2053197	2225	Children's Hospital of Pittsburgh	disease	draft	Hiller 2010	ABWB00000000.1	454 FLX	27x
ST2011v4-CGSSpBS455	NT	2011	2086050	2182	Children's Hospital of Pittsburgh	disease	draft	Hiller 2010	ADHN00000000.1	454 FLX Titanium	28x
SV35-T23	23F	81	2156885	2242	AIDS clinic of St. Vincent's Medical Center, Richmond, New York, US	disease	draft	Hiller,2011	ADNN01000000	454 FLX	26x
SV36-T3	3	81	2162633	2239	AIDS clinic of St. Vincent's Medical Center, Richmond, New York, US	disease	draft	Hiller,2011	ADNO01000000	454 FLX	26x
PN4595-T23	23F	81	2169192	2259	Lisbon, Portugal	carriage	draft	Hiller,2011	ABXO01	454 FLX	27.6x
CGSP14	14	15	2209198	2206	China, Beijing Institute of Genomics	disease	complete	Ref. [18]	CP001033	Sanger	Finished*
CCRI 1974	14	124	2005075	2074	McGill University, Canada	disease	draft	Ref. [20]	ABZC00000000	454 FLX	20x
CCRI 1974M2	14	124	2003231	2069	McGill University, Canada	disease	draft	Ref. [20]	ABZT00000000	454 FLX	20x
AP200	11A	62	2084139	2149	University of Siena, Italy	disease	draft	Unpublished	CP002121	454	10x

BASIC PROTOCOL 2: SpSGH Array to Determine Gene Content

The SpSGH array can be used to determine the gene content of isolates by employing DNA-DNA hybridizations. This has been described for the H. influenzae SGH array [160] *and is described here for the SpSGH.* The process has five steps: 1) grow bacterial strains; 2) extract genomic DNA (gDNA); 3) label Cy3 gDNA ; 4) hybridize gDNA on array, wash and scan 5) analyze the data. A schematic for these steps is represented in **Fig 6.2.2**.

We tested the SpSGH array by comparing array results with WGS data for 5 strains. We also investigated genome content for 2 non-sequenced strains isolated from a patient with a polyclonal upper respiratory infection.

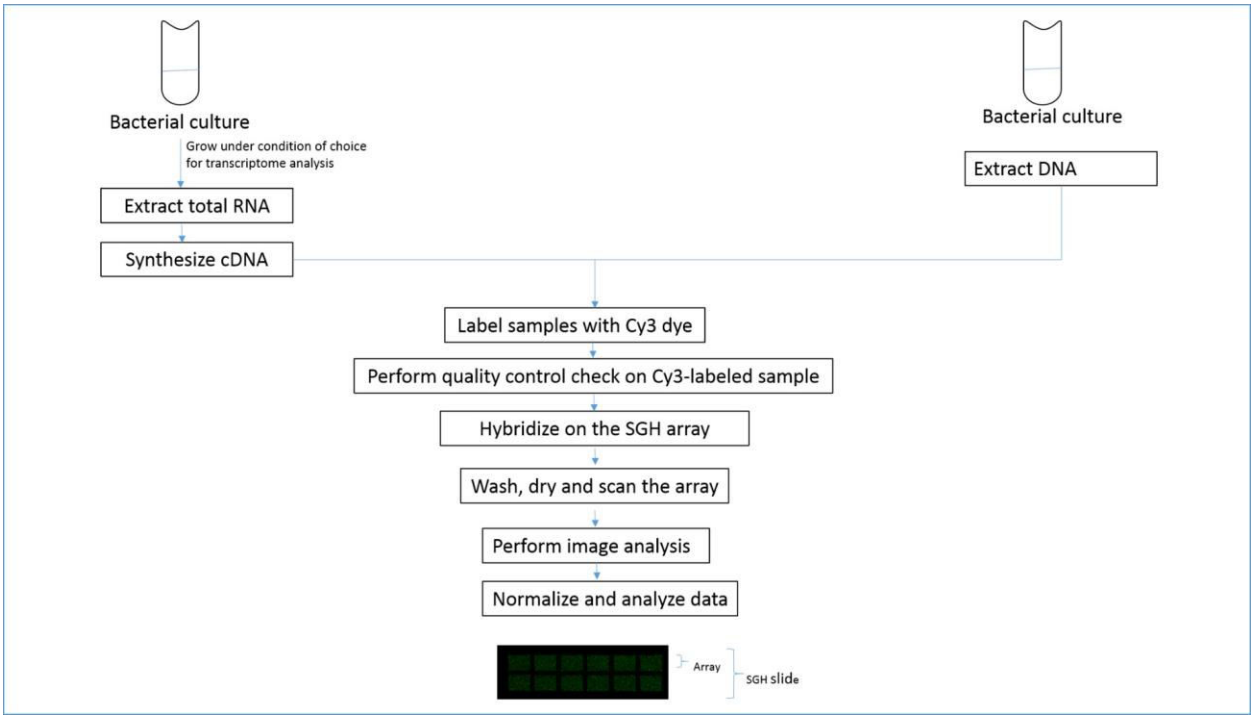


Fig 6.2.2 Schematic for processing of nucleic acid samples for SpSGH array

Materials

Bacterial strain of interest

Standard media for bacterial growth (Columbia broth is used for *S. pneumoniae*)

Pneumococcal cell lysis cocktail: lysozyme (15ml/mL), mutanolysin (30µg/mL), proteinase K (20mg/mL), 1x Tris EDTA Buffer

Chloroform:isoamylalcohol (24:1)

RNAseA (4mg/mL)

1% TAE agarose gel

NanoDrop 1000 UV spectrophotometer

Centrifuge that allows harvesting cells from a 15mL culture volume

Vacuum concentrator

SGH array

gDNA Cy-3 labeling reagents from the array manufacturer

Hybridization and washing reagents from the array manufacturer

Hybridization station from the array manufacturer

Fluorescent scanner

Thermocycler

Vortex

Strain and growth conditions

1. Set up 15mL bacterial cultures in duplicates and allow growth to mid-log phase or stationary phase, in standard media.

The goal is to determine gene content, thus the only condition for growth is that which provides sufficient DNA. For *S. pneumoniae* we use Columbia broth and continue culture until an $OD_{600\ of}$ 0.5 is achieved, as these conditions provide for high cell numbers yet limited Lyt-A-mediated autolysis. For *H. influenzae* we use overnight cultures grown in supplemented BHI broth.

2. Harvest the bacterial cells by centrifugation at 5000xg for 10mins, and freeze the pelleted cultures at -80°C so that the subsequent steps can be performed at a later time point, if desired. (We recommend frozen pellets be used within two weeks of freezing.)

3. Thaw the pelleted cells by incubating at room temperature for 15 minutes. Lyse the bacterial cells. To achieve *S. pneumoniae* cell lysis, resuspend pellets in a 220µL cocktail of lysozyme (15mg/ml), mutanolysin (30µg/mL), and proteinase K (20mg/mL, Qiagen) in 1X Tris-EDTA buffer for 15 minutes at room temperature, with intermittent vortexing every 2 minutes.

gDNA extraction

4. Perform a standard 24:1 chloroform/isoamyl alcohol method for gDNA extraction and store samples in 1X TE buffer [165].

5. Measure the concentration and purity of the gDNA using the ratio of absorbance at 260 and 280 nm on a UV spectrophotometer (NanoDrop 1000, Thermo Scientific), where pure gDNA has an $A_{260/280}$ ratio of 1.8.

6. Confirm the purity of the gDNA by running ~1 µg of DNA on a 1% TAE agarose gel. If the purity is low, the gDNA should be treated with RNaseA (4mg/mL) and/or Proteinase K (20mg/mL), then re-precipitated and re-analyzed to ensure purity of the DNA.

Cy3 gDNA labeling and quality control

7. Label gDNA samples with Cy3 dye. This step utilizes a nucleotide synthesis reaction which incorporates Cy3 labeled random nonamers into double stranded DNA using the NimbleGen One Color DNA Labeling Kit (NimbleGen Arrays User's Guide, Gene Expression Arrays version 6.0). To this end, heat the gDNA sample to 98˚C for 10 minutes in the presence of the Cy3 labeled random nonamers, and rapidly cool in an ice water bath. For DNA polymerization, add the dNTPs and Klenow fragment to the reaction and incubate for 2 hours at 37˚C.

8. Use isopropanol precipitation of the labeled gDNAs to eliminate any unincorporated nucleotides and primers from the labeling reaction.

9. Dry the DNA pellet in a vacuum concentrator (SpeedVac) and protect from light. Rehydrate the sample in nuclease-free water,

10. Measure the concentration and quality using 260/280 absorbance ratio on a spectrophotometer. *The final concentration and volume of Cy3-labeled gDNA for the next step, depends on the array design and manufacturer. A NimbleGen array, with 12 hybridization regions of 135K each (that is, 135,000 probe capacity), requires 2 □g of labeled gDNA per region.*

Hybridization, washing and slide scanning

SGH slides allow for parallel processing of multiple samples per slide, such that each sample is loaded onto a separate array on the slide. Our NimbleGen slide contains 12 arrays. Cross-reaction of different samples is monitored using sample tracking controls (STC) provided by the manufacturer. Each array has a different STC. For the NimbleGen 12 X 135K array, 2µg of Cy3-labeled DNA is lypophilized in a SpeedVac and resuspended in sample tracking solution. On each slide, probes specific to the STC are placed as repeating sets of 20 along the perimeter of each array and bordering their corners. The STCs assist later during imaging where, by performing a sample tracking control analysis and visually verifying the outlines of each array, the user can confirm that samples have not mixed with each other.

11. Prior to hybridization, mix the sample with the components of the Hybridization Kit (NimbleGen Hybridization buffer, component A, and alignment oligomer) and incubate at 95°C for 5 minutes.

12. *The array manufacturer supplies a proprietary mixing device that is designed to align and adhere to the surface of the array.* Place the adherence assembly (mixer device + SpSGH slide) on the hybridization station, and load 6□l of each sample into a fill port.

13. Set the hybridization station at 42˚C and incubate the slides for 18 hours.

14. After the incubation, disassemble the slide from the mixer and wash to remove unbound sample. Washing and drying involve a series of wash buffers from the NimbleGen Hybridization Wash Buffer kit and the Microarray slide dryer. *For best results, perform the steps leading from washing to scanning without any pauses.*

15. Measure the fluorescence using a fluorescent scanner with suitable resolution; we used the Molecular Devices Axon GenePix 4200AL for the one-color array scanning of the SpSGH slide. Process the images using the NimbleScan, or equivalent imaging software, to measure the intensity and the relative position of each fluorescent signal.

BASIC PROTOCOL 3: Data Analysis of SpSGH Array to Determine Gene Content

The analysis of gene content involves multiple steps: a) conversion of the fluorescent signals into quantitative intensity values and determination of data integrity; b) normalization of the values across all arrays on the same slide, or among slides; c) determination of the threshold for presence/absence of a gene to establish genome content for the strains of interest.

Materials

Slide scanning software provided by the array manufacturer (such as NimbleScan)

Generating quantitative intensity values and assessing data integrity

1. Following the slide scan, use the array software to burst the single multiplex image of the slide into separate array images based on the format of the slide. Each image will correspond to one strain/sample.

2. Align each of the separated images to the design file that contains information on the placement of the probes on the array. A grid setup in the software assists in aligning the images correctly. During this step, perform the sample tracking control check that verifies absence of cross-contamination among samples by indicating which sample tracking controls (STC) are present in each array. *Only one STC should be present per*

sample. The same analysis step also generates an experimental metrics report, consisting of a spreadsheet reporting signal density, signal range and uniformity.

3. Normalize the data across regions on the array as per the user's choice. Normalization uses a Robust Multichip Average (RMA) algorithm and quantile normalization (also available from the chip manufacturer, in this case via the NimbleScan software). In the raw data, an intensity value is available for each probe. The NimbleScan normalization process will combine the multiple probes for each subcluster using a median polish, generating a table that lists each subcluster and an associated intensity value for each probe set. At the end of this step, the user will have a tab-delineated sheet with one value integrating the multiple probes per subcluster (i.e per allele) (~9/ alleles for the SpSGH array), with triplicates for each value (since all probes were placed on the array in triplicate). An example sheet can be visualized in **Table S5** in Appendix I.

Determining present/absent genes

4. To determine which genes are present in the sample, select a hybridization value threshold that separates genes present versus genes absent. We recommend this threshold be 1.5 times the median background value.

5. Convert the normalized data into a $\log_2$ scale and determine the inter-slide consistency using the Student's *t* distribution analysis. These standard statistics can be applied from any program of choice; we use a python-based script. A gene is considered present if the signal for any of its subclusters is above the threshold and the p-value from the Student's *t* test is less than 0.05. Conversely, a gene is considered

absent if the signal for all its subclusters is below the threshold or if the p-value from the Student's *t* test is above 0.05 for all subclusters above threshold. Importantly, binding to subsets of subclusters is not used to investigate polymorphisms given that hybridization can occur even when small variations exist between the probe and the allele.

Genomic Content of Seven Pneumococcal Isolates Using the SpSGH Array

The gene possession profile of 5 pneumococcal isolates was interrogated for presence/absence of 3,027 clusters (out of 3,204 total) using the SpSGH array. The presence/absence profile was compared to WGS to calculate the sensitivity (determined by the number of false positives) and specificity (determined by the number of false negatives) of the SpSGH array output relative to WGS. The results are described in **Table 2**, where false positives varied from 3-5 genes/genome and false negatives between 21-37 genes/genome, suggesting that >98% of genes were accurately predicted by the SpSGH array.

An additional 2 pneumococcal genomes, ST13v3 and ST2011v5 [166] were assayed using the SpSGH array. These strains were isolated from a young child with a polyclonal infection, previously referred to as patient 19 [166]. MLST and serotype analyses suggested these strains were similar to a pair of strains isolated from patient 19 at different clinical visits, strains ST13v1 (ST13v1-CGSSp14BS292) and ST2011v4 (ST2011v4-CGSSpBS455), respectively. The SpSGH analysis confirms this prediction,

demonstrating that these strains were isolated from the same patient at separate time points consistent with chronic colonization (**Table 2**).

Table 2: Validation of the SpSGH using whole genome sequence. The supragenome analysis used for SpSGH design contained 3204 gene clusters, of which 3027 are represented on the SpSGH. The last two rows depict data from unsequenced genomes isolated during a polyclonal infection. The SpSGH results demonstrate they are almost identical to other strains isolated from the same patient (ST13v3 is compared to ST13v1 and ST2011v5 to ST2011v4)

strain	# CLUSTERS	# CLUSTERS w/PROBE ON CHIP	# CLUSTERS DETECTED BY CHIP	% CORR. PREDICTED	# FALSE POSITIVE	# FALSE NEGATIVE
ST13v1	2028	1918	1902	99.2	5	21
SP3	1996	1890	1857	98.3	4	37
SP14	2119	2002	1973	98.6	4	33
SV35	2059	1950	1927	98.8	3	26
SP23	2044	1933	1909	98.8	4	28
ST13v3	2028	1918	1902	99.2	5	21
ST2011v5	2009	1903	1868	98.2	8	43

BASIC PROTOCOL 4: SpSGH Array for Gene Expression Profiling

The SpSGH array data is quantitative in nature, thus it can be used for gene expression profiling. To this end, cDNA, instead of gDNA is analyzed.

Profiling of gene expression requires the following steps: 1) strain growth; 2) RNA extraction, 3) conversion of RNA to cDNA; 4) Cy3 labeling of the cDNA; 5) cDNA hybridization, washing, and scanning; and 6) data analyses. Steps 1, 4-5 are very similar to those described for gDNA, such that this section will focus on the differences only. As an example, we used the SpSGH array to investigate the relative levels of transcripts for two pneumococcal strains relative to housekeeping controls. To measure accuracy, the results were compared with data for 54 genes using the nCounter Analysis System by NanoString Technologies [144,167]. **Fig 6.2. 2** provides a schematic for these steps.

Materials

Bacterial strain of interest

Standard media for bacterial growth (Columbia broth is used for *S. pneumoniae*)

Centrifuge that allows harvesting cells from a 15mL culture volume

RNAProtect Bacterial Reagent (Qiagen)

RNeasy Mini Kit (Qiagen) for RNA extraction

Pneumococcal cell lysis cocktail: lysozyme (15ml/mL), mutanolysin (30µg/mL),

proteinase K (20mg/mL), 1x Tris EDTA Buffer

Lysis buffer RLTplus (Qiagen)

DNAse, 2units/µL (Turbo DNAse, Ambion)

gDNA eliminator column (Qiagen)

Agilent Bioanalyzer and RNA 6000 Nano Kit

SuperScript III First-Strand Synthesis SuperMix kit (Invitrogen)

SuperScript Double-Stranded cDNA Synthesis kit (Invitrogen)

RNaseA (4mg/mL)

Phenol:chloroform:isoamylalcohol (25:24:1)

Ammonium acetate, glycogen and ethanol

Vacuum concentrator (SpeedVac)

Nuclease-free water

NanoDrop1000 spectrophotometer

SGH array

gDNA Cy3 labeling reagents from the array manufacturer

Hybridization and washing reagents from the array manufacturer

Hybridization station from the array manufacturer

Fluorescent scanner

Thermocycler

Vortex

Strains and growth conditions

1. Grow bacteria under *in vitro* condition(s) of interest. cDNA profiles can be compared across multiple types of samples. For example, the same strain under different growth conditions, or a wild type (WT) strain versus its cognate mutant strain.

For the pneumococcal work presented here, we selected *S. pneumoniae* strains PN4595-T23 (ABXO01) and 3063-00 (AGQG01). The former is one of the 51 genomes used in the probe design, while the latter was not included in the probe design. PN4595-T23 is a member of the Pneumococcal Molecular Epidemiology Network clone 1 lineage (PMEN1) and 3063-00 is related to the Taiwan19F, thus both represent isolates from widespread and multidrug-resistant lineages. These strains are grown in 15mL Columbia broth in a 50mL tube to an $OD_{600\ of}$ 0.5. This mid-log phase OD is chosen as it yields high cell numbers while the Lyt-A mediated autolysis is absent.

2. Harvest bacterial cells by centrifugation at 5000xg for 10 minutes and *immediately* resuspend the pellet in RNAProtect Bacterial Reagent (Qiagen) to stabilize RNA before storing at -80˚C.

RNA extraction and quality check

It is critical that all tubes and water used for sample preparation are DNase/RNase-free.

3. Use the RNeasy Mini Kit (Qiagen) to extract total bacterial RNA. This process can be divided into 3 steps: cell lysis, RNA extraction, and elimination of any residual DNA. To achieve pneumococcal cell lysis, resuspend the cell pellets in a cocktail of lysozyme (15 mg/ml), mutanolysin (30 μg/mL), and proteinase K (20 mg/mL) in 1X Tris-EDTA buffer for 15 minutes at room temperature, with intermittent vortexing every 2 minutes to aid the lysis process.

4. Add lysis buffer RLTplus (Qiagen) to the preparation. Apply the lysate to the gDNA eliminator column (Qiagen) to remove genomic DNA, next apply to an RNeasy column for RNA isolation.

5. Treat the eluted RNA with 2 units/μL of DNAse (TurboDNase, Ambion) for 1.5h at 37˚C.

6. Assess the RNA integrity by running samples on the Agilent Bioanalyzer using an RNA 6000 Nano Kit. These RNA chips consist of micro-channels that separate nucleic acid fragments based on their electrophoretic mobility (Agilent Technologies, Inc.). Intact peaks corresponding to 16SrRNA and 23SrRNA and high RIN number in electropherograms are measures of RNA integrity.

7. Confirm the RNA purity using polymerase chain reaction (PCR) for the glyceraldehyde 3-phosphate dehydrogenase (GAPDH) housekeeping gene, such that no amplification would be observed in pure RNA samples, while amplicons would be observed in the pure genomic DNA control.

8. For an additional quality check, measure sample absorbance using a spectrophotometer. High quality RNA has an $A_{260/280}$ ratio of ≥ 2.0.

cDNA preparation and quality check

All RNA, cDNA and reagents should be maintained on ice.

9. Use SuperScript III First-Strand Synthesis SuperMix kit (Invitrogen) for synthesis of the first strand of cDNA. For this, start with 5 µg of good quality total RNA as a template, add the random hexamers primers supplied in the kit, and heat at 70˚C for 10 minutes. Next, add First Strand Buffer, DTT and dNTPs. These aid in the removal of any secondary structures in the RNA template. Add Superscript III reverse transcriptase as the last component and incubate the sample at 42 ˚C for 1 hour to synthesize an RNA:DNA hybrid.

10. Use SuperScript Double-Stranded cDNA Synthesis kit (Invitrogen) for second strand cDNA synthesis. Incubate the samples with kit components (DEPC water, 5x Second Strand Buffer, dNTPs, DNA Ligase, DNA Polymerase I and RNase H) at 16˚C for 2 hours.

11. Eliminate residual template RNA by treating the reaction with RNaseA (4mg/mL) and extract using phenol:chloroform:isoamylalcohol (25:24:1). Precipitate the cDNA in the upper aqueous layer using ammonium acetate, glycogen and ethanol, followed by concentrating the pellet in a vacuum concentrator (SpeedVac) until a gel-like consistency is reached.

12. Rehydrate the cDNA samples with nuclease-free water and assess their quality and quantity (we use the NanoDrop1000 spectrophotometer, A260/280 ≥ 1.8). If desired, check the cDNA samples by running on an agarose gel or Agilent Bioanalyzer. *It is important to ensure that degraded/poor quality samples are not carried through further steps.*

Cy3 cDNA labeling, hybridization, washing, and scanning

These steps are as described above in Basic Protocol 2.

BASIC PROTOCOL 5: Data Analysis of SpSGH Array for Gene Expression Profiling

The data analysis for transcriptional profiling can be subdivided into the following steps: A) selection of the relevant probes for the analysis; B) normalization within and among arrays; C) comparison of transcription levels between/among sample sets.

Materials

Slide scanning software provided by the array manufacturer (such as NimbleScan)

NCBI-BLAST

Software of choice for transcriptomic analysis (e.g. CyberT, http://cybert.microarray.ics.uci.edu/; SAM http://statweb.stanford.edu/~tibs/SAM/faq.html; TM4/MeV , http://www.tm4.org/mev.html)

If the probe set of choice in selected by hybridization, also refer to Basic Protocol 2.

Selection of the relevant probes for the analysis

The advantage of using the SpSGH array is that the same platform can be used to assay any isolate from the species. However, for each isolate, the relevant probe set must be selected. For this quantitative analysis, a distribution of the intensity values per subcluster should fit a normal distribution. If the analysis accounts for all the subclusters, the majority of probes will not hybridized and will have very low fluorescent values. Thus, the relevant subclusters should be singled out.

1. Perform subcluster selection using one of two methods described below. If the genome of the test strain is known, the relevant subclusters can be selected *in silico* (i). If the genome sequence is not known, the relevant clusters can be selected using hybridization of gDNA on the SpSGH (ii).

i. Clusters selection using WGS. Compare the representative sequence for each subcluster to the WGS of selected strain using BLAST (basic local alignment search tool) from NCBI [168]. Include the top hit for each query sequence into the relevant subcluster set. We recommend downloading BLASTn onto a Linux computer and running the program locally with the following command line:

Prepare database for blastn:

makeblastdb –in [multi-fasta file with the longest sequence for all subclusters where gene is labeled with the subcluster ID (i.e sequences submitted for probe design)] –dbtype nucl

Run blastn:blastn –evalue 1e-20 –query [multi-fasta with the CDSs for the WGS of selected strain] –db [multi-fasta file with the longest sequence for all subclusters where gene is labeled with the subcluster ID (i.e file submitted for probe design)] > output

Parse the output to select the top hit. We use a BioPerl script to parse the output, such that only the cluster in the top hit is included in the relevant set of subclusters. For any cluster, all subclusters are included.

ii. Cluster Selection using SpSGH array. Hybridize the gDNA to the SpSGH array, as described in Basic Protocol 2 above. Finalize Basic protocol 2. Select only the clusters with a signal intensity above threshold and include all subclusters for each of these positive clusters. This represents the subcluster set relevant for your strain of interest.

Data Normalization

The analysis involves multiple steps: A) conversion of the fluorescent signal into quantitative intensity values and integrity check of the data; B) normalization of the values across all arrays on the same slide, or between arrays.

1. Use the array manufacturer's software to convert fluorescence intensity to quantitative value for each probe. The user can follow the description in "Part II: data analyses" above that describes how to: burst the single multiplex image of the slide into separate array images; check for cross-contamination among samples;

acquire the signal density, signal range and uniformity; normalize the data across regions on the array; and generate a tab-delineated sheet with one value integrating the multiple probes per subcluster, each in triplicate. An example sheet can be visualized in **Table S5** in Appendix I.

Comparison of transcription between sample sets.

3**.** This analysis will differ depending on the samples being compared, and standard array methods can be employed (e.g. CyberT: http://cybert.microarray.ics.uci.edu/; SAM: http://statweb.stanford.edu/~tibs/SAM/faq.html; TM4/MeV: http://www.tm4.org/mev.html). Janto and colleagues provide a detailed analysis comparing wildtype and deletion mutant strains over multiple conditions and time points using the HiSGH [161]. In their analysis, a web-based microarray analysis tool, Cyber T [169] is used to obtain Bayesian corrected p-values, Bonferroni corrected p-values and Benjamini-Hochberg values. These data are then combined and filtered in the following order: 1) SAM FDR <10%, 2) Bayesian p-values < .05, 3) Benjamini-Hochberg FDR < 10%, 4) Bonferroni corrected p-value < .05. This pipeline, with progressively more stringent statistical parameters, generates a robust set of differentially regulated genes for transcriptomic analysis.

Transcriptional analysis of two pneumococcal Isolates using the SpSGH Array

The following section presents an example analysis that reflects on reproducibility of the SpSGH by comparing transcriptional values within a single

genome using the SpSGH array and an alternative transcriptomic profiling technique, the NanoString technology.

Cluster Selection

For the analysis of strains PN4595-T23 and 3063-00, the relevant clusters were selected using *in silico* analysis. 1,929 and 1,886 total clusters were analyzed for PN4595-T23 and 3063-00, respectively.

Reproducibility of the SpSGH Array

To assess the SpSGH array reproducibility we compared: 1) the signal intensity values across the triplicate probe sets within the same array (**Fig 6.2.3**); 2) the signal intensity values across biological replicates on the same slide (**Fig 6.2.4A**); and 3) the signal intensity values across biological replicates on two slides, hybridized and analyzed independently (**Fig 6.2.4B**).

The robustness of the array can be measured by the reproducibility of triplicate sets within each array. Each subcluster is represented by up to 10 unique probes, the probe values are condensed into a final hybridization value per subcluster. Given the triplicate probe sets, there are 3 final hybridization values per array. The values across the triplicate probe sets were analyzed using coefficient of variance, where a standard deviation is calculated for each set of triplicate probe sets and reported as a percentage of the average signal for that probe set. We find that over 94% of the probe sets have a coefficient of variance below 0.3 (**Fig 6.2.3**).

Each slide is manufactured with multiple arrays present on the same slide allowing multiple samples to be processed together. We compared the final

hybridization values for biological replicates hybridized on the same slide **Fig 6.2.4A** (A.1 and A.2); as well as biological replicates hybridized on separate slides and processed independently (**Fig 6.2.4B**). In both cases, the arrays showed good reproducibility as illustrated by an R^2 of 0.980 and 0.949, respectively.

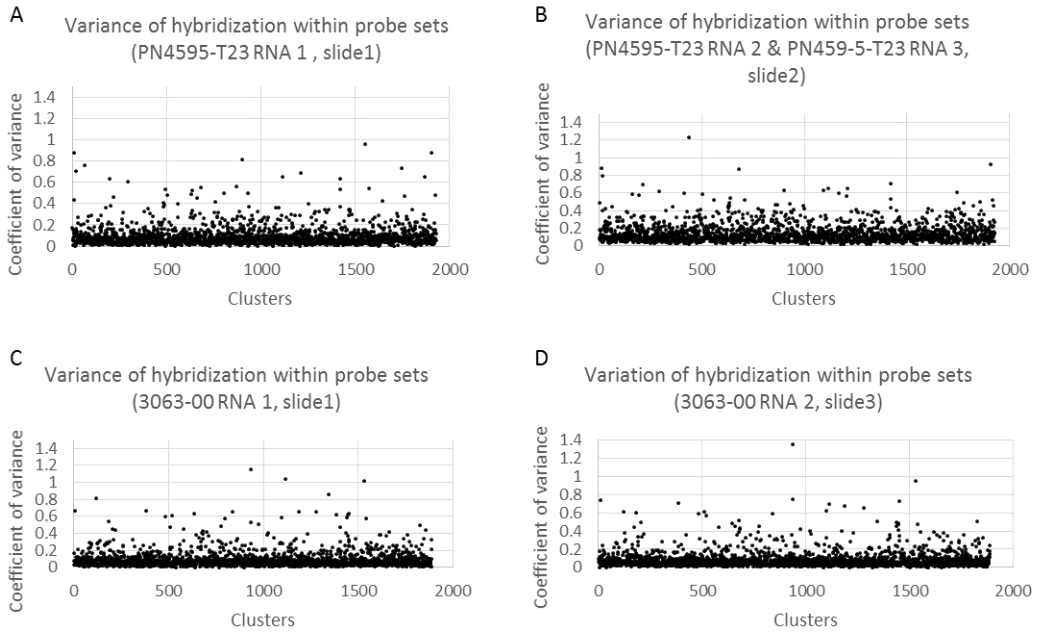


Fig 6.2.3 Comparison of probe specificity within each array based on coefficient of variance of hybridization of RNA samples to probe set.

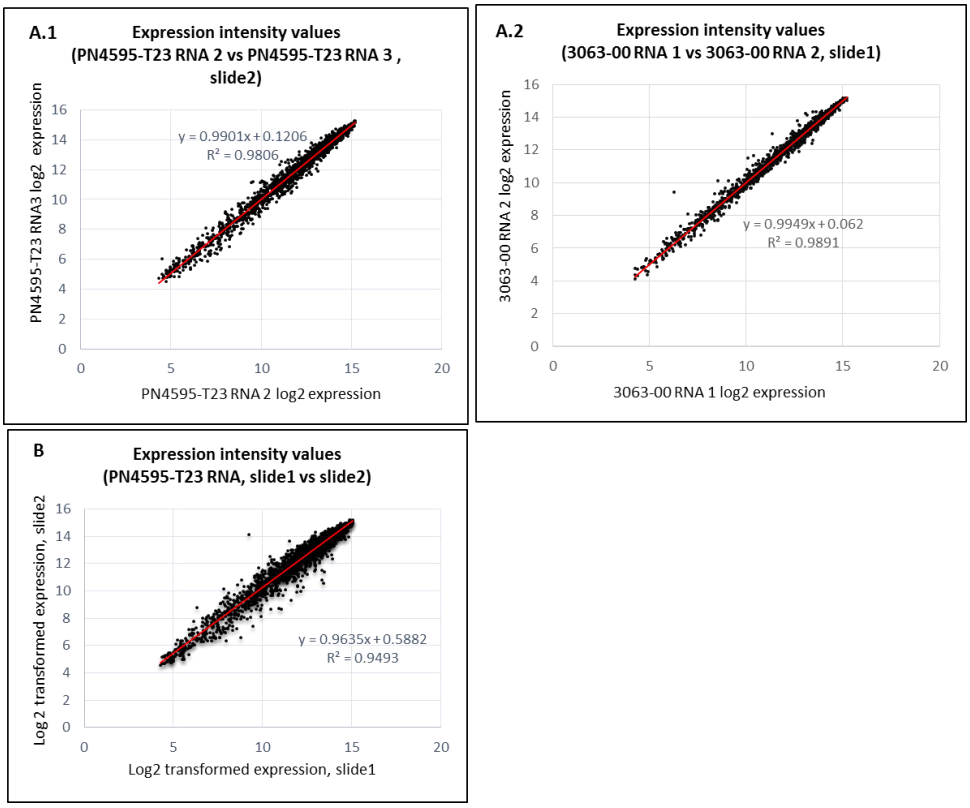


Fig 6.2.4 Comparison of hybridization values for biological RNA replicates. (A) within the same slide. A.1. slide, A.2. slide; and (B) across slides.

Measurement of relative transcriptional levels within a sample using the SpSGH array and nCounter Analysis System

The quantitative value of the SpSGH array was assessed by comparing the intensity values for RNA extracted from wildtype strains relative to two housekeeping genes: gyrase B (*gyrB*) and methionyl-tRNA synthetase (*metG*). We compared these fold changes to those obtained using another transcriptomics technology, the nCounter Analysis System (NanoString Technologies). The nCounter system directly captures mRNA with a sequence-specific DNA probe and quantifies the signal by single molecule imaging of unique transcripts (thus without any amplification step). The method is highly quantitative and reproducible, thus serves as a good method to verify the results from SpSGH array. The probe sets for the SpSGH array and the nCounter system were designed independently. Finally, all the RNA analyses, across SpSGH arrays and between the array and nCounter, were derived from different biological replicates (where cells were grown and RNA extracted independently).

In both the SpSGH array and nCounter analyses, all values were normalized against *gyrB* and *metG* using the geometric mean for their signal intensity. Next, the value of signal intensity for each cluster was divided by the geometric mean, yielding a relative intensity value which was converted to $\log_2$. Finally the results from each method were plotted against each other (**Fig 6.2.5**). The comparison between the

nCounter and SGH arrays reveals similar trends, suggesting that like the HiSGH array, the SpSGH arrays can also be used for transcriptomic studies.

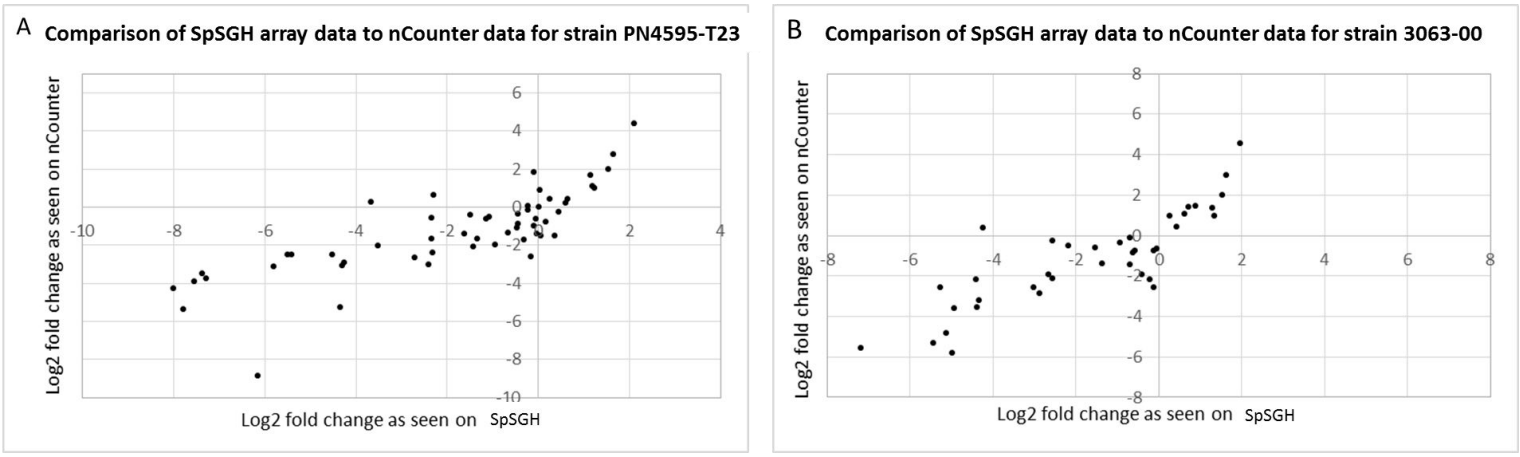


Fig 6.2.5 Validation of the SpSGH array, by comparing relative expression using SpSGH array and the nCounter from NanoString technologies. (A) PN4595-T23 . (B) 3063-00.

Commentary

Background Information

There can be extensive differences in allelic content and gene possession among strains of a single bacterial species. Our goal was to design a gene chip that quantitatively captured the genetic diversity in a bacterial population. The SpSGH array described in this unit: 1) captures >90% of non-rare genes allowing genomic analysis of any *S. pneumoniae* isolate, and 2) is quantitative, thereby allowing for gene expression profiling of *S. pneumoniae* strains under *in vitro* conditions. The methodology described can be applied to the construction and modification of an *S.pneumoniae* SGH array, as well as applied to other bacterial species as long as multiple WGSs are available.

Critical Parameters and Troubleshooting

This unit describes the design of a SGH array and its implementation for genomic and transcriptomic analyses. In the design step, it is important to ensure that the probe set capture the genetic diversity of the population of interest. The coverage of the probe set will depend on the number and the phylogenetic distance of the whole genome sequences in the input set. The final probes must capture the differences in gene possession as well as the allelic variations. For probe selection, the user may select any pangenome analysis tool. We describe in detail methods to organize the genomic

content of any number of strains into clusters of highly similar genes for probe design. The genomic analysis and/or transcriptomic profiling require multiple steps from cell growth and nucleic acid extraction to nucleic acid labeling, hybridization and washing. It is imperative that every step be carefully monitored by performing quality control of the output and adding additional control probes.

Anticipated Results

The SGH array can be used to analyze DNA and reveal the genetic content of an isolate, or to analyze cDNA and reveal the gene expression profile of an isolate.

Time Considerations

Once whole genome sequences are selected for the pangenome analysis, the clustering and selection of sequences for probe design can be finalized within 1-2 weeks. The rate- limiting step is the comparison of each sequence to all other sequences, which depends on the number of sequences in the set and the processing power available. If no problems are encountered, the processing of DNA or cDNA to genomic content or gene expression respectively can be achieved in 1 week. The number of samples that can be processed in parallel depends on the design of the chip.

6.3 General Materials & Methods used for experiments in Chapters 2, 3, and 4

Ethics statement

Laboratory animals were maintained in accordance with the applicable portions of the Animal Welfare Act and the guidelines prescribed in the DHHS publication, Guide for the Care and Use of Laboratory Animals. The Office of Laboratory Animal Welfare (OLAW) Assurance of Compliance number is A3693-01. All chinchilla experiments were conducted with the approval of the Allegheny-Singer Research Institute (ASRI) Institutional Animal Care and Use Committee (IACUC) A3693-01/1000. Research grade young adult chinchillas (Chinchilla lanigera) weighing 400-600 grams were acquired from R and R Chinchilla Inc., Ohio. Animals were maintained in BSL2 facilities and all experiments were done while chinchillas were under subcutaneously injected ketamine-xylazine anaesthesia (1.7mg/kg animal weight for each). For virulence studies, chinchillas (a minimum of 10 in each cohort) were infected with 100 CFUs/ear by transbullar inoculation within each middle ear. During the course of the experiment (10 days), animals with severe acute infection perished; animals showing prolonged signs of discomfort were administered with pain relief (Rimadyl, 0.1ml of 50mg/mL)). Animals with severe signs of pain and illness were euthanized by administering an intra-cardiac injection of 1mL potassium chloride after regular sedation. All experiments involving mice were performed with prior approval of and in accordance with guidelines of the St. Jude Institutional Animal Care and Use Committee. The St Jude laboratory animal facilities have been fully accredited by the American Association for Accreditation of Laboratory Animal Care. All mice were maintained in BSL2 facilities and all experiments were done while the mice were under inhaled isoflurane (2.5%) anesthesia. Mice were monitored daily for signs of infection.

This work was approved under the IACUC protocol number 538-100013-04/12 R1. Mice were monitored for disease progression and euthanized via CO2 asphyxiation.

Comparative Genomics

We performed a comparative genomic analysis of PMEN1 and non-PMEN1 strains to identify genes unique to the PMEN1 lineage [95]. To this end, we used a set of 60 curated pneumococcal whole-genome sequences (WGS), including four from the PMEN1 lineage (**S1 Table** Appendix I). The set of 60 genomes includes the 44 genomes used for the first large-scale pneumococcal pangenome study [85], additional genomes from PCV-7 immunized children [170], as well as genomes from non-encapsulated strains[171]. Together 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.

To determine the distribution of *tprA2* across pneumococcal strains we searched for this gene in the genome sequence of 215 PMEN1 isolates [35]. A few genomes displayed disruption in the *tprA2* locus, so the sequences were confirmed by PCR. Primers to *tprA2* and *gapdh* (positive control) were used to amplify these respective genes from genomic DNA. The genomes from strains 111 (ERS004810), 11933 (ERS005313) and HKP38 (ERS004775) display substantial differences in the locus encoding TprA2/PhrA2.

To search for *cre* sites we inspected the 190 basepairs upstream of phrA2 and before the start of tprA2. We searched for the cre site motif from *L. lactis*

(WGWAARCGYTWWMA), and allowed for up to three discrepancies as has been observed in a subset of *S. pneumoniae cre* [117,172].

Bacterial strains and growth conditions

Wild-type *S. pneumoniae* strains PN4595-T23 (GenBank ABXO01) and SV36 (GenBank ADNO01), graciously provided by Drs. Alexander Tomasz and Herminia deLancastre, were used as PMEN1 representatives [99]. Strains 111 (ERS004810), 11933 (ERS005313) and HKP38 (ERS004775) were shared by Drs. Julian Parkhill and Stephen Bentley, and originally obtained from Drs. Lesley McGee, Mark can der Linden, So Hyun Kim and Jae Hoon Song.

For growth on solid media, *S. pneumoniae* (PN4595-T23) and isogenic mutants were streaked on TSA II plates with 5% sheep blood (BD BBL, New Jersey, USA). For growth in liquid culture, colonies from a frozen stock were grown overnight on TSA plates, inoculated into Columbia broth (Remel Microbiology Products, Thermo Fisher Scientific, USA), and incubated at 37°C and 5% CO2 without shaking. Columbia broth contains 10mM glucose. Experiments in chemically defined media (CDM) were performed utilizing previously published recipe [117], and galactose was used at a final concentration of 55mM. Growth in CDM was initiated by growing a pre-culture for 9 hours and back dilution to OD_{600} 0.1 to initiate a culture.

Generation of deletion mutants and complement strains

All deletion mutant strains were generated by site-directed homologous recombination where the target region was replaced with the spectinomycin-resistance gene (*aadR*) or kanamycin-resistance gene, as previously described [95,173] . Briefly, ~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 (provided by Dr. Donald Morrison). Assembly of the transforming cassette was achieved either by sticky-end ligation of restriction enzyme-cut PCR products or by Gibson Assembly using NEBuilder HiFi DNA Assembly Cloning Kit. The resulting construct was transformed into PN4595-T23 and confirmed using PCR and DNA sequencing.

Complement strains were made by generating a cassette where ~100bp of the 5'UTR and the CDS of the gene to be complemented were fused at the 3' end of an antibiotic selection cassette lacking a transcription terminator. This cassette was introduced in the genome of the strain at one of the two regions: the intergenic region between the orthologues of spr_0515 and spr_0516, an inert genomic region that has been successfully employed in other constructs in the lab, or the *bga* region a commonly employed site for complementation [55]. After subsequent transformation, qRT-PCR (LightCycler480, Roche Life Sciences, USA) was done to verify the levels of expression of the complemented gene. Primers used to generate the constructs are listed in **S6 Table** in Appendix I.

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 the appropriate concentration of antibiotic for selection; spectinomycin, 100µg/mL; erythromycin 2µg/mL, kanamycin 150µg/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. The strains generated in this study are listed in **Table 2** in Chapter 2.

Treatment with synthetic peptides

Bacterial cultures were treated with synthetic peptides corresponding to the following sequences: 1) C-terminal PhrA2 heptamer (VDLGLAD); 2) C-terminal PhrA heptamer (LDVGKAD); and 3) scrambled peptide comprised of the same residues as the PhrA2 heptamer (DAGVLDL). These were custom ordered from GenScript, (NJ, USA) at 99.7% purity. 1µM peptide was added in the mid-log phase (OD_{600} of 0.5), cultures were incubated at 37°C, 5% CO2 for 1 hour, after which RNA later (Ambion®,

Thermo Fisher Scientific, USA) was added to the cultures to preserve RNA and subsequent RNA extraction and qRT-PCR were performed.

For experiments where different peptides were compared in parallel, the original culture was distributed into separate tubes, and each one was treated with the relevant peptide, in addition to a no-peptide control. Using a single parent culture for different peptide additions ensured minimal variation when comparing treatments.

Treatment with cell-free supernatant

To determine whether secreted peptides can stimulate gene expression in a recipient wild-type culture, recipient cultures and supernatant donor cultures were grown alongside to selected OD_{600}. To prepare cell-free supernatant, bacterial cells were pelleted and the supernatants were filtered (pore size 0.2 microns). At the desired OD_{600}, the wild-type recipient culture was distributed into separate tubes, cultures were centrifuged at 4000g for 7 minutes, and resuspended in the same volume of cell-free supernatant or media control. At 1 hour post-treatment, RNA later (Ambion®, Thermo Fisher Scientific, USA) was added to each culture, and samples were prepared for RNA extraction and qRT-PCR.

Preparation of cell lysates and RNA collection, extraction, and quality assessment

For experiments on *in vitro* transcriptional analysis, samples were collected for RNA extraction at an OD_{600} of 0.5 unless otherwise stated and preparation of RNA was performed as previously described in [174]. For RNA extraction from *in vivo* experiments, chinchillas were euthanized 48h post-inoculation of PN4595-T23, and a small opening was generated through the bulla to access the middle ear cavity. Effusions were siphoned out from the middle ear and flash frozen in liquid nitrogen to preserve the bacterial RNA. For bacterial cell lyses, the sample were re-suspended in an enzyme cocktail (2mg/mL proteinase K, 10mg/mL lysozyme and 20µg/mL mutanolysin), and submitted to bead beating with glass beads, acid-washed 425-600µm (Sigma) and 0.5mm Disruption Beads made by Zirconia/Silica in FastPrep®-24 Instrument (MP Biomedicals, USA). These cell lysates were frozen for microarray, qRT-PCR or NanoString analyses. The RNA concentration was measured by NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, USA) and its integrity was confirmed on gel electrophoresis.

Microarray analyses of gene expression levels

We utilized the Pneumococcal Supragenome Hybridization Array (SpSGH) to compare gene expression between the wild-type PN4595-T23 strain and the Δ*tprA2* [174]. The array provides coverage for ~85% of the PMEN1 open reading frames. Strains were grown to mid-log cultures (OD_{600} 0.5) in Columbia broth (note, that glucose in the

media will inhibit genes under catabolic repression). RNA extraction, cDNA preparation and cDNA labeling were performed as previously described [174]. Cyber T was used for data analysis [169,175]. Genes with at least a 10-fold difference between strains and Bayesian P values < 0.05, Benjamini-Hochberg FDR < 10%, and Bonferroni-corrected P value < 0.05 are displayed in **Table 2.1**. The complete dataset is deposited in GEO web storage (under submission).

qRT-PCR analyses of gene expression levels

High quality RNA (DNA free and A260/280 ~ 2.1) was used as template for the synthesis of first strand of cDNA using SuperScript VILO synthesis kit (Invitrogen). After first strand cDNA synthesis, the product was directly used for qRT-PCR using LighCycer480 Master Mix SYBRGreen in a LightCycler480 Instrument (Roche Life Sciences, USA). For normalization, we used 16S rRNA, as well as gyrB (DNA gyrase subunit B) and/or gapdh (glyceraldehyde-3-phosphate dehydrogenase). The raw data was converted using LC480 Conversion: conversion of raw LC480 data" software (available at http://www.hartfaalcentrum.nl/index.php?main=files&sub=0) and LinregPCR for expression data analysis [176,177], where the output expression data is displayed in arbitrary fluorescence units (N_0) that represent the starting RNA amount for the test gene in that sample . Statistical significance was determined by performing Student t-test (unpaired samples, one tailed), using GraphPad Prism 6 tool.

Virulence studies in the chinchilla OM model

All chinchilla experiments were conducted with the approval of the Allegheny-Singer Research Institute (ASRI) Institutional Animal Care and Use Committee (IACUC) A3693-01/1000. Research grade young adult chinchillas (Chinchilla lanigera) weighing 400-600 grams were acquired from R and R Chinchilla Inc., Ohio. Animals were maintained in BSL2 facilities and all experiments were done while chinchillas were under subcutaneously injected ketamine-xylazine anaesthesia (1.7mg/kg animal weight for each). For virulence studies, chinchillas (a minimum of 10 in each cohort) were infected with 100 CFUs/ear by transbullar inoculation within each middle ear. During the course of the experiment (10 days), animals with severe acute infection perished; animals showing prolonged signs of discomfort were administered with pain relief (Rimadyl, 0.1ml of 50mg/mL)). Animals with severe signs of pain and illness were euthanized by administering an intra-cardiac injection of 1mL potassium chloride after regular sedation. 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. Additionally, we assessed local diseases using visual otoscopic inspection (VetDock, USA). Otologic disease ranged from no disease to a ruptured tympanic membrane, where a score of '1' is given for animals with mild or no disease, '2' with moderate disease (where pus and air are present), '3' with frank purulence, and '4' with tympanic membrane rupture [84,178].

Virulence studies in the murine lung model

All experiments involving mice were performed with prior approval of and in accordance with guidelines of the St. Jude Institutional Animal Care and Use Committee. The St Jude laboratory animal facilities have been fully accredited by the American Association for Accreditation of Laboratory Animal Care. Laboratory animals were maintained in accordance with the applicable portions of the Animal Welfare Act and the guidelines prescribed in the DHHS publication, Guide for the Care and Use of Laboratory Animals. All mice were maintained in BSL2 facilities and all experiments were done while the mice were under inhaled isoflurane (2.5%) anesthesia. Mice were monitored daily for signs of infection. This work was approved under the IACUC protocol number 538-100013-04/12 R1. For bacterial burden and survival studies, strains were grown in C+Y media to an OD_{620} of 0.4 and diluted according to a previously determined standard curve. Bacteria were enumerated to assure that the proper amount of bacteria was used in infection. Bacteria were introduced into 7-week-old female BALB/c mice (Jackson Laboratory) via intranasal administration of 5 x104 CFU of bacteria in PBS (100 µL). Mice were monitored for disease progression and euthanized via CO_2 asphyxiation. Blood for titer determination was collected via tail snip at 24 and 48 hours post-infection and subsequent serial dilution and plating. Bacteria colonizing the nasopharynx were collected by insertion and removal of PBS (20 µL) into the nasal cavity. One cohort was used for Δ*phrA2-ABC*, Δ*lcpAMT*, and Δ*tprA2*Δ*lcpAMT*, while two cohorts were used for WT and Δ*tprA2* (**Fig. 2.9A, 2.9B**).

Survival data were analyzed using the Mann-Whitney U test in Prism 6. Bacterial titers were compared using nonparametric Mann-Whitney U t test in Prism 6.

Generation of Streptococcal Species Tree

Fifty-five streptococcal strains were selected for phylogenetic analysis (**Table S1**, labeled "Distribution within Streptococcus sp." in Appendix I). The 33 pneumococcal strain were selected to capture the major sequence clusters within this species, including 4 PMEN1 genomes given the focus of this manuscript on this lineage. The *S. mitis* and *S. pseudopneumoniae* strains represented the available genomes for these species at the time this study was initiated. The *S. tigurinus* were selected as a potential novel species related to *S. mitis* [179]. According to our analysis, the *S. tigurinus* genomes and a subset of the S. mitis genomes cluster with *S. oralis*. The whole genome sequence (WGS) for all 55 strains were aligned using MAUVE [180,181] and the core region corresponding to 995531 total sites and 352,371 informative sites, was extracted from the Mauve output files. Alignment of the core region was performed using MAFFT (FFT-NS-2) [182] and model selection was performed using MODELTEST [183]. The phylogenetic tree was built with PhyML 3.0 [184], model GTR+I(0.63) using maximum likelihood and 100 bootstrap replicates.

Gene Distribution Analysis and Generation of TprA2/TprA Gene Tree

To identify genes that are highly enriched within the PMEN1 lineage relative to other pneumococcal lineages we clustered the coding sequences from 60 highly curated pneumococcal whole genome sequences (WGS), and selected clusters unique to the PMEN1 genomes. The 60 genomes are listed in **Table S1** and marked as "To establish PMEN1 enrichment", and the analysis has been previously described in detail [95]. Briefly, it involved CDS prediction by RAST [26], CDS clustering by utilizing tfasty36 (FASTA v.3.6 package) [164] and parsing the output to assemble genes that share at least 70% identity over 70% of their length into clusters of homologous sequences, and selecting clusters that are present in all PMEN1 genomes while absent in all other lineages.

To establish the gene presence/absence profiles within the 215 PMEN1 WGSs we performed an in silico PCR on the genomes previously published by Croucher and colleagues at the Sanger Center (listed in **Table S1** [35]). In cases where the in silico analysis was inconclusive, we performed experimental PCR using forward and reverse primers to *tprA2*. To establish the gene presence/absence profiles within the 55 Streptococcal WGSs (**Table S1**, strains labeled as "Distribution within Streptococcus sp."), as displayed in **Fig 2.1B**, we employed the basic local alignment search tool (Blastn) using an e-value threshold of 1e-20 [168]. All of the *tprA2* CDSs displayed >= 95% similarity. The Lan locus is represented by three CDSs downstream of TprA2/PhrA2, and the Lan* locus is represented by seven CDSs downstream of

TprA/PhrA; the genes with Lan and Lan* display exactly the same phylogenetic distribution in the 55 samples (i.e all present or all absent). In the vast majority of the genomes, the lantibiotic genes were neighboring the associated QS systems; the exceptions are genomes with contig breaks or low sequence coverage in these regions (these are noted in **Fig 2.2**).

The phylogenetic tree of *tprA2/tprA* was generated on the 48 sequences extracted in the analysis of the 55 streptococcal genomes. The nucleotide sequences were aligned using MAFFT (G-INS-i), and model selection was performed using MODELTEST. The phylogenetic tree was build with PhyML 3.0, model HKY+I(0.39) using maximum likelihood and 100 bootstrap replicates. Logos were generated from the C-terminal heptapeptides of (i) 6 PhrA2 sequences and (ii) 36 PhrA peptides using WebLogo [73] (**Fig 4.1A and 4.1B**).

Gene tree construction and reconciliation analysis

TprA gene sequences were extracted from the 55 genomes using BLASTn; all *tprA* orthologues displayed a minimum of 86% similarity. A maximum likelihood tree was generated using RAxML. The tree branches with bootstrap support value less than 65, were rearranged using Notung-2.8.1.2-beta [113,114] The trees were then reconciled under DTL (duplication, transfer, and loss) model using Notung-2.8.1.2-beta, with a cost of transfer, duplication, and loss of 3,1.5 and 1 respectively.

Chapter 7. Discussion

Bacterial strategies mediating survival and propagation inside the complex environment of the human host continue to be a subject of intrigue and active investigation. The center stage in this problem is occupied by the binary lifestyles of opportunistic pathogens that are equipped with genomic assets to colonize the host as commensals or to afflict damage and fatality to the host. In this manner, opportunistic pathogens encounter decision-making checkpoints during an infection, such as the choice between local attachment and dissemination to peripheral tissues that ultimately results in colonization or systemic disease, respectively [185,186]. The outcome depends on factors such as available host sugars and achievement of a reliable bacterial quorum, to enumerate a few. A well-organized bacterial command center would therefore comprise of sensory components that efficiently integrate multiple input signals, executors that modulate gene expression, and effectors that translate the input into appropriate molecular and phenotypic responses.

Several Gram positive bacteria employ peptide-regulated transcriptional factor systems to perform these functions. These regulators belong to the RRNPP superfamily. Small peptides (pheromones) perform signaling function by transmitting information

from the extracellular milieu and interacting with a cognate transcriptional regulator that leads to modification of gene expression to best support the need of the bacteria. Recently, the wealth of genomic data has been tapped to reveal the identity and the evolutionary history of novel members of the RRNPP family [70].

TprA/PhrA-like systems in *Streptococcus pneumoniae*

Recent reports on *Streptococcus pneumoniae* have shed light on two related peptide-regulated systems in pneumococcus the TprA/PhrA and the TprA2/PhrA2 [74,187]. Both regulatory systems consist of transcriptional regulators TprAs and activating peptides PhrAs. TprAs are negative regulators of gene expression, while PhrAs reverse TprA-mediated gene inhibition presumably by binding to and withdrawing TprAs from the DNA. PhrAs are quorum sensing peptides with active C-terminal portions and are autoinducing in nature. The architecture of TprA/PhrA and TprA2/PhrA2 genomic loci are similar in that both encode for lanthionine containing peptides downstream. A key difference between TprA/PhrA and TprA2/PhrA2 systems is their phylogenetic distribution. TprA/PhrA is encoded in almost all pneumococcal strains. In stark contrast, TprA2/PhrA2 is rare across the species; it is present almost exclusively in the clinically important PMEN1 lineage. A distribution that is consistent with a horizontal gene transfer from outside the species into the ancestral strain of the PMEN1 lineage. With the acquisition of TpA2/PhrA2 the PMEN1 strains possess both of these quorum sensing systems. This raises questions on the

function and molecular mechanism ofTprA2/PhrA2, and provides a model to study interactions between related QS systems.

Role of TprA/PhrA in nutrition sensing

Like most pathogens, pneumococcus encounters variable microenvironments in the host regarding carbon sources used to derive energy for survival. The *in vivo* concentration and the chemical composition of sugar sources changes based on tissue sites, vicinity to circulation, as well as the diet and lifestyle of the host. For example, galactose is the primary sugar in the nasopharynx and glucose and mannose can be abundant sugars in the circulation [93,109,188]. An interesting aspect of TprA/PhrA control is its regulation by sugars. Yesilkaya and colleagues show that TprA is regulated by CcpA and GlnR, the regulators for catabolite control through catabolite response element (*cre*) sites and glutamine synthesis respectively. CcpA inhibits TprA while GlnR activates TprA. In addition, they found that PhrA is activated by galactose and mannose and repressed in glucose and N-acetylglucosamine while TprA inhibits its own expression under galactose and mannose conditions. The crucial role of sugars in regulation of TprA and PhrA is further supported by the observation that *tprA* and *phrA* deletion mutants are significantly attenuated in galactose and mannose conditions. In this way, a busy interactome with input from multiple regulators and multiple nutritional components reveals a complex TprA/PhrA regulatory circuit that plays a central role in utilization of multiple derivatives of sugars inside the host, a decision of

grave importance for the bacteria. Together, an assessment of the combined regulatory landscape governing the activity of the TprA/PhrA system supports the role of this system as a nutrition sensor and responder.

Role of TprA/PhrA in virulence

Availability of different simple and complex sugars inside the host is an important determinant of bacterial survival and propagation. Experiments demonstrating regulatory dependence of TprA/PhrA on sugars strongly suggest that this system plays a role during host infection. D39 parental strains with mutation in TprA/PhrA components (D39Δ*phrA* or D39Δ*tprA*) display a highly attenuated phenotype in murine models of pneumococcal septicemia and pneumonia infection (Motib and Yesilkaya, personal communication), as well as the chinchilla model of middle ear disease (Chapter 3). Thus TprA and PhrA are virulence factors.

A second noteworthy observation, is that neither Δ*phrA* or Δ*tprA* survive in blood in both direct inoculation into the bloodstream or bacteraemia developed from intranasal infections. Thus, TprA and PhrA appear to play key role in blood infection and initiation or maintenance of systemic infection. Given that TprA and PhrA are differentially expressed in response to various host sugars, it is likely that this reflect the particular nutrient composition of the murine blood.

TprA2/PhrA2 system in PMEN1 strains

The TprA2/PhrA2 is a peptide-regulated system, unique to the PMEN1 strains. As characterized in Chapter 2, TprA2 plays a role in promoting commensalism over tissue dissemination. It also regulates a novel lanthionine-containing peptide, LcpA, which is a virulence determinant as demonstrated using the mouse model.

Epidemiologically, PMEN1 strains do not exhibit heightened virulence potential, instead their success is attributed to wide spread distribution and moderate virulence. [37,40,41,156]. Person-to-person spread of pneumococcal infection occurs from the upper respiratory tract. Thus, stabilization of nasopharyngeal colonization by TprA2 may contribute to widespread transmission of PMEN1 strains.

It is intriguing that TprA2 and LcpA appear to have seemingly contradictory functions, yet they were acquired by the PMEN1 strain on the same mobile element. This suggests that the unique ability of PMEN1 strains in fact lies in carrying out strategic balance of the activities of the commesalism promoter TprA2, its activator peptide PhrA2 and virulence factor LcpA.

TprA2/PhrA2 and TprA/PhrA in PMEN1 strains

The TprA2/PhrA2 and TprA/PhrA systems co-occur in the PMEN1 strains. Our data show that PhrA2 induces gene expression of TprA/PhrA however, PhrA does not effect levels of TprA2 or PhrA2. Additionally, unlike the TprA/PhrA system of strain

D39 that is inhibited under glucose, neither TprA/PhrA nor TprA2/PhrA2 of PMEN1 exhibit glucose inhibition. The difference in TprA/PhrA regulation between strains could be due to direct interaction of TprA2/PhrA2 on the TprA/PhrA system or other changes in the regulatory circuits in PMEN1 strains. The contribution of PhrA2 into the phenotype of TprA/PhrA in PMEN1 strains is further suggested in virulence studies in chinchilla otitis media model where Δ*tprA* of PMEN1 has only a moderate and non-significant influence on virulence (42% survival in WT vs 60% survival in Δ*tprA* cohort) whereas Δ*tprA* of D39 abrogates virulence in this model (0% survival in WT vs 90% survival in Δ*tprA* cohort). We propose that, PMEN1 strains exhibit a novel regulatory circuit wherein TprA/PhrA system occurs downstream of TprA2/PhrA2 and each of the components may be regulated by additional factors.

The acquisition of the TprA2/PhrA2 system may have contributed to the ecological success of the PMEN1 strains through multiple ways. One possible mechanism could be to perform the attenuation of virulence caused by solely TprA/PhrA system, thereby dissuading extreme invasive diseases in PMEN1 strains and supporting commensalism. Additionally, TprA2/PhrA2 may act to relieve the stringency of the TprA/PhrA system towards specific sugars, as is suggested by the sugar-independence of the TprA/PhrA system of the PMEN1 strains. TprA/PhrA plays a role in utilization of host sugars and therefore, a relaxed regulatory control in the PMEN1 strains could prove beneficial.

Inter-strain communication via PhrA2

PhrA2 peptide from PMEN1 can induce gene expression of TprA/PhrA system in the non-PMEN1 strain D39 where TprA/PhrA system is a virulence determinant (Chapter 3). This opens the door for inter-strain communication mediated by PhrA2 in multi-strain infections. Can PhrA2 from PMEN1 increase the virulence of non-PMEN1 strains in multi-strain infections? The differential regulatory effect of glucose on TprA/PhrA of non-PMEN1 and TprAs/PhrAs of PMEN1 presents a physiological condition where peptide from PMEN1 can modify the transcriptional landscape of the pneumococcal community inside the host and influence the net outcome of disease.

The etiology of various human diseases such as diabetes, stress and inflammatory responses lead to elevation of glucose levels [189–191]. Therefore, I hypothesize that multi-strain pneumococcal infections that include PMEN1 strain may acerbate the outcome of pneumococcal infection under systemic or local hyperglycaemic conditions. The mechanism of multi-strain infections and their impact on the outcome of infection and clinical interventions is understudied, as is the mechanism of interactions between strains. This work provides strains and molecules to test the exciting hypothesis that inter-strain interactions can lead to disease outcomes from multi-strain infections that differ from those of the single-strain infections

Appendix

Supplementary information from manuscripts

Table S1. *S. pneumoniae* strains utilized for pangenome analysis.Bold: PMEN1 strains.				
Strain ID	**Species**	**GenBank Accession No.**	**Analyses performed**	
SK1076	*Streptococcus infantis*	AFNN00000000	Distribution within Streptococcus sp.	
SK970	*Streptococcus infantis*	AFUT00000000	Distribution within Streptococcus sp.	
ATCC 6249	*Streptococcus mitis*	AEEN00000000	Distribution within Streptococcus sp.	
B6	*Streptococcus mitis*	FN568063	Distribution within Streptococcus sp.	
bv. 2 str. F0392	*Streptococcus mitis*	AFUO00000000	Distribution within Streptococcus sp.	
bv. 2 str. SK95	*Streptococcus mitis*	AFUB00000000	Distribution within Streptococcus sp.	
NCTC 12261	*Streptococcus mitis*	AEDX00000000	Distribution within Streptococcus sp.	
SK1073	*Streptococcus mitis*	AFQT00000000	Distribution within Streptococcus sp.	
SK1080	*Streptococcus mitis*	AFQV00000000	Distribution within Streptococcus sp.	
SK321	*Streptococcus mitis*	AEDT00000000	Distribution within Streptococcus sp.	
SK564	*Streptococcus mitis*	AEDU00000000	Distribution within Streptococcus sp.	
SK569	*Streptococcus mitis*	AFUF00000000	Distribution within Streptococcus sp.	
SK575	*Streptococcus mitis*	AICU01000001	Distribution within Streptococcus sp.	
SK597	*Streptococcus mitis*	AEDV00000000	Distribution within Streptococcus sp.	
Uo5	*Streptococcus oralis*	FR720602	Distribution within Streptococcus sp.	
PN4595-T23	*Streptococcus pneumoniae*	ABXO00000000	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
70585	*Streptococcus pneumoniae*	CP000918	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
ATCC 700669	*Streptococcus pneumoniae*	FM211187	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
CDC0288_04	*Streptococcus pneumoniae*	ABGF00000000	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
CDC1087-00	*Streptococcus pneumoniae*	ABFT00000000	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
CDC1873-00	*Streptococcus pneumoniae*	ABFS00000000	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
CDC3059-06	*Streptococcus pneumoniae*	ABGG00000000	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
CGSP14	*Streptococcus pneumoniae*	CP001033	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
D39	*Streptococcus pneumoniae*	CP000410	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
G54	*Streptococcus pneumoniae*	CP001015	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
GA13494	*Streptococcus pneumoniae*	AGOZ01000001	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
Hungary19A-6	*Streptococcus pneumoniae*	CP000936	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
ICE44	*Streptococcus pneumoniae*	AUYF00000000	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
INV104	*Streptococcus pneumoniae*	FQ312030	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
JJA	*Streptococcus pneumoniae*	CP000919	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
MNZ14	*Streptococcus pneumoniae*	ASJO00000000	Distribution within Streptococcus sp.	
MNZ85	*Streptococcus pneumoniae*	ASJF00000000	Distribution within Streptococcus sp.	

SP11-BS70	*Streptococcus pneumoniae*	ABAC00000000	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
SP14-BS69	*Streptococcus pneumoniae*	ABAD00000000	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
SP18-BS74	*Streptococcus pneumoniae*	ABAE00000000	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
SP19-BS75	*Streptococcus pneumoniae*	ABAF00000000	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
SP23-BS72	*Streptococcus pneumoniae*	ABAG00000000	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
SP3-BS71	*Streptococcus pneumoniae*	AAZZ00000000	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
SP6-BS73	*Streptococcus pneumoniae*	ABAA00000000	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
SP9-BS68	*Streptococcus pneumoniae*	ABAB00000000	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
SPN1041	*Streptococcus pneumoniae*	CACE00000000	Distribution within Streptococcus sp.	
SPNA45	*Streptococcus pneumoniae*	CACG00000000	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
SV35-T23	*Streptococcus pneumoniae*	ADNN00000000	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
SV36-T3	*Streptococcus pneumoniae*	ADNO00000000	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
Taiwan19F-14	*Streptococcus pneumoniae*	CP000921	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
TIGR4	*Streptococcus pneumoniae*	AE005672	Distribution within Streptococcus sp.	To establish PMEN1 enrichment
WL400	*Streptococcus pneumoniae*	AVFA00000000	Distribution within Streptococcus sp.	
WL677	*Streptococcus pneumoniae*	AUWZ00000000	Distribution within Streptococcus sp.	
ATCC BAA-960	*Streptococcus pseudopneumoniae*	AICS00000000	Distribution within Streptococcus sp.	
IS7493	*Streptococcus pseudopneumoniae*	CP002925	Distribution within Streptococcus sp.	
SK674	*Streptococcus pseudopneumoniae*	AJKE00000000	Distribution within Streptococcus sp.	
2426	*Streptococcus tigurinus*	ASXA00000000	Distribution within Streptococcus sp.	
AZ_3a	*Streptococcus tigurinus*	AORU00000000	Distribution within Streptococcus sp.	
INV200	*Streptococcus pneumoniae*	FQ312029	Distribution within pneumococcus	To establish PMEN1 enrichment
ST13v1	*Streptococcus pneumoniae*	NZ_ABWQ00000000	Distribution within pneumococcus	To establish PMEN1 enrichment
ST13v12	*Streptococcus pneumoniae*	NZ_ABWU00000000	Distribution within pneumococcus	To establish PMEN1 enrichment
ST13v6	*Streptococcus pneumoniae*	NZ_ABWB01000000	Distribution within pneumococcus	To establish PMEN1 enrichment
AP200	*Streptococcus pneumoniae*	CP002121	Distribution within pneumococcus	To establish PMEN1 enrichment
MLV016	*Streptococcus pneumoniae*	ABGH00000000	Distribution within pneumococcus	To establish PMEN1 enrichment
ICE59	*Streptococcus pneumoniae*	AUYE00000000	Distribution within pneumococcus	To establish PMEN1 enrichment
Sp647	*Streptococcus pneumoniae*	AUYL00000000	Distribution within pneumococcus	To establish PMEN1 enrichment
Sp6706B	*Streptococcus pneumoniae*	CP002176	Distribution within pneumococcus	To establish PMEN1 enrichment
SPAIN6B	*Streptococcus pneumoniae*	AUYK00000000	Distribution within pneumococcus	To establish PMEN1 enrichment
CCRI_1974M2	*Streptococcus pneumoniae*	ABZT00000000	Distribution within pneumococcus	To establish PMEN1 enrichment
CCRI_1974	*Streptococcus pneumoniae*	ABZC00000000	Distribution within pneumococcus	To establish PMEN1 enrichment
SP195	*Streptococcus pneumoniae*	ABGE00000000	Distribution within pneumococcus	To establish PMEN1 enrichment

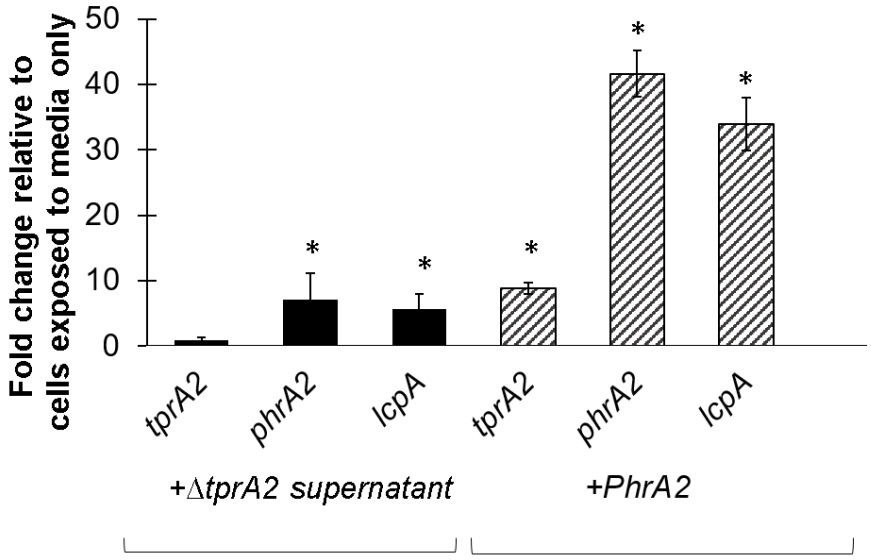
Spain1417	*Streptococcus pneumoniae*	AUYH00000000	Distribution within pneumococcus	To establish PMEN1 enrichment
Sp88-1	*Streptococcus pneumoniae*	AUYI00000000	Distribution within pneumococcus	To establish PMEN1 enrichment

Gene ID	Annotation	Fold change (Δ*tprA2*/WT)	*P*-value	Bonferroni	Benjamini-Hochberg

Sp439-1	*Streptococcus pneumoniae*	AUYJ00000000	Distribution within pneumococcus	To establish PMEN1 enrichment
SPAIN9V	*Streptococcus pneumoniae*	AUYG00000000	Distribution within pneumococcus	To establish PMEN1 enrichment
N034156	*Streptococcus pneumoniae*	FQ312045	Distribution within pneumococcus	To establish PMEN1 enrichment
N034183	*Streptococcus pneumoniae*	FQ312043	Distribution within pneumococcus	To establish PMEN1 enrichment
N994038	*Streptococcus pneumoniae*	FQ312041	Distribution within pneumococcus	To establish PMEN1 enrichment
N994039	*Streptococcus pneumoniae*	FQ312044	Distribution within pneumococcus	To establish PMEN1 enrichment
OXC141	*Streptococcus pneumoniae*	FQ312027	Distribution within pneumococcus	To establish PMEN1 enrichment
SPN021198	*Streptococcus pneumoniae*	CACH01000000	Distribution within pneumococcus	To establish PMEN1 enrichment
SPN072838	*Streptococcus pneumoniae*	CACI01000000	Distribution within pneumococcus	To establish PMEN1 enrichment
SPN1041	*Streptococcus pneumoniae*	CACE01000000	Distribution within pneumococcus	To establish PMEN1 enrichment
P1031	*Streptococcus pneumoniae*	CP000920	Distribution within pneumococcus	To establish PMEN1 enrichment
N032672	*Streptococcus pneumoniae*	FQ312039	Distribution within pneumococcus	To establish PMEN1 enrichment
N033038	*Streptococcus pneumoniae*	FQ312042	Distribution within pneumococcus	To establish PMEN1 enrichment
SPN061370	*Streptococcus pneumoniae*	CACJ01000000	Distribution within pneumococcus	To establish PMEN1 enrichment
R6	*Streptococcus pneumoniae*	AE007317	Distribution within pneumococcus	To establish PMEN1 enrichment
N7465	*Streptococcus pneumoniae*	CACF01000000	Distribution within pneumococcus	To establish PMEN1 enrichment
ST2011v4	*Streptococcus pneumoniae*	NZ_ADHN01000000	Distribution within pneumococcus	To establish PMEN1 enrichment

SPN23F_12701	*lcpA*	62.11	0	0	0
CGSSp4595_1258	putative ABC transporter, permease protein	58.08	0	0	0
CGSSp4595_1257	*lcpM*	45.39	0	0	0
CGSSp4595_1256	*lcpT*	40.38	0	0	0
CGSSp4595_1260	ABC transporter, ATPase	34.70	0	0	0
CGSSp4595_1255	FIG01114468: hypothetical protein	33.68	0	0	0
CGSSp4595_1261	*phrA2*	32.56	0	0	0
CGSSp4595_1259	hypothetical protein	31.72	0	0	0
CGSSp4595_1253	unknown	31.62	0	0	0
CGSSp4595_1254	FIG01116415: hypothetical protein	28.36	0	0	0
CGSSp4595_1267	SNF2 family protein	6.99	0	0	0
CGSSp4595_1263	FIG01118149: hypothetical protein	6.69	0	0	0
CGSSp4595_1266	Retron-type RNA-directed DNA polymerase (EC 2.7.7.49)	4.48	1.06E-13	2.04E-10	9.71E-12
CGSSp4595_1264	unknown	4.24	0	0	0
CGSSp4595_1270	conserved domain protein	2.76	1.89E-06	0.0036369	5.87E-05
CGSSp4595_0698	hypothetical protein	2.28	0.00012	0.2259078	0.003094627
CGSSp4595_1274	FIG01114970: hypothetical protein	2.27	4.82E-05	0.0930484	0.001348527
CGSSp4595_1698	Transcriptional regulator, GntR family	2.24	3.28E-10	6.32E-07	1.41E-08
CGSSp4595_1552	FIG01114360: hypothetical protein	2.21	9.73E-12	1.88E-08	5.69E-10
CGSSp4595_1549	DNA-cytosine methyltransferase (EC 2.1.1.37)	2.20	2.88E-11	5.56E-08	1.54E-09
CGSSp4595_0276	Transposase of IS657	2.18	0.00022	0.4150853	0.005321607
CGSSp4595_0297	putative ATP-dependent Clp proteinase (ATP-binding subunit)	2.11	1.05E-05	0.0201752	0.000296694
CGSSp4595_0367	Macrophage infectivity potentiator protein	2.07	5.61E-07	0.0010831	1.84E-05
CGSSp4595_1674	probable beta-D-galactosidase	2.04	0.00273	1	0.039885863
CGSSp4595_1273	Tn5252, Orf23	2.03	0.00128	1	0.02271382
CGSSp4595_1531	putative large terminase subunit	2.02	1.12E-10	2.16E-07	5.14E-09
CGSSp4595_1550	Single-stranded DNA-binding protein	2.01	3.13E-11	6.05E-08	1.59E-09
CGSSp4595_0800	Type I restriction-modification system, specificity subunit S (EC 3.1.21.3)	-4.23	0	0	0
CGSSp4595_1262	transcriptional regulator, *tprA2*	-18.10	0	0	0

Table S2. List of genes with at least 2 fold difference in expression levels between wild-type PN4595-T23 and the isogenic Δ*tprA2* strains



S1 Fig. PhrA2 modulates the expression levels of the TprA2 regulon. qRT-PCR measurements in gene expression of QS-lcp genes in WT strain PN4595-T23. Data was normalized to 16S rRNA expression. Y-axis displays fold change in gene expression, upon exposure to supernatant from□□*tprA2* cultures or synthetic PhrA2, relative to media-only control. Error bars represent standard deviations for biological replicates (n=3). Mid-log WT cells where split into three groups, and were submitted to treatment with media alone, cell-free supernatant from □*tprA2* cultures or, as a positive control, PhrA2 C-terminal heptapeptide (VDLGLAD). On the left, dark bars represent the fold change between addition of cell-free supernatant from □*tprA2* cultures relative to addition of media only. On the right side, striped bars represent the fold change between addition of PhrA2 C-terminal heptapeptide (VDLGLAD) relative to addition of media only. Both the culture supernatant and the PhrA2 heptapeptide lead to upregulation of *phrA2* and *lcpA*. * Statistically significant difference in gene expression (*P*-value<0.05).

Table S3. *In vivo* phenotype of PN4595-T23 WT and isogenic mutants in a chinchilla model of pneumococcal disease

	Strains & Phenotype			*P* -values		
Strain	PN4595-T23 (WT)	Δ*phrA2-ABC*	Δ*tprA2*	WT vs Δ*phrA2-ABC*	WT vs Δ*tprA2*	Δ*tprA2* vs Δ*phrA2-ABC*
Mortality	14/19 (74%)	7/10 (70%)	7/10 (70%)	NS	NS	NS
Bacteria in the Brain	18/19 (95%)	8/10 (80%)	8/10 (80%)	NS	NS	NS
Bacteria in the Lung	10/19 (52%)	2/10 (20%)	8/10 (80%)	0.048	NS	0.0041

Table S4: A list of the longest sequence from each pneumococcal subcluster. These sequences were used to design upto 10 probes/subcluster. This is a preview table, due to limited space. The final table has 4452 subcluster rows.

ClusterID	Type of Clus	Sequence
cluster100	core	ATGGTGACCTTAGTAGATAAATTTGTAACACATGTCATTTCTGAAAGTTCATTTGAGGAAATGGATCGAATCTACCTGACCAATCGTGTTTTGGCACGAGTGGGAGAA
cluster1014a	core	ATGAAACATTTTTTTGCAGGAATTGGTGAAATAAACTTTGTAAGTTATTTATTGTATATATGTGTAGGAATAGCCCCTTTTTTTCACGTATATATAATAGGTTCTGAAAT
cluster1014b	core	ATGAAACATTTTTTTGCAGGAATTGGTGAGATAAGATTTAGTAGTTATCTATTGTCTCTATGTGTAGGACTAGCTCCTCTTTTTCATATCTATGTAATAGGCTCTGAAAT
cluster1019	core	GTGACGGAGGAATTTATGAATGTAAATCAGATTGTACGGATTATTCCTACTTTAAAAGCTAATAATAGAAAATTAAATGAAACATTTTATATTGAAACCCTTGGAATGA
cluster1020	core	ATGAAAGACTTGTTTTTAAAGAGAAAGCAGGCCTTTCGTAAGGAGTGTCTTGGTTATCTGCGCTATGTGCTCAATGACCACTTTGTCTTGTTCCTGCTTGTCCTGTTGG
cluster1021	core	ATGGCCCAATCACTTAACAAAACAGTACTCCTCAGTACGACAGGCACTTCCTACCTCTCTATCGCTGGGAAAGTTGGGAAATTCCTTGTCGGAGATCAGGCTTTGGAA
cluster1022	core	GTGTTCACCTCCGTAATTATTTTGCTTGGGGAATTTTTCAAAAAGAAAAACGCGCCCAAACCTAGACACGAAAGTACAATACGCTTCTTTTTACATGATACGTCTTGTC
cluster1025	core	ATGAAAAAATTAGTCTTTGTCTGTCTGGGAAATATTTGCCGTAGCCCTATGGCCGAGTTTGTTATGAAATCAATGACAGATAACTACGAAATCCAAAGTCGAGCAACTT
cluster1026	core	gtggaaaatcttacgaattttacgaaaagtatcgtgtctatctgactcgtccacgtttagagctttggcagtagttaccattgtttttctgtgctgtactcgtctttttttctaaatattccaggaaaaggtgtcttaaaactc
cluster1027a	core	ATGAAGGAACTTTTAAACAAGGCGTTTTTCAACAAAAATAAAGCCTCCTTATCAAAGGAGGTATTATTAGAATCGCAAGGTAAGCGCTTACTTGCAAATCCCTTTCTTT
cluster1027b	core	ATGAAGGAACTTTTAAACAAGGCGTTTTTCAACAAAAATAAAGCCTCCTTATCAAAGGAGGTATTATTAGAATCACAAGGTAAGCGCTTACCTGTAAATCCCTTTCTTT
cluster1028	core	ATGGAGGAAAATATGGCTGATAAAAAAACTGTGACACCAGAGGAAAAGAAACTCGTTGCTGAAAAACACGTAGATGAGTTGGTTCAAAAAGCTCTAGTTGCCCTTGA
cluster1029	core	ATGCAACAAAGTGCGACTGAACGTGGGATTGAAATCTCTATCCAAGCAAAATCAATGACAGAAGCTAAAAAGAATATCCATGAAGCAGATGTCATTCTAATTGGGCCA
cluster1030	core	ATGAAAAACTCAAAATTTATAGATCAATTTGCCACCTTTGCTGGTAAACTAGGGAACCAAATTCATTTAAAAACCCTAAGAGATGCATTCGTAACAGTAATGCCATTAT
cluster1031	core	ATGAATCCAAATATTACTTTTTTAATCATGCTTGTAGGTATGATGGCCTTGATGTTCTTTATGCAACGTTCTCAAAAGAAACAAGCTCAAAAGCGTATGGAAAGCTTAA
cluster1032	core	ATGGCTACTATTCAATGGTTTCCTGGTCACATGTCTAAAGCTCGTCGACAGGTGCAGGAGAATTTAAAATTTGTTGATTTTGTGACGATTTTAGTAGATGCACGCTTGC
cluster1033a	core	ATGGCGACGATTAAAGAAATCAAAGAACTACTTGCTACAGTCAAGGACCTTGATAACCCTATTTTTTTAGATCTTGAAAAGGACAATCGCTCTGGAGTTCAAAAGGAA
cluster1033b	core	ATGGCAACGATTAAAGAAATCAAAGAACTCCTTGTGACAGTCAAGGAGTTAGAAAGCCCTATTTTTTTAGAGCTTGAAAAGGATAATCGCTCAGGAGTTCAAAAGGAA
cluster1034	core	TTGAAAGGAAATACCATGGAGGAACAGTCGGAAATAGTCCGTTCTAAGAAAGAATTCGCCTTTGCATCCAGCACTATACTATCCCAAGTTGGTCGAGGAATCATTGTC
cluster1035	core	ATGAAACGTGAGATTTTACTGGAACGAATCGACAAACTAAAACAACTCATGCCCTGGTATGTTCTGGAATACTACCAATCTAAGCTGGCTGTACCCTACAGTTTTACAA
cluster1040a	core	ATGGTTAAATACGGTGTTGTTGGAGCAGGGTATTTTGGAGCTGAATTGGCTCGCTACATGCAAAAGAATGATGGAGCAGAGATTACTCTTCTCTATGATCCAGATAAT
cluster1040b	core	TTGAAAATATATAAAAAGAGGATAAAAATTATGGTAAATTACGGTATTGTTGGAGCTGGATATTTTGGAGCTGATTTAGCTCGCTCAATGAACAAAATTGAAGATGCAA
cluster1041a	core	ATGCTTGAAAGACTGAAAAGCATACATTATATGTTTTGGATCAGTTTAATTTTTATGATTTTCCCCATCCTACCTGTAGTGACTGGGTGGCTTTCTGCCTGGCATTTATT
cluster1041b	core	ATGCTTGAAAGACTGAAGAGCATAGACTATATGTATTGGGCCAGTTTGATTTTTATGCTTTTTCCCATCCTTCCTGTAGTGATTGGGGAGCTTGCTGTCTGGCATTTAT
cluster1041c	core	ATGCTTGAAAGACTGAAAAGAATACATTATATGTTTTGGATCAGTTTAATTTTTATGATTTTCCCCATCCTGCCTGTAGTGACTGGGTGGCTTTCTGCCTGGCATTTATT
cluster1041d	core	ATGCTTGAAAGACTGAAAAGCATAGACTATATGTATTGGACCAGTTTGATTTTTATGGTTTTTCCCATCGTTCCTGTAGTGACTGGGGAGCTTCCTAGCTGGCATTTGT
cluster1041e	core	ATGCTTGAAAGACTGAAAAGCATACACTATATGTTTTGGGCCAGTTTGATTTTTATGCTTTTTCCCATCCTACCTGTAGTGATTGGGGAGCTTCCTGCTTGGCATTTATT
cluster1041f	core	ATGCTTGAAAGACTGAAAAGCATACACTATATGTTTTGGGCCAGTTTAATTTTTATGATTTTCCCCATCCTACCTGTAGTGACTGGGTGGCTTTCTGCCTGGCATTTATT
cluster1042	core	TTGAAGTCAATAATCGTAGCTTGTTGGATACAATCTTTGAAGAAACACAAAAGGGAGATGACTAAGATGAAACGATGGATCGCACTAAACAAGATAGAATTTCTATTG
cluster1044	core	TTGCTTTACTTCTATAGTGAAAAGGTTTACAATTATAGTAAGAAAATTTCAGGAGGTTTCATTATGGCACAACGTTACCAAAATATCATGGTCGCAATCGATGGTTCTA
cluster1045	core	ATGTGGTATAATCATTGTATCACAAAAGGGGAGATTGGGATAGGATGGAAGTCAAAGCTGTTTTTTTTTGATATCGATGGAACCTTGGTCAACAATCGCAAGAGTGTT
cluster1046	core	ATGAAAGAATATAACAAGTCTAGTAAGTTAGAGCATGTTGCTTATGATATCCGTGGTCCTGTATTGGAAGAAGCCATGCGGATGCGAGCGAATGGAGAAAAGATTTTA
cluster1047	core	ATGATGAAAAACTCTTTTCAAAAATCAAATTTCCTTTATTATGGAATTATCCTAGTTTCAGTCTTGGTAGAAGTGATCATGATCTGTCAATTAGAGAGTCTGGTACCGCT
cluster1048a	core	ATGTGTGAGATAGAAAGGACAGGAAAAGGAATGAATGCAGATGATACAGTAACCATTTATGATGTCGCTCGTGAAGCAGGTGTTTCCATGGCGACGGTCAGCCGTGT
cluster1048b	core	ATGCATNTTGATCTTGGTGCCATTAGTATGCGTATGTTGACCAAGATTATGCATAAGGAAGAGTTGGAAGAACGTGAAGTTCTCTTACCTCATGGTTTGACAGAACGT
cluster1049	core	ATGAAAGTATTAGTCGCAGAAGATCAAAGTATGTTGCGAGATGCCATGTGTCAATTGCTCACGCTTCAACCAGATGTGGAGTCTGTCCTTCAAGCCAAGAATGGGCAA
cluster105	core	ATGTTAAAAATTGGTACAGCTTGTGGTTCAGGATTAGGTTCAAGTTTTATGGTACAGATGAATATTGAATCTGTATTGAGTGATTTGAATGTTTCGGATGTAGAAGTTG
cluster1050	core	ATGCCTAAGAAAATCCTTGTTTTACATACGGGTGGAACTATTTCCATGCAGGCCGATGCTTCTGGCGCTGTTGTGACGAGTTCAGATAATCCCATGAACCATGTGTCC
cluster1052	core	GTGAAATCAATGCATCCCAACATGGAAAGGAGGACTAGAATGAAACGACAAACTGTAAACCAGACGCTCAAACGTTTAGCCGTAGATTTAGCAAGCCATCCTTTCCTC
cluster1053	core	ATGAAACACCTATTATCTTACTTCAAACCCTACATCAAGGAATCAATTTTAGCCCCCTTGTTCAAGCTGTTAGAAGCTGTTTTTGAGCTCTTGGTTCCCATGGTGATTGC
cluster1055	core	TTGGTATATGCCTCTTATTTAAGAGTAACTGAAGCTCCAGCTTCTTCCAATTTAGCTTTGATTTCTTCAGCTTCTGCAGTTGCAACGCCTTCTTTAACAAGTGCTGGTGC
cluster1056	core	TTGGGCGCGTTTTTCTTTTTGAAAAATTCCCCAAGCAAAATAATTACGGAGGTGAACACACTAATGAGTGAAGCAATTATTGCTAAAAAAGCGGAACTAGTTGACGTAC
cluster1059	core	ATGAAAAAAATTGTCTTTACAGGTGGGGGGACGGTTGGACACGTTACCCTCAATCTTTTGTTAATGCCTAAGTTCATCGAAGATGGTTGGGAAGTCCACTATATCGGG
cluster106	core	TTTTGCAGGAAGTCTAATAAAGAAACTAGCAAGGTCAGGCCTTACTTTATTGGTGGACATCCAGGATTTAATTGTCTTCTACTTGATGATGAAGTCTATGAAGATTATT
cluster1060a	core	ATGATACCGCAAGAAATGTGGAATAAATACAAGCAAATCAATCCTTTGATCGGAGATGAAATCGATGCCTGGGCTTTTGGAGTTGAACCAGACCTTTTAGCGGATTTG
cluster1060b	core	ATGACACCGCAAGAAATGTGGAATAAATACAAGCAAATCAATCCTTTGATCGGAGATGAAATCGATGCTTGGGCTTTTGGAGTTGAACCAGACCTTTTAGCGGATTTG
cluster1061	core	ATGCGAGAACATCGTCCAATCATTGCTCTTGATTTTCCTAATTTTGAGGCGGTCAAGGAATTTTTAGCTCTTTTCCCAGCAGAAGAAAGCCTTTATCTCAAGGTAGGGA
cluster1066	core	ATGACTAAACGTGTACTCATTACAGGAGTGAGTTCAGGGATCGGATTGGCTCAAGCTCGCCTCTTTTTAGAGAAGGGCTATCAAGTTTATGGAGTTGACCAAGGTGAA

Table S5. A preview of the table generated after data normalization. After software normalization step, each subcluster receives 3 hybridization intensity values, the average of which translates to gene expression (or content) based on the type of the experiment. RANDOM indicates the condensed value of the 1000 random control probes placed by the manufacturer on each array for background hybridization correction. If a cluster is divided in more than one subcluster, these are labeled alphabetically.

SLIDE NO.	SUBCLUSTER ID	NO. OF PROBES/SUBCLUSTER	HYBRIDIZATION INTENSITY VALUES	STD. ERROR IN HYB. INTENSITY VALUES
555464A02_40020_532	RANDOM_GC42_STRAIGHT	1000	22.8957	0.7802
555464A02_40020_532	cluster1000	10	10758.374	0.9914
555464A02_40020_532	cluster1000	10	10135.3688	0.9907
555464A02_40020_532	cluster1000	5	12998.6896	0.9931
555464A02_40020_532	cluster1002	4	70.9433	0.863
555464A02_40020_532	cluster1002	4	55.7854	0.8109
555464A02_40020_532	cluster1002	4	77.1737	0.8416
555464A02_40020_532	cluster1004	10	3438.8074	0.9858
555464A02_40020_532	cluster1004	10	3003.1397	0.992
555464A02_40020_532	cluster1004	5	3757.5763	0.9953
555464A02_40020_532	cluster1005a	10	6499.1932	0.9876
555464A02_40020_532	cluster1005a	10	6454.1628	0.9881
555464A02_40020_532	cluster1005a	5	6278.6896	0.9938
555464A02_40020_532	cluster1006	10	10459.9151	0.9865
555464A02_40020_532	cluster1006	10	9603.448	0.9946
555464A02_40020_532	cluster1006	5	10124.0563	0.9919
555464A02_40020_532	cluster1007	10	12536.6767	0.9927
555464A02_40020_532	cluster1007	10	12473.2585	0.9889
555464A02_40020_532	cluster1007	5	11711.1329	0.985
555464A02_40020_532	cluster1008b	7	3787.4663	0.9911
555464A02_40020_532	cluster1008b	7	3746.1696	0.993
555464A02_40020_532	cluster1008b	4	4019.7742	0.9854

Table S6. Primers and NanoString probes used in this study

Pimer name	Primer Sequence	Application
Sp23F_12750_Flank1F	ACCGACCAAATTCCCAACTATCTC	Deletion of *tprA2*
Sp23F_12750_Flank1R_XhoI	ATATATCTCGAGTGGTCATCTCTAACCCTCTTTAAA	Deletion of *tprA2*
Sp23F_12750_Flank2F_NheI	ATATATGCTAGCGATAGGTGACTTCGATTGTAATAAG	Deletion of *tprA2*
Sp23F_12750_Flank2R	CGGCGAAATATTTTAGTCGAACA	Deletion of *tprA2*
Sp23F_12750_F1SpecF2F	CCATGGGAATTAAGATACCACTACTCA	Deletion of *tprA2*
Sp23F_12750_F1SpecF2R	GAAATCCATCATGTGAGAAAACTAAAGG	Deletion of *tprA2*
F1_ermB_no_promo_fwd	GCGCTTCTATGATTTTCAG	Overexpression of *tprA2*
F1_ermB_no_promo_rev	aaccctctTTATTTCCTCCCGTTAAATAATAG	Overexpression of *tprA2*
plcR_no_promo_fwd	gaggaaataaAGAGGGTTAGAGATGACC	Overexpression of *tprA2*
plcR_no_promo_rev	taactttccTCACCTATCCTTATCTTTCAAAAAATG	Overexpression of *tprA2*
F2_no_promo_fwd	gataggtgaGGAAAGTTACACGTTACTAAAG	Overexpression of *tprA2*
F2_no_promo_rev	GCCCACGAGTAAAGAGAC	Overexpression of *tprA2*
F1_ermB_promo_fwd	CCCATTCATGCAGGAATTATG	Overexpression of *tprA2*
F1_ermB_promo_rev	tttctaatacTTATTTCCTCCCGTTAAATAATAG	Overexpression of *tprA2*
plcR_with_promo_fwd	gaggaaataaGTATTAGAAAGTTGAAAAATAGAGATTAAAAG	Overexpression of *tprA2*

Bibliography

1. 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: e103293. doi:10.1371/journal.pone.0103293

2. 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;374: 893–902. doi:10.1016/S0140-6736(09)61204-6

3. Rudan I, O'Brien KL, Nair H, Liu L, Theodoratou E, Qazi S, et al. Epidemiology and etiology of childhood pneumonia in 2010: estimates of incidence, severe morbidity, mortality, underlying risk factors and causative pathogens for 192 countries. J Glob Health. 2013;3: 010401. doi:10.7189/jogh.03.010401

4. Usuf E, Bottomley C, Adegbola RA, Hall A. Pneumococcal carriage in sub-Saharan Africa--a systematic review. PloS One. 2014;9: e85001. doi:10.1371/journal.pone.0085001

5. Kamng'ona AW, Hinds J, Bar-Zeev N, Gould KA, Chaguza C, Msefula C, et al. High multiple carriage and emergence of Streptococcus pneumoniae vaccine serotype variants in Malawian children. BMC Infect Dis. 2015;15: 234. doi:10.1186/s12879-015-0980-2

6. Ercibengoa M, Arostegi N, Marimón JM, Alonso M, Pérez-Trallero E. Dynamics of pneumococcal nasopharyngeal carriage in healthy children attending a day care center in northern Spain. influence of detection techniques on the results. BMC Infect Dis. 2012;12: 69. doi:10.1186/1471-2334-12-69

7. 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;50 Suppl S2: 75–81.

8. Kadioglu A, Weiser JN, Paton JC, Andrew PW. The role of Streptococcus pneumoniae virulence factors in host respiratory colonization and disease. Nat Rev Microbiol. 2008;6: 288–301. doi:10.1038/nrmicro1871

9. Branger S, Casalta JP, Habib G, Collard F, Raoult D. Streptococcus pneumoniae Endocarditis: Persistence of DNA on Heart Valve Material 7 Years after Infectious Episode. J Clin Microbiol. 2003;41: 4435–4437. doi:10.1128/JCM.41.9.4435-4437.2003

10. 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 Pathog. 2014;10: e1004383. doi:10.1371/journal.ppat.1004383

11. CDC - Pinkbook: Pneumococcal Disease Chapter - Epidemiology of Vaccine-Preventable Diseases [Internet]. [cited 30 Mar 2014]. Available: http://www.cdc.gov/vaccines/pubs/pinkbook/pneumo.html

12. Antibiotic Resistance Threats in the United States, 2013 | Antibiotic/Antimicrobial Resistance | CDC [Internet]. [cited 26 Aug 2017]. Available: https://www.cdc.gov/drugresistance/threat-report-2013/

13. Schumann B, Hahm HS, Parameswarappa SG, Reppe K, Wahlbrink A, Govindan S, et al. A semisynthetic Streptococcus pneumoniae serotype 8 glycoconjugate vaccine. Sci Transl Med. 2017;9: eaaf5347. doi:10.1126/scitranslmed.aaf5347

14. Shapiro ED, Berg AT, Austrian R, Schroeder D, Parcells V, Margolis A, et al. The protective efficacy of polyvalent pneumococcal polysaccharide vaccine. N Engl J Med. 1991;325: 1453–1460. doi:10.1056/NEJM199111213252101

15. Douglas RM, Paton JC, Duncan SJ, Hansman DJ. Antibody response to pneumococcal vaccination in children younger than five years of age. J Infect Dis. 1983;148: 131–137.

16. Nuorti JP, Whitney CG, Centers for Disease Control and Prevention (CDC). Prevention of pneumococcal disease among infants and children - use of 13-valent pneumococcal conjugate vaccine and 23-valent pneumococcal polysaccharide vaccine - recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Recomm Rep Morb Mortal Wkly Rep Recomm Rep. 2010;59: 1–18.

17. van der Linden M, Falkenhorst G, Perniciaro S, Fitzner C, Imöhl M. Effectiveness of Pneumococcal Conjugate Vaccines (PCV7 and PCV13) against Invasive Pneumococcal Disease among Children under Two Years of Age in Germany. PLoS ONE. 2016;11. doi:10.1371/journal.pone.0161257

18. Lee H, Choi EH, Lee HJ. Efficacy and effectiveness of extended-valency pneumococcal conjugate vaccines. Korean J Pediatr. 2014;57: 55–66. doi:10.3345/kjp.2014.57.2.55

19. Daniels CC, Rogers PD, Shelton CM. A Review of Pneumococcal Vaccines: Current Polysaccharide Vaccine Recommendations and Future Protein Antigens. J Pediatr Pharmacol Ther JPPT. 2016;21: 27–35. doi:10.5863/1551-6776-21.1.27

20. Spijkerman J, van Gils EJM, Veenhoven RH, Hak E, Yzerman F, van der Ende A, et al. Carriage of Streptococcus pneumoniae 3 Years after Start of Vaccination Program, the Netherlands. Emerg Infect Dis. 2011;17: 584–591. doi:10.3201/eid1704101115

21. van Hoek AJ, Sheppard CL, Andrews NJ, Waight PA, Slack MPE, Harrison TG, et al. Pneumococcal carriage in children and adults two years after introduction of the thirteen valent pneumococcal conjugate vaccine in England. Vaccine. 2014;32: 4349–4355. doi:10.1016/j.vaccine.2014.03.017

22. Pilishvili T, Bennett NM. Pneumococcal disease prevention among adults: Strategies for the use of pneumococcal vaccines. Vaccine. 2015;33 Suppl 4: D60–65. doi:10.1016/j.vaccine.2015.05.102

23. Desai AP, Sharma D, Crispell EK, Baughman W, Thomas S, Tunali A, et al. Decline in Pneumococcal Nasopharyngeal Carriage of Vaccine Serotypes After the Introduction of the 13-Valent Pneumococcal Conjugate Vaccine in Children in Atlanta, Georgia. Pediatr Infect Dis J. 2015;34: 1168–1174. doi:10.1097/INF.0000000000000849

24. Weinberger DM, Malley R, Lipsitch M. Serotype replacement in disease following pneumococcal vaccination: A discussion of the evidence. Lancet. 2011;378: 1962–1973. doi:10.1016/S0140-6736(10)62225-8

25. Bonten MJM, Huijts SM, Bolkenbaas M, Webber C, Patterson S, Gault S, et al. Polysaccharide Conjugate Vaccine against Pneumococcal Pneumonia in Adults. N Engl J Med. 2015;372: 1114–1125. doi:10.1056/NEJMoa1408544

26. Isaacman DJ, McIntosh ED, Reinert RR. Burden of invasive pneumococcal disease and serotype distribution among Streptococcus pneumoniae isolates in young children in Europe: impact of the 7-valent pneumococcal conjugate vaccine and considerations for future conjugate vaccines. Int J Infect Dis IJID Off Publ Int Soc Infect Dis. 2010;14: e197–209. doi:10.1016/j.ijid.2009.05.010

27. Avery OT, Dubos R. THE PROTECTIVE ACTION OF A SPECIFIC ENZYME AGAINST TYPE III PNEUMOCOCCUS INFECTION IN MICE. J Exp Med. 1931;54: 73–89.

28. 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;28: 871–899. doi:10.1128/CMR.00024-15

29. Beckler E, MacLeod P. THE NEUFELD METHOD OF PNEUMOCOCCUS TYPE DETERMINATION AS CARRIED OUT IN A PUBLIC HEALTH LABORATORY: A STUDY OF 760 TYPINGS 1. J Clin Invest. 1934;13: 901–907.

30. Henrichsen J. Six newly recognized types of Streptococcus pneumoniae. J Clin Microbiol. 1995;33: 2759–2762.

31. Enright MC, Spratt BG. A multilocus sequence typing scheme for Streptococcus pneumoniae: identification of clones associated with serious invasive disease. Microbiol Read Engl. 1998;144 (Pt 11): 3049–3060. doi:10.1099/00221287-144-11-3049

32. 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;39: 2565–2571. doi:10.1128/JCM.39.7.2565-2571.2001

33. Pletz MWR, McGee L, Jorgensen J, Beall B, Facklam RR, Whitney CG, et al. Levofloxacin-resistant invasive Streptococcus pneumoniae in the United States: evidence for clonal spread and the impact of conjugate pneumococcal vaccine. Antimicrob Agents Chemother. 2004;48: 3491–3497. doi:10.1128/AAC.48.9.3491-3497.2004

34. Reinert RR, Ringelstein A, van der Linden M, Cil MY, Al-Lahham A, Schmitz F-J. Molecular epidemiology of macrolide-resistant Streptococcus pneumoniae isolates in Europe. J Clin Microbiol. 2005;43: 1294–1300. doi:10.1128/JCM.43.3.1294-1300.2005

35. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, Linden M van der, et al. Rapid Pneumococcal Evolution in Response to Clinical Interventions. Science. 2011;331: 430–434. doi:10.1126/science.1198545

36. 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;13: R103. doi:10.1186/gb-2012-13-11-r103

37. Corso A, Severina EP, Petruk VF, Mauriz YR, Tomasz A. Molecular characterization of penicillin-resistant Streptococcus pneumoniae isolates causing respiratory disease in the United States. Microb Drug Resist Larchmt N. 1998;4: 325–337.

38. Kang L-H, Liu M-J, Xu W-C, Cui J-J, Zhang X-M, Wu K-F, et al. Molecular epidemiology of pneumococcal isolates from children in China. Saudi Med J. 2016;37: 403–413. doi:10.15537/smj.2016.4.14507

39. Muñoz R, Coffey TJ, Daniels M, Dowson CG, Laible G, Casal J, et al. Intercontinental spread of a multiresistant clone of serotype 23F Streptococcus pneumoniae. J Infect Dis. 1991;164: 302–306.

40. Roberts RB, Tomasz A, Corso A, Hargrave J, Severina E, PRP Collaborative Study Group. Penicillin-resistant Streptococcus pneumoniae in metropolitan New York hospitals: case control study and molecular typing of resistant isolates. Microb Drug Resist Larchmt N. 2001;7: 137–152. doi:10.1089/10766290152045011

41. Sjöström K, Spindler C, Ortqvist A, Kalin M, Sandgren A, Kühlmann-Berenzon S, et al. Clonal and capsular types decide whether pneumococci will act as a primary or opportunistic pathogen. Clin Infect Dis Off Publ Infect Dis Soc Am. 2006;42: 451–459. doi:10.1086/499242

42. 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: R107. doi:10.1186/gb-2010-11-10-r107

43. 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;189: 8186–8195. doi:10.1128/JB.00690-07

44. Tettelin H, Hollingshead SK. Comparative Genomics of Streptococcus pneumoniae: Intrastrain Diversity and Genome Plasticity. 2004; 15–29. doi:10.1128/9781555816537.ch2

45. 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: e1001108. doi:10.1371/journal.ppat.1001108

46. Andam CP, Hanage WP. Mechanisms of genome evolution of Streptococcus. Infect Genet Evol. 2015;33: 334–342. doi:10.1016/j.meegid.2014.11.007

47. Croucher NJ, Walker D, Romero P, Lennard N, Paterson GK, Bason NC, et al. Role of conjugative elements in the evolution of the multidrug-resistant pandemic clone Streptococcus pneumoniaeSpain23F ST81. J Bacteriol. 2009;191: 1480–1489. doi:10.1128/JB.01343-08

48. Dowson CG, Hutchison A, Brannigan JA, George RC, Hansman D, Liñares J, et al. Horizontal transfer of penicillin-binding protein genes in penicillin-resistant clinical isolates of Streptococcus pneumoniae. Proc Natl Acad Sci U S A. 1989;86: 8842–8846.

49. 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;296: 202–211. doi:10.1001/jama.296.2.202

50. 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;118: 292–298.

51. 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;15: 347–351. doi:10.1097/MOO.0b013e3282b97327

52. Sanderson AR, Leid JG, Hunsaker D. Bacterial biofilms on the sinus mucosa of human subjects with chronic rhinosinusitis. The Laryngoscope. 2006;116: 1121–1126. doi:10.1097/01.mlg.0000221954.05467.54

53. Hall-Stoodley L, Costerton JW, Stoodley P. Bacterial biofilms: from the Natural environment to infectious diseases. Nat Rev Microbiol. 2004;2: 95–108. doi:10.1038/nrmicro821

54. Gilley RP, Orihuela CJ. Pneumococci in biofilms are non-invasive: implications on nasopharyngeal colonization. Front Cell Infect Microbiol. 2014;4. doi:10.3389/fcimb.2014.00163

55. 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. doi:10.3389/fcimb.2014.00194

56. 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;237: 223–234.

57. 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;35: 566–576.

58. Blue CE, Mitchell TJ. Contribution of a Response Regulator to the Virulence of Streptococcus pneumoniae Is Strain Dependent. Infect Immun. 2003;71: 4405–4413. doi:10.1128/IAI.71.8.4405-4413.2003

59. 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;51: 1071–1086.

60. Dawid S, Roche AM, Weiser JN. The blp bacteriocins of Streptococcus pneumoniae mediate intraspecies competition both in vitro and in vivo. Infect Immun. 2007;75: 443–451. doi:10.1128/IAI.01775-05

61. Halfmann A, Kovács M, Hakenbeck R, Brückner R. Identification of the genes directly controlled by the response regulator CiaR in Streptococcus pneumoniae: five out of 15 promoters drive expression of small non-coding RNAs. Mol Microbiol. 2007;66: 110–126. doi:10.1111/j.1365-2958.2007.05900.x

62. Håvarstein 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;92: 11140–11144.

63. Ibrahim YM, Kerr AR, McCluskey J, Mitchell TJ. Role of HtrA in the Virulence and Competence of Streptococcus pneumoniae. Infect Immun. 2004;72: 3584–3591. doi:10.1128/IAI.72.6.3584-3591.2004

64. Maricic N, Anderson ES, Opipari AE, Yu EA, Dawid S. Characterization of a Multipeptide Lantibiotic Locus in Streptococcus pneumoniae. mBio. 2016;7: e01656–15. doi:10.1128/mBio.01656-15

65. Mascher T, Heintz M, Zähner D, Merai M, Hakenbeck R. The CiaRH system of Streptococcus pneumoniae prevents lysis during stress induced by treatment with cell wall inhibitors and by mutations in pbp2x involved in beta-lactam resistance. J Bacteriol. 2006;188: 1959–1968. doi:10.1128/JB.188.5.1959-1968.2006

66. Mascher T, Zähner D, Merai M, Balmelle N, de Saizieu AB, Hakenbeck R. The Streptococcus pneumoniae cia regulon: CiaR target sites and transcription profile analysis. J Bacteriol. 2003;185: 60–70.

67. McKessar SJ, Hakenbeck R. The Two-Component Regulatory System TCS08 Is Involved in Cellobiose Metabolism of Streptococcus pneumoniae R6. J Bacteriol. 2007;189: 1342–1350. doi:10.1128/JB.01170-06

68. 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;187: 3969–3979. doi:10.1128/JB.187.12.3969-3979.2005

69. Zähner D, Kaminski K, van der Linden M, Mascher T, Meral M, Hakenbeck R. The ciaR/ciaH regulatory network of Streptococcus pneumoniae. J Mol Microbiol Biotechnol. 2002;4: 211–216.

70. Perez-Pascual D, Monnet V, Gardan R. Bacterial Cell–Cell Communication in the Host via RRNPP Peptide-Binding Regulators. Front Microbiol. 2016;7. doi:10.3389/fmicb.2016.00706

71. 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;87: 913–923. doi:10.1007/s00253-010-2651-y

72. 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;36: 3791–3801. doi:10.1093/nar/gkn149

73. Slamti L, Lereclus D. A cell-cell signaling peptide activates the PlcR virulence regulon in bacteria of the Bacillus cereus group. EMBO J. 2002;21: 4550–4559.

74. 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; doi:10.1111/mmi.13029

75. Dubois T, Faegri K, Perchat S, Lemy C, Buisson C, Nielsen-LeRoux C, et al. Necrotrophism Is a Quorum-Sensing-Regulated Lifestyle in Bacillus thuringiensis. PLoS Pathog. 2012;8. doi:10.1371/journal.ppat.1002629

76. 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;6: e00393–15. doi:10.1128/mBio.00393-15

77. 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;289: 22427–22436. doi:10.1074/jbc.M114.583989

78. 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;155: 4123–4134. doi:10.1099/mic.0.028282-0

79. 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;7: e1002190. doi:10.1371/journal.ppat.1002190

80. Rutherford ST, Bassler BL. Bacterial Quorum Sensing: Its Role in Virulence and Possibilities for Its Control. Cold Spring Harb Perspect Med. 2012;2: a012427. doi:10.1101/cshperspect.a012427

81. Susi H, Barrès B, Vale PF, Laine A-L. Co-infection alters population dynamics of infectious disease. Nat Commun. 2015;6: ncomms6975. doi:10.1038/ncomms6975

82. Balmer O, Tanner M. Prevalence and implications of multiple-strain infections. Lancet Infect Dis. 2011;11: 868–878. doi:10.1016/S1473-3099(11)70241-9

83. Birger RB, Kouyos RD, Cohen T, Griffiths EC, Huijben S, Mina MJ, et al. The potential impact of coinfection on antimicrobial chemotherapy and drug resistance. Trends Microbiol. 2015;23: 537–544. doi:10.1016/j.tim.2015.05.002

84. 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. Ahmed N, editor. PLoS ONE. 2008;3: e1969. doi:10.1371/journal.pone.0001969

85. 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: R107. doi:10.1186/gb-2010-11-10-r107

86. Chancey ST, Agrawal S, Schroeder MR, Farley MM, Tettelin H, Stephens DS. Composite mobile genetic elements disseminating macrolide resistance in Streptococcus pneumoniae. Antimicrob Resist Chemother. 2015;6: 26. doi:10.3389/fmicb.2015.00026

87. Galperin MY, Nikolskaya AN, Koonin EV. Novel domains of the prokaryotic two-component signal transduction systems. FEMS Microbiol Lett. 2001;203: 11–21.

88. Agaisse H, Gominet M, Økstad OA, Kolstø A-B, Lereclus D. PlcR is a pleiotropic regulator of extracellular virulence factor gene expression in Bacillus thuringiensis. Mol Microbiol. 1999;32: 1043–1053. doi:10.1046/j.1365-2958.1999.01419.x

89. Gohar M, Faegri K, Perchat S, Ravnum S, Økstad OA, Gominet M, et al. The PlcR Virulence Regulon of Bacillus cereus. PLoS ONE. 2008;3: e2793. doi:10.1371/journal.pone.0002793

90. Riedel CU, Monk IR, Casey PG, Waidmann MS, Gahan CGM, Hill C. AgrD-dependent quorum sensing affects biofilm formation, invasion, virulence and global gene expression profiles in Listeria monocytogenes. Mol Microbiol. 2009;71: 1177–1189. doi:10.1111/j.1365-2958.2008.06589.x

91. Lee MS, Morrison DA. Identification of a New Regulator in Streptococcus pneumoniae Linking Quorum Sensing to Competence for Genetic Transformation. J Bacteriol. 1999;181: 5004–5016.

92. Dawid S, Roche AM, Weiser JN. The blp bacteriocins of Streptococcus pneumoniae mediate intraspecies competition both in vitro and in vivo. Infect Immun. 2007;75: 443–451. doi:10.1128/IAI.01775-05

93. 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: e0121042. doi:10.1371/journal.pone.0121042

94. 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;42: D206–D214. doi:10.1093/nar/gkt1226

95. 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;6: e00173–15. doi:10.1128/mBio.00173-15

96. Li B, Yu JPJ, Brunzelle JS, Moll GN, van der Donk WA, Nair SK. Structure and mechanism of the lantibiotic cyclase involved in nisin biosynthesis. Science. 2006;311: 1464–1467. doi:10.1126/science.1121422

97. Zhang Q, Yu Y, Vélasquez JE, Donk WA van der. Evolution of lanthipeptide synthetases. Proc Natl Acad Sci. 2012;109: 18361–18366. doi:10.1073/pnas.1210393109

98. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, et al. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. 2011;39: D225–229. doi:10.1093/nar/gkq1189

99. 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: e28850. doi:10.1371/journal.pone.0028850

100. LeMessurier KS, Ogunniyi AD, Paton JC. Differential expression of key pneumococcal virulence genes in vivo. Microbiol Read Engl. 2006;152: 305–311. doi:10.1099/mic.0.28438-0

101. 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. doi:10.1002/9780471729259.mc09d04s36

102. Rosch JW, Mann B, Thornton J, Sublett J, Tuomanen E. Convergence of Regulatory Networks on the Pilus Locus of Streptococcus pneumoniae. Infect Immun. 2008;76: 3187–3196. doi:10.1128/IAI.00054-08

103. Mann B, Opijnen T van, Wang J, Obert C, Wang Y-D, Carter R, et al. Control of Virulence by Small RNAs in Streptococcus pneumoniae. PLOS Pathog. 2012;8: e1002788. doi:10.1371/journal.ppat.1002788

104. Brueggemann AB, Griffiths DT, Meats E, Peto T, Crook DW, Spratt BG. Clonal relationships between invasive and carriage Streptococcus pneumoniae and serotype- and clone-specific differences in invasive disease potential. J Infect Dis. 2003;187: 1424–1432. doi:10.1086/374624

105. Hanage WP, Kaijalainen TH, Syrjänen RK, Auranen K, Leinonen M, Mäkelä PH, et al. Invasiveness of Serotypes and Clones of Streptococcus pneumoniae among Children in Finland. Infect Immun. 2005;73: 431–435. doi:10.1128/IAI.73.1.431-435.2005

106. Zemlickova H, Jakubu V, Urbaskova P, Motlova J, Musilek M, Adamkova V. Serotype-specific invasive disease potential of Streptococcus pneumoniae in Czech children. J Med Microbiol. 2010;59: 1079–1083. doi:10.1099/jmm.0.018390-0

107. Gominet M, Slamti L, Gilois N, Rose M, Lereclus D. Oligopeptide permease is required for expression of the Bacillus thuringiensis plcR regulon and for virulence. Mol Microbiol. 2001;40: 963–975. doi:10.1046/j.1365-2958.2001.02440.x

108. Kozlowicz BK, Bae T, Dunny GM. Enterococcus faecalis pheromone-responsive protein PrgX: genetic separation of positive autoregulatory functions from those involved in negative regulation of conjugative plasmid transfer. Mol Microbiol. 2004;54: 520–532. doi:10.1111/j.1365-2958.2004.04286.x

109. Kobayashi K, Morimoto K, Kataura A, Akino T. Sugar components of glycoprotein fractions in middle ear effusions. Arch Otorhinolaryngol. 1985;242: 177–182.

110. Juhn SK, Huff JS. Biochemical characteristics of middle ear effusions. Ann Otol Rhinol Laryngol. 1976;85: 110–116.

111. Giebink GS. Otitis media: the chinchilla model. Microb Drug Resist Larchmt N. 1999;5: 57–72.

112. Fulghum RS, Brinn JE, Smith AM, Daniel HJ, Loesche PJ. Experimental otitis media in gerbils and chinchillas with Streptococcus pneumoniae, Haemophilus influenzae, and other aerobic and anaerobic bacteria. Infect Immun. 1982;36: 802–810.

113. Durand D, Halldórsson BV, Vernot B. A hybrid micro-macroevolutionary approach to gene tree reconstruction. J Comput Biol J Comput Mol Cell Biol. 2006;13: 320–335. doi:10.1089/cmb.2006.13.320

114. Chen K, Durand D, Farach-Colton M. NOTUNG: a program for dating gene duplications and optimizing gene family trees. J Comput Biol J Comput Mol Cell Biol. 2000;7: 429–447. doi:10.1089/106652700750050871

115. 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. 2014;5. doi:10.1038/ncomms6471

116. 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. doi:10.1038/ncomms6411

117. Carvalho SM, Kloosterman TG, Kuipers OP, Neves AR. CcpA Ensures Optimal Metabolic Fitness of Streptococcus pneumoniae. PLoS ONE. 2011;6: e26707. doi:10.1371/journal.pone.0026707

118. 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;112: 5177–5182. doi:10.1073/pnas.1500357112

119. 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;7: e1002190. doi:10.1371/journal.ppat.1002190

120. Guder A, Wiedemann I, Sahl HG. Posttranslationally modified bacteriocins--the lantibiotics. Biopolymers. 2000;55: 62–73. doi:10.1002/1097-0282(2000)55:1<62::AID-BIP60>3.0.CO;2-Y

121. Bierbaum G, Sahl H-G. Lantibiotics: Mode of Action, Biosynthesis and Bioengineering. Curr Pharm Biotechnol. 2009;10: 2–18. doi:10.2174/138920109787048616

122. Asaduzzaman SM, Sonomoto K. Lantibiotics: Diverse activities and unique modes of action. J Biosci Bioeng. 2009;107: 475–487. doi:10.1016/j.jbiosc.2009.01.003

123. Chatterjee C, Paul M, Xie L, van der Donk WA. Biosynthesis and mode of action of lantibiotics. Chem Rev. 2005;105: 633–684. doi:10.1021/cr030105v

124. Siezen RJ, Kuipers OP, de Vos WM. Comparison of lantibiotic gene clusters and encoded proteins. Antonie Van Leeuwenhoek. 1996;69: 171–184.

125. Willey JM, van der Donk WA. Lantibiotics: peptides of diverse structure and function. Annu Rev Microbiol. 2007;61: 477–501. doi:10.1146/annurev.micro.61.080706.093501

126. Rogers LA, Whittier EO. LIMITING FACTORS IN THE LACTIC FERMENTATION. J Bacteriol. 1928;16: 211–229.

127. McAuliffe O, Ryan MP, Ross RP, Hill C, Breeuwer P, Abee T. Lacticin 3147, a Broad-Spectrum Bacteriocin Which Selectively Dissipates the Membrane Potential. Appl Environ Microbiol. 1998;64: 439–445.

128. Brötz H, Bierbaum G, Markus A, Molitor E, Sahl HG. Mode of action of the lantibiotic mersacidin: inhibition of peptidoglycan biosynthesis via a novel mechanism? Antimicrob Agents Chemother. 1995;39: 714–719.

129. Begley M, Cotter PD, Hill C, Ross RP. Identification of a Novel Two-Peptide Lantibiotic, Lichenicidin, following Rational Genome Mining for LanM Proteins. Appl Environ Microbiol. 2009;75: 5451–5460. doi:10.1128/AEM.00730-09

130. Brötz H, Bierbaum G, Reynolds PE, Sahl H-G. The Lantibiotic Mersacidin Inhibits Peptidoglycan Biosynthesis at the Level of Transglycosylation. Eur J Biochem. 1997;246: 193–199. doi:10.1111/j.1432-1033.1997.t01-1-00193.x

131. Chugunov A, Pyrkova D, Nolde D, Polyansky A, Pentkovsky V, Efremov R. Lipid-II forms potential "landing terrain" for lantibiotics in simulated bacterial membrane. Sci Rep. 2013;3. doi:10.1038/srep01678

132. Wiedemann I, Bottiger T, Bonelli RR, Schneider T, Sahl H-G, Martinez B. Lipid II-Based Antimicrobial Activity of the Lantibiotic Plantaricin C. Appl Environ Microbiol. 2006;72: 2809–2814. doi:10.1128/AEM.72.4.2809-2814.2006

133. Phelps HA, Neely MN. SalY of the Streptococcus pyogenes Lantibiotic Locus Is Required for Full Virulence and Intracellular Survival in Macrophages. Infect Immun. 2007;75: 4541–4551. doi:10.1128/IAI.00518-07

134. Hurst A. Nisin. Adv Appl Microbiol. 1981;27: 85–123. doi:10.1016/S0065-2164(08)70342-3

135. Paik SH, Chakicherla A, Hansen JN. Identification and Characterization of the Structural and Transporter Genes for, and the Chemical and Biological Properties of, Sublancin 168, a Novel Lantibiotic Produced by Bacillus subtilis 168. J Biol Chem. 1998;273: 23134–23142. doi:10.1074/jbc.273.36.23134

136. Kamarajan P, Hayami T, Matte B, Liu Y, Danciu T, Ramamoorthy A, et al. Nisin ZP, a Bacteriocin and Food Preservative, Inhibits Head and Neck Cancer Tumorigenesis and Prolongs Survival. PLoS ONE. 2015;10. doi:10.1371/journal.pone.0131008

137. Kodani S, Lodato MA, Durrant MC, Picart F, Willey JM. SapT, a lanthionine-containing peptide involved in aerial hyphae formation in the streptomycetes. Mol Microbiol. 2005;58: 1368–1380. doi:10.1111/j.1365-2958.2005.04921.x

138. Biswas S, Biswas I. A Conserved Streptococcal Membrane Protein, LsrS, Exhibits a Receptor-Like Function for Lantibiotics. J Bacteriol. 2014;196: 1578–1587. doi:10.1128/JB.00028-14

139. Liu Y, Ames B, Gorovits E, Prater BD, Syribeys P, Vernachio JH, et al. SdrX, a Serine-Aspartate Repeat Protein Expressed by Staphylococcus capitis with Collagen VI Binding Activity. Infect Immun. 2004;72: 6237–6244. doi:10.1128/IAI.72.11.6237-6244.2004

140. McCrea KW, Hartford O, Davis S, Eidhin DN, Lina G, Speziale P, et al. The serine-aspartate repeat (Sdr) protein family in Staphylococcus epidermidis. Microbiol Read Engl. 2000;146 (Pt 7): 1535–1546. doi:10.1099/00221287-146-7-1535

141. Askarian F, Uchiyama S, Valderrama JA, Ajayi C, Sollid JUE, Sorge NM van, et al. Serine-Aspartate Repeat Protein D Increases Staphylococcus aureus Virulence and Survival in Blood. Infect Immun. 2017;85: e00559–16. doi:10.1128/IAI.00559-16

142. Kimaro Mlacha SZ, Romero-Steiner S, Dunning Hotopp JC, 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. doi:10.1186/1471-2164-14-383

143. Pettigrew MM, Marks LR, Kong Y, Gent JF, Roche-Hakansson H, Hakansson AP. Dynamic Changes in the Streptococcus pneumoniae Transcriptome during Transition from Biofilm Formation to Invasive Disease upon Influenza A Virus Infection. Infect Immun. 2014;82: 4607–4619. doi:10.1128/IAI.02225-14

144. Xu W, Solis NV, Filler SG, Mitchell AP. Pathogen Gene Expression Profiling During Infection Using a NanoString nCounter Platform. Methods Mol Biol Clifton NJ. 2016;1361: 57–65. doi:10.1007/978-1-4939-3079-1_3

145. Gualdi L, Hayre JK, Gerlini A, Bidossi A, Colomba L, Trappetti C, et al. Regulation of neuraminidase expression in Streptococcus pneumoniae. BMC Microbiol. 2012;12: 200. doi:10.1186/1471-2180-12-200

146. 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;51: 835–840. doi:10.1128/JCM.02920-12

147. Ahmed A, Earl J, Retchless A, Hillier SL, Rabe LK, Cherpes TL, et al. Comparative genomic analyses of 17 clinical isolates of Gardnerella vaginalis provide evidence of multiple genetically isolated clades consistent with subspeciation into genovars. J Bacteriol. 2012;194: 3922–3937. doi:10.1128/JB.00056-12

148. Boissy R, Ahmed A, Janto B, Earl J, Hall BG, Hogg JS, et al. Comparative supragenomic analyses among the pathogens Staphylococcus aureus, Streptococcus pneumoniae, and Haemophilus influenzae using a modification of the finite supragenome model. BMC Genomics. 2011;12: 187. doi:10.1186/1471-2164-12-187

149. Ehrlich GD, Ahmed A, Earl J, Hiller NL, Costerton JW, Stoodley P, et al. The distributed genome hypothesis as a rubric for understanding evolution in situ during chronic bacterial biofilm infectious processes. FEMS Immunol Med Microbiol. 2010;59: 269–279. doi:10.1111/j.1574-695X.2010.00704.x

150. 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: R103. doi:10.1186/gb-2007-8-6-r103

151. 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." Proc Natl Acad Sci U S A. 2005;102: 13950–13955. doi:10.1073/pnas.0506758102

152. Davie JJ, Earl J, de Vries SP, Ahmed A, Hu FZ, Bootsma HJ, et al. Comparative analysis and supragenome modeling of twelve Moraxella catarrhalis clinical isolates. BMC Genomics. 2011;12: 70. doi:10.1186/1471-2164-12-70

153. He M, Sebaihia M, Lawley TD, Stabler RA, Dawson LF, Martin MJ, et al. Evolutionary dynamics of Clostridium difficile over short and long time scales. Proc Natl Acad Sci. 2010;107: 7527–7532. doi:10.1073/pnas.0914322107

154. Conlan S, Mijares LA, NISC Comparative Sequencing Program, Becker J, Blakesley RW, Bouffard GG, et al. Staphylococcus epidermidis pan-genome sequence analysis reveals diversity of skin commensal and hospital infection-associated isolates. Genome Biol. 2012;13: R64. doi:10.1186/gb-2012-13-7-r64

155. Erdos G, Sayeed S, Antalis P, Hu FZ, Hayes J, Goodwin J, et al. Development and characterization of a pooled Haemophilus influenzae genomic library for the evaluation of gene expression changes associated with mucosal biofilm formation in otitis media. Int J Pediatr Otorhinolaryngol. 2003;67: 749–755.

156. Coffey TJ, Dowson CG, Daniels M, Zhou J, Martin C, Spratt BG, et al. Horizontal transfer of multiple penicillin-binding protein genes, and capsular biosynthetic genes, in natural populations of Streptococcus pneumoniae. Mol Microbiol. 1991;5: 2255–2260.

157. Engelmoer DJP, Rozen DE. Competence increases survival during stress in Streptococcus pneumoniae. Evol Int J Org Evol. 2011;65: 3475–3485. doi:10.1111/j.1558-5646.2011.01402.x

158. Prudhomme M, Attaiech L, Sanchez G, Martin B, Claverys J-P. Antibiotic stress induces genetic transformability in the human pathogen Streptococcus pneumoniae. Science. 2006;313: 89–92. doi:10.1126/science.1127912

159. 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;207: 439–449. doi:10.1093/infdis/jis703

160. Eutsey RA, Hiller NL, Earl JP, Janto BA, Dahlgren ME, Ahmed A, et al. Design and validation of a supragenome array for determination of the genomic content of Haemophilus influenzae isolates. BMC Genomics. 2013;14: 484. doi:10.1186/1471-2164-14-484

161. Janto BA, Hiller NL, Eutsey RA, Dahlgren ME, Earl JP, Powell E, et al. Development and Validation of an Haemophilus influenzae Supragenome Hybridization (SGH) Array for Transcriptomic Analyses. PLoS ONE. 2014;9: e105493. doi:10.1371/journal.pone.0105493

162. 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: R103. doi:10.1186/gb-2007-8-6-r103

163. Boissy R, Ahmed A, Janto B, Earl J, Hall BG, Hogg JS, et al. Comparative supragenomic analyses among the pathogens Staphylococcus aureus, Streptococcus pneumoniae, and Haemophilus influenzae using a modification of the finite supragenome model. BMC Genomics. 2011;12: 187. doi:10.1186/1471-2164-12-187

164. Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. Proc Natl Acad Sci U S A. 1988;85: 2444–2448.

165. Sambrook J, Russell DW. Molecular Cloning: A Laboratory Manual. CSHL Press; 2001.

166. 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: e1001108. doi:10.1371/journal.ppat.1001108

167. 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;26: 317–325. doi:10.1038/nbt1385

168. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215: 403–410. doi:10.1016/S0022-2836(05)80360-2

169. Kayala MA, Baldi P. Cyber-T web server: differential analysis of high-throughput data. Nucleic Acids Res. 2012;40: W553–559. doi:10.1093/nar/gks420

170. 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: e74867. doi:10.1371/journal.pone.0074867

171. 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;1. doi:10.1128/genomeA.00520-13

172. Zomer AL, Buist G, Larsen R, Kok J, Kuipers OP. Time-resolved determination of the CcpA regulon of Lactococcus lactis subsp. cremoris MG1363. J Bacteriol. 2007;189: 1366–1381. doi:10.1128/JB.01013-06

173. 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;7: 43587. doi:10.1038/srep43587

174. 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. doi:10.1002/9780471729259.mc09d04s36

175. Baldi P, Long AD. A Bayesian framework for the analysis of microarray expression data: regularized t -test and statistical inferences of gene changes. Bioinforma Oxf Engl. 2001;17: 509–519.

176. Ruijter JM, Lorenz P, Tuomi JM, Hecker M, van den Hoff MJB. Fluorescent-increase kinetics of different fluorescent reporters used for qPCR depend on monitoring chemistry, targeted sequence, type of DNA input and PCR efficiency. Mikrochim Acta. 2014;181: 1689–1696. doi:10.1007/s00604-013-1155-8

177. Ramakers C, Ruijter JM, Deprez RHL, Moorman AFM. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci Lett. 2003;339: 62–66.

178. Buchinsky FJ, Forbes ML, Hayes JD, Shen K, Ezzo S, Compliment J, et al. Virulence phenotypes of low-passage clinical isolates of nontypeable Haemophilus influenzae assessed using the chinchilla laniger model of otitis media. BMC Microbiol. 2007;7: 56. doi:10.1186/1471-2180-7-56

179. Zbinden A, Mueller NJ, Tarr PE, Eich G, Schulthess B, Bahlmann AS, et al. Streptococcus tigurinus, a novel member of the Streptococcus mitis group, causes invasive infections. J Clin Microbiol. 2012;50: 2969–2973. doi:10.1128/JCM.00849-12

180. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. Genome Res. 2004;14: 1394–1403. doi:10.1101/gr.2289704

181. Darling AE, Mau B, Perna NT. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. PLoS ONE. 2010;5: e11147. doi:10.1371/journal.pone.0011147

182. 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;30: 3059–3066.

183. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. Bioinformatics. 1998;14: 817–818. doi:10.1093/bioinformatics/14.9.817

184. 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;59: 307–321. doi:10.1093/sysbio/syq010

185. Brown SP, Cornforth DM, Mideo N. Evolution of virulence in opportunistic pathogens: generalism, plasticity, and control. Trends Microbiol. 2012;20: 336–342. doi:10.1016/j.tim.2012.04.005

186. Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ. Within-host evolution of bacterial pathogens. Nat Rev Microbiol. 2016;14: 150–162. doi:10.1038/nrmicro.2015.13

187. 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;13: e1006339. doi:10.1371/journal.ppat.1006339

188. 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; IAI.00277–16. doi:10.1128/IAI.00277-16

189. Peiró C, Romacho T, Azcutia V, Villalobos L, Fernández E, Bolaños JP, et al. Inflammation, glucose, and vascular cell damage: the role of the pentose phosphate pathway. Cardiovasc Diabetol. 2016;15: 82. doi:10.1186/s12933-016-0397-2

190. Marik PE, Bellomo R. Stress hyperglycemia: an essential survival response! Crit Care. 2013;17: 305. doi:10.1186/cc12514

191. Umpierrez GE, Isaacs SD, Bazargan N, You X, Thaler LM, Kitabchi AE. Hyperglycemia: An Independent Marker of In-Hospital Mortality in Patients with Undiagnosed Diabetes. J Clin Endocrinol Metab. 2002;87: 978–982. doi:10.1210/jcem.87.3.8341