DISCRIMINATION OF VACCINE-PREVENTABLE SEROTYPES OF *STREPTOCOCCUS PNEUMONIAE* BY PCR AND SEQUENCING

by

Hayley Gillis

Submitted in partial fulfillment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
June 2017

Table of Contents

List of Tables

List of Figures

Abstract

Streptococcus pneumoniae is a bacterium that causes significant morbidity and mortality worldwide, and vaccines have been developed against a certain number of its serotypes. Serotyping of *S. pneumoniae* is important to monitor disease epidemiology and assess the impact of pneumococcal vaccines. Traditionally, the Quellung reaction used serotype-specific antibodies to classify isolates based on differences in capsular antigens. More recently, serotype deduction using multiplex PCR (cmPCR) has been applied broadly for pneumococcal surveillance, and relies on differences in the capsule biosynthesis genes (*cps* loci). However, PCR lacks discrimination between certain serotypes, including some vaccine-preventable serotypes. For example, the vaccine-preventable serotypes 6A and 6B would need to be discriminated for the non-vaccine serotypes 6C and 6D, and similarly, 7F from 7A; 9V from 9A; 9N from 9L; 11A from 11D; 12F from 12A, 12B, 44 and 46; 15B from 15C; 18C from 18F, 18A, 18B; 22F from 22A, and 33F from 33A and 37.

This study evaluated two new molecular approaches that could resolve vaccine-preventable serotypes of *S. pneumoniae*: 1) next generation sequencing and comparative genomic was used to identify novel serotype-specific PCR targets outside the *cps* loci; and 2) PCR targeted sequencing was used for specific molecular signature inside the *cps* loci. In both approaches, specificity was tested using all 92 serotypes of *S. pneumoniae* previously characterized by Quellung reaction, as well as 32 other members of the *Streptococcaceae* family. Reproducibility was evaluated using multiple isolates of each *S. pneumoniae* serotype under evaluation, which differed temporally and geographically. Overall, this study showed that the select PCR targets outside the *cps* loci could not reproducibly discriminate *S. pneumoniae* serotypes, whereas PCR and sequencing targets inside the *cps* loci showed 82% accuracy compared to traditional serotyping Quellung. While discrepant analyses are required and underway, the sequence-based approach shows much promise for discrimination of vaccine-preventable serotypes of *S. pneumoniae*. Since serotyping is important to monitor *S. pneumoniae* epidemiology and determine the proportion of disease is vaccine-preventable, this study represents a significant technological advance for pneumococcal disease surveillance.

List of Abbreviations Used

°	Degrees
µ	Micro
Ac	Acetate
ACIP	Advisory Committee on Immunization Practices
AOM	Acute Otitis Media
BLASTn	Nucleotide Basic Local Alignment Search Tool
bp	Base pairs
C	Celcius
CNDSS	Canadian Notifiable Disease Surveillence System
CAP	Community Acquired Pneumonia
CAPITA	Community Acquired Pneumonia Trial in Adults
CSBN	Canadian Bacterial Surveillance Network
CDC	Center for Disease Control and Prevention
ChoP	Phosphorylcholine
CIRN	Canadian Immunization Research Network
cmPCR	Conventional multiplex PCR
CO_2	Carbon dioxide
cps	Capsular polysaccharide (locus)
Ct	Threshold cycle
CTV	Capsule Type Variant
DNA	Deoxyribonucleic Acid
EB	Elution Buffer
EIA	Enzyme Immunoassay
f	Furanose
FDA	Food and Drug Administration
FucNAc	N-acetylfucosamine
g	Gram
Gal	Galactose
GalNAc	N-acetylgalactosamine
Glc	Glucose
GlcNAc	N-acetylglucosamine
GlcUA	Glucuronic acid
Gro	Glycerol
h	Hour
HIV-1	Human Immunodeficiency Virus 1
IgG	Immunoglobulin G
IPD	Invasive Pneumococcal Disease
l	Litre
m	Metre, mili-
MALDI-TOF	Matrix-Assisted Laser Desorption Ionization-Time of Flight
ManNacA	N-acetylmannosaminuronic acid
MCM	Manual of Clinical Microbiology

MDR	Multi-drug resistant
$MgCl_2$	Magnesium chloride
MGE	Mobile genetic elements
min	Minute
N/A	Not applicable
NACI	National Advisory Committee on Immunization
NGS	Next generation sequencing
NML	National Microbiology Laboratory
NMR	Nuclear Magnetic Resonance
NR	No Rox
NP	Nasopharyngeal
NVT	Non-Vaccine Type
p	Phosphate
P	Pyranose
PCR	Polymerase chain reaction
PCV7	7-valent polysaccharide conjugate vaccine
PCV13	13-valent polysaccharide conjugate vaccine
PHAC	Public Health Agency of Canada
PG	Phosphatidyl glycerol
PneumoCaT	Pneumococcal Capsule Typing
PPV23	23-valent polysaccharide vaccine
Rha	Rhamnose
Rib-ol	Ribitol
rPAF	Receptor for platelet activating factor
rpm	Revolutions per minute
rmPCR	Real-time multiplex polymerase chain reaction
s	Second
SB	Sodium borate
SBA	Sheep blood agar
SNP	Single nucleotide polymorphism
SOS	Serious Outcomes Surveillance
SSI	Statens Serum Institute
STI	Sexually transmitted infection
TIBDN	Toronto Invasive Bacterial Diseases Network
TNA	Total nucleic acid
TSA	Trypticase Soy Agar
UDP	Uridine diphosphate
Und-P	Undecaprenyl-phosphate
U.K.	United Kingdom
U.S.	United States
V	Volts
WHO	World Health Organization

Acknowledgements

I would like to acknowledge and express my appreciation to everyone that helped me complete my graduate education. These individuals provided unwavering support and helped me on the path to a future career as a physician. First and foremost, a sincere thank you to my supervisor, Dr. Jason LeBlanc. Dr. LeBlanc's passion for scientific research is inspiring. His kindness, support and guidance during this project were above and beyond expectations. I could not have asked for a greater mentor. I would also like to thank my co-supervisor, Dr. Shelly McNeil, and my committee members, Drs. Ross Davidson and Dan Gaston for the expertise and advice they contributed to this project. Finally, this project could not have taken place without the funding support from the Canadian Immunization Research Network, the Nova Scotia Health Research Foundation, the Nova Scotia Health Authority Research Fund, and the Department of Pathology at Dalhousie University.

Chapter 1. Introduction

1.1. *Streptococcus pneumoniae*: An Overview

Streptococcus pneumoniae (or pneumococcus) is a bacterium of public health concern, and causes significant morbidity and mortality worldwide. Fortunately, pneumococcal vaccines have been developed to protect us against the most prevalent "serotypes" causing disease. The following section will describe a brief history of *S. pneumoniae*, its ability to colonize the human oro- and nasopharynx, the pneumococcal diseases that could ensue, serotype-specific pneumococcal vaccines to help alleviate the burden of pneumococcal diseases, and a review of the current knowledge on serotyping methods for *S. pneumoniae*.

1.1.1. A Historical Perspective on *S. pneumoniae* Isolation

S. pneumoniae has a very interesting history involving many reputable scientists (Watson *et al*., 1993; Geno *et al*., 2015). This bacterium was discovered independently in 1881 by two researchers, the U.S. Army physician, George M Stenberg, and a French chemist, Louis Pasteur (Sternberg, 1881; Pasteur, 1881). Each investigator injected human saliva into rabbits; Pasteur used saliva from a child who died of rabies, and Sternberg used his own saliva. In both cases, a lancet-shaped diplococcic was isolated from rabbit blood. While developing the Gram stain in 1880, Christian Gram was experimenting with staining techniques to visualize bacteria in a lung tissue from a patient who had died of pneumonia, and described an organism called "*the cocci of croupous pneumoni*" (Gram, 1884). By 1886, this organism was being referred to as *Pneumococcus* due to its ability to cause pulmonary disease (Fraenkel, 1886). Shortly after, a number of reports demonstrated this organism to cause diseases such as meningitis and otitis media (reviewed by Austrian, 1981a and b). In 1920, the organism was called *Diplococcus pneumoniae* (Winslow *et al*., 1920), but based on its ability to grow in chains of cocci in liquid media, it was renamed *Streptococcus pneumoniae* in 1974 (Deibel and Seeley, 1974).

Today, *S. pneumoniae* is a well recognized human pathogen, and member of the *Streptococcus* genus [Phylum: Firmicutes; Order: *Lactobacilliales;* and Family: *Streptococcaceae*] (Spellerberg and Brandt, 2015). Microscopically, it is described as oval/lancet-shaped, Gram-positive cocci in pairs (diplococci) or short chains. For culture, sheep blood agar (SBA) incubated in 5% CO_2 fulfills its nutrient requirements and *S. pneumoniae* is considered a facultative anaerobe. If grown aerobically on SBA, *S. pneumoniae* produces a green coloration (α-hemolysis), and the colonies often have a moist appearance from expression of capsular polysaccharide. Another distinguishing characteristic visible in older cultures is a central navel-like indention in the colonies that occur due to autolysis (Spellerberg and Brandt, 2015). Despite these distinctive features, up to 20% of *S. pneumoniae* appear indistinguishable from other related α-hemolytic bacteria found in respiratory flora, the viridans group streptococci (Richter *et al*., 2008). In the early 1900s, a German physician and biologist Fred Neufeld discovered that addition of a small amount of ox bile to a *S. pneumoniae* culture quickly resulted in complete lysis of the organism (Neufeld, 1900). Morganroth and Levy (1911) showed that susceptibility to a quinine derivative, ethylhydrocupreine (more commonly known as optochin), could also be used to differentiate *S. pneumoniae* from viridians group streptococci. Both bile solubility and susceptibility to optochin became a staple in clinical and research laboratories worldwide (Spellerberg and Brandt, 2015).

1.1.2. *S. pneumoniae* Colonization

The natural reservoir of *S. pneumoniae* is primarily humans, where it generally resides in the oro- or nasopharynx (Bogaert *et al*., 2004a; Kadioglu *et al*., 2008; Song *et al*., 2012; Trzciński *et al*., 2013; Wyllie *et al*., 2016). Following acquisition of *S. pneumoniae* through respiratory droplets or person-to-person contact (Public Health Agency of Canada (PHAC), 2012; CDC, 2015a), the bacterium can colonize the upper respiratory tract (Bogaert *et al*., 2004a; Simell *et al*., 2012). The carriage rates of the organism show considerable variation amongst different age groups, health status,

geographical location, vaccine history, and environmental factors (Bogaert *et al*., 2004a; Orsi *et al*., 2016; Simell *et al*., 2012). Although reports vary, detection rates of nasopharyngeal (NP) colonization of children tend to range from 30 to 70%, compared to 1.5 to 30% in adults (Bogaert *et al*., 2004a; Orsi *et al*., 2016; Regev-Yochay *et al*., 2004; Simell *et al*., 2012). Carriage rates of adults increase when living with preschool age children, suggesting transmission within close contact in households (Hendley *et al*., 1975). In fact, transmission of *S. pneumoniae* is favoured by crowding, especially in settings of institutionalization such as hospitals, long-term care facilities, day-care centers and prisons (Orsi *et al*., 2016).

S. pneumoniae is well suited for the colonization of the human nasopharynx due to its numerous virulence factors and ability of its capsule to undergo phase-variation, which is evident in culture by the appearance of transparent and opaque colony morphologies (AlonsoDeVelasco *et al*., 1995; Kadioglu *et al*., 2008; Li *et al*., 2016; Manso *et al*., 2014; Nelson *et al*, 2007; Weiser *et al*., 1994). Bacterial pathogens are generally trapped in a mucous layer of the human upper respiratory tract, however, most pneumococcal capsules are negatively charged which helps repel mucous and allows them to gain access to the epithelium (Nelson *et al*, 2007). Once near the epithelium, changes in gene expression through a process called phase-variation prepares the organism for attachment (Kadioglu *et al*., 2008; Li *et al*., 2016; Manso *et al*., 2014). Phase variation roughly correlates to the change from an avirulent (transparent) form that expresses a thinner layer of capsule to reduce steric hindrance and aid in adhesion, and the virulent (opaque) form of *S. pneumoniae* that produces a thicker capsule to evade host immune responses during invasive disease (Li *et al*., 2016; Manso *et al*., 2014; Weiser *et al*., 1994). Virulence factors other than capsule also play an important role in colonization, including phosphorylcholine (ChoP), which binds to the receptor for platelet activating factor (rPAF) in the nasopharynx (AlonsoDeVelasco *et al*., 1995; Cundell *et al*., 1995).

1.1.3. Spectrum of Pneumococcal Diseases

Despite colonization with *S. pneumonaie*, most individuals remain asymptomatic. In some cases, the organism is able to gain access to normally sterile sites and cause disease. The factors that allow for the progression from colonization to disease are not fully understood (Kadioglu *et al*., 2008; Li *et al*., 2016; Manso *et al*., 2014), however a clear link has been established between recent colonization of *S. pneumoniae* and subsequent disease by the same serotype (Syrjanen *et al*., 2005). With the exception of intraocular infections, *S. pneumoniae* colonization of the upper respiratory tract is thought to be a prerequisite for pneumococcal disease (Bogaert *et al*., 2004a).

S. pneumoniae is a significant cause of morbidity and mortality worldwide (Drijokoningen *et al*., 2014; Jain *et al*., 2015; LeBlanc *et al*., 2017; McNeil *et al*., 2016; O'Brien *et al*., 2009). For example, in 2012, The World Health Organization (WHO) estimated that 1.2 million children under the age of five years die each year from *S. pneumoniae* infections (WHO, 2012). *S. pneumoniae* infections range in severity from non-invasive infections such as acute otitis media (AOM) and sinusitis, to more severe manifestations like community acquired pneumonia (CAP) and invasive pneumococcal diseases (IPD). AOM is the most common manifestation of disease caused by *S. pneumonie*; however this illness is generally limited to children and is usually self-limiting (Simell *et al*., 2012). CAP and IPD pose a much larger burden on the health care systems, with high morbidity and mortality, particularly in individuals with risk factors (Jain *et al*., 2015; LeBlanc *et al*., 2017; McNeil *et al*., 2016; O'Brien *et al*., 2009).

1.1.3.1. Pneumococcal Community Acquired Pneumonia

CAP is an infection of the lungs that presents on a spectrum from mild to severe disease in all age groups. While the etiology often remains undefined, *S. pneumoniae* can cause up to 50% and 23% of all-cause pneumonia in children and adults, respectively (LeBlanc *et al*., 2017). CAP caused by *S. pneumoniae* often presents as a lobular pneumonia (Watson *et al*., 1993), and causes difficulty breathing, cough and

fever (CDC, 2015a). In adults, CAP is the most common form of *S. pneumoniae* disease (CDC, 2015b) resulting in approximately 400,000 hospitalizations in the US annually (Huang *et al*., 2011). In Canada in 2001 prior to conjugate vaccine introduction, an estimated 565,077 cases of *S. pneumoniae* disease, and 3002 deaths occurred that are mainly attributed to pneumonia in the elderly (Morrow *et al*., 2007). In a retrospective study investigating the burden of CAP disease in Canada, a decrease in pneumococcal pneumonia in adults ≥ 65 years of age was noted for years 2004 to 2010, and was attributed to herd immunity from childhood immunization programs with pneumococcal vaccines (Kim *et al*., 2016; McNeil *et al*., 2016). More recently, following active CAP surveillance, LeBlanc *et al*., (2017) demonstrated decreasing trends of vaccine-type pneumococcal disease over time, and ongoing studies are underway. Prior to these studies, herd effects from pneumococcal vaccines had only been documented pertaining to IPD in Canada, through the voluntary, passive surveillance (Leal *et al*., 2012; PHAC, 2016; Rudnick *et al*., 2013).

1.1.3.2. Invasive Pneumococcal Disease

IPD encompasses all *S. pneumoniae* isolated from normally sterile sites, including blood (bacteremia), cerebral spinal fluid (meningitis), and pleural fluid (empyema). IPD can occur in all age groups; however, individuals at greatest risk include those at the extremes of age (young or elderly), those with immunocompromizing conditions, or other co-morbidities such as chronic lung disease, asthma, or smoking (Kellner *et al*., 2009; LeBlanc *et al*., 2017). In the United States, prior to the introduction of polysaccharide-conjugate vaccines, the rates of IPD were as high as 100 cases per 100,000 individuals (CDC, 2016). Following childhood immunization programs, rates of IPD dropped to as low as 22 cases per 100,000 (CDC, 2016). Similar decreases in IPD were reported in Canada following childhood immunization (Bettinger *et al*., 2010; Kellner *et al*., 2009; PHAC, 2016; Rudnick *et al*., 2013). Recent reports suggests that the overall incidence rate of IPD was stable since 2013, averaging at approximately 9 cases per 100,000 population with the highest rates of disease seen in infants <1 year of age

(14.5 cases per 100,000) and in the ≥60 years of age group (20.3 cases per 100,000) (PHAC, 2015 and 2017).

1.1.3.3. Risk Factors for Pneumococcal Disease

Young children, the elderly, and individuals that are immunocompromised or with medical co-morbidities are at the greatest risk for severe outcomes associated with pneumococcal disease. For the young and elderly individuals, a common feature is reduced immunocompetency due to underdeveloped and waning immune systems, respectively (PHAC, 2016). The fact that children 6-11 months of age are at high risk for pneumococcal should not be surprising since this group most frequently colonized by *S. pneumoniae* (Butler *et al*., 2004; O'Brien *et al.*, 2009). In all age groups, but particularly in the elderly, the risk of pneumococcal disease is greatest in individuals with underlying medical or immunocompromizing conditions. Chronic conditions (lung, liver, kidney, and heart disease), diabetes mellitus, sickle-cell disease, anatomic or functional asplenia, human-immunodeficiency virus (HIV-1) infection and underlying malignant neoplasms have all been associated as risk factors for pneumococcal disease (Jain *et al*., 2015; LeBlanc *et al*., 2017; O'Brien *et al*., 2009; PHAC, 2016; Simell *et al*., 2012). Finally, social habits such as homelessness, smoking, drug-use, and alcoholism are also known risk factors (LeBlanc *et al*., 2017; PHAC, 2016).

The risk of developing IPD has also been related to capsular serotype, as suggested by a variation in the frequency of different serotypes isolated from cases of IPD (Simell *et al*., 2012). The "invasiveness" of a serotype is essentially its ability to cause disease per colonization event (Simell *et al*., 2012). However, the most invasive serotypes are not necessarily the most predominant causes of IPD, since serotypes that are more commonly found to colonize individuals are provided a greater "temporal opportunity" to cause disease (Song *et al*., 2012). In other words, the largest proportion of disease may be caused by less invasive serotypes if these serotypes are more frequently found in the population (Brueggemann *et al*., 2003). This concept is

important with regards to epidemiology post-vaccine implementation, as vaccines are designed to cover the serotypes most frequently recovered from cases of IPD.

1.1.4. *S. pneumoniae* Virulence with a Focus on Capsule

The capsule is the primary virulence factor of *S. pneumoniae*. It was first described by Louis Pasteur and has since been studied for over a century (Pasteur, 1881). Studies of culture supernatants led to the discovery that the capsule is composed of polysaccharides (Heidelberger and Avery, 1923), which are covalently attached to the peptidoglycan of the cell wall in most serotypes (Sorensen *et al*., 1990). The majority of pneumococcal capsules are anionic (Kamerling *et al*., 2000) which is thought to play a role in electrostatic repulsion of phagocytes and prevent clearance by the mucus in the upper respiratory tract (Nelson *et al*, 2007). The capsule also provides a shielding-effect from both Fc-mediated phagocytosis and complement activation (Musher, 1992; Winkelstein, 1981). *S. pneumoniae* that lack capsule are rarely isolated from individuals with invasive disease (Kostyukova *et al*., 1995). In the nasopharynx *S. pneumoniae* express reduced amounts of capsule to aid in their adherence to the host's cells to allow for colonization (Kim and Weisner, 1998; Magee and Yother, 2001). In the lung or in the blood stream, larger amounts of capsule are expressed to reduce recognition and clearance by the immune system (Li *et al*., 2016; MacLeod and Krauss, 1950; Magee and Yother, 2001; Manso *et al*., 2014; Weiser *et al*., 1994).

Early experiments, including the famous bacterial transformation study conducted by Frederick Griffith suggested that *S. pneumoniae* capsule genetic material (genes) was arranged in close proximity (Griffith, 1928) and Oswald Avery later showed this material was DNA (Avery *et al*., 1944). Later experiments confirmed a cassette-like arrangement of genes (Ianneli *et al*., 1999; Kolkman *et al*., 1998). The majority of genes involved in capsule formation are found in a region unknown as the capsular polysaccharide (*cps*) locus (Moscoco and Garcia, 2009). Sequencing of 90 serotypes by the Sanger Institute confirmed a similar organization for the *cps* loci in all *S. pneumoniae*

serotypes (Bentley *et al*., 2006). The *cps* loci are flanked by genes *dexB* and *aliA* on the bacterial chromosome and begin with four common genes ordered *cpsABCD* (or *wzg, wzh, wzd*, and *wze*) that are involved in regulation of capsule formation (Figure 1) (Geno *et al*., 2015). The gene encoding one of two different types of initiating glycosyltransferases follows. The remaining downstream regions of the *cps* locus are serotype-specific. This serotype-specific region of *cps* loci contains genes whose products are involved in the assembly of specific sugars, their polymerization, and modification (Bentley *et al*., 2006; Geno *et al*., 2015; Yother, 2011).

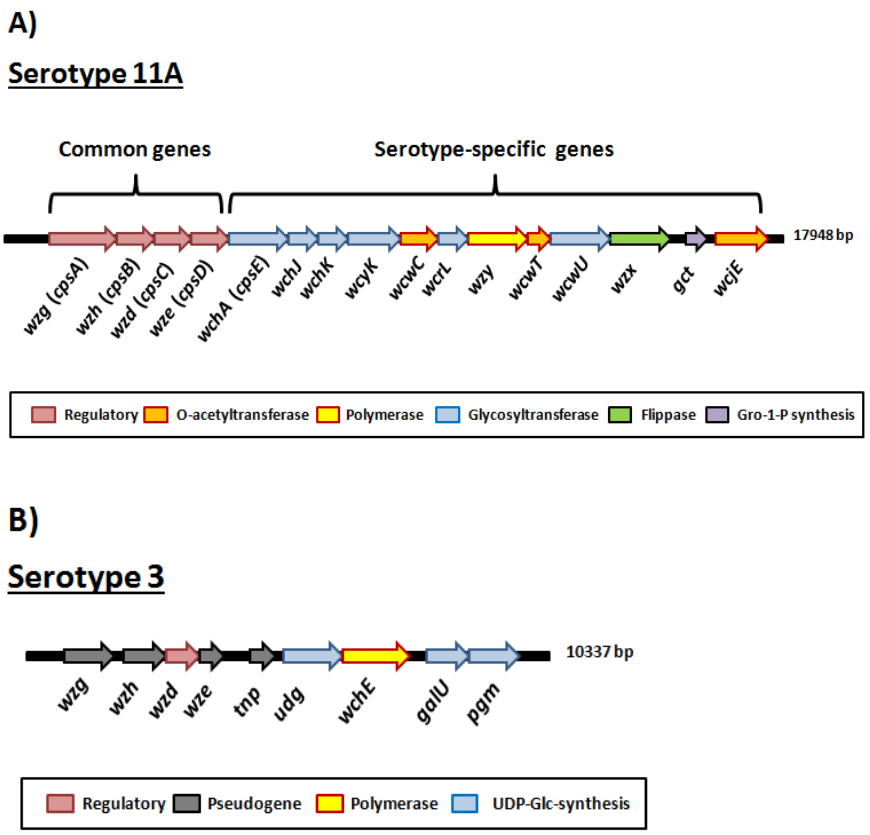


Figure 1. Genetic arrangement of the capsule biosynthesis (*cps*) loci. A) *cps* locus of serotype 11A demonstrating common genes for all serotypes as well as serotype-specific genes; B) *cps* locus of serotype 3. In both A) and B) the inset represents a color-coded legend for gene function. This figure was inspired from Bentley *et al*., 2006.

In *S. pneumoniae,* two capsular biosynthesis pathways exist and are named according to the mechanisms of action of their sugar polymerization: the Wzy-dependent (Figure 2) and synthase-dependent pathways (Figure 3) (Geno *et al*., 2015; Yother, 2011). The majority of serotypes use the Wzy-dependent pathway for capsular biosynthesis, and share a *cps* loci structures similar to 11A (Figure 1A). In contrast, serotypes 3 and 37 used the synthase-dependent pathway. Although serotype 3 shares some similarities in *cps* locus structure with serotypes that use the Wzy-dependent pathway, only some of its genes are functional (Figure 1B). This serotype uses a synthase encoded by *wchE,* Cps3S to generate its capsule (Garcia *et al*., 1997). Serotype 37 contains a similar *cps* locus structure to that of 33F, however it is completely defective. In contrast to other serotypes, serotype 37 uses a single gene located outside the *cps* loci called *tts* to encode its synthase enzyme (Llull *et al.,* 1999).

The Wzy-dependent pathway usually generates capsule polysaccharide structures that are branched and contain multiple different sugars, whereas the synthase-dependent pathway produces linear structures with only one or two different sugars (Chaffin *et al*., 2000; Geno *et al*., 2015; Yother, 2011). The Wzy-dependent mechanism involves the assembly of repeat-units consisting of nucleotide-charged sugars linked to an undecaprenyl-phosphate (Und-P) on the inner face of cell membrane (Figure 2). The repeat-units are flipped to the outer leaflet of the cell membrane by a Wzx flippase and assembled by Wzy polymerase in a non-processive manner (Geno *et al*., 2015). In contrast, the synthase-dependent mechanism involves a single integral membrane enzyme (WchE for serotype 3 or Tts in serotype 37) to initiate, polymerize, and transport the polymer (Figure 3). An individual sugar is transferred to a lipid acceptor to initiate the assembly of sugars. The polysaccharide chain is extended by the processive addition of sugars, and is then transported across the cell membrane (Geno *et al*., 2015). Despite *S. pneumoniae* using only two mechanisms of capsule assembly, unique structures of capsular polysaccharides is essentially unlimited by using a variety of sugars, linkages, and modifications.

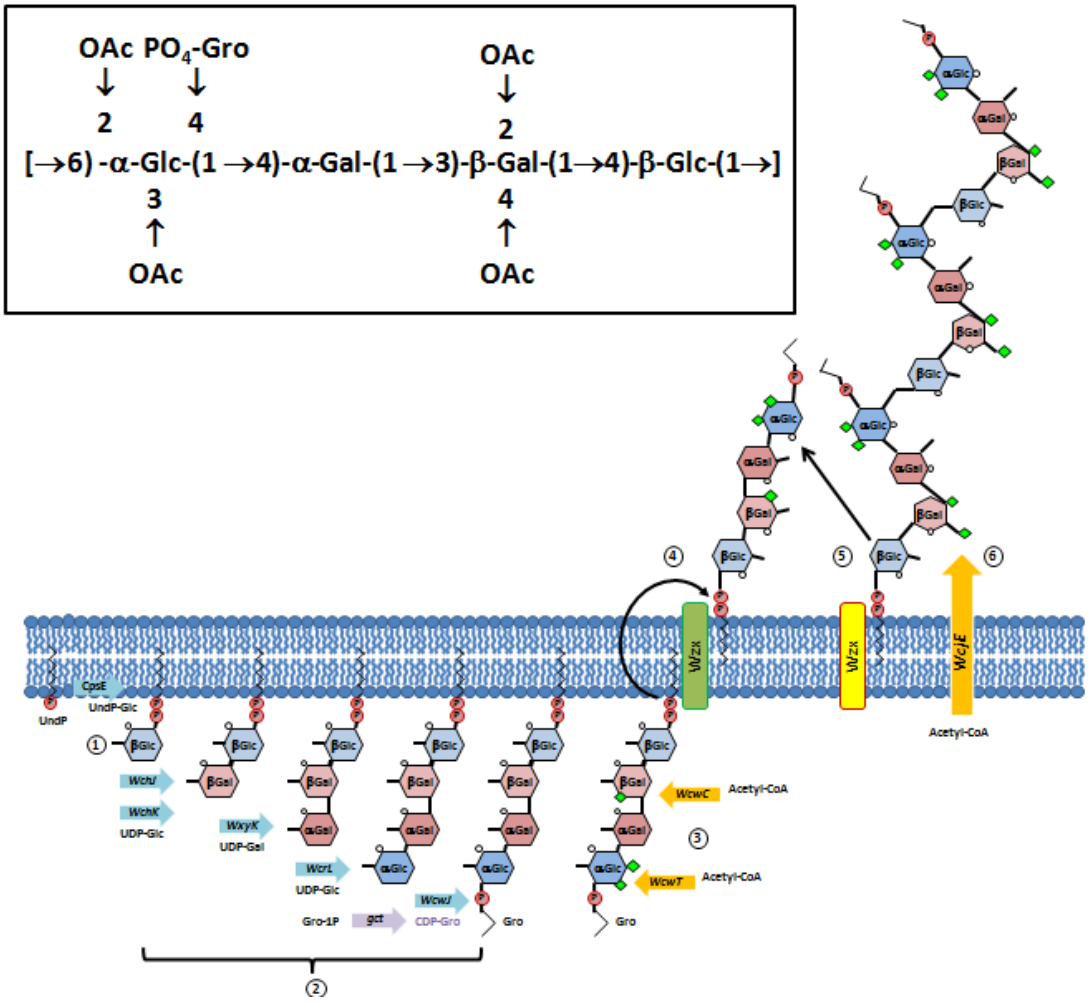


Figure 2. Wzy-dependent polysaccharide synthesis for serotype 11A. Steps involved include: 1) transfer of glucose-1-phosphate to an undecaprenyl phosphate (UndP) acceptor; 2) the repeat unit is sequentially assembled by glycosyltransferases; 3) cytoplasmic acetyltransferases decorate some sugars with acetyl groups; 4) acetylation can also occur before export; and (5 and 6) after polymerization. The inset shows the biochemical structure of the repeating unit in serotype 11A polysaccharide. Abbreviations: Gro, glycerol; Gal, galactose; Glc, glucose. This figure was inspired by Geno *et al.*, 2015.

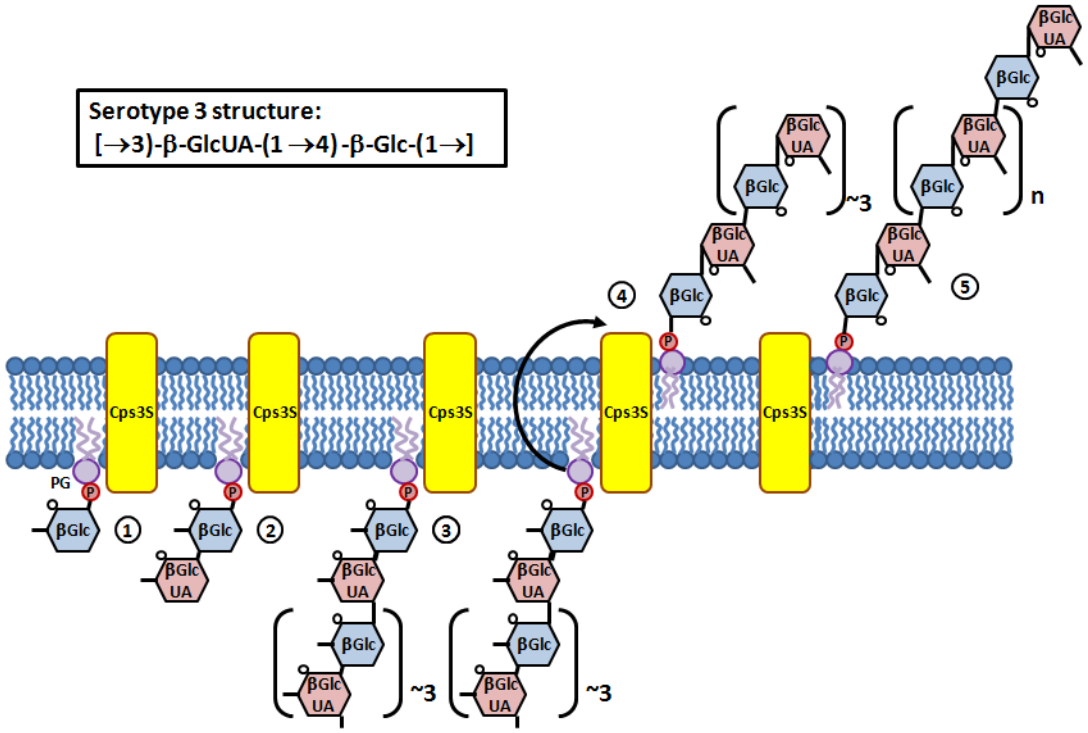


Figure 3. Synthase-dependent assembly of capsular polysaccharide for serotype 3. The Cps3S synthase synthesizes serotype 3 polysaccharide, and its structure is shown in the inset. Cps3S initiates synthesis by: 1) transfer of glucose (Glc) from UDP-glucose to a phosphatidyl glycerol (PG) acceptor; 2) transfer of glucuronic acid (GlcUA) from UDP-GlcUA to the PG-linked Glc; 3) extension of the capsule to approximately 8 saccharide units; 4) Translocation by Cps3S of the polysaccharide chain to the external face of the membrane; and 5) increases chain length by a processive capsular synthesis mechanism. This figure was inspired by Geno *et al*., 2015.

1.1.5. Pneumococcal Vaccines

The epitopes of the different capsular polysaccharides are immunogenic, however, immunity is generally not cross protective between serotypes (Bogaert *et al*., 2004b; Janeway *et al*., 2001). This means that each serotype is essentially seen as a separate organism by the adaptive immune system (Janeway *et al*., 2001). From both a practicality and cost standpoint, the diversity of pneumococcal capsular polysaccharides has made high valency vaccine development a challenge. For this reason, pneumococcal vaccines have been designed to protect against the serotypes that cause the majority of invasive disease. There are two types of pneumococcal vaccines: polysaccharide and polysaccharide conjugate vaccines. Over time, different vaccine formulations have been developed with different coverage for *S. pneumonaie* serotypes (Figure 4).

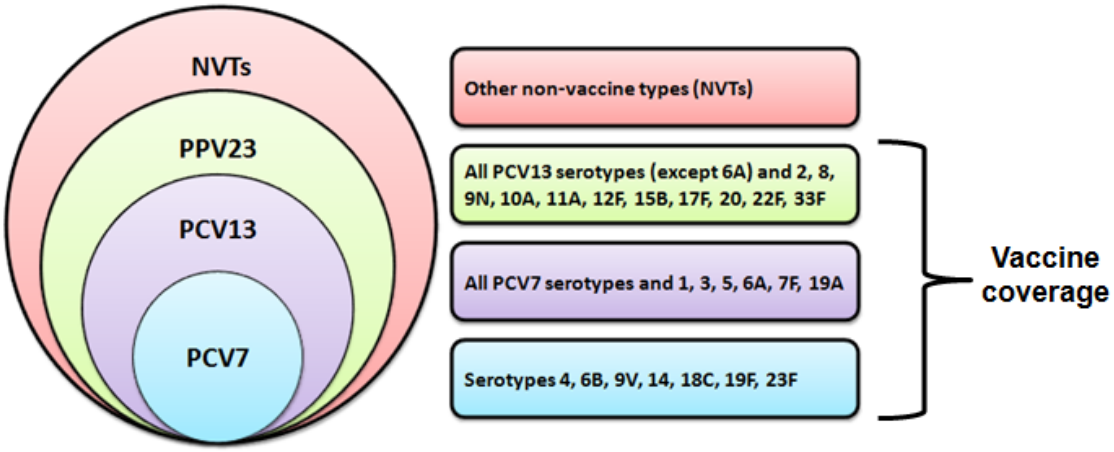


Figure 4. Vaccine coverage by pneumococcal polysaccharide and conjugate vaccines. Serotype coverage for the 7-valent and 13-valent pneumococcal conjugate vaccine (PCV7 and PCV13, respectively), and for the 23-valent pneumococcal polysaccharide vaccine (PPV23) are depicted. Since vaccines only provide protection against the most predominant serotypes causing IPD, other serotypes are also depicted as non-vaccine types (NVT) but are not all listed.

1.1.5.1. Pneumococcal Polysaccharide Vaccines

Studies in 1930s and 1940s led to clinical uses of two hexavalent pneumococcal polysacccharide vaccines (PPV) in 1946 (one for children and one for adults) but they were withdrawn due to therapeutic successes with antibiotics (Robbins *et al*., 1983). A PPV-based vaccine was again introduced in 1977 as a 14-valent formulation, but was later replaced in 1983 by the current 23-valent pneumococcal polysaccharide vaccine (PPV23) called Pneumovax 23 (Merck & Co., Inc., Kenilworth, NJ) (Grabenstein and Klugman, 2012; Robbins *et al*., 1983). The 23 serotypes against which PPV23 provides protection are depicted in Figure 4.

In Canada, the National Advisory Committee on Immunization (NACI) provides recommendations for immunization. NACI recommends PPV23 for children over the age of two at risk for IPD or for individuals aged ≥65 years regardless of risk factors (NACI, 2015). A recent meta-analysis and review on the vaccine effectiveness of PPV23 in preventing both IPD and CAP in adults has recently been published (Kraicer-Melamed *et al*., 2016). This study reported significant PPV23 vaccine effectiveness in preventing IPD in adults aged ≥65 years across studies reviewed. Evidence for the prevention of CAP, however, was not shown to be significant (Kraicer-Melamed *et al*., 2016).

One limitation of PPV23 is due to the polysaccharide composition of the vaccine, which leads to poor immune responses in children under 24 months. For this reason the vaccine was not recommended for routine childhood immunization (NACI, 2015). Moreover, polysaccharide vaccines fail to induce immunological memory for individuals of all ages (Heilmann, 1990). This reduced immunogenicity has been shown to result from a T-cell-independent immune response, which results in B-cells that produce immature IgM antibodies and fail to differentiate into long-living plasma cells. This is why re-vaccination is often needed following PPV23 (NACI, 2015).

1.1.5.2. Pneumococcal Conjugate Vaccines

To overcome the lack of immunogenicity of PPV23, polysaccharide conjugate vaccines were developed. Conjugating a protein carrier to a polysaccharide provides the immunologic signals needed to stimulate a T-cell response, which in turn promotes the production of B-cell memory and capulse-specific antibodies (NACI, 2016; Pichichero, 2013; Stein, 1992). In the U.S. in 2000, a 7-valent pneumococcal conjugate (PCV7) became available called Prevnar7 (previously Wyeth, now Pfizer Inc., New York, NY). Serotype coverage included the seven most predominant serotypes causing IPD at the time (Figure 4). Polysaccharides for each of the seven serotypes were conjugated to a variant of diphtheria toxoid protein, CRM_{197} (Pichichero, 2013). In 2002, NACI recommended the use of PCV7 in all childhood vaccination programs for the prevention of IPD, and it was implemented in all Canadian provinces by 2006 (Bettinger, 2010; NACI, 2002). Several Canadian studies have since demonstrated that PCV7 was effective in reducing the burden of IPD caused by vaccine-type serotypes in children, and indirectly in adults through herd immunity (Bettinger *et al*., 2010; Kellner *et al*., 2009; NACI, 2010; Rudnick *et al*., 2013; Tyrrell *et al*., 2009). For example, the Calgary Area *Streptococcus pneumoniae* Epidemiology Research (CASPER) study measured the impact of PCV7 infant vaccination implementation on IPD in Calgary, Alberta from 1998-2007. The study found a 86% decrease in IPD incidence caused by PCV7 serotypes in children less than 2 years old, and 78% decrease in IPD incidence in individuals aged 65 years or older (Kellner *et al*., 2009).

Despite the encouraging decrease in incidence of vaccine-type IPD after the introduction of PCV7, the successes were short lived. Other IPD surveillance studies began to notice a rise in non-PCV7 serotypes including 7F and 19A (CDC, 2005; Hicks *et al*., 2007; Kellner *et al*., 2009; Moore *et al*., 2008; Pilishvili *et al*., 2010). Such a shift in serotype distribution is called “serotype replacement” and was reported in many countries including Canada and the U.S. (CDC, 2005, Kellner *et al*., 2009; Kim *et al*., 2016; Tyrell *et al*., 2009). Additionally, an increase in antibiotic resistance was correlated

with the emergence of serotype 19A in the post-PCV7 era (Beall *et al*., 2011; Moore *et al*., 2008).

In 2009, a 10-valent pneumococcal, *Haemophilus influenzae*, and diphtheria vaccine (PHiD10), consisting of all the PCV7 serotypes plus serotypes 1, 5 and 7F, was used in Québec, Ontario and Newfoundland and Labrador. However, interest for this vaccine waned quickly since a 13-valent conjugate vaccine (PCV13) called Prevnar-13 (Pfizer Inc., New York, NY) was introduced in 2010. Compared to PCV7, PCV13 contained six additional serotype polysaccharides conjugated to the same CRM_{197} protein as PCV7 (Figure 4). Both the U.S. Advisory Committee on Immunization Practices (ACIP) and NACI subsequently made recommendations to replace PCV7 with PCV13 in childhood immunization programs for children under 24 months of age, but immunization schedules varied by jurisdiction (CDC, 2005; NACI, 2010; Nuorti and Whitney, 2010). Immune responses to PCV13 were not inferior to that of PCV7 based anti-polysaccharide antibody responses and opsonophagocytic responses (Bryant *et al*., 2010; Kieninger *et al*., 2008). Since the implementation of PCV13 introduction in the U.S., IPD in children less than 5 years of age caused by the six non-PCV7 serotypes included in PCV13 has declined by approximately 93% from its 2010 implementation to June 2013 (Kim *et al*., 2016; Moore *et al*., 2015). Similar data from the PHAC in Canada for IPD cases showed a decreasing trend of PCV13 serotypes since its implementation in the childhood immunization schedule in 2010 (PHAC, 2015).

While PCV13 was implemented into childhood immunization programs and recommended for adults and children of any age who are at risk for pneumococcal disease (NACI, 2010), recommendations for immunocompetent adults are still under evaluation (LeBlanc *et al.,* 2017; NACI, 2016). The recent landmark study for PCV13 in adults was the Community-Acquired Pneumonia Immunization Trial in Adults (CAPITA) in the Netherlands (Bonten *et al*., 2015). The trial was conducted from January 2008 to September 2010, and was the first to assess PCV13 effectiveness for the prevention of

vaccine-type CAP and IPD in immunocompetent adults aged 65 years or older. The study was a randomized, placebo-controlled and double blind trial that recruited 84,496 individuals, which were randomized to 42,240 who received PCV13 and 42,256 who received placebo. This study found that PCV13 was ~75% effective in preventing pneumococcal CAP caused by PCV13 serotypes, and ~45% effective for vaccine-type IPD (Bonten *et al*., 2015). Shortly after, and largely based on the result of the CAPITA trial, the U.S. ACIP recommended that all adults 65 years or older should be given PCV13 followed by PPV23 for the prevention of CAP and IPD (Tomczyk, 2014). Recently, NACI released a similar interim recommendation where PCV13 may be considered on an "individual basis" for adults aged 65 years or older (NACI, 2016).

The implementation of PCV13-based childhood immunization has drastically reduced the incidence of IPD worldwide (Kim *et al*., 2016; Loharikar *et al*., 2016; NACI, 2016). The Centers for Disease Control and Prevention (CDC) reported an overall decrease in IPD from 100 in 100,000 cases in 1998 to 9 in 100,000 cases in 2015 (CDC, 2016). However, it should be noted that the serotypes covered by conjugate vaccines were chosen because these were predominant serotypes causing IPD in the U.S. at the time of their design. As such, it may not be an ideal match for the circulating serotypes in all geographic regions (Lin *et al*., 2010), and may lead to varying levels of effectiveness (Bogaert *et al*., 2004b). As demonstrated with the increase in IPD caused by serotype 19A following implementation of PCV7-based infant immunization, bacterial vaccines can lead to serotype replacement. The benefits of PCV13 in preventing IPD (or CAP) will only be sustained if the PCV13 serotypes are circulating. Recently, decreases have been noted for PCV13 serotypes and increases noted for non-PCV13 serotypes such as 6C, 22F, and 33F (PHAC, 2015; Ritcher *et al*., 2013). For these reasons, ongoing surveillance of pneumococcal disease including serotype distribution is essential worldwide to make informed recommendation for pneumococcal vaccines.

1.1.6. Canadian Surveillance of Pneumococcal Disease and Role for Serotyping

Canada has been collecting aggregate data on IPD nationally by volunteer passive surveillance through the Canadian Notifiable Diseases Surveillance System (CNDSS) (PHAC, 2015). With help from the National Microbiology Laboratory (NML) in Winnipeg, Manitoba, who performed Quellung serotyping of all IPD isolates, the Public Health Agency of Canada (PHAC) releases summary reports to describe incidence and serotype trends annually (PHAC 2015 and 2017). The most recent reports describe a decreasing trend (9.5% to 4.9%) for IPD cases caused by PCV7 serotypes rates for all age groups combined over the 5-year period from 2010 to 2014 (PHAC, 2015), and PCV7 increased slightly to 7.3% in 2015 (PHAC 2017). In the same time period, the proportion of PCV13 serotypes in Canada has also decreased from 45.6% in 2010 to 26.0% in 2014 (PHAC, 2015) and continued to decline in 2015 to 23.4% (PHAC 2017). In contrast, the proportion of isolates representing PPV23 serotypes have increased from 24.7% to 38.0% from 2010 to 2014, and non-vaccine serotypes increased from 20.2% to 31.1% (PHAC, 2015). In 2015, PPV23 and NVT serotypes have remained relatively unchanged at 38.6% and 30.7%, respectively (PHAC 2017). While there are clear trends of PCV13 strains decreasing following its use in childhood immunization programs, the data presented by PHAC is limited to IPD. Similar IPD surveillance is being performed solely for the province on Ontario through the Toronto Invasive Bacterial Diseases Network (TIBDN).

Since the incidence of CAP is much greater than IPD for adults, and no CAP surveillance was being performed in Canada, pneumococcal CAP surveillance became a research priority for NACI. With NACI requiring data to support an evidence-based recommendation for use of PCV13 for the prevention of CAP in immunocompetent adults following the CAPITA trial, there was a need for Canadian data to assess the contributions of PCV13 serotypes to pneumococcal CAP over time (NACI, 2016). To cope with this gap, the Serious Outcomes Surveillance (SOS) Network of the Canadian Immunization Research Network (CIRN), formerly PHAC /Canadian Institutes of Health

Research Influenza Research Network Surveillance (PCIRN), began performing active surveillance for CAP and IPD in 2010 (LeBlanc *et al.,* 2017; McNeil *et al*., 2014). While the number of hospitals has varied over the years, active surveillance for CAP and IPD is still underway and spans five Canadian provinces. From 2010 to 2013, 4769 cases of all-cause CAP cases were identified, and 23.2% were identified as pneumococcal CAP (LeBlanc *et al.*, 2017). Pneumococcal CAP cases increased with age, and the burden of illness was evident in terms of requirement for mechanical ventilation, intensive care unit admission, and 30-day mortality. Of the serotypeable results, there was a predominance for PCV13 serotypes (3, 7F, and 19A), and the proportion of hospitalized CAP cases caused by a PCV13-type *S. pneumoniae* was 14.8%. Over time, there seemed to be a decrease in PCV13 serotypes, which would corroborate recent IPD data suggesting possible herd immunity effects (LeBlanc *et al.*, 2017; PHAC, 2015). Serotyping will continue to play an essential role to monitor pneumococcal disease epidemiology and the effectiveness of pneumococcal vaccines over time.

1.2. Laboratory Techniques for Serotyping *S. pneumoniae*

Soon after *S. pneumoniae* was recognized as a cause of pneumonia, immune serum was investigated as a therapeutic option. Throughout the 20th century, it became well recognized that immune protection against pneumococci was serotype-specific, and a large number of serotypes and serogroups had been identified (Watson *et al*., 1993). A serotype was defined as a *S. pneumoniae* strain producing a polysaccharide capsule with a unique chemical structure and generating a distinct serologic property (Geno *et al.,* 2015). A serogroup was a group of serotypes that share some serologic properties (i.e. cross-reacting antibodies). The distinction between serogroup and serotype is important. Antiserum 9L (from Lederle Laboratories Inc.) and 9N (from New York State Laboratory) were ineffective at treating pneumococcal pneumonia caused by serogroup 9 in the Danish Prince Valdemar. After his death in 1939, the *S. pneumoniae* that caused his disease was identified as a new serotype of serogroup 9, named 9V for

Valdermar (Vammen, 1939). The concept of serotype-specific protection provided a strong stimulus for researchers to distinguish all serotypes within each serogroup.

To facilitate antiserum therapy, many serotyping procedures were evaluated but only one was widely accepted and used as the reference method worldwide, called the Quellung reaction (Neufeld, 1902; Sabin, 1933; Beckler and MacLeod, 1934). The Quellung reaction was developed Dr. Fred Neufeld, the assistant of Robert Koch, who used specific antisera to categorize *S. pneumoniae* serotypes (Neufeld, 1902). As the years went by, many laboratories used the Quellung reaction and new serotypes were continuously reported. Over time, two different serotyping systems co-evolved: the Danish (Kauffman *et al*. 1940, Lund, 1970) and the American systems (Eddy, 1944). The Danish system groups serotypes on the basis of antigenic similarities (for instance serogroup 18 consists of 18F, 18A, 18B, and 18C), whereas in the American system the serotype numbers are in the order in which they were first identified (18, 44, 55, and 56, respectively) (Table 1) (Geno *et al*., 2015). The Danish nomenclature for *S. pneumoniae* serotyping has proved to be more practical, and is now commonly used in both research and reference laboratories worldwide. To date, Quellung serotyping or molecular methods have been used to characterize *S. pneumoniae* into 97 serotypes, but commercial sera are only available for 92 (Geno *et al*., 2015).

All of the modern serotyping techniques for *S. pneumoniae* are either derived from, or validated against *S. pneumoniae* strains characterized by Quellung reaction. Specific capsule structures can now be associated to specific genetic signatures (e.g. *cps* loci), which can be assessed by immunologic or molecular methods (Geno *et al*., 2015). While Quellung reaction and PCR-based serotyping are well-established tools for pneumococcal disease surveillance, other surveillance tools have been proposed. The following sections describe methods used for capsule structure determination, the evolution of serotyping techniques using immunologic or genetic-based approaches, as well as some of their advantages and limitations.

Table 1. Comparison of the Danish and US typing systems for *S. pneumoniae*.

Serogroup	Danish type	U.S. type
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6A	6
	6B	26
	6C	6C
	6D	6D
	6E*	ND
	6F	6F
	6G	6G
	6H	6H
7	7F	51
	7A	7
	7B	48
	7C	50
8	8	8
9	9A	33
	9L	49
	9N	9
	9V	68
10	10F	10
	10A	34
	10B	ND
	10C	ND
11	11F	11
	11A	43
	11B	76
	11C	53
	11D	ND
	11E	ND
12	12F	12
	12A	83
	12B	ND
13	13	13
14	14	14
15	15F	15
	15A	30
	15B	54
	15C	77
16	16F	16
	16A	ND
17	17F	17
	17A	78
18	18F	18
	18A	44
	18B	55
	18C	56
19	19F	19
	19A	57
	19B	58
	19C	59
20	20A	20
	20B	ND
21	21	21
22	22F	22
	22A	63
23	23F	23
	23A	46
	23B	64
24	24F	24
	24A	65
	24B	60
25	25F	25
	25A	ND
27	27	27
28	28F	28
	28A	79
31	31	31
32	32F	32
	32A	67
33	33F	70
	33A	40
	33B	42
	33C	39
	33D	ND
34	34	41
35	35F	35
	35A	47/62
	35B	66
	35C	61
36	36	36
37	37	37
38	38	71
39	39	69
40	40	45
41	41F	38
	41A	74
42	42	80
43	43	75
44	44	81
45	45	72
46	46	73
47	47F	52
	47A	84
48	48	82

Note: 6E (*) is a potential new serotype. ND signifies “not defined”.

1.2.1. Capsule Structure Determination

Studies on capsular structure began shortly after the discovery of *S. pneumoniae*, with work on the “soluble specific substance” in culture supernatants (Heidelberger and Avery, 1923). Understanding *S. pneumoniae* epidemiology and the immune responses to specific serotypes requires a much more detailed analysis of capsule structure. There is a need not only to purify the capsular polysaccharides but also to precisely characterize structures by defining any repeating sugar subunits, their order, linkages, and modifications (Figure 6) (Calix *et al*., 2011a and 2012a; Geno *et al*., 2015). Traditional chemical analyses and gas-liquid chromatography were limited as they only describe the capsule sugar composition (Reeves and Goebel, 1941). As such, the capsule structures of

some historically circulating *S. pneumoniae* serotypes still remain undefined (Bentley *et al*., 2006; Geno *et al*., 2015). With more recent technologies, such as matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectroscopy and nuclear magnetic resonance (NMR) spectroscopy, atomic level resolution of capsules can be elucidated (Cai *et al*., 2012; Calix *et al*., 2012a; Geno *et al*., 2015).

Despite the advancements in these technologies to further our understanding of *S. pneumoniae* capsule structures, these techniques have challenges and can be misleading. MALDI-TOF and NMR spectroscopy require sophisticated instrumentation, a high level of technical expertise, and most importantly, purified capsular polysaccharide that remains in its native state. This can be a particular challenge since the purification of capsular polysaccharide is prone to contamination with cell wall components like teichoic acid and peptidoglycan, and modifications like acetylation are labile and can be lost during purification (Lewis et al., 2004; Rajam et al., 2007; Sorensen *et al*., 1990). From a practical perspective, and with the noted limitations, these techniques would not easily be applied for pneumococcal surveillance. However, these techniques have greatly advanced our knowledge of *S. pneumoniae* capsule structures and understanding of serotype-specific differences. Examples of serotype-specific capsule structures are depicted in Figure 5.

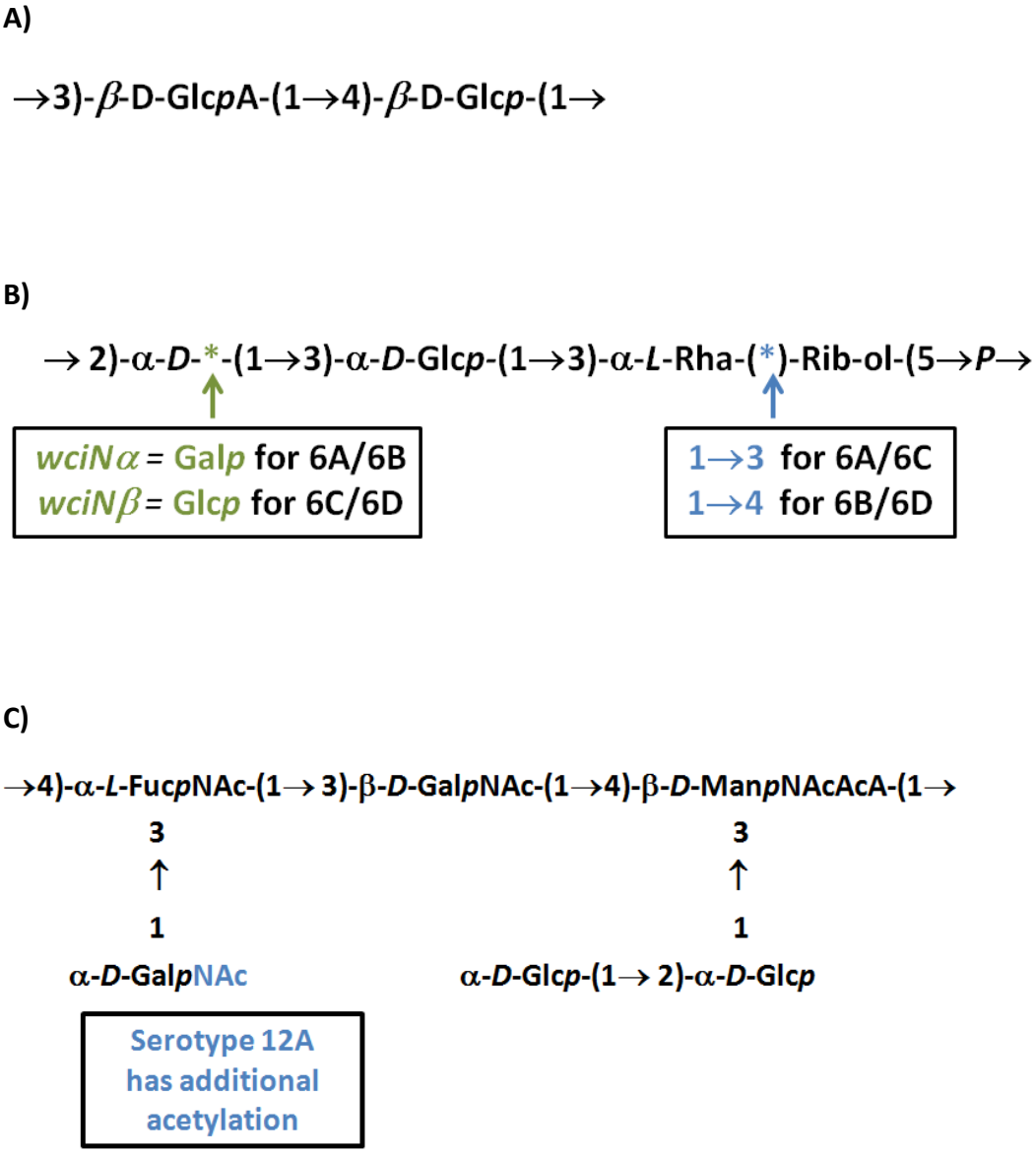


Figure 5. Examples of capsule structure differences. In A) a simple structure for serotype 3 which used the synthase-dependent assembly pathway; B) serotypes 6A, 6B, 6C, and 6D show similar structures, but differences in sugar composition and linkages; C) Acetylation is the only difference between serotypes 12F and 12A. Abbreviations: acetate (Ac); N-acetylfucosamine (FucNAc); galactose (Gal); N-acetylgalactosamine (GalNAc); glucose (Glc); N-acetylglucosamine (GlcNAc); N-acetylmannosaminuronic acid (ManNacA); rhamnose (Rha); ribitol (Rib-ol); phosphate (*p*); pyranose (*P*).

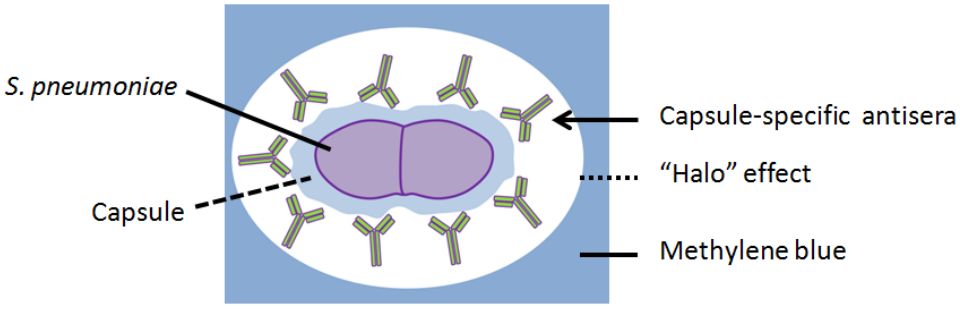
1.2.2. Immunological Methods for *S. pneumoniae* Serotyping

Immunologic methods for *S. pneumoniae* serotyping are phenotypic and rely on antisera directed against specific epitopes present in the pneumococcal capsule. The following subsections describe the well-established reference methods for serotyping (i.e. the Quellung reaction), and a Luminex multiplex bead-based antigen capture technique that has been used for pneumococcal surveillance in recent clinical trials (Bonten *et al*., 2015; LeBlanc *et al*., 2017). Other immunologic techniques that have been developed but not widely used are also summarized, including other agglutination reactions and enzyme immunoassays (EIAs).

1.2.2.1. Traditional Serotyping Using the Quellung Reaction

The Quellung reaction (or Neufeld test) was developed by Fred Neufeld in the early 1900s, and remains the gold standard for *S. pneumoniae* serotyping (Neufeld *et al*., 1902). With availability of commercial reagents through the Statens Serum Institute (SSI) in Denmark, it is commonly used in research and reference laboratories worldwide. The term Quellung refers to the German word for swelling, which is a misnomer for the principle of the test, which relies on microscopic examination of viable *S. pneumoniae* suspensions in presence of pneumococcal capsule-specific antisera (Austrian *et al.,* 1976; Neufeld *et al.,* 1902; Sorensen *et al.,* 1993) (Figure 6A). In a positive reaction, the capsular antigen binds the specific antibody resulting in a change in the refractive index of the capsule, so that it appears swollen and more visible. After the addition of a counter stain like methylene blue, the *S. pneumoniae* cells stain dark blue and are surrounded by a sharply demarcated halo, which represents the outer edge of the capsule. The light transmitted through the capsule appears brighter than either the stained pneumococcal cell or the background. With a negative Quellung reaction, no halo is observed (Austrian *et al.,* 1976; Sorensen *et al.,* 1993). To assign a serotype, different antisera (i.e. pool, group, type, and factor-specific sera) are used in a defined testing algorithm, and serotypes can be deduced from the combined results from positive and negative reactions (Austrian *et al.,* 1976; Sorensen *et al.,* 1993) (Figure 6B).

A)



B)

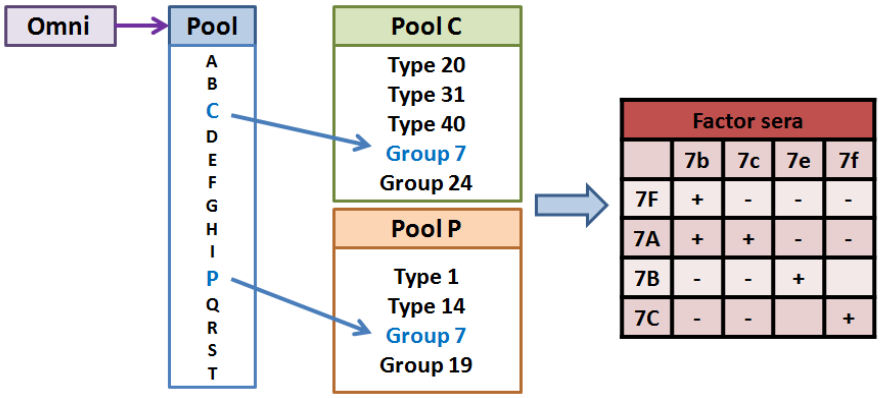


Figure 6. Principle of the Quellung reaction for *S. pneumoniae* serotyping. A) Capsule-specific antibody is added to a suspension of *S. pneumoniae*, and after microscopic examination, exclusion of the methylene blue is observed as a halo for positive reactions. B) The differentiation of serotypes 7F, 7A, 7B, and 7C, relies on pool sera (C and P), group sera (7, 19, and 24), type sera (1, 14, 20, 30, 31, and 40), and factor sera (7b, 7c, 7e, and 7f). Omni serum was used as a pan-pneumococcal positive control. Figure 7B was inspired by Skovsted, 2015.

While Quellung serotyping is simple to perform, and 92 serotypes can be differentiated with the Quellung reaction and commercially available antisera, this method does have limitations such as the subjective interpretation of results, its laborious nature with large surveillance studies, and the requirement for live *S. pneumoniae* producing capsule. Over the years, a number of other immunologic methods have been attempted to streamline serotyping of *S. pneumoniae*, including other agglutination reactions, enzyme immunoassays (EIAs), and flow cytometry.

1.2.2.2. Capsule-Specific Agglutination Reactions

Agglutination reactions like latex- or co-agglutination reactions have been developed for *S. pneumoniae* serotyping using capsule-specific antisera; however, unlike Quellung reactions, microscopy is not required (Kirkman *et al*., 1970). For latex agglutination, each specific serum is coated on latex beads, whereas in co-agglutination, the serum is bound to an organism through a surface protein like staphylococcal protein A (Congeni *et al*., 1984; Lafong and Crothers, 1988; Slotved et al., 2004; Smart and Henrichsen, 1986). In presence of a specific *S. pneumoniae* serotype, a positive agglutination reaction is noted visually by the presence of clumping which is mediated through antibody-antigen recognition (Kirkman *et al*., 1970). When applied to cultured *S. pneumoniae* isolates, the performance characteristics of latex- or co-agglutination are similar to Quellung; however, the sensitivity is poor in clinical specimens, as these methods require a significant amount of antigen for agglutination to occur (Porter *et al*., 2014). The specificity in clinical specimens is also poor due to the many possible cross-reactions with other respiratory flora (Perkins, 1995). The advantage of these methods is that it facilitates specimen workflow, which is more feasible for large-scale surveillance studies (Porter *et al*., 2014). Agglutination reactions can be performed simultaneously on multi-well cards or in multi-well microplates (Geno *et al*., 2015). Some laboratories have further streamlined testing by using only sera for vaccine-preventable serotypes, but this may not be prudent given the possibility of serotype replacement (Sorensen *et al*., 1993).

1.2.2.3. Enzyme Immunoassays

EIA were developed to capture pneumococcal capsular polysaccharide in various bodily fluids (Lankinen *et al*., 2004; Leeming *et al*., 2005; Schaffner *et al*., 1991). These methods been shown to be highly sensitive for the detection of *S. pneumoniae* capsular antigens. On the other hand, serotype-specific EIAs require higher antibody concentrations than commercially available antisera panels for Quellung reactions, and these must be raised prior to use (Jauneikatie *et al*., 2015). The number of detectable serotypes is currently limited to available monoclonal antibodies, and with no commercially available reagents, standardization between laboratories is challenging. However, these methods are appealing since they are amenable to automation, which would be of significant value for large-scale surveillance of pneumococcal disease.

1.2.2.4. Multiplexed Immunoassays

To further automate the detection of serotype-specific pneumococcal antigens, liquid bead-based multiplexed immunoassays were developed (Huijts *et al*., 2013; Pride *et al*., 2012). This technology relies on antigen capture technology using serotype-specific monoclonal antibodies and detection using flow cytometry. Each capsule-specific antibody is coupled to a color-coded microsphere (or bead), and combinations of differentially colored beads can be added to a clinical specimen to simultaneously assess for the presence of different serotypes. A sample of the specimen is added to a flow cytometer with a dual laser, and each bead is analyzed independently. Each bead is subjected to two lasers: one to identify positive reactions if the target capsular antigen is present, and the second to identify the color-coded bead. With thousands of beads are analyzed per specimen, serotypes are assigned if the number of fluorescent counts per serotype is above validated cutoff values for positivity (Huijts *et al*., 2013; Pride *et al*., 2012). Such technology has been applied for the detection of PCV13 capsular antigens in urine (Bonten *et al*., 2015; LeBlanc *et al*., 2017). The disadvantages are the high cost of instrumentation and reagents, and the limited availability. This method has been validated and used in large clinical trials for the effectiveness of PCV13 [*i.e* the

CAPiTA trial], but it is only available through industry collaboration by the manufacturer of PCV13, Pfizer Industries (Bonten *et al*., 2015; LeBlanc *et al*., 2017).

1.2.3. Molecular Methods for *S. pneumoniae* Serotype Deduction

Dr. Neufeld's work on the Quellung reaction paved the way for other important discoveries; such as those by Fred Griffith who showed that *S. pneumoniae* could transfer genetic information and change serotypes, and by Oswald Avery who proved this transformation was due to DNA (Griffith, 1928; Avery *et al*., 1944). With only one exceptions known to date (*tts* in serotype 37), all the genes required for capsular biosynthesis in *S. pneumoniae* are located in the *cps* locus, which is always flanked by the *dexB* and *aliA* genes (Figure 1). In 2006, the Sanger Institute published the DNA sequences of the *cps* loci for the 90 *S. pneumoniae* serotypes known at the time (Bentley *et al*., 2006). Each *S. pneumoniae* serotype has genes that are highly conserved (e.g. *cpsA*), and others that are serotype-specific (Kolkman *et al*., 1998) (Figure 1). As the genotype dictated by the *cps* loci could be linked to phenotype with traditional Quellung serotyping, the molecular era of pneumococcal serotyping had begun. Since molecular methods for *S. pneumoniae* serotype deduction could be used directly on clinical specimens, without the need for culture or live organism, interest for these methods grew rapidly. Many molecular methods for *S. pneumoniae* serotyping have been developed which include many variations of polymerase chain reaction (PCR), and sequencing-based reactions such as sequetyping and next generation sequencing (NGS) (Jauneikatie *et al*., 2015).

1.2.3.1. Polymerase Chain Reaction

Over the years, few molecular methods have been able to provide feasible solutions for pneumococcal disease surveillance, but the success includes conventional and real-time multiplex PCR technologies, and some sequence-based approaches.

1.3.2.1.1. Conventional Multiplex PCR

Conventional multiplex PCR (cmPCR) has been widely used for serotype deduction in *S. pneumoniae.* It relies on serotype-specific genes within the *cps* loci, and is feasible in any laboratory able to perform PCR and agarose gel electrophoresis. Using well-established protocols from the CDC, many research studies have used cmPCR for pneumococcal surveillance (Dias *et al.,* 2007; Jourdain *et al.,* 2011; Lang *et al.,* 2015 and 2016; Morais *et al.,* 2007; Pai *et al.,* 2006). The CDC protocols describe a series of eight cmPCR reactions in which 40 targets that detect 72 different *S. pneumoniae* serotypes (CDC, 2014; Dias *et al.,* 2007). To facilitate workflow, cmPCR reactions are normally performed in sequentially, and most (~75%) of *S. pneumoniae* serotypes can be identified within the first three reactions (Dias *et al*., 2007; Jourdain *et al*., 2011). Due to differences in circulating serotypes over time and in different countries, the order and composition of cmPCR reactions have been modified (CDC, 2014). Lang *et al*., (2016) recently defined a series of cmPCR reactions using oligonucleotides permutations to focus primarily on vaccine-preventable serotypes (Figure 7).

PCR-based serotype deduction has gained popularity over the last few years since it is easy to perform, and does not require culture and isolation of live *S. pneumoniae* (Lang *et al*., 2015; Saha *et al*., 2008). However, PCR-based serotype deduction has limitations. For example, cmPCR reactions were designed to target the most prevalent serotypes causing IPD. While the remaining serotypes could still be detected as *cpsA*-positive by the internal control (incorporated into each of the eight cmPCR reactions), the identity of these serotypes would not be resolved. Secondly, some *S. pneumoniae* serotypes can be detected with cmPCR, but cannot be discriminated from other serotypes (Figure 7). For example, the vaccine-preventable serotype 7F cannot be distinguished from the closely related non-vaccine serotype 7A, and is termed 7F/A by cmPCR. This poses challenges for surveillance programs monitoring serotype changes following implementation of pneumococcal vaccine (Lang *et al.,* 2017).

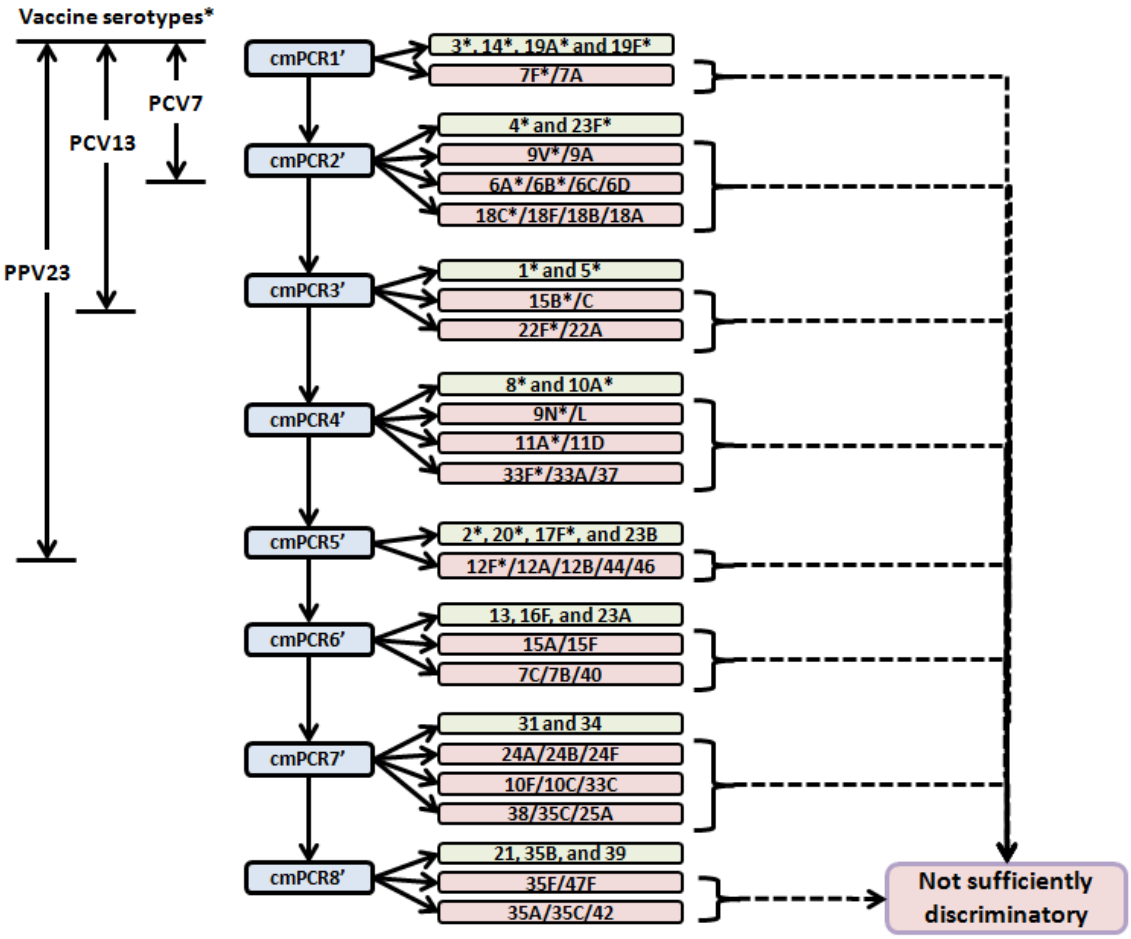


Figure 7. Lack of serotype discrimination with cmPCR. While the eight cmPCR reactions could resolve some serotypes, others were detected but could not be discriminated, including serotypes: 6A/6B/6C/6D, 7F/A, 7C/7B/40, 9A/V, 9N/L, 10F/10C/33C, 11A/11D, 12F/12A/12B/44/46, 15A/15F, 15B/15C, 18C/18F/18B/18A, 22F/22A, 24F/24A/24B, 33F/33A/37, 35A/35C/42, 35F/47F, and 38/25F/25A. The asterisk (*) indicates a vaccine-preventable serotype. Serotype target order for cmPCR reactions are as described by Lang *et al.,* 2017.

1.2.3.1.2. Real-Time Multiplex PCR

When applied directly on clinical specimens, a proportion of cmPCR reactions were unable to assign a serotype in specimens confirmed to be *S. pneumoniae*-positive (Lang *et al*., 2015). In fact, cmPCR was shown to be less sensitive than the detection methods using real-time PCR for *lytA* and *cpsA* (Lang *et al*., 2015). Real-time multiplex PCR (rmPCR) reactions were shown to be 10 to 100-fold more sensitive than cmPCR for detection of target serotypes, and for most serotypes showed equivalent sensitivity as the detection methods (Lang *et al.,* 2015). The rmPCR reactions were designed as seven triplex real-time PCR reactions, with serotype-specific primers and probe sequences (Pimenta *et al*., 2013). While rmPCR eliminated the need for gel electrophoresis and were more sensitive than cmPCR, it was limited by the cost of instrumentation, lack of serotype discrimination, and fewer targets could be detected compared to cmPCR. In particular, being unable to identify all PPV23 serotypes limits its use alone for epidemiological studies (Pimenta *et al*., 2013).

In order to increase the number of detectable targets to 50 (including all PPV23 serotypes), as well as simplify specimen processing; a nanofluidic rmPCR assay was developed for *S. pneumoniae* serotype deduction. With low DNA input volumes, the ability to detect 50 targets simultaneously in a single reaction module and its high sensitivity, nanofluidic rmPCR showed some promise for large epidemiological studies (Dhoubhadel *et al*., 2014). The main limitations are the high cost of custom reagents and instrumentation, single-specimen processing, and lack of serotype discrimination like cmPCR (Jauneikatie *et al*., 2015).

1.2.3.2. Sequence-Based Technologies for *S. pneumoniae* Serotype Deduction

The lack of serotype discrimination for PCR-based methods is obvious when looking at the target sequences in the *cps* loci. In some cases, only single point mutation throughout the entire *cps* loci differentiates closely related serotypes (Kapatai *et al*.,

2016; Mavroidi *et al.,* 2007). Two sequence-based techniques for *S. pneumoniae* serotyping are presented below.

1.2.3.2.1. Sequetyping

In 2012, Leung *et al*. (2012) developed an assay called sequetyping, which involved use of a single primer pair to target a conserved region of the *cps* loci spanning the *cpsA* and *cpsB* gene sequences. Within this region, the sequence variation between *S. pneumoniae* serotypes allowed for differentiation when compared to a validated database. Although the technique is relatively simple and cost-effective, it was shown to have limited specificity and could only partially discriminate 48 serotypes (Leung *et al*., 2012). In a more recent study, sequetyping failed to amplify any sequence in 18% of isolates, and the accuracy was shown to be poor (Dube *et al*., 2015). While initially showing promise, the poor performance of sequetyping precludes its use for pneumococcal surveillance.

1.2.3.2.2. Next Generation Sequencing

In 2016, Kapatai *et al*. developed and validated an automated bioinformatics pipeline for *S. pneumoniae* serotype deduction based on next generation sequencing (NGS) data and comparative genomics. The software is freely accessible online (https://github.com/phe-bioinformatics/PneumoCaT) and is called Pneumococcal Capsule Typing (or PneumoCaT). Briefly, DNA is extracted from the *S. pneumoniae* isolates, and is subjected to NGS. Data quality checks were performed and if valid (>90% coverage for length, a minimal depth of five reads per base pair, and a mean depth of >20 reads per base pair), the genomic sequences are compared to a FASTA file containing serotype-specific signatures within the *cps* loci of 92 different serotypes of *S. pneumoniae* and two molecular types (6E and 23B1). If the query sequences match a single serotype, the result was released. If multiple serotypes could not be distinguished, a second step of the analysis was performed, where the *cps* loci was aligned to a Capsule Type Variant (CTV) database, which contains unique serotype

defining sequences such as presence or absence of capsule biosynthesis genes, inactivation mutations such as early stop codons, single nucleotide polymorphisms (SNPs), and frameshift insertions or deletions.

Overall, the combination of PneumoCaT with the CTV database was able to distinguish 87 of the 92 serologically different serotypes, as well as two molecular types 6E and 23B1. This bioinformatics analysis pipeline provides a valuable tool for pneumococcal surveillance, and was shown to be accurate, and expandable. When tested against 5010 isolates from the U.K., U.S., Iceland, and Thailand, there was over 99% concordance with traditional Quellung serotyping (Kapatai *et al*., 2016). The limitations of NGS and comparative genomics are lack of data to resolve four serotypes in serogroups 24 and 33, the high cost of specimen processing, and the need for a high level of bioinformatic expertise. Future avenues of research are exploring how NGS analyses might become more automated, user friendly, and applied to clinical specimens.

Chapter 2. Rationale, Hypothesis, and Objectives

2.1. Rationale

S. pneumoniae serotypes are defined by antibody-based (Quellung) reactions targeting specific saccharides in their capsules. While serotype-specific genes within the *cps* loci are responsible for differences in the capsule composition, PCR-based serotyping for *S. pneumoniae*, using serotype-specific genes within the *cps* loci, were not sufficiently discriminatory to differentiate pneumococcal serotypes (Lang *et al*., 2015, 2016, and 2017). Serotype-specific phenotypes could arise by loss or gain of gene functions, through insertion, deletion, or modification of genes encoding surface proteins, encoding enzymes involved in sugar metabolism, or even capsule structure modification such as acetylation (Geno *et al*., 2015; Kapatai *et al*., 2016). With the possibility that genetic signatures contributing to serotype specificity could occur inside or outside the *cps* loci, two approaches were undertaken: validation of novel PCR targets outside the *cps* loci, or validation of sequence-based targets inside the *cps* loci for the 28 *S. pneumoniae* serotypes that fell into cmPCR groups which included a vaccine-preventable serotype that could not be discriminated from closely related serotypes. For *S. pneumoniae* serotype 37, the *tts* gene solely responsible for capsule production, is located outside a non-functions 33F *cps* loci (Llull *et al.,* 1999), supporting search for novel PCR targets outside the *cps* loci. Similarly, mutations within the *cps* loci have recently been reported for all serotypes (Kapatai *et al*., 2016).

2.2. Hypothesis

This study hypothesizes that novel PCR targets located outside the *cps* loci or sequencing targets within the *cps* loci may be able to discriminate vaccine-preventable serotypes of *S. pneumoniae* that were undistinguishable by previous multiplex PCR technologies.

2.3. Objectives

2.3.1. Objective 1

The first objective of the study was to identify and validate novel PCR targets located outside the *cps* loci that enable the discrimination of vaccine-preventable serotypes of *S. pneumoniae* (Figure 8).

2.3.2. Objective 2

The second objective of the study was to identify and validate PCR and sequencing targets within the *cps* loci that enable the discrimination of vaccine-preventable serotypes of *S. pneumoniae* (Figure 8).

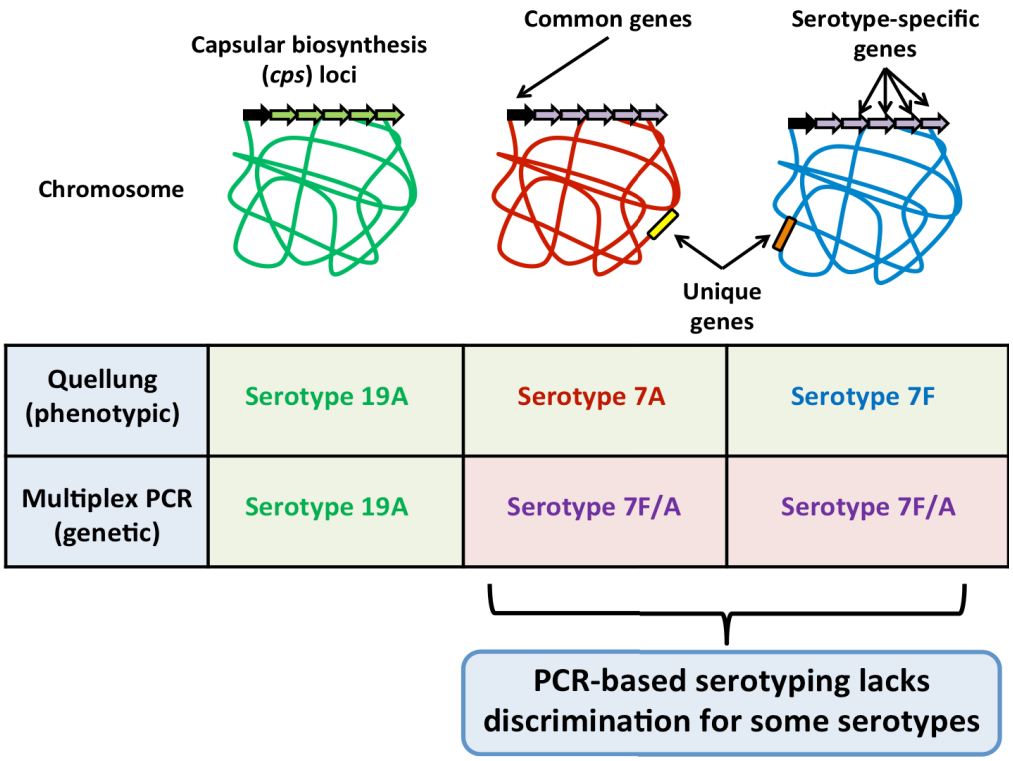


Figure 8. Diagram illustrating molecular discrimination of *S. pneumoniae* serotypes from inside or outside the *cps* loci.

Chapter 3. Methods

3.1. Next Generation Sequencing and Comparative Genomics

Next generation sequencing (NGS) and comparative genomics were used to identify serotype-specific PCR targets outside the *cps* loci for cmPCR groups containing a vaccine-preventable serotype (Figure 7 and 8). Briefly, a representative of each of the 28 *S. pneumoniae* serotypes previously characterized by Quellung serotyping by the Streptococcus and STI Unit at NML (Winnipeg, MB) were subjected to NGS using Illumina-based technology on a HiSeq platform (Illumina, San Diego, CA) at the NML.

In order to identify unique DNA fragments for each *S. pneumoniae* serotypes, NGS data was analyzed and a pipeline was developed by AccuraSciences Inc. (Johnston, IA). First, quality control was performed on the sequence data using FastQC, version 0.11.2 (Andrews, 2014). For mapping, sequencing reads from each strain were aligned to the reference genome using Bowtie version 1.1.2, an ultrafast alignment program for aligning short DNA sequence reads to large genomes that is freely available online (http://bowtie.cbcb.umd.edu) (Langmead *et al*., 2009). The reference genome used was a Canadian isolate of a multi-drug resistance (MDR) *S. pneumoniae* serotype 19A (Genbank accession number ACNU00000000) (Pillai *et al*., 2009). As unique sequences for each serotype were sought, only unmapped reads entered the assembly stage. Reads mapping to the reference genome, or the "mate reads" of the serotypes within the same PCR group that mapped to the reference genome, were eliminated. The remaining unmapped reads were assembled as contigs using SOAPdenovo, which is designed for *de novo* genome assembly using NGS short reads (Luo *et al*., 2012). Velvet was used to manipulate de Bruijn graphs, a compact representation based on short words (k-mers) that is ideal for high coverage, very short read (25-50 bp) data sets for genomic sequence assembly (Zerbino and Birney, 2008). The assembled contigs for each *S. pneumoniae* serotype were subject to nucleotide Basic Local Alignment Search Tool (BLASTn) analysis against contigs identified for the other serotypes within their PCR

group. Unique contigs ranging in length from 250 to 1000 base pairs (bp), and with a depth of coverage >20, were retained for evaluation.

3.2. Classification of Unique Serotype-Specific Sequences

Using NGS and comparative genomics, serotype-specific DNA sequences ("contigs") were identified for 28 selected targets. Of the many potential PCR targets that were possible for each serotype, only a subset could be evaluated due to time and budgetary restrictions. Four targets per serotype were selected for evaluation (but up to seven were tested if the first four did not yield desired results). With only a limited number of DNA sequences that could be evaluated, these contigs were classified to favor sequences that could best enable differentiation of the desired *S. pneumoniae* serotype, without cross-reactions with other serotypes or other organisms. First, a BLASTn analysis was used to classify each candidate PCR target into classes A to D. "Class A" represented DNA sequences with no known similarity to other any other *S. pneumoniae* serotypes, streptococci, or organisms. "Class B" represented DNA sequences with similarity to a limited number of other organisms, but not *S. pneumoniae*. "Class C" represented DNA sequences with similarity to a limited number of *S. pneumoniae* serotypes, but not found in the target PCR group. "Class D" represented DNA sequences with similarity to a large number of organisms and/or *S. pneumoniae,* or a target that is unable to differentiate the desired serotype from the PCR groups. All PCR targets in class D were excluded from further analyses. Since serotype-specific DNA sequences should ideally be highly conserved, any targets with similarity to mobile genetic elements (MGE) such as phages, prophages, or mobile integration and conjugation elements, were also excluded (Croucher *et al.*, 2014).

3.3. Primer Design and Optimization for Serotype-Specific PCR Targets

For each serotype under evaluation, four DNA sequences were selected as potential PCR targets, and forward and reverse oligonuclotides (i.e. primers) sequences for each were designed (Table 2) using default parameters of Primer-BLAST

(http://www.ncbi.nlm.nih.gov/tools/primer-blast/) with melting temperatures ranging between 59°C and 62°C. Following primer pair selection using Primer-BLAST, a BLASTn analysis was repeated with the expected DNA sequence obtained for each serotype (Bentley *et al*., 2006). If BLASTn analyses revealed any similarity hits to sequences from other *S. pneumoniae* serotype, other streptococci, or other organisms (which could potential interact with the oligonucleotides), the primers were re-designed ("optimized") to target areas showing sequences mismatches against these organisms (Figure 9). This process was repeated for all unique contigs. Following primer re-design and a final verification with BLASTn, the selected PCR targets were re-classified following the definitions above (Figure 10 and Table 2).

3.4. Bacterial Culture

Validation of PCR- (or sequence based) serotyping targets required testing with DNA extracted from *S. pneumoniae* and other organisms (Tables 3 and 4). *S. pneumoniae* stock cultures and other streptococci isolates were provided by the freezer strain biorepository in the Division of Microbiology, Nova Scotia Health Authority (NSHA), the American Type Culture Collection (ATCC) (Manassas, VA), the NML, The Canadian Immunization Research Network (CIRN) (Halifax, NS), the Canadian Bacterial Surveillance Network (CSBN) (Toronto, ON), the Toronto Invasive Bacterial Disease Network (TIBDN) (Toronto, ON), the Global Pneumococcal Strain Bank at the CDC (Atlanta, GA), or the Statens Serum Institute (SSI) (Copenhagen, Denmark) (Table 3). All *S. pneumoniae* strains were serotyped by Quellung reaction at the NML, CDC, TIBDN, CSBN, or SSI using commercial omni, pool, group, type, and factor antisera (SSI Diagnostica, Statens Serum Institute, Copenhagen, Denmark). Each isolate stock was maintained at -80°C in skim milk, and cultured onto tripticase soy agar (TSA) plates with 5% sheep blood (Becton Dickinson), or chocolate agar for *Abiotrophia defectiva* and *Granulicatella adjacens*), and all incubated overnight at 35°C in 5% CO_2. Bacterial growth was suspended in PCR-grade water (Invitrogen) to a McFarland value of approximately 2.0, and 200 µl was subjected to a nucleic acid extraction.

Table 2. Classification and oligonucleotides used for the selected serotype-specific PCR targets.

PCR group	Serotype	Target	Class	Oligonucleotides				Size (bp)
				Name	Forward (5' to 3')	Name	Reverse (5' to 3')	
6A/6B/6C/6D	6A*	1	A	JL6AF1	CAATCAACAAGACGCAACAG	JL6AR1	TAAAAGATGGACTGAGCCAAC	105
		2	A	JL6AF2	GCTTGATCATAGTGCGACTCC	JL6AR2	ATTCCATCTGCATTTAATTGTAATAC	146
		3	A	JL6AF3	TGTTGGAAATTTTGAGTAATTTGG	JL6AR3	CTTCAAAAGAAAAATTCCCATTG	285
		4	C	JL6AF4	CAAATGCGCTAGGCTAATCC	JL6AR4	TTGCCAGTTTATTGTCTATGGGTA	201
	6B*	1	A	JL6BF1	AGCCGTTATATCACCTTTTGC	JL6BR1	CCGCACAATATGATGCAATGAG	177
		2	A	JL6BF2	TCTATCGAGCACTTATTTGCCC	JL6BR2	TTCAAAAAGAAGCTAAAGAGAATC	268
		3	A	JL6BF3	CGCCCCTCTAAATAGGTCG	JL6BR3	GGGAATATCTATGGGATCTTG	238
		4	A	JL6BF4	GTAATGTTAGATTTGGAAATTCC	JL6BR4	ATTTCAGATCATGATTCGTTAGAAG	112
	6C	1	A	JL6CF1	GCTTTTCAATACAATTGTTCTTGTGT	JL6CR1	AATTGTAGTTGGAGTCGAAATTGA	101
		2	A	JL6CF2	TCCCAAAATCGCCTTTTTAGT	JL6CR2	GCCAATGTTAACAGGGAAGG	145
		3	A	JL6CF3	GCAATACCTTTCCCAACAGC	JL6CR3	GTTGGTGAAAAATTAATCGCA	258
		4	A	JL6CF4	TGGTTTGTATATCTGGGCCTTT	JL6CR4	TCCAGTTAGAGTAGGATAGGAATTC	195
	6D	1	A	JL6DF1	CTAATATTATCTAATTAAT	JL6DR1	AAGAATCGACTCCAGTGTAAG	86
		1	A	JL6DF1B	GGATGAGTATCGCGAGTCAGG	JL6DR1B	TCGACTCCAGTGTAAGTCCCT	148
		2	A	JL6DF2	ATCCTAGGCTGACAGTTTTAC	JL6DR2	GATTAGAGGATTTAACACGG	219
		3	A	JL6DF3	GATACATTAAGAGAATTATATGAT	JL6DR3	CATTAAAGGGGTCTACTTGGGA	118
		3B	A	JL6DF3B	GGGGATGGACAAGGAGTACG	JL6DR3B	TTGCCCTCTCGTGTTCCTAC	226
		4	A	JL6DF4	AACCAGACGTTGGCGTCTT	JL6DR4	CGAATAAAGAAGTGACAGGCG	142
		4B	A	JL6DF4B	ACCAGACGTTGGCGTCTTTG	JL6DR4B	AGGGTTCAAGTTGGCTCCAT	166
		5	A	JL6DF5	ATTGTCACCGTACTCTATCGTG	JL6DR5	GATGTTGAAGTTCCCGATTGTCT	158
7F/7A	7F*	1	A	JL7FF1	TGAATTTTTCGAAAGCAAAGG	JL7FR1	TCAATGAATCCAAATCTCTCG	102
		1B	A	JL7FF1B	AGCAGTGACTGAATCTGACTGT	JL7FR1B	TCCGTCAATTCTTCTGCTGACA	144
		1C	A	JL7FF1C	CCGGTTAAGAGTGTTATCGTGTG	JL7FR1C	AGTCCGTCAATTCTTCTGCTG	297
		2	A	JL7FF2	TGACAAATTGTTGCATTCTTAAA	JL7FR2	TTTCCATGGCTTTTTGTTGTC	182
		2B	A	JL7FF2B	ACGACAACAAAAAGCCATGGAA	JL7FR2B	CGGGTTTTATGGCTTTAACCCT	421
		3	B	JL7FF3	CAGTTACGATTGTATAATTCTCTT	JL7FR3	AACTTTCTGGCTTAAGTTGGGC	299
		4	C	JL7FF4	CTGTATCGGAACACCTGACG	JL7FR4	AAATACAAGTATGCCCAATTTGTCT	351
		5	C	JL7FF5	AGAACGTTGGATACGGCAGA	JL7FR5	GCGCTAACAATTTCTTCTTCGCT	221
	7A	1	A	JL7AF1	GCTGTTGAGTATCTTGAGAGTTTCC	JL7AR1	TCCGTACATAGGGAGCAAAA	238
		1B	A	JL7AF1B	ACCGTGCCTGTACCATCTTT	JL7AR1B	CAACGTTGATAGCGTGAAGCG	102
		2	B	JL7AF2	TTTCGCCTGCAGTTTCAATGA	JL7AR2	CATTCAATCCGGCGGTACG	371
		3	C	JL7AF3	ATCAATTATCCCGCTAACTAAC	JL7AR3	GTGACATTAAAAGGAGTGAAAAC	117
		4	C	JL7AF4	CCGAGCCGAACATAAGTTGT	JL7AR4	CACCAACTGCAACTGATGAAA	189
		4B	C	JL7AF4B	AATCGGTGCGCACTTTTTCA	JL7AR4B	ACTTATGTTCGGCTCGGTTGA	156
		4C	C	JL7AF4C	ACCGAGCCGAACATAAGTTGT	JL7AR4C	ACCATCCGAACCCTAATACCA	217
		5	B	JL7AF5	AGACTTCACACTCAAGCGCA	JL7AR5	ACGGCATAAAAGCCGTCTGA	193
		6	B	JL7AF6	GCTTACCGGCCTCATCCTAC	JL7AR6	TGGAAATCAGGGTAACGCCC	176
9V/9A	9V*	1	A	JL9VF1	CCGTCAATGCACAATAATCG	JL9VR1	AGTCAAAGCACCAAGAAATCA	250
		1B	A	JL9VF1B	CACTGGATCATCGTGCGACT	JL9VR1B	CGTCAGCATGCCAAAGTCAAA	155
		2	A	JL9VF2	GTGTTTGACAACAGGTTCTGG	JL9VR2	AATTTGGCAAGATGGCGTTG	109
		3	A	JL9VF3	CAATCATCAACCGATCGTTAG	JL9VR3	AAATGAGCATGGAAACATTGGT	140
		4	A	JL9VF4	ATATCGGTGTTAGATGAACGC	JL9VR4	TGAAATAACCGATTTTGGAAAC	201
	9A	1	A	JL9AF1	CGAGAGCAACTAATGTTTTTG	JL9AR1	TATACTTAACACAACCAGCAG	238
		1	A	JL9AF1B	TCTGACGGCGTTATAGTTCGT	JL9AR1B	AATCAAACAAGCACGCACACA	141
		2	A	JL9AF2	TAGAGGGAGTGTTTTCCGG	JL9AR2	TAGCTTCAGTATCCGTATCTA	201
		3	A	JL9AF3	TTAAAAGAGTTCTTCTGACTTTCA	JL9AR3	TCTGAGCAATGAGTTAGGGAG	173
		4	A	JL9AF4	AGTTGTTGTGCGCGCAGTG	JL9AR4	GGAAACGTGGGGTTTGCTG	134

PCR group	Serotype	Target	Class	Oligonucleotides				Size (bp)
				Name	Forward (5' to 3')	Name	Reverse (5' to 3')	
9N/9L	9N*	1	A	JL9NF1	TAAGTCAATGGGTAGGTTTCC	JL9NR1	CTGGTAATTCTTCTAAAAGATA	118
		1B	A	JL9NF1B	AGTCAATGGGTAGGTTTCCCT	JL9NR1B	CCGGTAGAGAATTACTTGTTTTGCC	142
		2	A	JL9NF2	TATGTACTTTGGTTTATGATTG	JL9NR2	TAGTACTAGGAGATGAATTACG	166
		3	A	JL9NF3	GTTTGATAGGTTGCCGTGCT	JL9NR3	CTTGCCATCTTTAACTAGTAGG	138
		3B	A	JL9NF3B	ACTTTGGTTTATGATTGCGATGT	JL9NR3B	GCGTTACCCTCATTAGTTGTGC	212
		4	A	JL9NF4	CATTTCTGCTTTATTTTGTTACTT	JL9NR4	AGTGAAGGAGATAAAGACGG	99
		5	A	JL9NF5	ATGAAGAGTAGATGCACACGC	JL9NR5	TTTTCGAACCAAGCACGGCA	100
		6	A	JL9NF6	TAGCACTCATTTTGACCTTGGA	JL9NR6	TGGAAAGCAGTAGGGATGGGA	221
	9L	1	C	JL9LF1	TGTTTTCATGAATTGGTAAATTGA	JL9LR1	ATGGTTATACAATCTTTGAAGA	278
		2	C	JL9LF2	GCGTTTATTCACAGAAGAAACT	JL9LR2	TTTACCATGTTCACCATTGTTC	154
		3	C	JL9LF3	TAGCTCGCATTGTGAGAGTC	JL9LR3	GATAAAAATGCCTGACTGTCTT	242
		4	C	JL9LF4	TTGGTGATATGCGTGACTTTG	JL9LR4	TTCACATGCTTTTCCTACCTCA	203
11A/11D	11A*	1	A	JL11AF1	AACAGACTTCGATGGTGATAC	JL11AR1	GTAGGACTACGTCTAAG	152
		1B	A	JL11AF1B	TACGGTGTTAAATTCGGTGCTG	JL11AR1	TCACCGAATGCGTCTAATGT	155
		2	A	JL11AF2	TTTCCTTCATTGTAACAATGTG	JL11AR2	TAGTATTGCATGTAACACAAGT	107
		2B	A	JL11AF2B	CTCGTAACGGCCTTCATCAAT	JL11AR2	ACTCGCAGAAGTTAAAGATGCAGA	151
		3	A	JL11AF3	CTGCACAATCTCCCAGTGA	JL11AR3	GGAGTTTATAGAGAAATGGGG	193
		3B	A	JL11AF3B	CACAATCTCCCAGTGAAGCCA	JL11AR3	GGGGTGGTCAGAGAGTGCAT	173
		4	A	JL11AF4	CACGCTTTTTTCCGCCAATTT	JL11AR4	TTTCAGATTCCGACGAGAGA	124
		5	A	JL11AF5	CCGTTGCTTCATCCGTACCA	JL11AR5	ACCATGGTGACCCTACGACT	114
	11D	1	A	JL11DF1	CCTGTATTCTATATAGACTCAG	JL11DR1	TTTAAAACCTGCCCCGCACA	142
		2	A	JL11DF2	CGTCAATAGCACCATCATCT	JL11DR2	GGGGTCGCACCGGAAAG	112
		3	A	JL11DF3	CCTTCTGAAGGTTTTTCATTTATAC	JL11DR3	GGTTTAAAGATTATATGATTGCAG	202
		4	A	JL11DF4	ATGGAGACTAGACGGATTTCCC	JL11DR4	CAGATTGCTCTTGATGGTGGT	234
12F/12A/12B/44/46	12F*	1	A	JL12FF1	GAGGCTGGTGGCTTCACTA	JL12FR1	CGAACCCCGTTAACAACATA	173
		2	A	JL12FF2	CTTTAAAACAAATAACAAATCCCC	JL12FR2	TTTTCAATTTATTTGGCAAATGAC	130
		3	A	JL12FF3	CTTCCAAGTAGCATAACTGCT	JL12FR3	AGTCTTTGGGGACAGTGCTA	328
		4	C	JL12FF4	CATCGTCATCATCGTTATAGCT	JL12FR4	AAAGGCTGATATTGCGT	215
	12A	1	A	JL12AF1	AAAATGGGGACTCTGCAAAA	JL12AR1	CTTGTTTTAGGTGCGTTGAGC	104
		2	A	JL12AF2	TGGATGCTAGTGATGATGTGC	JL12AR2	TCCTCAATCGTTACCAACTCTG	133
		3	A	JL12AF3	CTGCAAAATTTAGCGGAAGC	JL12AR3	CATCCAATTCGCCTTTGACT	192
		4	A	JL12AF4	TCTTCAGTTTATGGCGGACA	JL12AR4	AGATATCAAGAGAGCCCTCG	373
	12B	1	A	JL12BF1	TCTCCTCCATATCTGCTTCAG	JL12BR1	TCACATGTGGATTGATTTGTCA	236
		2	A	JL12BF2	GGGTTGAAAGTAGGAAAGAACG	JL12BR2	TGTCTTGACGATTATTAGATGTTCC	201
		3	A	JL12BF3	CCCATTGCCCTCTGTATGTAA	JL12BR3	TGAAGCAAAGTTCAGAGAAACA	106
		4	B	JL12BF4	GGTTTTTCTTAGAAAACTCTATC	JL12BR4	TGCATTATCTCTTTGTAATACC	139
	44	1	A	JL44F1	CCAAATTAATGGATGGTGCAG	JL44R1	TGATTTTTCACAAACTGCCCTAC	251
		2	A	JL44F2	CATTTGAAATTCTTGCCGTGT	JL44R2	CAACGTTGTTGTATCGACCA	216
		3	A	JL44F3	TCGTTGGCTCAGTCAATCTTT	JL44R3	GAGTCGCACTATGATCAAGCA	295
		4	A	JL44F4	CCTTTGAGTCATTCAATATACC	JL44R4	ACATTGATTTTAATAGTGCAGG	149
	46	1	A	JL46F1	ACGGCTCGTTCTCAACCG	JL46R1	ATATATGACCATATACTTGAGC	191
		2	C	JL46F2	CAACTTCAATCTTTATCATCTTC	JL46R2	GACAGAGACGGGAAGAGTG	165
		3	C	JL46F3	AACTCACGAATTGGTGTCGC	JL46R3	CACCTGTATGGGATAATCAAGC	97
		4	C	JL46F4	GAATCAATGCCGGAAAAATC	JL46R4	CAGAATAAGCGACTCGTTTGC	142
15B/15C	15B*	1	A	JL15BF1	CTTTGATAGAAAGGATGATATAG	JL15BR1	GATAACATTATATTCAACATTACC	199
		2	A	JL15BF2	TAGATTTCGCACGATTACTTAA	JL15BR2	TCTCTCCTGATACTTAGT	109
		2B	A	JL15BF2B	TCGCACGATTACTTAACCAACTG	JL15BR2	GCCGCCTCTACATCTGAACA	124
		3	A	JL15BF3	CAAAACATTTTTCAATCATTGGGA	JL15BR3	ATCTACCGAACATTCACTATG	242
		3B	A	JL15BF3B	GTTTACAATGCGCTCCGACC	JL15BR3	ACCAAGTTTAGGATGCTCGTGA	158
		4	A	JL15BF4	GTTTGAATTTTAGGAGTATACATT	JL15BR4	AGCAAAATGTTTGAGTCCAGC	142
		4B	A	JL15BF4B	GGCGATCAAAAAGTTCCGCC	JL15BR4	CATGACAAAGTTGGTCTCGAAG	152
		5	A	JL15BF5	ACTTCACTTAGAACTGATGCAGG	JL15BR5	AGTGCGTTATCTTTACACTGCT	143
	15C	1	A	JL15CF1	ACGTATAGTAGACATATCTGCC	JL15CR1	ATTGATTATTCTAAGTATGGTCTGTC	111
		2	A	JL15CF2	GAAACTGGTGCTGTTGTCCG	JL15CR2	TCGTCTAGCATAGGGTACAG	203
		2B	A	JL15CF2B	ACTGGTGCTGTTGTCCGATT	JL15CR2	TCGTCTCGCATAGGGTACAG	200
		3	A	JL15CF3	AATCTGTACTAATTCCTCTATC	JL15CR3	GTTAGCAAAATCCACAACGTGC	157
		4	B	JL15CF4	TGATGGTGATTACGAAAGTAAA	JL15CR4	CCTTATATTATCAGTTTCCATTTG	247
		5	C	JL15CF5	TCATAATAACGCAAACGGCG	JL15CR5	TCGGCTTTGTAATGAGTTGTTTCT	105
		6	B	JL15CF6	AGCCTCATTGTAACCACGCA	JL15CR6	GCTGAAGCATTTGCTCAAGGT	189
		7	C	JL15CF7	AAATCACCATCCTCGCTTAACCA	JL15CR7	AATTGAAATCCTTCTCGTTGGACA	100

PCR group	Serotype	Target	Class	Oligonucleotides				Size
				Name	Forward (5' to 3')	Name	Reverse (5' to 3')	(bp)
18C/18F/18B/18A	18C*	1	A	JL18CF1	CAAATTATTTCCCTAATCTTCGC	JL18CR1	ACTGTCGACAAATTCTGAAACT	179
		2	C	JL18CF2	TCTGAGGGGGATGCGTATAA	JL18CR2	CGGTACTACGACTGTAAGG	135
		3	C	JL18CF3	TTAACCTCTTCAGATTGTTGC	JL18CR3	CATGGCAGATGGCGTCACACA	99
		4	C	JL18CF4	AATACCAAACGCTCGCCTTA	JL18CR4	TGAAAGTTTCGCTTCCAAGG	247
	18F	1	A	JL18FF1	TGAAACAAAGCGGGAAATAAAG	JL18FR1	TTGATTCATATAAAGAAATCGGAT	161
		2	A	JL18FF2	ATTCATCTGTAGAAATTGCAGC	JL18FR2	GTCCGAGCGGCTAATTGTTGGG	116
		3	A	JL18FF3	ATGGCCTACGCGCAAATTGC	JL18FR3	GGGCAATGTTTGCATTAATCAA	201
		4	A	JL18FF4	AAATATGGGGATTGGCTTGA	JL18FR4	AACTGCTTAGTGACGACAGC	300
	18B	1	A	JL18BF1	TTTTTCTCCTCGGCCTGATA	JL18BR1	TCCATGGAGAAGCTCGACTT	125
		1B	A	JL18BF1B	TTCTCCTCGGCCTGATATGAT	JL18BR1	GGAGAAGCTCGACTTTTACTCA	115
		2	A	JL18BF2	AACGGAAATTTGTTGCGGTTC	JL18BR2	AAGACGACGACATCCGAACT	156
		3	B	JL18BF3	GTGAGAAATGGCTTATAACAAAT	JL18BR3	AGATGTTTCTGATGTTGAATCA	246
		4	B	JL18BF4	AAAATCTCCATTTTTGATGGCG	JL18BR4	CCTATCTATTAGAAATGGTAAT	280
		5	C	JL18BF5	GAGTAAGTGCGGGTGTGCT	JL18BR5	ACGCCTTACTACATACATCAGG	171
		6	C	JL18BF6	CACAGTCCCAAAATTGTCAGCA	JL18BR6	TTTTGCCGGATATTTTAGCGGG	270
		7	B	JL18BF7	GGTGCTCAGTAGAGGGAAAGT	JL18BR7	CCATTCAATCCGTACTGCTGAC	145
	18A	1	A	JL18AF1	GGTAATAATCAGATAGCAGGGAA	JL18AR1	CTTTCTAGTAAAAGAGTTTTCATC	113
		2	A	JL18AF2	CGCGACAGGTGGCTGATA	JL18AR2	GACAAATCTCTGTGAGTATTG	252
		2B	A	JL18AF2B	GCAAACTTTCTGGTCTGCCA	JL18AR2	AGGAGGAAGGCCTAAATCTATACG	158
		2	A	JL18AF2C	CGTGTCATATCACAGACCGT	JL18AR2	GGCATCAGCATTAGTTGTTCCA	135
		3	A	JL18AF3	CGTATTGCAATAAGCATCTAC	JL18AR3	CGATTATCTAAGGATTGGAGTTG	151
		3	A	JL18AF3B	ACGCAGGCCGCTGAAATAGA	JL18AR3	GGAAAACCCTCAGCCGCAA	167
		4	A	JL18AF4	ATTATTAGGGGTATCGTGCT	JL18AR4	ACCGCGATACTACTAACAAC	200
		5	A	JL18AF5	GGTGACGCGTACCTAAGTAGA	JL18AR5	CGACAGCGTAAGAATCTGGGAT	106
22F/22A	22F*	1	A	JL22FF1	AGATTGCTAGACTAGAGAGAAC	JL22FR1	TGTTGATGATGGTTGTTGACG	140
		1B	A	JL22FF1B	CACAAGCGCGGCAGTAAGTA	JL22FR1	TCTCTTTTACGGCGTTTGGCA	195
		2	C	JL22FF2	GCTCAGGAAACTTTGCATTC	JL22FR2	TGGAAGTGTGAGATTTTAGAATT	162
		2B	C	JL22FF2B	GGGCGAGAGTAACGGTTGAT	JL22FR2	AGAGGGTGTATCAAATGCGGA	256
		3	C	JL22FF3	TGTCAGTAGGGCCATCTCCT	JL22FR3	GGAATCGTTTAACTCCCACCT	328
		3B	C	JL22FF3B	TGTCAGTAGGGCCATCTCCT	JL22FR3	GGGAATCGTTTAACTCCCACCT	364
		4	C	JL22FF4	AGCAATTCGTTACCGACAGG	JL22FR4	AGATGAAAGGGATCGCAAAA	252
		5	C	JL22FF5	GGTAACTTTCTTGGAAGTGAAGCA	JL22FR5	ATGTATCCTGCTCCACCTGTT	135
	22A	1	A	JL22AF1	AATCTGTACCAATTCCTCTATC	JL22AR1	ACTACACAAGACCTAATATGTTA	176
		1B	A	JL22AF1B	ACCATTGAGCTTGCAATATCCG	JL22AR1	AAAGGCATAGATTCCCTGATGGG	185
		1C	A	JL22AF1C	GTTGTACACTCCCATCAGGGAA	JL22AR1	GGTTGTCGAAACTTACGAGGAAA	101
		2	C	JL22AF2	TGGAAGGTGAATTGCCTGAT	JL22AR2	CCTTCCGGTCATCACTTCAG	328
		2B	C	JL22AF2B	ATCTCCGTACCTGAGAAACCA	JL22AR2	GGGAGGGATAGATGCAGGTTC	220
		3	C	JL22AF3	GTTCTGGAATTTCTCCGCTG	JL22AR3	ACGAGTTATTGTGCCCGAAC	416
		4	C	JL22AF4	TGCAGGAACAAGTTCGACAC	JL22AR4	GTTAAAAGAGGACGTTCAATGTA	111
		4B	C	JL22AF4B	GTGCAGGAACAAGTTCGACA	JL22AR4	GGAGCACCTACGCCATCAAAA	145
33F/33A/37	33F*	1	B	JL33FF1	CAGCTTGAGCTACTTGAGCG	JL33FR1	ATAAAAGGAGGACTGGAATGTT	141
		2	A	JL33FF2	TTTTCCAATTTAAGCACTTGCC	JL33FR2	GCCATTGGTTTCTTGTTGGCA	197
		3	A	JL33FF3	GGGATATATTACTTTGATGTTAAA	JL33FR3	AAACATGATGGTACTTCTCAGA	163
		4	A	JL33FF4	ATTAAATCGCTTTGTTGGATTAAA	JL33FR4	TAAACAAATACTAATAAAATGAATC	222
	33A	1	A	JL33AF1	AGCAAGCGTCTTTGACGCG	JL33AR1	CCATAATCAACTAATTCTTTTAAC	101
		1B	A	JL33AF1B	TTAATGTTGCGAACCCTAGTGA	JL33AR1	ATAACCCTACTTCAACGCGTCA	100
		1C	A	JL33AF1C	TGACGCGTTGAAGTAGGGTT	JL33AR1	ACAACTTGCTAATATCGACATTGTT	117
		2	A	JL33AF2	GATTTCAAATGGAACCGTGGC	JL33AR2	TTGCTATTACGTGTACTTCCG	181
		3	A	JL33AF3	ATGGATTAAAGAAACAGACTGC	JL33AR3	AAAGCTGTGACAAGAGACCA	112
		4	A	JL33AF3B	TGGATTAAAGAAACAGACTGCGT	JL33AR3	AGCTGTGACAAGAGACCACA	109
		4B	A	JL33AF4	AGTTTTTGGTGAATGTTGTCAG	JL33AR4	TATAAGCTAGTTAACAAATGGTC	137
		5	C	JL33AF5	TCGGAATGTTGTTGCCTGCTT	JL33AR5	AAGACCTGCCTAGCTTTTGGA	119
	37	1	A	JL37F1	TCATTGTTATTAGCATTTGATCAGT	JL37R1	AAAAGGAGAAAAACCTGAAGAAA	203
		1B	A	JL37F1B	TTTTCTGCTTCTCTATGCTCACA	JL37R1B	AGATCTTTAAAAGGAGAAAAACCT	100
		2	A	JL37F2	GAAATTTTTCTGTTTTACTTTGTCCA	JL37R2	TGTCAATCATACTACGAAATTGTAAC	148
		3	A	JL37F3	AAGAGCGTGTTTAGGTGGCTA	JL37R3	AACATAATCACCTAAAATTAGCAAA	259
		3B	A	JL37F3B	AAAGAGCGTGTTTAGGTGGC	JL37R3B	GCCATATTCAAGTAGTATGTTCCAA	109
		4	A	JL37F4	TTTCTAAGTGACACTACTACG	JL37R4	GTCAGCAACGCACGCGAT	96
		4B	A	JL37F4B	AGCGGTTTTCTAAGTGACACTAC	JL37R4B	CAGCAACGCACGCGATAAA	100
		5	A	JL37F5	TCAGCCGTTTTTGTATTACATCAGT	JL37R5	AACCCATATACTCAATTGGATGGA	116
		6	A	JL37F6	TGAACGGCTGAGATAAACGGG	JL37R6	GGGAGAACGCAGAGAGAATGA	209

Table 3. *S. pneumoniae* strains used in this study.

PCR group	Serotype	No.	Source/Strains
6A/B/C/D	6A	8	CIRN (ON02-3575, BC05-5319); NML (SC022*); NSHA (25-02, 38-76, 66-08, 68-08, 68-32)
	6B	12	CDC (2231, 4132, 5009, 2240, 1564); NML (SC0023*, SC12-1240, SC12-1803, SC15-1221, SC15-2638, SC15-3272); NSHA (37-12)
	6C	20	CIRN (BC05-5238, NB01-5782, NS01-3580, ON06-1083, ON06-1090, QC07-5135); NML (SC0262*), NSHA (25-28, 25-58, 37-67, 37-68, 38-13, 38-14, 38-58, 66-60, 66-73, 68-66, 68-44, 68-47, 68-77)
	6D	13	CDC (5457, 3390, 5766, 3493); CIRN (4-19, 4-19, 4-20); NML (SC0129*, SC12-0507, SC13-1651, SC13-4981, SC14-2976, SC15-3073)
7F/A	7F	22	CIRN (BC05-7717, NB01-5341, NS01-1162, NS01-1178, NS01-5156, ON02-5044, ON06-3030, ON06-3431, ON06-5125, ON06-7300, ON06-7335, ON06-10008, QC07-3042, QC07-3118, QC07-5099, QC07-5354, QC07-7251), NML (SC0218*); NSHA (37-34, 37-44, 37-45, 68-22)
	7A	4	NML (SC0025*, SC0217, SC10-1737, SC12-2699)
	7A(7F)**	10	CBSN (2544, 27134, 26923, 26902, 43767); TIBDN (31349, 32197, 36029, 38137, 38137)
9V/A	9V	11	CDC (99, 2545, 4167, 4908, 5886); NML (SC0172*, SC12-0585, SC12-3457, SC13-4234, SC14-0735, SC15-1614)
	9A	8	CDC (3677, 3747, 3399, 5874, 7280); NML (SC0029*, SC10-0658, SC10-1771, SC13-0662, SC13-2978)
9N/L	9N	17	CIRN (NB01-7003, NB01-7089, ON02-3678; ON06-3119; ON06-5067; 0N06-5984, ON06-10032); NML (SC0031*); NSHA (4-21, 25-16, 25-37, 25-32, 68-04, 37-30, 37-37, 68-19, 68-25)
	9L	21	CDC (566, 667, 1864, 4273, 5479); CSBN (15084, 20014, 21574, 23824, 28582, 27650); NML (SC0011*, SC10-0659, SC11-1928, SC16-0928); TIBDN (22074, 24642, 25164, 27221, 27194, 27225)

PCR group	Serotype	No.	Source/Strains
11A/D	11A	18	CIRN (BC05-3283, NS01-5169, ON02-1183, ON04-1171, ON06-13510, QC09-7535, QC09-11405); NML (SC0035*); NSHA (25-17, 37-32, 37-69, 38-11, 38-19, 38-70, 66-80, 68-46, 68-63, 68-70)
	11D	4	CARA (CARA GZ-11D); CDC (2462); NML (SC0271*); TIBDN (21411)
12F/A/B/44/46	12F	13	CDC (1199, 2292, 4438, 5617, 6566); CIRN (4-5, ON02-1240); NML (SC0199*, SC14-0150, SC14-1506, SC14-3562, SC15-2072, SC15-2485)
	12A	8	CDC (1706, 3762, 4170, 4233, 4508); NML (SC0066*); TIBDN (37347, 40719)
	12B	7	CIRN (6-25), CSBN (27651, 29419, 29423, 40453); NML (SC0268*); TIBDN (40454)
	44	2	CDC (2473); NML (SC0212*)
	46	3	CDC (2474); NML (SC0096*, SC0214)
15B/C	15B	11	CIRN (NB01-3484, NB01-5484, NB01-7369, ON02-3563, ON06-5963, ON06-5968, ON06-7326); NML (SC0044*); NSHA (37-60, 38-80, 68-
	15C	19	CDC (1822, 2015, 3973, 4976, 5022); CIRN (BC05-7070; ON06-7418); NML (SC0045*, SC14-3041, SC14-4014, SC15-0862, SC15-2794, SC15-3462); NSHA (25-35, 25-57, 25-60, 37-07, 38-03, 63-57, 68-75, 69-1)
18C/F/A/B	18C	11	CDC (151, 4612, 4977, 7077); NML (SC0050*, SC13-0168, SC14-2419, SC15-1822, SC15-3157); NSHA (38-36, 38-55)
	18F	7	CDC (2277, 2312, 4311, 4599); CIRN (ON02-1227, ON02-3071); NML (SC0051*)
	18A	12	CDC (1102, 1905, 3394, 3625, 5050); CSBN (30377, 30969); NML (SC0049*, SC16-0700); TIBDN (16485, 46792, 48770)
	18B	22	CDC (269, 3185, 4598, 5042, 5122); NML (SC0009*, SC10-2572, SC11-3743, SC12-0474, SC13-2464); TIBDN (29122, 34132, 34360, 41217, 41383); CSBN (15437, 20631, 21738, 25503, 25547, 28141, 28457)

PCR group	Serotype	No.	Source/Strains
22F/A	22F	20	CIRN (NB01-5143, NB01-5479, NS01-3332, NS01-10032, NS01-10043, ON02-5394, ON06-3281, ON06-5202, ON06-5228, ON06-5414, ON06-7426, QC07-7213;); NML (SC0291*); NSHA (38-15, 38-31, 38-51, 66-05, 66-66, 68-43, 68-51)
	22A	10	CDC (500, 1850, 4257, 4349, 5796); NML (SC0059*, SC10-2480, SC12-0917, SC14-2981, SC15-1795)
33F/33A/37	33F	16	CIRN (BC05-10100, ON06-1035, ON06-1174, ON06-10805, ON06-13122, QC07-10484, QC09-7966,); NML (SC0190*); NSHA (32-23, 37-59, 37-61, 37-65, 66-03, 66-22, 66-46, 68-14)
	33A	6	CDC (1754, 6987, 7015); NML (SC0077, SC0078, SC0080*)
	37	10	CDC (1858, 3598, 4265, 4763, 5897); NML (SC0086*, SC12-0885, SC14-2402, SC14-4028, SC15-2412)

*The asterisks denote reference strains used for each *S. pneumoniae* serotype.

Table 4. Streptococci used in the specificity analysis.

Organism	ATCC No.	Lancefield group
Streptococcus pyogenes	19615	A
Streptococcus agalactiae	12386	B
Streptococcus dysgalactiae subsp. equisimilis	12388	C
Streptococcus sp.	9933	L
Streptococcus dysgalactiae subsp. equisimilis	12394	G
Streptococcus equi subsp. zooepidemicus	700400	C
Streptococcus equi subsp. equi	33398	C
Streptococcus salivarius subsp. thermophilus	19258	
Streptococcus sp.	9328	F
Enterococcus faecium	19434	D
Enterococcus faecalis	19433	D
Streptococcus sp.	9936	N
Streptococcus pseudopneumoniae	BAA-960	
Streptococcus infantarius subsp. infantarius	BAA-102	
Streptococcus mitis	49456	
Streptococcus oralis	35037	
Streptococcus gordonii	12396	
Streptococcus sanguinis	49298	
Streptococcus mutans	35668	
Streptococcus salivarius subsp. salivarius	7073	
Streptococcus gallolyticus subsp. gallolyticus	43143	D
Streptococcus gallolyticus subsp. pasteurianus	700338	D
Streptococcus anginosus	33397	G
Streptococcus constellatus subsp. constellatus	27513	
Streptococcus intermedius	27335	
Streptococcus porcinus	35650	E
Abiotrophia defectiva	49176	
Granulicatella adjacens	49175	
Streptococcus suis	43765	R
Streptococcus sp.	9935	M
Streptococcus sp.	11843	O

*Abbreviations: American Type Culture Collection (ATCC)

3.5. Nucleic Acid Extraction

Total nucleic acids (TNA) were extracted from a 200 µL bacterial suspension using a MagNA Pure LC Total Nucleic Acid Isolation Kit (Roche Diagnostics, Indianapolis, IA) on a Roche MagNA Pure LC extractor, according to manufacturer instructions. TNA was eluted in a final elution volume of 100 µL, and 5 µl was used as template for all PCR reactions. TNA was stored at 4°C for a maximum of 24h, or at -20°C for long-term storage. Of note, PCR-grade water (Invirogen) was used as a negative control (NC) for all nucleic acid extractions, and the resulting TNA was use and the NC for all PCR reactions.

3.6. Quantitative Real-Time PCR

Pneumococcal DNA obtained from nucleic acid extraction was quantified using *lytA* real-time PCR (Lang *et al*., 2015). Real-time PCR for the detection of *lytA* was performed using a Life Technologies Taqman Universal PCR Master Mix kit (ThermoFisher Scientific, Waltham, MA) in 25 µl reactions as follows: 1× Master Mix, 200 nM of primers LytA-F (ACG CAA TCT AGC AGA TGA AGC A) and LytA-R (TCG TGC GTT TTA ATT CCA GCT), and 200 nM of probe (LytA-pb) (FAM-TGC CGA AAA CGC TTG ATA CAG GGA G-BHQ1).

Amplification was performed using a Life Technologies ABI 7500 Fast instrument under the following thermocycling conditions: initial activation at 95°C for 10 minutes, followed by 45 cycles of denaturation at 95°C for 15 seconds and a combined annealing/extension step at 60°C for 60 seconds. Fluorescence was captured at 520 nanometers (nm) during the annealing stage, and threshold cycle (Ct) values were determined by the manufacturer software. Results are expressed as copies/µl by comparison to a previously generated standard curve (Lang *et al*., 2015). The *lytA* standard curve represented an inverse linear relationship ($y = -3.258x + 37.38$; $R^2 = 0.9981$) between Ct values and concentration (log copies/µl).

3.7. Conventional PCR

Pneumococcal DNA was subjected to conventional PCR to confirm Quellung serotyping results (Tables 5 and 6) (Lang *et al.,* 2017), to evaluate novel serotype-specific PCR targets (Table 2), or to amplify serotype-specific targets prior to sequencing (Table 7). All oligonucleotides were synthesized by Sigma Genosys (Sigma Aldrich, St. Louis, MO). Each of the cmPCR reactions were performed in 25 µl reactions using 1x Quantitect Multiplex Master Mix No Rox (NR) (Qiagen Inc, Hilden, Germany), and 500 nM primers (Table 2 or 7), with the exception of the traditional cmPCR in which concentrations varied by serotype (Table 5 and 6) (Lang *et al.,* 2017). DNA amplifications were performed in 96-well plates on a C1000 Touch Thermal Cycler (BioRad Laboratories Inc., Mississauga, ON). The timing and temperatures of the amplification cycles are as follows: initial denaturation at 95°C for 15 min; 35 cycles of denaturation at 94°C for 30s, annealing 54°C for 90s, and extension at 72°C for 60s; and a final extension at 72°C for 10 min. Amplicon DNA was held at stored at 4°C until use.

Of note, the novel PCR targets were assessed with DNA extracted from the reference serotypes (Table 3) at concentration of 10^6 and 10^9 copies/ml, which represents concentrations 10× and 10000× the limit of detection of traditional cmPCR (Lang *et al.*, 2015). If no targets were amplified, the $MgCl_2$ concentration was increased from 1.5 mM to 2.5 mM, or the annealing temperature was decreased to 50°C. If the target was amplified, it was tested for reproducibility, and against DNA extracted from other *S. pneumoniae* serotypes (Table 3), and other streptococci (Table 4).

Table 5. Oligonucleotides used in the conventional multiplex PCRs by Lang *et al*., 2017.

Name	Sequence (5' to 3')	Concentration (nM)	Serotypes detected
cmCpsA-F	GCA GTA CAG CAG TTT GTT GGA CTG ACC	100	All but 25A, 25F, and 38
cmCpsA-R	GAA TAT TTT CAT TAT CAG TCC CAG TC	100	
cm1-F	CTC TAT AGA ATG GAG TAT ATA AAC TAT GGT TA	300	1
cm1-R	CCA AAG AAA ATA CTA ACA TTA TCA CAA TAT TGG C	300	
cm2-F	TAT CCC AGT TCA ATA TTT CTC CAC TAC ACC	300	2
cm2-R	ACA CAA AAT ATA GGC AGA GAG AGA CTA CT	300	
cm3-F	ATG GTG TGA TTT CTC CTA GAT TGG AAA GTA G	300	3
cm3-R	CTT CTC CAA TTG CTT ACC AAG TGC AAT AAC G	300	
cm4-F	CTG TTA CTT GTT CTG GAC TCT CGA TAA TTG G	300	4
cm4-R	GCC CAC TCC TGT TAA AAT CCT ACC CGC ATT G	300	
cm5-F	ATA CCT ACA CAA CTT CTG ATT ATG CCT TTG TG	300	5
cm5-R	GCT CGA TAA ACA TAA TCA ATA TTT GAA AAA GTA TG	300	
cm6A/6B/6C/6D-F	AAT TTG TAT TTT ATT CAT GCC TAT ATC TGG	300	6A, 6B, 6C, 6D
cm6A/6B/6C/6D-R	TTA GCG GAG ATA ATT TAA AAT GAT GAC TA	300	
cm6C/6D-F	CAT TTT AGT GAA GTT GGC GGT GGA GTT	500	6C, 6D
cm6C/6D-R	AGC TTC GAA GCC CAT ACT CTT CAA TTA	500	
cm7C/7B-F	CTA TCT CAG TCA TCT ATT GTT AAA GTT TAC GAC GGG A	300	7B, 7C, 40
cm7C/7B-R	GAA CAT AGA TGT TGA GAC ATC TTT TGT AAT TTC	300	

Name	Sequence (5’ to 3’)	Concentration (nM)	Serotypes detected
cm7F/7A-F	CCT ACG GGA GGA TAT AAA ATT ATT GAG	400	7A, 7F
cm7F/7A-R	CAA ATA CAC CAC TAT AGG CTG TTG AGA CTA AC	400	
cm8-F	GAA GAA ACG AAA CTG TCA GAG CAT TTA CAT	200	8
cm8-R	CTA TAG ATA CTA GTA GAG CTG TTC TAG TCT	200	
cm9N/9L-F	GAA CTG AAT AAG TCA GAT TTA ATC AGC	500	9L, 9N
cm9N/9L-R	ACC AAG ATC TGA CGG GCT AAT CAA T	500	
cm9V/9A-F	GGG TTC AAA GTC AGA CAG TGA ATC TTA A	500	9A, 9V
cm9V/9A-R	CCA TGA ATG AAA TCA ACA TTG TCA GTA GC	500	
cm10A-F	GGT GTA GAT TTA CCA TTA GTG TCG GCA GAC	500	10A
cm10A-R	GAA TTT CTT CTT TAA GAT TCG GAT ATT TCT C	500	
cm10F/10C/33C-F	GGA GTT TAT CGG TAG TGC TCA TTT TAG CA	300	10C, 10F, 33C
cm10F/10C/33C-R	CTA ACA AAT TCG CAA CAC GAG GCA ACA	300	
cm11A/11D-F	GGA CAT GTT CAG GTG ATT TCC CAA TAT AGT G	300	11A, 11D
cm11A/11D-R	GAT TAT GAG TGT AAT TTA TTC CAA CTT CTC CC	300	
cm12F/12A/44/46-F	GCA ACA AAC GGC GTG AAA GTA GTT G	500	12A, 12B, 12F, 44, 46
cm12F/12A/44/46-R	CAA GAT GAA TAT CAC TAC CAA TAA CAA AAC	500	
cm13-F	TAC TAA GGT AAT CTC TGG AAA TCG AAA GG	400	13
cm13-R	CTC ATG CAT TTT ATT AAC CGC TTT TTG TTC	400	
cm14-F	GAA ATG TTA CTT GGC GCA GGT GTC AGA ATT	300	14
cm14-R	GCC AAT ACT TCT TAG TCT CTC AGA TGA AT	300	

Name	Sequence (5’ to 3’)	Concentration (nM)	Serotypes detected
cm15A/15F-F	ATT AGT ACA GCT GCT GGA ATA TCT CTT C	300	15A, 15F
cm15A/15F-R	GAT CTA GTG AAC GTA CTA TTC CAA AC	300	
cm15B/15C-F	TTG GAA TTT TTT AAT TAG TGG CTT ACC TA	300	15B, 15C
cm15B/15C-R	CAT CCG CTT ATT AAT TGA AGT AAT CTG AAC C	300	
cm16F-F	CTG TTC AGA TAG GCC ATT TAC AGC TTT AAA TC	400	16F
cm16F-R	CAT TCC TTT TGT ATA TAG TGC TAG TTC ATC C	400	
cm17F-F	TTC GTG ATG ATA ATT CCA ATG ATC AAA CAA GAG	500	17F
cm17F-R	GAT GTA ACA AAT TTG TAG CGA CTA AGG TCT GC	500	
cm18C/18F/18B/18A-F	CTT AAT AGC TCT CAT TAT TCT TTT TTT AAG CC	300	18A, 18B, 18C, 18F,
cm18C/18F/18B/18A-R	TTA TCT GTA AAC CAT ATC AGC ATC TGA AAC	300	
cm19A-F	GAG AGA TTC ATA ATC TTG CAC TTA GCC A	300	19A
cm19A-R	CAT AAT AGC TAC AAA TGA CTC ATC GCC	300	
cm19F-F	GTT AAG ATT GCT GAT CGA TTA ATT GAT ATC C	500	19F
cm19F-R	GTA ATA TGT CTT TAG GGC GTT TAT GGC GAT AG	500	
cm20-F	GAG CAA GAG TTT TTC ACC TGA CAG CGA GAA G	300	20
cm20-R	CTA AAT TCC TGT AAT TTA GCT AAA ACT CTT ATC	300	
cm21-F	CTA TGG TTA TTT CAA CTC AAT CGT CAC C	200	21
cm21-R	GGC AAA CTC AGA CAT AGT ATA GCA TAG	200	
cm22F/22A-F	GAG TAT AGC CAG ATT ATG GCA GTT TTA TTG TC	500	22A , 22F
cm22F/22A-R	CTC CAG CAC TTG CGC TGG AAA CAA CAG ACA AC	500	

Name	Sequence (5’ to 3’)	Concentration (nM)	Serotypes detected
cm23A-F	TAT TCT AGC AAG TGA CGA AGA TGC G	500	23A
cm23A-R	CCA ACA TGC TTA AAA ACG CTG CTT TAC	500	
cm23B-F	CCA CAA TTA GCG CTA TAT TCA TTC AAT CG	200	23B
cm23B-R	GTC CAC GCT GAA TAA AAT GAA GCT CCG	200	
cm23F-F	GTA ACA GTT GCT GTA GAG GGA ATT GGC TTT TC	500	23F
cm23F-R	CAC AAC ACC TAA CAC TCG ATG GCT ATA TGA TTC	500	
cm24F/24A/24B-F	GCT CCC TGC TAT TGT AAT CTT TAA AGA G	200	24A, 24B, 24F
cm24F/24A/24B-R	GTG TCT TTT ATT GAC TTT ATC ATA GGT CGG	200	
cm31-F	GGA AGT TTT CAA GGA TAT GAT AGT GGT GGT GC	500	31
cm31-R	CCG AAT AAT ATA TTC AAT ATA TTC CTA CTC	500	
cm33F/33A/37-F	GAA GGC AAT CAA TGT GAT TGT GTC GCG	300	33A, 33F, 37
cm33F/33A/37-R	CTT CAA AAT GAA GAT TAT AGT ACC CTT CTA C	300	
cm34-F	GCT TTT GTA AGA GGA GAT TAT TTT CAC CCA AC	300	34
cm34-R	CAA TCC GAC TAA GTC TTC AGT AAA AAA CTT TAC	300	
cm35A/35C/42-F	ATT ACG ACT CCT TAT GTG ACG CGC ATA	300	35A, 33C, 42
cm35A/35C/42-R	CCA ATC CCA AGA TAT ATG CAA CTA GGT T	300	
cm35B-F	GAT AAG TCT GTT GTG GAG ACT TAA AAA GAA TG	500	35B
cm35B-R	CTT TCC AGA TAA TTA CAG GTA TTC CTG AAG CAA G	500	
cm35F/47F-F	GAA CAT AGT CGC TAT TGT ATT TTA TTT AAA GCA A	300	35F, 47F
cm35F/47F-R	GAC TAG GAG CAT TAT TCC TAG AGC GAG TAA ACC	300	

Name	Sequence (5’ to 3’)	Concentration (nM)	Serotypes detected
cm38/25F/25A-F	CGT TCT TTT ATC TCA CTG TAT AGT ATC TTT ATG	300	25A, 25F, 38
cm38/25F/25A-R	ATG TTT GAA TTA AAG CTA ACG TAA CAA TCC	300	
cm39-F	TCA TTG TAT TAA CCC TAT GCT TTA TTG GTG	200	39
cm39-R	GAG TAT CTC CAT TGT ATT GAA ATC TAC CAA	200	

Table 6. Serotypes detected by the cmPCR reactions and expected band sizes.

Reaction	Serotype (Expected size in bp)					
A	*cpsA* (160)	14 (189)	19F (304)	3 (371)	19A (566)	7F/A (826)
B	*cpsA* (160)	6A/B/C/D (250)	23F (384)	4 (430)	18C/F/B/A (573)	9V/A (816)
C	*cpsA* (160)	1 (280)	5 (362)	15B/C (496)	22F/A (643)	6C/D (727)
D	*cpsA* (160)	8 (201)	33F/A/37 (338)	11A/D (463)	9N/L (516)	10A (628)
E	*cpsA* (160)	23B (199)	2 (290)	12F/A/B/44/46 (376)	20 (514)	17F (693)
F	*cpsA* (160)	7C/B/40 (260)	15A/F (434)	13 (655)	23A (722)	16F (988)
G	24A/B/F (99)	*cpsA* (160)	10F/C/33C (248)	34 (408)	38/25F/A (574)	31 (701)
H	39 (98)	*cpsA* (160)	21 (192)	35A/C/42 (280)	35F/47F (517)	35B (677)

Table 7. Oligonucleotides used for sequence-based serotyping of *S. pneumoniae.*

PCR group	Name	Target	Sequence (5' to 3')	Expected size (bp)
6AB	F6ABWciN	*wciN α*	GAATACAAGCAGGTGTTGGGT	375
	R6ABWciN		ACTGCCTTCTTGCCAAGGTC	
	6CD-f	*wciN β*	CATTTTAGTGAAGTTGGCGGTGGAGTT	727
	6CD-r		AGCTTCGAAGCCCATACTCTTCAATTA	
	Fsp6WciP	*wciP*	CAGAGGTATTCTAAGATAGATG	609
	Rsp6WciP		AGCATGTGTAAGTGCCAATTC	
7FA	F7IC	*wzy*	TCCAAACTATTACAGTGGGAATTACGG	599
	R7IC		ATAGGAATTGAGATTGCCAAAGCGAC	
	Fsp7WcwD	*wcwD*	AGACCTGCTTATGGGCTTATCT	486
	Rsp7WcwD		ATGTGCGTATAGAGGACGCTT	
9AV	F9AVIC	*wzy*	GAACTGAATAAGTCAGATTTAATCAGC	816
	R9AVIC		ACCAAGATCTGACGGGCTAATCAAT	
	Fsp9WcjE	*wcjE*	TGTTGTTTTCAAAAATCGTAAGTCC	511
	Rsp9WcjE		ATACCAGCCTAGCACAATCCC	
9NL	Fsp9WcjA	*wcjA*	TTAGATGTTGACAGTAAGAATCC	886
	Rsp9WcjA		CTCCCATTAATTTTCTATAAAGTC	
	F9NwcjB	*wcjB*	TTCTAGCTGATGATGTATGGTTG	329
	Rsp9WcjB		CTACAGGTGAGAGAATCCGGC	
	Fsp9wchA	*wchA*	AGCGGAACGGTTTAGTATCTCT	884
	R9NWchA		CTTGCATCGTATTTTGATCCATCAAC	
	Fsp9Wzy	*wzy*	TTTTAACACAATCCAGTACAGCG	550
	Rsp9Wzy		GAGGCTAACATAGCCAACAG	
11AD	F11WcwC	*wcrL*	GGGGAGTATCCTATAGGCAT	911
	Rsp11Wcrl		GCCATCACCCAAAGAAATCGG	
	F11WcwR	*wcwR*	AAGACATGTAGGACGGTCAAC	243
	R11WcwR		CCCACGTCCCTTTCCATCC	
12FAB4446	Fsp12WcxD	*wcxD*	AGCTAGCCTGTCGTTTTGCT	632
	Rsp12WcxD		TGTTGACAATTCTTCTGCTGGA	
	Fsp12WcxF	*wcxF*	CCATAAGGAGGCGCGTGA	547
	RI12WcxF		CAAGAAGGCTTCAAGCAGTTC	
	FI12WcxF	*wcxF*	GTGTTCCGATACAGACTGTC	795
	Rsp12WcxF		AGTCTCTCCCAAAACTCGGATA	
	Fsp12Wzy	*wzy*	TGGCTATTTTGGCTTGGACT	507
	Rsp12Wzy		ACGTCAGTTGCTCCTCCAAT	
15BC	15B-f	*wzy*	TTGGAATTTTTTAATTAGTGGCTTACCTA	496
	15B-r		CATCCGCTTATTAATTGAAGTAATCTGAACC	
	Fsp15WciZ	*wciZ*	GCACGATTTGCAGTTCCAATG	606
	Rsp15WciZ		TGGTTGCTAGCACAGAACTT	
18FBCA	Fsp18WciX	*wzy/wciZ*	TTCAGATGCTGATATGGTTTACAGA	678
	Rsp18WciX		AGTGGTGTTGCTAACGAGCC	
	Rev18A	*wzy/wciY*	GGCGTAAGACCATAAATTCAAG	155
	Fsp18WcxM	*wcxM*	GGACTCTTTCGTGGGGTTGT	345
	Fsp18WcxM		ACGAACATTCTGACCTGCGA	

PCR group	Name	Target	Sequence (5' to 3')	Expected size (bp)
22FA	22F-f	*wcwV*	GAGTATAGCCAGATTATGGCAGTTTTATTGTC	643
	22F-r		CTCCAGCACTTGCGCTGGAAACAACAGACAAC	
	F22FWcwA	*wcwA* (22F)	TTCCTTCCACCTACTCGGCA	250
	R22FWcwA		TCGTGTTAGTTCCGGTGAGA	
	F22FWcwC	*wcwC* (22F)	GGATGGGATCCAAATTGCGG	148
	R22FWcwC		CCCACCTGCAGAACCGAATA	
	F22AWcwC	*wcwA/wcwC* (22A)	AACTTCCCTTGATCTGGCACG	399
	R22AWcwC		CAACTCGCCACTCTTCACCA	
33FA37	33F-f	*wzy*	GAAGGCAATCAATGTGATTGTGTCGCG	338
	33F-r		CTTCAAAATGAAGATTATAGTACCCTTCTAC	
	F37tts	*tts*	CGCATTGGCCAGTGTATCAG	185
	R37tts		CGTCAGCAAATCCTACATGGT	
	Fsp33WcjE	*wcjE*	TCGTTGTTGATAGCACAAGTCT	593
	Rsp33WcjE		CCCACACCCCTGGGTTACTA	

3.8. Agarose Gel Electrophoresis

Amplicons produced during PCR were resolved by 1% (w/v) agarose gel electrophoresis in 1× sodium borate (SB) buffer [16 g/l sodium hydroxide and 94 g/l boric acid (Sigma Aldrich, Mississauga, ON)], and 10 µg/ml ethidium bromide (Invitrogen, Waltham, MA). Low-melt electrophoresis grade agarose (BioShop Canada Inc., Burlington, ON) was used to facilitate gel extraction, if needed. A 100 base pair (bp) DNA Ladder (Invitrogen, Waltham, MA) was used every 15-20 wells to estimate amplicon sizes (Thermo Fisher Scientific, Waltham, MA). Agarose in 1× SB buffer were subjected to a constant voltage of 250 V. After the desired resolution, a Gel Doc XR$^+$ Imaging System (BioRad Laboratories Inc., Hercules, CA) with Image Lab software, version 5, was used for DNA visualization. Expected bp sizes for each PCR reaction are denoted in Tables 2, 6, and 7. For DNA sequencing, the bands of interest were excised from the agarose gels, and stored in sterile DNase/RNase-free microtubes at 4°C until purification.

3.9. Purification of Amplicon DNA

Amplicon DNA was purified from agarose gels using a QIAquick Gel Extraction Kit (Qiagen Inc, Hilden, Germany). Extractions performed according to manufacturer instructions with the following exceptions: the additional optional wash steps with 500 μL buffer PE were included, and DNA was eluted with 50 μL of PCR-grade water (Invitrogen) instead of the elution buffer (EB). All centrifugations were performed in a Legend Micro 21 microcentrifuge (ThermoFisher Scientific, Waltham, MA).

3.10. DNA Sequencing and Analysis

All sequencing was carried out at McLab Molecular Cloning Laboratories (www.mclab.com) (San Francisco, CA) using proprietary DNA sequencing technologies and reagents on a Life technologies ABI 3730XL sequencer (ThermoFisher Scientic). Sequence chromatograms were visualized and converted to FASTA format using Chromas Lite version 2.1 (http://chromas-lite.software.informer.com/2.1/). Overlapping sequences from the forward and reverse strands were obtained with BLASTn (https://blast.ncbi.nlm.nih.gov). The resulting sequences were downloaded as FASTA files, and compared to references sequences extracted from the Genbank databases of the National Center for Bioinformatical Information (NCBI) (https://www.ncbi.nlm.nih.gov/) using accession numbers defined by Bentley *et al*. (2006) or serotype specific sequences (Kapatai *et al*., 2016). Clustal Omega was used for a multiple sequence alignment (http://www.ebi.ac.uk/Tools/msa/clustalo/). Aligned segments of sequence were searched for serotype-specific signature mutations in the PneumoCaT database (https://github.com/phe-bioinformatics/PneumoCaT) (Kapatai *et al*, 2016).

Chapter 4. Results

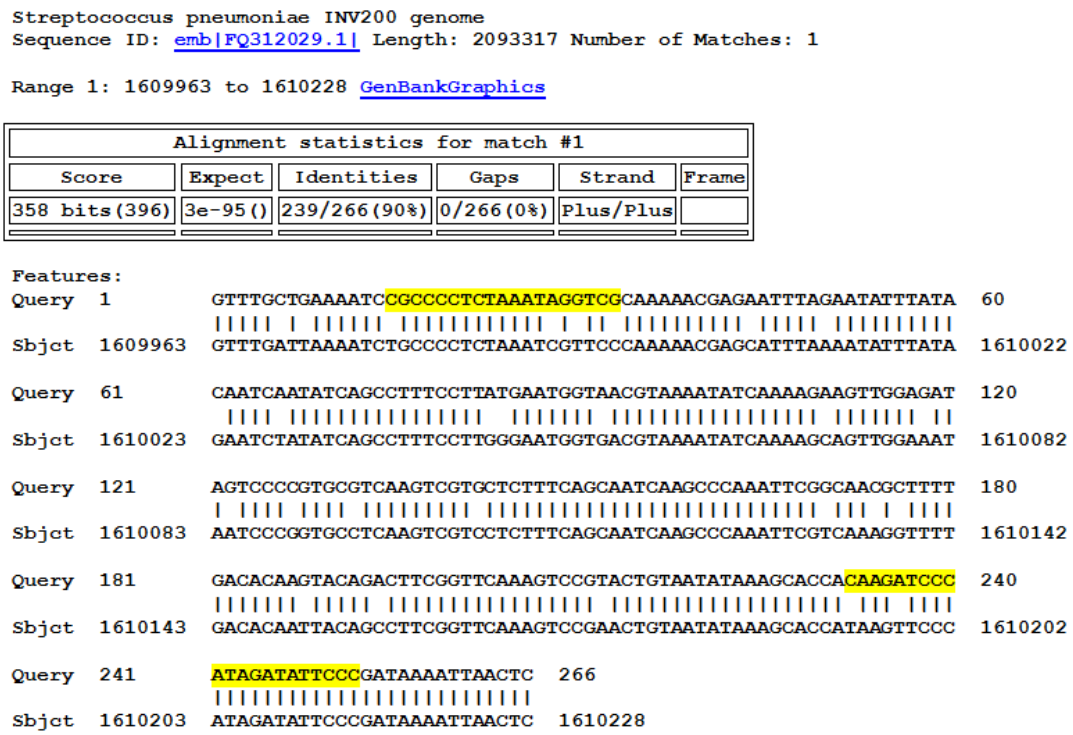
4.1. Identification of Serotype-Specific PCR Targets

Following NGS, high quality data was provided for comparative genomics. DNA sequences reads were compared to the reference genome of *S. pneumoniae* serotype 19A, and rigorous bioinformatics analyses revealed a number of unique unmatched sequence reads for each serotype (Table 8). Following BLASTn analyses and exclusion of sequences that showed high similarity within PCR groups, unique DNA sequences for each *S. pneumoniae* serotype were identified (Table 8). While a large number of potential targets were obtained (Table 8), only four targets for each serotype were initially chosen for subsequent analysis due to time and budgetary restrictions (Table 2). To help select the best candidates PCR targets, a classification system was used to avoid any MGEs or cross-reactions with *S. pneumoniae* serotypes within the PCR groups, other *S. pneumoniae* serotypes, other streptococci, or other organisms. Following target select and classification, primers were designed and optimized to increase serotype specificity. For example, target 3 of serotype 6B was designed to target a region of sequence mismatch with *S. pneumonaie* serotype 14 (Figure 9), which would otherwise have cross-reacted during PCR. The classification and optimization of each of the four selected PCR targets reduced the number of class B and C targets, and enriched target classification as class A (Figure 10A). For example, the overall proportions for target 1 (all serotypes combined) consisted of 32% class A targets, 25% class B, and 43% class C; however, following optimization, the proportions of class A targets increased to 96% and only a small amount (4%) of class C remained (Figure 10B). A similar approach was taken for the other PCR targets, and similar but less impressive trends were noted (Figure 10A and B). While bioinformatic approaches to identify serotype-specific PCR targets were successful, it is important to validate each selected target through experimental testing.

Table 8. Summary of NGS, comparative genomics data, and target classifications.

PCR Groups	Serotype	NML Ref. No.	Mapped reads (%)[1]	Unmapped reads eliminated (%)[2]	Assembled contigs[3]	Unique contigs[4]	Mean depth coverage	Classification				
								A	B	C	D	MGE
6A/B/C/D	6A*	SC0022	44.6% (716121/1606402)	48.1% (427821/890281)	93	77	38	3	0	23	33	18
	6B*	SC0023	43.4% (660499/1521373)	60.5% (520721/860874)	62	54	27	0	5	8	29	12
	6C	SC0262	44.9% (651957/1453636)	59.5% (477008/801679)	143	112	58	1	13	15	62	21
	6D	SC0129	55.8% (646477/1157602)	90.2% (461245/511125)	41	32	61	1	0	3	17	11
7F/A	7F*	SC0218	51.2% (557305/1106960)	67.7% (372128/549655)	68	59	49	1	4	18	16	20
	7A	SC0025	50.4% (640291/1266104)	69.0% (431542/625813)	31	26	48	1	5	2	4	13
9N/L	9N*	SC0031	39.2% (547405/1396440)	54.4% (461715/849035)	54	44	34	0	0	5	31	8
	9L	SC0011	34.0% (509177/1498713)	54.8% (542659/989536)	61	46	27	0	0	12	27	7
9V/A	9V*	SC0172	44.2% (687663/1554182)	54.1% (468457/866519)	43	27	48	1	5	0	11	10
	9A	SC0029	33.8% (476535/1409622)	53.7% (501404/933087)	68	55	69	0	1	25	18	11
11A/D	11A*	SC0035	35.6% (474357/1333743)	50.6% (434973/859386)	79	61	31	0	0	22	17	22
	11D	SC0271	48.5% (596899/1229783)	76.2% (482319/632884)	78	57	72	0	0	8	27	22
12F/A/B/44/46	12F*	SC0199	48.0% (614675/1279247)	84.2 (559343/664572)	96	49	92	1	1	9	30	8
	12A	SC0066	41.8% (530911/1270571)	80.7% (596724/739660)	65	44	53	5	1	24	10	4
	12B	SC0268	31.3% (458545/1464218)	41.2% (414103/1005673)	141	129	66	3	3	28	89	6
	44	SC0212	38.4% (427850/1113277)	81.9% (561524/685427)	77	57	53	3	5	7	35	7
	46	SC0096	44.3% (558918/1261758)	68.7% (482992/702840)	41	26	37	0	0	3	15	8
15B/C	15B*	SC0044	51.2% (679523/1326977)	64.7% (418714/647454)	59	44	43	1	0	14	15	14
	15C	SC0045	29.5% (461286/1563810)	51.3% (565680/1102524)	55	44	62	0	2	10	24	8
18C/F/B/A	18C*	SC0050	41.9% (523716/1249950)	57.1% (414508/726234)	38	25	51	0	0	10	8	7
	18F	SC0051	45.4% (566801/1249496)	63.5% (433611/682695)	54	45	81	0	0	17	14	14
	18B	SC0049	29.0% (440355/1519211)	51.3% (553680/1078856)	103	82	47	0	15	26	23	18
	18A	SC0009	31.1% (489831/1577129)	38.4% (418025/1087298)	21	18	62	0	0	7	9	2
22F/A	22F*	SC0059	51.4% (571824/1113298)	69.1% (373862/541474)	45	29	41	0	0	14	7	8
	22A	SC0291	47.7% (671211/1406292)	71.4% (525104/735081)	27	16	38	0	0	3	7	6
33F/A/37	33F*	SC0190	65.4% (719571/1100975)	76.6% (292313/381404)	62	56	31	0	1	5	18	32
	33A	SC0082	47.3% (672448/1420770)	79.7% (596332/748322)	78	57	36	0	0	14	26	17
	37	SC0086	31.1% (482400/1549524)	77.6% (828158/1067124)	86	65	28	3	0	11	37	14

[1]Mapped to the reference genome; [2]Unmapped reads were eliminated if a "mate read" in a serotype within the same PCR group mapped the reference genome; [3]Contigs assembled by both Velvet and SOAPdenovo; [4]Unique contigs represent the ones not found in other serotypes within the PCR groups; The asterisks (*) indicates vaccine-preventable serotypes.



Streptococcus pneumoniae INV200 genome
Sequence ID: emb|FQ312029.1| Length: 2093317 Number of Matches: 1

Range 1: 1609963 to 1610228 GenBankGraphics

Alignment statistics for match #1					
Score	Expect	Identities	Gaps	Strand	Frame
358 bits(396)	3e-95()	239/266(90%)	0/266(0%)	Plus/Plus	

Features:

```
Query  1        GTTTGCTGAAAATCCGCCCCTCTAAATAGGTCGCAAAAACGAGAATTTAGAATATTTATA  60
               ||||| | |||||| |||||||||||| | || |||||||||| ||||| ||||||||||
Sbjct  1609963  GTTTGATTAAAATCTGCCCCTCTAAATCGTTCCCAAAAACGAGCATTTAAAATATTTATA  1610022

Query  61       CAATCAATATCAGCCTTTCCTTATGAATGGTAACGTAAAATATCAAAAGAAGTTGGAGAT  120
                ||||  ||||||||||||||||  ||||||| |||||||||||||||| ||||||| ||
Sbjct  1610023  GAATCTATATCAGCCTTTCCTTGGGAATGGTGACGTAAAATATCAAAAGCAGTTGGAAAT  1610082

Query  121      AGTCCCCGTGCGTCAAGTCGTGCTCTTTCAGCAATCAAGCCCAAATTCGGCAACGCTTTT  180
               | |||| |||| ||||||||| |||||||||||||||||||||||||| ||| | ||||
Sbjct  1610083  AATCCCGGTGCCTCAAGTCGTCCTCTTTCAGCAATCAAGCCCAAATTCGTCAAAGGTTTT  1610142

Query  181      GACACAAGTACAGACTTCGGTTCAAAGTCCGTACTGTAATATAAAGCACCACAAGATCCC  240
               ||||||| ||||| |||||||||||||||| |||||||||||||||||| ||| ||||
Sbjct  1610143  GACACAATTACAGCCTTCGGTTCAAAGTCCGAACTGTAATATAAAGCACCATAAGTTCCC  1610202

Query  241      ATAGATATTCCCGATAAAATTAACTC  266
                |||||||||||||||||||||||||
Sbjct  1610203  ATAGATATTCCCGATAAAATTAACTC  1610228
```

Figure 9. Example of primer optimization for *S. pneumoniae* serotype 6B. While performing a BLASTn analysis on the unique contig generated for PCR target 3 of serotype 6B (query sequence), a similarity hit was observed to the *S. pneumoniae* INV200 genome, serotype 14 (subject sequence). To avoid cross-reactions with serotype 14, the primers chosen for serotype 6B target 3 (highlighted in yellow) spanned regions of sequence mismatch with the INV200 genome.

A)

Classification of PCR targets prior to optimization

Target	Serotype																									
	6A	6B	6C	6D	7F	7A	9V	9A	9N	9L	11A	11D	12F	12A	12B	44	46	15B	15C	18C	18F	18B	18A	22F	22A	33F	33A	37
1	A	B	B	C	A	A	A	B	C	C	C	C	B	A	A	A	C	A	B	C	C	B	C	C	C	B	C	A
2	A	B	B	C	B	B	B	C	C	C	C	C	B	A	A	A	C	C	C	C	C	B	C	C	C	C	C	A
3	A	C	B	C	B	C	B	C	C	C	C	C	C	A	A	B	C	C	C	C	C	B	C	C	C	C	C	C
4	C	C	B	C	C	C	B	C	C	C	C	C	C	C	B	B	C	C	C	C	C	B	C	C	C	C	C	C

Classification of PCR targets following optimization

Target	Serotype																									
	6A	6B	6C	6D	7F	7A	9V	9A	9N	9L	11A	11D	12F	12A	12B	44	46	15B	15C	18C	18F	18B	18A	22F	22A	33F	33A	37
1	A	A	A	A	A	A	A	A	A	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
2	A	A	A	A	A	B	A	A	A	C	A	A	A	A	A	A	C	A	A	C	A	A	A	C	C	A	A	A
3	A	A	A	A	B	C	A	A	A	C	A	A	A	A	A	A	C	A	A	C	A	B	A	C	C	A	A	A
4	C	A	A	A	C	C	A	A	A	C	A	A	C	A	B	A	C	A	B	C	A	B	A	C	C	B	A	A

B)

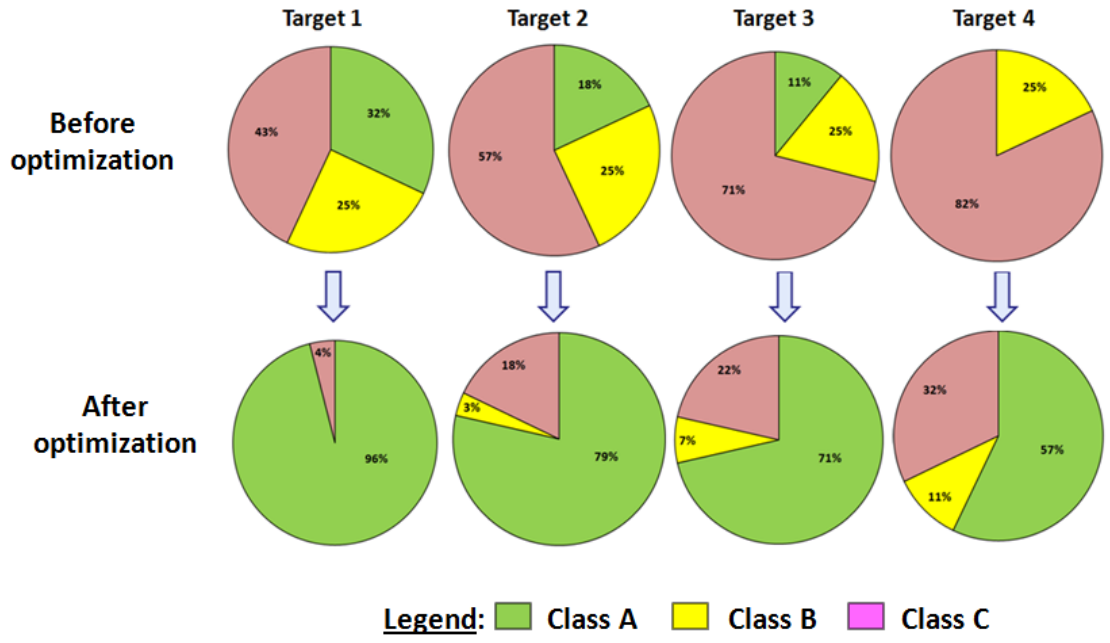


Figure 10. Benefits of primer optimization. A) For each *S. pneumoniae* serotype, four unique DNA sequences (contigs) as potential serotype-specific PCR targets. Each target was classified based on potential cross-reactions with other *S. pneumoniae* serotypes, other streptococci, or other organisms. B) Overall class change for targets 1 to 4 after optimization.

4.2. Evaluation of Serotype-Specific PCR Targets

For some PCR targets identified by NGS and comparative genomics (Table 2), an amplicon was observed with the expected size (Table 9). For example, PCR group 12F/B/A/44/46 could be discriminated using serotype-specific primers used as a multiplex (Figure 11). However, for all serotypes, target amplification was either inconsistent, insensitive or cross-reaction occurred with *S. pneumoniae* serotypes within PCR groups, between other *S. pneumoniae* serotypes, or with other streptococci (Table 9). Even targets that were detected with no cross-reaction failed to be reproducible when tested with different strains of the same serotypes (Table 9). It should be noted that in many cases, no target was amplified despite high concentrations used for testing (with Ct values in *lytA* real-time PCR ranging between 20 and 25). When the DNA was used in traditional cmPCR reactions, the desired amplicon were observed, suggesting the DNA quality and concentration was adequate. Overall, while the proof-of-principle that serotype-specific targets outside the *cps* loci could be generated, the lack of reproducibility, or problems with cross-reactivity precluded their used for serotype discrimination.

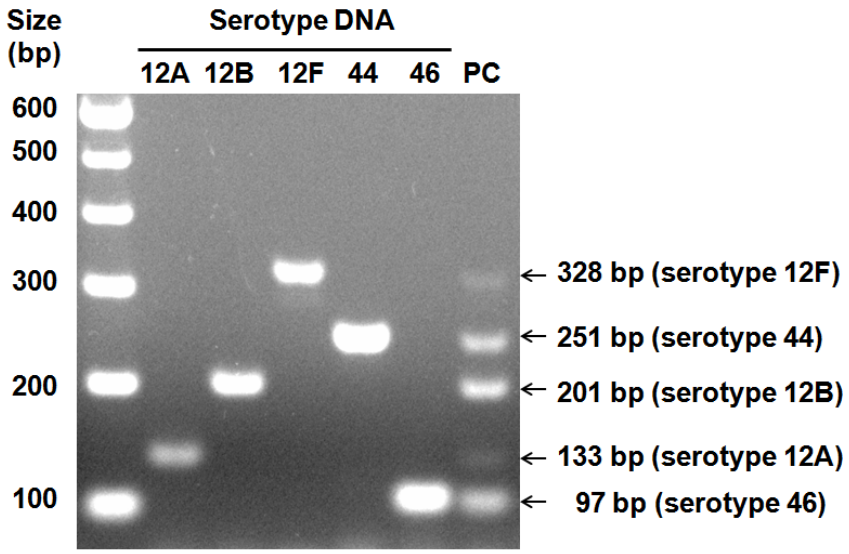


Figure 11. Multiplex PCR with novel targets for discrimination of *S. pneumoniae* serotypes in the PCR group 12F/B/A/44/46. The five serotype-specific targets were tested using DNA extracted from representative *S. pneumoniae* serotypes 12A, 12B, 12F, 44, and 46. The expected band sizes for each serotype-specific PCR target are noted. PC signifies positive control.

Table 9. Summary of the performance characteristics for the novel serotype-specific PCR targets.

PCR group	Serotype	Target	Test criteria met								Comments (exclusion criteria)
			10^9 copies/ml	10^6 copies/ml	$MgCl_2$	Anneal 50°C	PCR group	Other serotypes	Other strep.	Reproducibility	
6A/B/C/D	6A	6A1	1/1	2/2	N/A	N/A	Y	N	Y	5/8	Cross-reactions (44) and not reproducible
		6A2	1/1	2/2	N/A	N/A	Y	Y	Y	2/8	Not reproducible
		6A3	1/1	2/2	N/A	N/A	N	N	N	N/A	Cross-reactions (6B, 15B, 17, 24A, 47F, and *S. equi*)
		6A4	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
	6B	6B1	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		6B2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		6B3	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		6B4	1/1	3/3	N/A	N/A	N	N	N	N/A	Cross-reactions (6A, 18C, 22F, 32A, 41A, GBS)
	6C	6C1	1/1	2/2	N/A	N/A	Y	Y	Y	7/20	Not reproducible
		6C2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		6C3	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		6C4	1/1	3/3	N/A	N/A	Y	N	Y	5/20	Cross-reactions (9V, 46) and not reproducible
	6D	6D1	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		6D1b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		6D2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		6D3	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		6D3b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		6D4	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		6D4b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		6D5	2/2	2/2	N/A	N/A	N	N	Y	N/A	Cross-reactions (6B, 7F, 44, 18A, 18B, 18F, 33A, 33F)
7F/A	7F	7F1	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		7F1b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		7F1c	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		7F2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		7F2b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		7F3	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		7F4	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		7F5	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
	7A	7A1	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		7A1b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		7A2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		7A3	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		7A4	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		7A4b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		7A4c	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		7A5	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		7A6a	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
9V/A	9V	9V1	2/2	2/2	N/A	N/A	Y	N	Y	2/10	Cross-reactions (6C, 23F, 35A, 46) and not reproducible
		9V1b	2/2	0/2	N/A	N/A	N	N	N	4/10	Insensitive, and not reproducible
		9V2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		9V3	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		9V4	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
	9A	9A1	1/1	3/3	N/A	N/A	N	Y	N	N/A	Cross-reactions (9V and Group B strep) and not reproducible
		9A2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		9A3	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		9A4	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
9N/L	9N	9N1	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		9N1b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		9N2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		9N3	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		9N3b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		9N4	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
	9L	9L1	2/2	0/2	N/A	N/A	N	Y	N	N/A	Insensitive and cross-reaction (19C, 22F, 41A)
		9L2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		9L3	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		9L4	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain

PCR group	Serotype	Target	Test criteria met								Comments (exclusion criteria)
			10^9 copies/ml	10^6 copies/ml	$MgCl_2$	Anneal 50°C	PCR group	Other serotypes	Other strep.	Reproducibility	
11A/D	11A*	11A1	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		11A1b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		11A2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		11A2b	1/1	0/2	N/A	N/A	N/A	N/A	N/A	N/A	Insensitive
		11A3	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		11A3b	0/2	0/3	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		11A4	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		11A5	2/2	0/2	N/A	N/A	Y	N	Y	N/A	Insensitive and cross-reactions (7F)
	11D	11D1	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		11D2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		11D3	1/1	1/2	N/A	N/A	Y	N	Y	N/A	Cross-reactions (7A, 9A, 9V, 9L, 12B, 33A, and 33F)
		11D4	0/2	0/3	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
12F/A/B/44/46	12F*	12F1	1	1/1	N/A	N/A	N	N/A	N/A	N/A	Cross-reaction within the PCR group (12B and 44)
		12F2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		12F3	1	1/1	N/A	N/A	Y	N	Y	2/10	Cross-reactions (9L and 28F)
		12F4	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
	12A	12A1	0/2	2/2	N/A	N/A	Y	N	Y	N/A	Cross-reaction (9L)
		12A2	0/2	2/2	N/A	N/A	Y	Y	Y	1/8	Not reproducible
		12A3	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		12A4	1	3/3	N/A	N/A	Y	N	Y	N/A	Cross-reaction (9N)
	12B	12B1	1	3/3	N/A	N/A	Y	N	Y	N/A	Cross-reactionS (15B and 15C)
		12B2	1	4/4	N/A	N/A	Y	Y	Y	2/6	Not reproducible
		12B3	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		12B4	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
	44	44-1	1	2/2	N/A	N/A	Y	Y	Y	1/3	Not reproducible
		44-2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		44-3	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		44-4	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
	46	46-1	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		46-2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		46-3	2/2	0/3	N/A	N/A	Y	N	Y	1/2	Insensitive and cross-reactions (6C, 7F, 28A, 28F, 36)
		46-4	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
15B/C	15B*	15B1	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		15B2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		15B2b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		15B3	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		15B3b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		15B4	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		15B4b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		15B5	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
	15C	15C1	1/1	2/2	N/A	N/A	N	N/A	N/A	N/A	Cross-reaction within PCR group (15B)
		15C2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		15C2b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		15C3	1/2	1/2	0/2	0/2	N/A	N/A	N/A	N/A	Inconsistent results
		15C3	1/1	2/2	N/A	N/A	N	N/A	N/A	N/A	Cross-reaction within PCR group (15B)
		15C4	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		15C5	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		15C8	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain

PCR group	Serotype	Target	Test criteria met								Comments (exclusion criteria)
			10^9 copies/ml	10^6 copies/ml	$MgCl_2$	Anneal 50°C	PCR group	Other serotypes	Other strep.	Reproducibility	
18C/F/B/A	18C*	18C1	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		18C2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		18C3	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		18C4	1/1	2/2	N/A	N/A	Y	N	Y	N/A	Cross-reactions (9L, 12A, 44, 19C, and 35C)
	18F	18F1	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		18F2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		18F3	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		18F4	1/1	2/2	N/A	N/A	Y	N	Y	N/A	Cross-reactions (9L and 11A)
	18B	18B1	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		18B1b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		18C2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		18C3	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		18C4	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		18C5	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		18C6	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		18C7	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
	18A	18A1	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		18A2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		18A2b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		18A2c	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		18A3	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		18A3b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		18A4	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		18A5a	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
22F/A	22F	22F1	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		22F1b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		22F2a	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		22F2b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		22F3a	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		22F3b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		22F4	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		22F5	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
	22A	22A1	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		22A1b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		22A1c	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		22A2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		22A3	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		22A4	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		22A4b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
33F/A/37	33F*	33F1	1/1	3/3	N/A	N/A	Y	Y	N	2/10	Cross-reaction (*Granulicatella adiacens*)
		33F2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		33F3	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		33F4	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
	33A	33A1	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		33A1b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		33A1c	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		33A2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		33A3	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		33A4	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		33A5	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		33A6	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
	37	37-1	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		37-1b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		37-2	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		37-3	1/1	0/2	N/A	N/A	Y	Y	Y	N/A	Insensitive
		37-3b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		37-4	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		37-4b	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		37-5	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain
		37-6	0/2	0/2	0/2	0/2	N/A	N/A	N/A	N/A	Not detectable in the reference strain

4.3. Sequence-Based Discrimination of *S. pneumoniae* Serotypes

Following the release of the NGS pipeline for *S. pneumoniae* serotyping (i.e. PneumoCaT), serotype-specific sequence signatures within the *cps* loci became publically available for all serotypes (Kapatai *et al*., 2016). In the second part of this study, serotype-specific mutations within the *cps* loci were evaluated by PCR and Sanger sequencing to ensure correlation of *S. pneumoniae* strains previously characterized by Quellung serotyping. For each of the 28 *S. pneumoniae* serotypes under evaluation (serotypes within the cmPCR groups containing a vaccine-preventable serotype), PCR targets were designed to amplify sequences flanking serotype-specific mutations. In some cases, a PCR-based approach was sufficient for serotype discrimination. The approach and results for each serotype within the select PCR groups are described in the following sub-sections, and use primers described in Table 7.

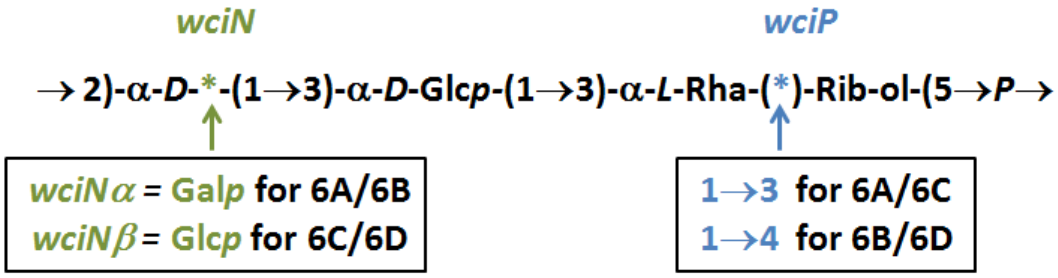
4.3.1. PCR and Sequencing-Based Discrimination of Serotypes 6A, 6B, 6C, and 6D

After the initial detection of serogroup 6A/B/C/D using the modified CDC cmPCR methods (Lang *et al.,* 2017), discrimination of each serotypes was pursued using a two-tiered approach focussing on two genes, *wciN* and *wciP* (Figure 12A and B) (Kapatai *et al*., 2016). These genes encode for enzymes that generate different sugar composition or linkages within the repeating capsular polysaccharide unit (Figure 12A) (Bentley *et al*., 2006; Geno *et al*., 2016). The *wciN* gene is found in two forms: *wciN*α that encodes a galactosyl-transferase in serotypes 6A and 6B, or *wciN*β that encodes a glycosyl-transferase in serotypes 6C and 6D (Figure 12A) (Park *et al*., 2007). In this study, primers (Table 7) were designed to differentiate serotypes 6A/B from 6C/D, using a 375 bp amplicon targeting *wciN*α or a 727 bp amplicon targeting *wciN*β, respectively (Figure 12A). For differentiation of serotypes 6A from 6B or serotypes 6C from 6D, sequencing of *wciP* is required. WciP is a rhamnosyl transferase responsible for the rhamnose-ribitol linkage in the repeating capsular saccharide unites of serogroup (Aanensen *et al*., 2007; Mavroidi *et al*., 2004). A guanine (G) to adenine (A) substitution at nucleotide position 583 results in a serine (S or Ser) to asparagine (N or Asn) at amino acid position 195, and

as such modifies the activity of WciP from generating 1→ 3 linkages between rhamnose and ribitol in serotypes in serotypes 6A and 6C to 1→4 linkages in serotypes 6B and 6D (Figure 12B) (Baek *et al*., 2014; Mavroidi *et al*., 2007; Sheppard *et al*., 2010). In this study, primers targeting a 609 bp fragment of *wciP* are used for PCR amplification and sequencing (Figure 13A). Since detection of serogroup 6 (serotypes 6A/B/C/D) using the CDC based cmPCR is also based on detection of *wciP*, the 250 bp CDC target was not included as an PCR internal control study (Pai *et al.*, 2006).

The reference strains for *S. pneumoniae* serotypes 6A, 6B, 6C, and 6D (Table 3) displayed expected amplicon sizes and sequences (Figure 13B and Table 10). For each serotype in the reproducibility analysis, *wciP* was amplified, with no cross-reactions observed between 32 other streptococci, or the other 88 *S. pneumoniae* serotypes (Table 10). With one exception, *wciN*α was specific for serotypes 6A and 6B, and *wciN*β for 6C and 6D. The *wciP* sequences were accurate for all serotypes evaluated, and protein translation had the expected amino acid at position 195. Overall, most results were concordant with Quellung serotyping: 8/8 strains accurately identified as serotypes 6A, 11/11 as serotype 6B, 20/20 as serotype 6C, and 10/13 as serotypes 6D. Overall, only three discordant results were observed for serotypes previously identified as 6D (Table 11). One of the three 6D serotypes was identified by PCR and sequencing as serotype 6B based on presence of the *wciN*α, and WciP with Asn at position 195. The other two discordant results were likely false negatives, due to low concentrations of pneumococcal DNA (*lytA* Cp values of 31.07 and 32.33 (corresponding to pneumococcal DNA concentrations of approximately 237 and 183 copies/ml, respectively). These were characterized as non-typeable (NT) until resolution can be found.

A)



B)

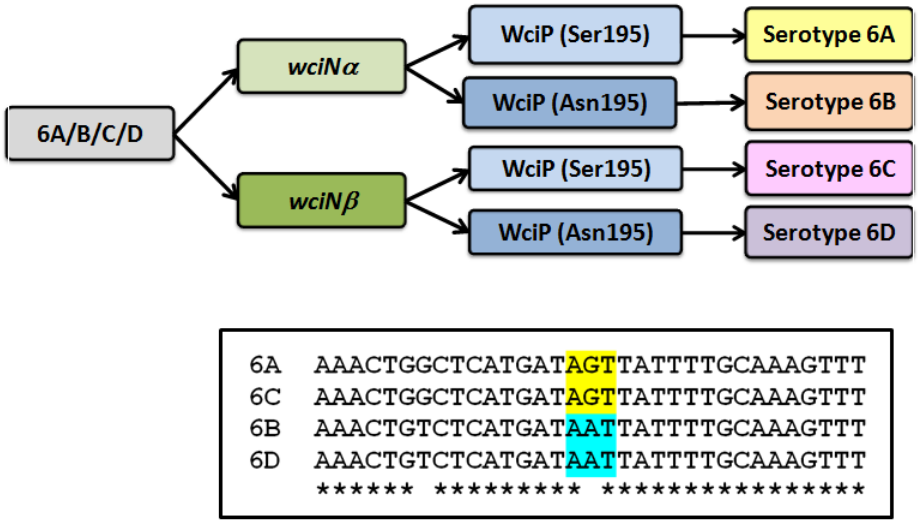
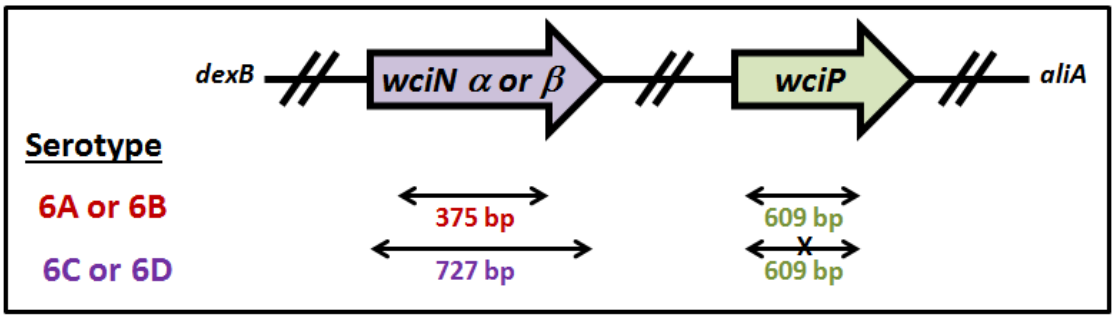


Figure 12. Rationale for the targets used for PCR and sequenced-based discrimination of *S. pneumoniae* serotypes 6A, 6B, 6C, and 6D. A) Diagram illustrating the repeating capsular unit of *S. pneumoniae* serogroup 6. B) Proposed algorithm for discrimination of serotypes within serogroup 6, relying on PCR for *wciN*α or *wciN*β, followed by PCR and sequencing of *wciP*. The inset shows the G to A substitution conferring the amino acid Ser to Asn change in WciP at position 195. Abbreviations: acetate (Ac); galactose (Gal); glucose (Glc); rhamnose (Rha); ribitol (Rib-ol); phosphate (p); pyranose (P).

A)



B)

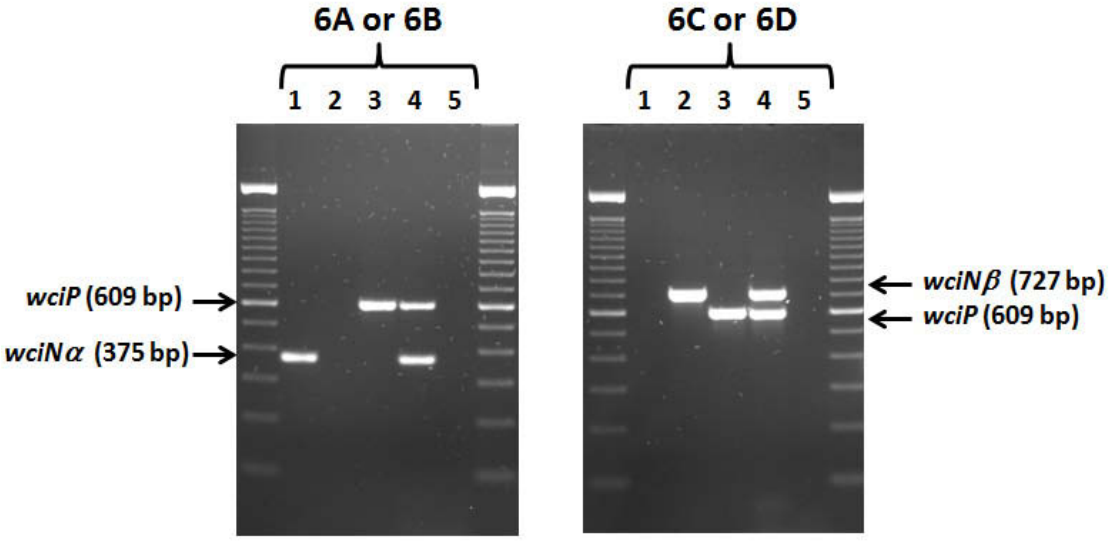


Figure 13. PCR amplification plan and results for serogroup 6. A) Summary of the desired PCR targets. B) Representative PCR reactions for serotypes 6A and 6B or 6C and 6D. Monoplex PCR reactions for *wciN*α (lane 1); *wciN*β (lane 2); *wciP* (lane 3); triplex PCR for all targets (lane 4); and the negative control for the triplex PCR (lane 5). Amplicon sizes were compared to a 100 bp DNA ladder (Invitrogen).

Table 10. Summary of the PCR and serotyping results for *S. pneumoniae* serogroup 6.

Target	Serotype				Other *S. pneumoniae* serotypes	Other streptococci
	6A (n=8)	6B (n=11)	6C (n=20)	6D* (n=13)		
wciNα	100% (8/8)	100% (11/11)	0% (0/20)	7.7% (1/13)	0% (0/90)	0% (0/32)
wciNβ	0% (0/8)	0% (0/11)	100% (20/20)	84.6% (11/13)	0% (0/90)	0% (0/32)
wciP	100% (8/8)	100% (11/11)	100% (20/20)	84.6% (11/13)	0% (0/88)	0% (0/32)
WciP-Ser195	100% (8/8)	0% (0/11)	100% (20/20)	9.1% (1/11)	N/A	N/A
WciP-Asn195	0% (0/8)	100% (11/11)	0% (0/20)	90.1% (10/11)	N/A	N/A

*** 2 strains of serotype 6D were not amplified by either *wciN* or *wciP***

Table 11. Discordant results between Quellung and molecular serotyping of *S. pneumoniae* serogroup 6.

Serotype	Strains	cmPCR result	Sequencing result	Repeat Quellung result	Result resolution
6D	NML SC12-0507	6ABCD	6B	pending	pending
6D	NML SC13-1651 NML SC13-4981	6ABCD	NT	6D	pending

4.3.2. PCR and Sequencing-Based Discrimination of Serotypes 7F and 7A

Following the detection of serotype 7F/A using A 826 bp segment of *wcwH* in the modified CDC cmPCR methods (Lang *et al.,* 2017), the discrimination of serotypes 7F and 7A was pursued using PCR and sequence. Differentiation relied on a frameshift thymine (T) insertion at position 587 of *wcwD* in serotype 7A (Mavroidi *et al*., 2007). Compared to serotype 7F, 7A lacks a side chain in the repeating capsular polysaccharide unit, which is believed to be added by the glycosyltransferase WcwD (Figure 14). In this study

primers were designed to target a 486 bp fragment of *wcwD*, as well as a 599 bp internal control gene *wzy* that is used in the CDC cmPCR reaction for 7F/A (Figure 15B) (Da Gloria Carvalho *et al.* 2010). The reference strains for *S. pneumoniae* serotypes 7F and 7A (Table 3) displayed expected amplicon sizes and sequences (Figure 15B). For each serotype in the reproducibility analysis, both *wzy* and *wcwD* were amplified, with no cross-reactions observed between 32 other streptococci, or the other 90 *S. pneumoniae* serotypes (Table 12). Monoplex reactions were slightly more sensitive than the duplex reactions (Figure 15B). In the reproducibility analyses, all 22 *S. pneumoniae* strains previously identified as serotype 7F by Quellung reactions were concordant with PCR and sequencing results (Table 12). In contrast, only 1/14 serotype 7A strains was concordant, with the expected T frameshift insertion at position 587. The discrepant results for serotype 7A are noted in Table 13. With the exception of the serotype 7A reference strain, 13 strains previously identified as serotype 7A had wild-type *wcwD* sequences in the target area, suggesting a 7F serotype signature. Following retesting using Quellung at the NML, 10 were resolved as serotype 7F (Table 13). While 10 strains of serotype 7A were now resolved as 7F, an additional three strains previously identified as serotype 7A by the NML repeated the same. Additional investigations are underway.

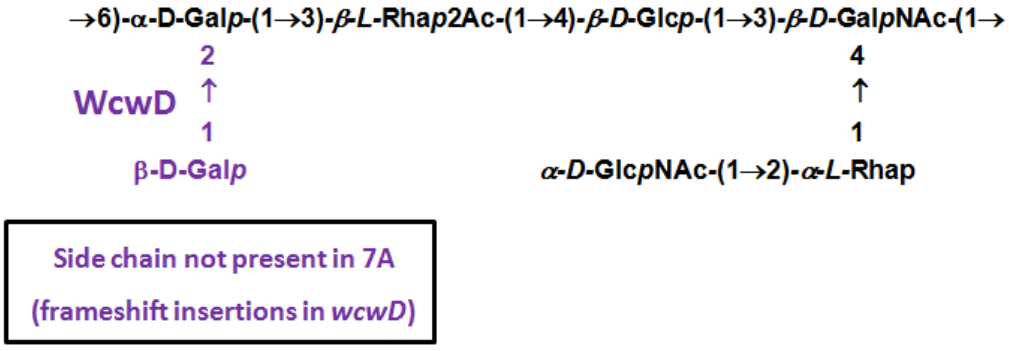
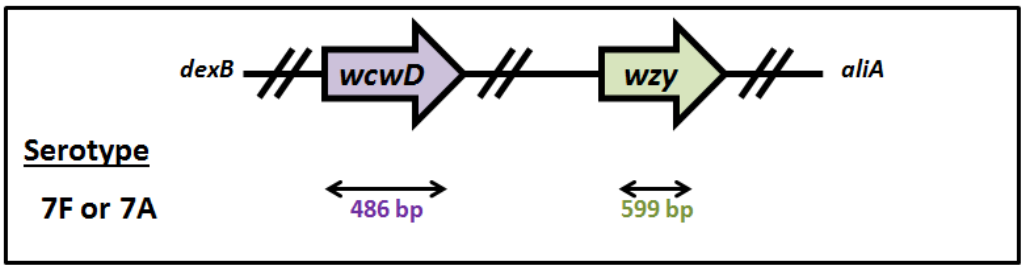


Figure 14. Rationale for the targets used for PCR and sequenced-based discrimination of *S. pneumoniae* serotypes 7F and 7A. The diagram illustrates the repeating unit of *S. pneumoniae* serogroup 7F and the differences noted for serotype 7A. Abbreviations: acetate (Ac); galactose (Gal); N-acetylgalactosamine (GalNAc); glucose (Glc); N-acetylglucosamine (GlcNAc); rhamnose (Rha); pyranose (*p*).

A)



B)

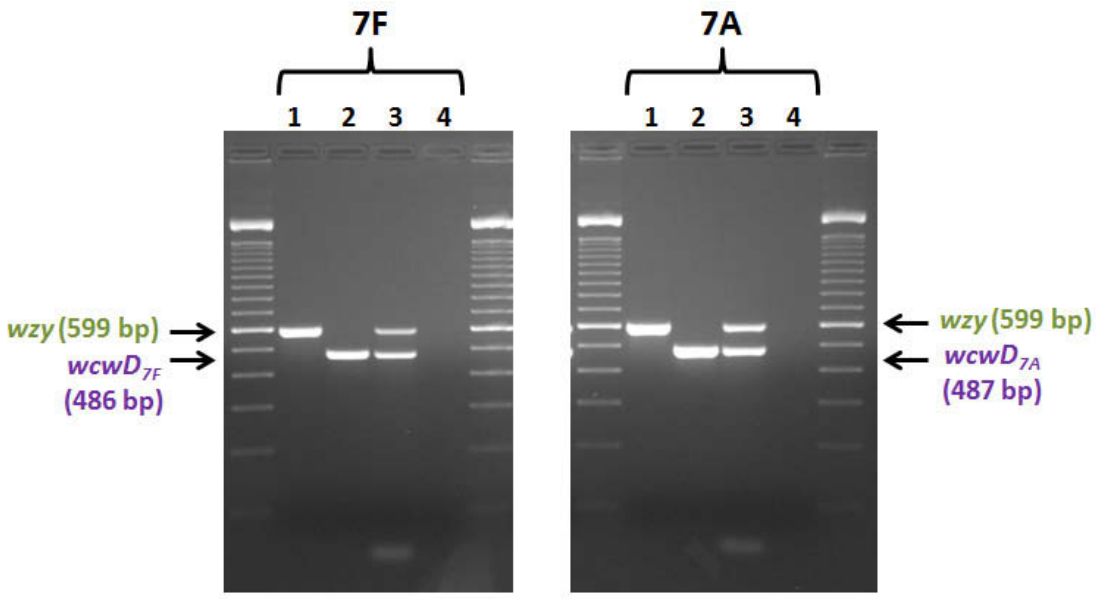


Figure 15. PCR amplification plan and results for serotypes 7F and 7A. A) Diagram illustrating the desired PCR targets. B) Representative PCR reactions for serotypes 7F and 7A. Monoplex PCR reactions for *wzy* (lane 1); *wcwD* (lane 2); duplex PCR (lane 3); and the negative control for the duplex PCR (lane 4). Amplicon sizes were compared to a 100 bp DNA ladder (Invitrogen).

Table 12. Summary of the PCR and serotyping results for *S. pneumoniae* serotypes 7F and 7A.

Target	Serotype		Other *S. pneumoniae* serotypes	Other streptococci
	7F (n=33*)	7A (n=4*)		
wzy	100% (33/33)	100% (4/4)	0% (0/90)	0% (0/32)
wcwD	100% (33/33)	100% (4/4)	0% (0/90)	0% (0/32)
wcwD "T" insertion at position 587	100% (33/33)	25% (1/4)	N/A	N/A

*After resolution of 10 discrepant results as serotype 7F

Table 13. Discordant results between Quellung and sequenced-based serotyping for *S. pneumoniae* serotypes 7F and 7A.

Serotype	Strains	cmPCR result	Sequencing result	Repeat Quellung result	Result resolution
7A	CSBN 43767	7F/A	7F	7F	YES
7A	CBSN 25544 CSBN 27134 CSBN 26923 CSBN 26902 TIBDN 31349 TIBDN 32197 TIBDN 36029 TIBDN 38137 TIBDN 38137	7F/A	7F	7F	YES
7A	NML SC0217 NML SC10-1737 NML SC12-2699	7F/A	7F	7A	pending

4.3.3. PCR and Sequencing-Based Discrimination of Serotypes 9V and 9A

Following the detection of serotype 9V/A using the *wzy* target in the modified CDC cmPCR methods (Lang *et al.,* 2017), the discrimination of serotypes 9A from 9V relied on detection of a frameshift deletion of a "G" at position 722 of *wcjE* in 9A. Loss of function of the WcjE O-acetyltranferase leads to differences in acetylation, and a 9A serotype structure (Figure 16) (Kapatai *et al*., 2016). Both *wcjD* and *wjcE* encode for O-acetyltransferase, and these can be used to differentiate serotypes 9V/A from 9N/L. As such, this study design primers to amplify an 1103 bp fragment of *wcjE* was used for differentiation of serotypes 9A and 9V, and initially, a 145 bp segment of *wcjD* was used as a PCR internal control (Figure 17A and Table 7). When *wcjD* and *wcjE* monoplex PCR reactions were compared to a duplex PCR containing both targets, a significant reduction in *wcjE* band intensity was observed (Figure 17B). While both targets could be tested independently, an alternative approach was evaluated. The *wcjE* PCR target was combined to a different internal control, the 816 bp fragment of *wzy* used for detection of 9V/A in the CDC cmPCR (Figure 17C) (Da Gloria Carvalho *et al.* 2010). The reference strains for *S. pneumoniae* serotypes 9V and 9A (Table 3) displayed expected amplicon sizes and sequences (Figure 17B and C). For each serotype in the reproducibility analysis, both *wzy* and *wcjE* showed no cross-reactions with 32 other streptococci, or the other 90 *S. pneumoniae* serotypes (Table 14). Only 5/11 serotype 9V and 4/10 serotype 9A were concordant. In six of the discordant results, *wzy* failed to be detected (three each for serotypes 9A and 9V). Similarly, *wcjE* failed to be detected in six 9A and six 9V serotypes (Table 15).

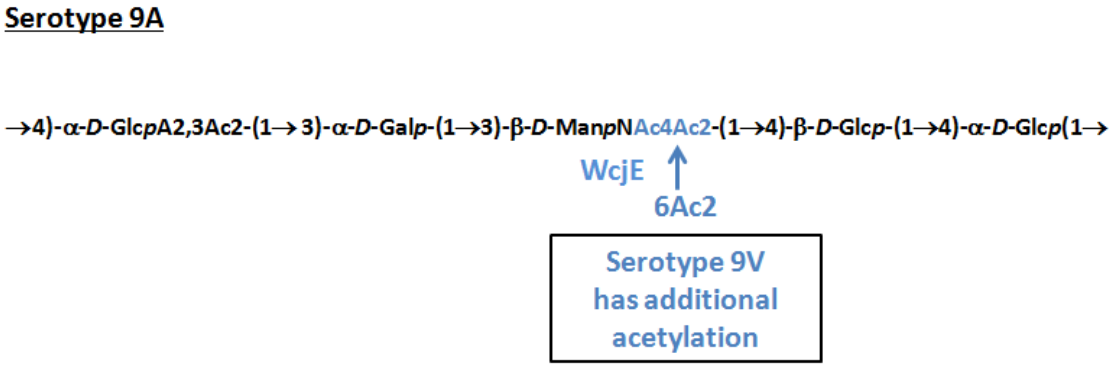


Figure 16. Rationale for the targets used for PCR and sequenced-based discrimination of *S. pneumoniae* serotypes 9V and 9A. The diagram illustrates the repeating unit of *S. pneumoniae* serotype 9A and the differences in acetylation noted for serotype 9V. Abbreviations: acetate (Ac); galactose (Gal); glucose (Glc); *N*-acetylmannosamine (ManNAc); pyranose (*p*).

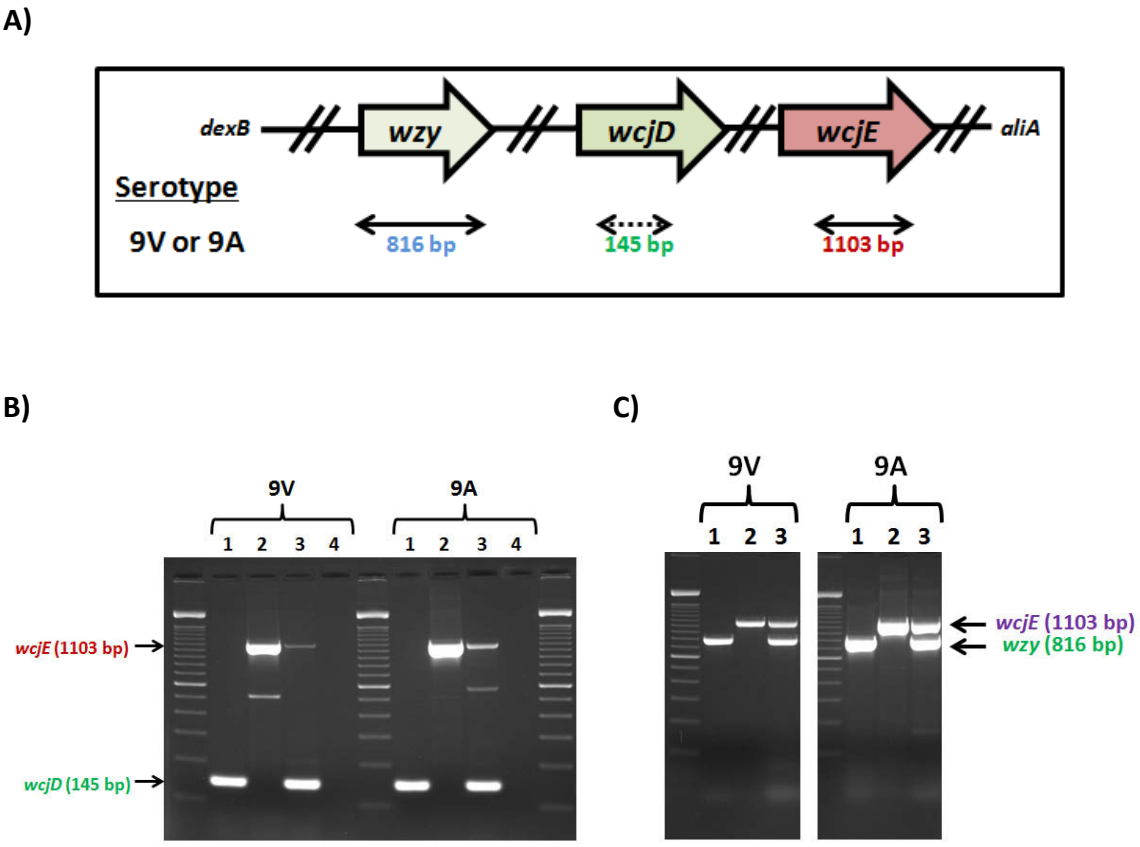


Figure 17. PCR amplification plan and results for serotypes 9V and 9A. A) Diagram illustrating the desired PCR targets. B and C) Representative PCR reactions for serotypes 9V and 9A. In lane 1, the monoplex PCR reactions for *wcjD* (in B) or *wzy* (in C); lane 2 corresponds to *wcjE*; and the duplex PCRs are in lane 3; the negative control for the duplex PCR is in lane 4. Amplicon sizes were compared to a 100 bp DNA ladder (Invitrogen).

Table 14. Summary of the PCR and serotyping results for *S. pneumoniae* serotypes 9V and 9A.

Target	Serotype		Other *S. pneumoniae* serotypes	Other streptococci
	9V (n=11)	9A (n=10)		
wzy	72.7% (8/11)	70.0% (7/10)	0% (0/90)	0% (0/32)
wcjE	45.5% (5/11)	40.0% (4/10)	0% (0/90)	0% (0/32)
wcjE ("G" deletion at position 722)	0% (0/5)	100% (4/4)	0% (0/90)	0% (0/32)

Table 15. Discordant results between Quellung and sequenced-based serotyping for *S. pneumoniae* serotypes 9V and 9A.

Serotype	Strains	cmPCR result	Sequencing result	Repeat Quellung result	Result resolution
9A	NML SC13-0662 NML SC13-2978 CDC 7280	*wzy* not detected	N/A	pending	pending
9V	NML SC12-3451 NML SC13-4234 CDC 4908	*wzy* not detected	N/A	pending	pending
9A	NML SC10-0658 CDC 5874 CDC 3747	9A/V	*wcjE* not detected	pending	pending
9V	NML SC15-1614 CDC 2545 CDC 4167	9A/V	*wcjE* not detected	pending	pending

4.3.4. PCR and Sequencing-Based Discrimination of Serotypes 9N and 9L

Following the detection of serotype 9N/L using the 516 bp *wzx* target in the modified CDC cmPCR methods (Lang *et al.,* 2017), the discrimination of serotypes 9N from 9L relied on detection of a number of single nucleotide polymorphisms across four genes: *wzy* (encoding the capsular unit polymerase), *wchA* (UDP-glucosyl-1-phosphate transferase), and two glycosyltransferase genes, *wcjA* and *wcjB* (Figure 19A). Amino acid changes in WcjA and WcjB are consistent with presence of glucose in 9N instead of galactose present in the polysaccharide repeat unit (Figure 18) (Kapatai *et al*., 2016; Mavroidi *et al*., 2007). As such, this study design two PCR reactions: the first contained primer pairs to amplify a 886 bp fragment of *wcjA* and 329 bp sequence spanning the *wzx*/*wcjB* genes; and the second used primers targets a 451 bp and 743 bp fragments of *wzy* and *wchA*, respectively (Table 7). No differences were noted between monoplex reactions for each target and either of the duplex PCR reactions. The reference strains for *S. pneumoniae* serotypes 9N and 9L (Table 3) displayed expected amplicon sizes in both PCR reactions (Figure 19B), and showed expected sequences in each of the four genes (Table 16). For each serotype in the reproducibility analysis, none of the genes (*wcjA*, *wcjB*, *wchA*, or *wzy*) showed any cross-reactions with 32 other streptococci, or the other 90 *S. pneumoniae* serotypes (Table 16). All 9N serotype strains (17/17) and 9L serotypes strains (21/21) amplified all four PCR targets. Sequencing results were as expected for all 17 of the 9N serotypes, and 12 of the 9L serotypes (Table 17). However, 9 *S. pneumoniae* strains previously identified as 9L by Quellung displayed all 10 SNPs across the four genes, which would be consistent with serotypes 9N (Table 17). These 9 discordant results are currently under investigation (Table 18).

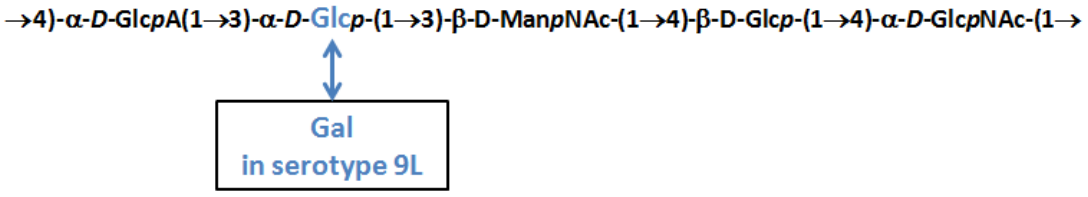
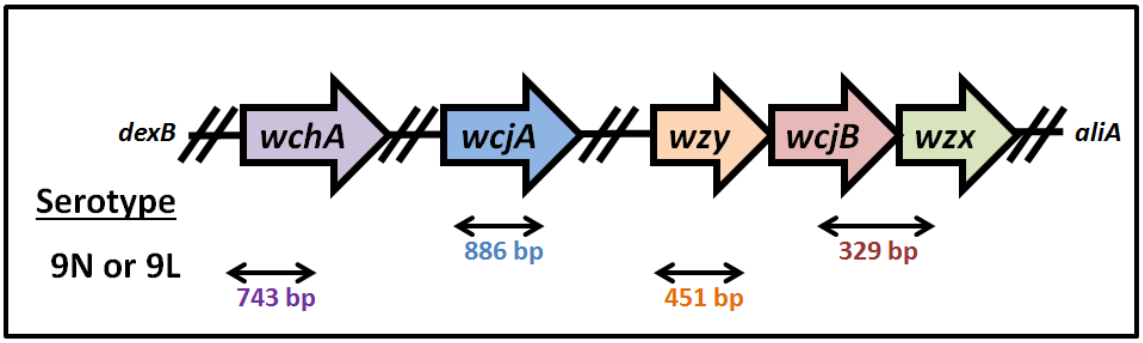


Figure 18. Rationale for the targets used for PCR and sequenced-based discrimination of *S. pneumoniae* serotypes 9N and 9L. The diagram illustrates the repeating unit of *S. pneumoniae* serogroup 9N and the differences in serotype 9L. Abbreviations: acetate (Ac); galactose (Gal); glucose (Glc); *N*-acetylglucosamine (GlcNAc); *N*-acetylmannosamine (ManNAc); pyranose (*p*).

A)



B)

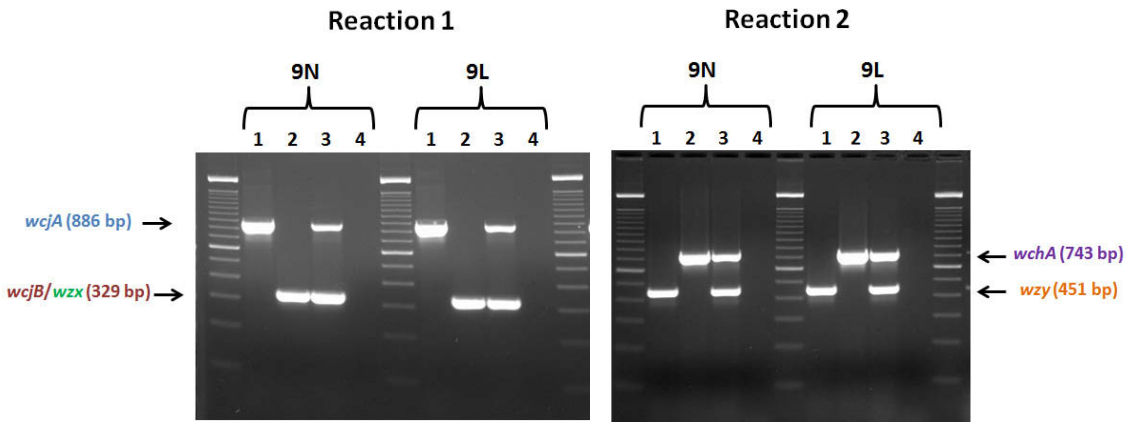


Figure 19. PCR amplification plan and results for serotypes 9N and 9L. A) Diagram illustrating the desired PCR targets for PCR reactions 1 and 2. B) Representative PCR reactions for serotypes 9N and 9L. In PCR reaction 1, the monoplex PCR for *wcjA* (lane 1), monoplex PCR for *wcjB/wzx* (lane 2); duplex PCRs (lane 3), and the negative control for the duplex PCR (lane 4). In PCR reaction 2, the monoplex PCR for *wzy* (lane 1), monoplex PCR for *wchA* (lane 2); duplex PCRs (lane 3), and the negative control for the duplex PCR (lane 4). Amplicon sizes were compared to a 100 bp DNA ladder (Invitrogen).

Table 16. Summary of the PCR results for *S. pneumoniae* serotypes 9N and 9L.

Target	Serotype		Other *S. pneumoniae* serotypes	Other streptococci
	9N (n=17)	9L (n=21)		
wchA	100% (17/17)	100% (21/21)	0% (0/91)	0% (0/32)
wcjA	100% (17/17)	100% (21/21)	0% (0/91)	0% (0/32)
wcjB/wzx	100% (17/17)	100% (21/21)	0% (0/91)	0% (0/32)
wzy	100% (17/17)	100% (21/21)	0% (0/91)	0% (0/32)

Table 17. Summary of concordant and discordant sequencing results for *S. pneumoniae* serotypes 9N and 9L.

Gene	Position	Ref. Seq.		Ref. strains		Concordant 9N isolates (n=16)	Concordant 9L isolates (n=11)	Discordant 9L isolates (n=9)
		9N	9L	9N	9L			
wchA	504	TCT	TAT	TCT	TAT	TCT	TAT	TCT
	879	TCA	CCA	TCA	CCA	TCA	CCA	TCA
wcjA	414	TAT	CAT	TAT	CAT	TAT	CAT	TAT
	429	AGT	GGT	AGT	GGT	AGT	GGT	AGT
	528	GGT	GAT	GGT	GAT	GGT	GAT	GGT
	636	AAT	GAT	AAT	GAT	AAT	GAT	AAT
	852	TCA	GCA	TCA	GCA	TCA	GCA	TCA
	957	ACT	ATT	ACT	ATT	ACT	ATT	ACT
wcjB	789	ACC	GCC	ACC	GCC	ACC	GCC	ACC
wzy	846	AAC	GAC	AAC	GAC	AAC	GAC	AAC

The green corresponds to expected results, whereas red signifies discordant results.

Table 18. Discordant results between Quellung and sequenced-based serotyping for *S. pneumoniae* serogroup 9N and 9L.

Serotype	Strains	cmPCR result	Sequencing result	Repeat Quellung result	Result resolution
9L	TIBDN 22074 TIBDN 24642 TIBDN 25164 TIBDN 27221 TIBDN 27225 TIBDN 27194 CSBN 20014 CSBN 15084 CSBN 21574	9N/L	9N	pending	pending

4.3.5. PCR and Sequencing-Based Discrimination of Serotypes 11A and 11D

Following the detection of serotype 11A/D using the 463 bp *wzy* target in the modified CDC cmPCR methods (Lang *et al.,* 2017), a strategy was developed to discriminate serotypes 11A from 11D, as well as other closely related serotypes in serogroup 11 (11B, 11C, and 11F). The *wcwC*, *wcjE,* and *wcwR* genes encode for acetyltransferases; *wcwC* and *wcjE* are only present in 11A, 11D, and 11F, whereas *wcwR* is only present in 11B and 11C. The donor sugar for the glysosyltransferase WcrL is Glc*p*NAc in serotypes 11F, 11B, and 11C, but Glc*p* in 11A (Figure 20). In this study, primers were designed to target a 243 bp segment of *wcwR* (for detection of serotypes 11B/C), coupled to a second primer pair spanning a 911 bp segment of *wcwC* and *wcrL*. Second, sequencing of the *wcrL* gene was used for differentiation of serotype 11A, D, and F, as differences in a codon at position 334. At this position, codon differences include AAT (encoding Asn) for serotype 11A, ACT (encoding Ser) for serotype 11D, and GCT (encoding Ala) for serotype 11F (Figure 21A) (Kapatai *et al*., 2016). For the PCR, no

differences were noted between monoplex reactions for each target and either of the multiplex PCR reaction. The reference strains for *S. pneumoniae* serotypes 11A and 11D (Table 3) displayed expected amplicon sizes, as did reference strains for serotypes 11B, 11B, and 11F (Figure 21B). In the reproducibility analysis, no targets showed cross-reactions with 32 other streptococci, or the other 90 *S. pneumoniae* serotypes (Table 19). All serotype 11A (18/18) showed the expected *wcrL* sequence, and for serotypes 11D, 2/4 were concordant with expected results (Table 19). The two discordant results were strains previously identified as 11D, but show a genetic signature consistent with 11A (Table 20). These two strains are currently under investigation.

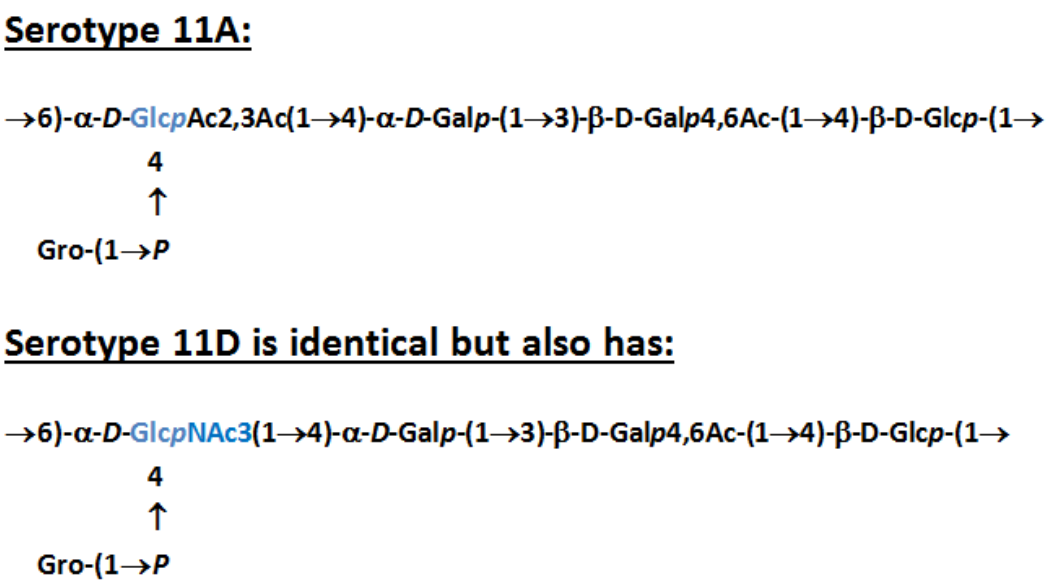
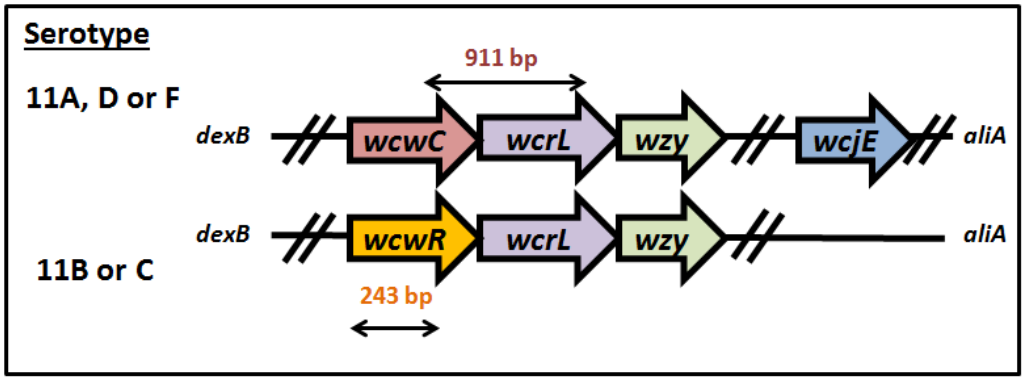


Figure 20. Rationale for the targets used for PCR and sequenced-based discrimination of *S. pneumoniae* serotypes 11A and 11D. The diagram illustrates the repeating unit of *S. pneumoniae* serogroup 11A and 11D. Abbreviations: acetate (Ac); galactose (Gal); glucose (Glc); glycerol (Gro); *N*-acetylglucosamine (GlcNAc); pyranose (*p*).

A)



B)

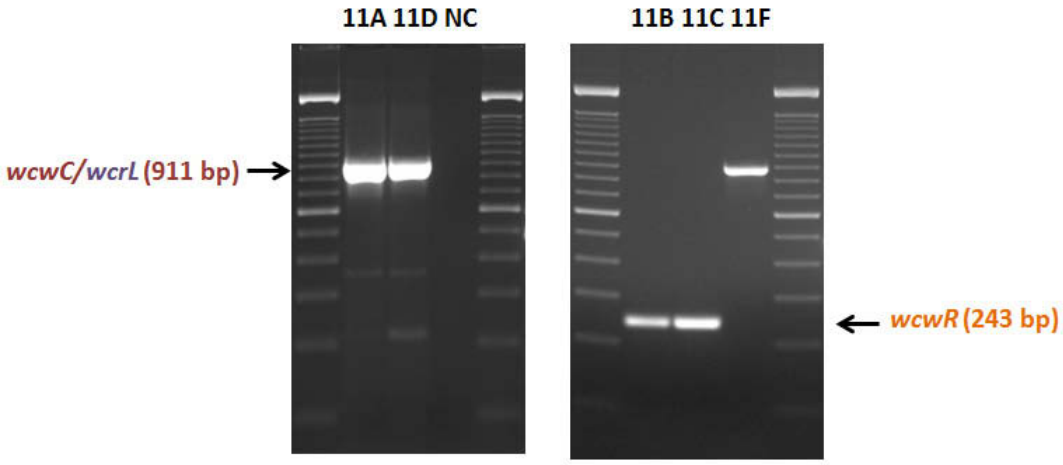


Figure 21. PCR amplification plan and results for serotypes 11A and 11D. A) Diagram illustrating the desired PCR targets. B) Representative PCR reactions for serotypes 11A and 11D, as well as 11B, 11C, and 11F. For serotypes 11A, 11D, and 11F, the PCR target spanning the *wcwC*/*wcrL* genes show an expected 911 bp amplicon. For serotypes 11B and 11C, the expected 243 bp amplicon for *wcwR* is present. The negative control (NC) is also shown. Amplicon sizes were compared to a 100 bp DNA ladder (Invitrogen).

Table 19. Summary of the PCR and sequencing results for *S. pneumoniae* serotypes 11A and 11D.

Target	Serotype		Other *S. pneumoniae* serotypes	Other streptococci
	11A (n=18)	11D (n=4)		
wcwC/wcrL	100% (18/18)	100% (4/4)	0% (0/90)	0% (0/32)
wcrL position 334: AAT	100% (18/18)	50% (2/4)	N/A	N/A
wcrL position 334: ACT	0% (18/18)	50% (2/4)	N/A	N/A

Table 20. Discordant results between Quellung and sequenced-based serotyping for *S. pneumoniae* serotype 11A and 11D.

Serotype	Strains	cmPCR result	Sequencing result	Repeat Quellung result	Result resolution
11D	TIBDN 21411 CARA GZ-11D	11A/D	11A	pending	pending

4.3.6. PCR and sequencing-based discrimination of serotypes 12F, 12A, 12B, 44, and 46

Following the detection of serotype 12F/A/B/44/46 using the 376 bp *wzx* target in the modified CDC cmPCR methods (Lang *et al.,* 2017), the discrimination of each serotypes relied on detection of a number of single nucleotide polymorphisms in three genes: two glycosyltransferase gene unique for the genogroup, *wcxD* and *wcxF*, as well as the gene encoding the capsular unit polymerase, *wzy* (Kapatai *et al*., 2016). Since limited structural information is available for serotypes 12B, 44, and 46, the effect of the glycosyltransferases is unknown (Mavroidi *et al*., 2007). However, the known constituents of serotype 46 (*D*-Gal, *D*-GalNAc, *D*-GlcNAc, and *L*-FucNAc) are consistent with the known structures for 12F and 12A (Figure 22) (Geno *et al*., 2015). For detection of the mutations conferring capsule differences, this study designed two PCR reactions.

The first triplex PCR reactions targeted a 632 bp fragment of *wcxD*, a 547 bp fragment of *wcxF*, and the 376 bp *wzx* gene from the CDC cmPCR reactions (used as an internal control). The second PCR reaction detected a 795 bp portion of the *wcxF* gene, and a 507 bp fragment of *wzy* (Figure 23A). No differences were noted between monoplex reactions for any of the PCR target and either of the multiplex PCR reactions. The reference strains for *S. pneumoniae* serotypes 12F, 12A, 12B, 44, and 46 (Table 3) displayed expected amplicon sizes in both PCR reactions (Figure 23B), and showed expected sequences in each of the three genes (Table 21). In the reproducibility analysis, no targets (*wcxD, wcxF, wzy,* and *wzx*) cross-reacted with 32 other streptococci, or the other 87 *S. pneumoniae* serotypes (Table 21). For each strains and serotype, the expected PCR amplicons were amplified with the appropriate sizes (Table 21). For the sequencing results, concordant results were obtained for all strains of serotype 44 (2/2) and 46 (3/3), some strains of serotype 12F (7/13), but only a limited number of serotypes 12A (1/8) and 12B (2/7) (Table 22). The discordant results are summarized in Tables 22 and 23 and are currently under investigation.

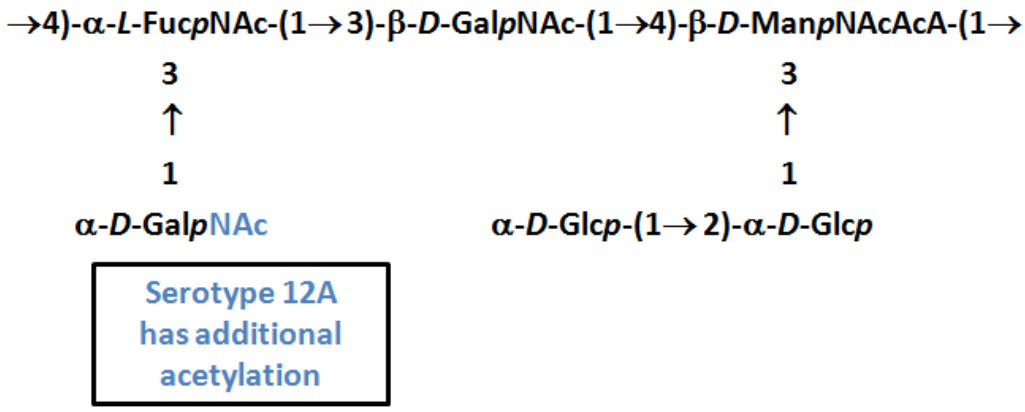
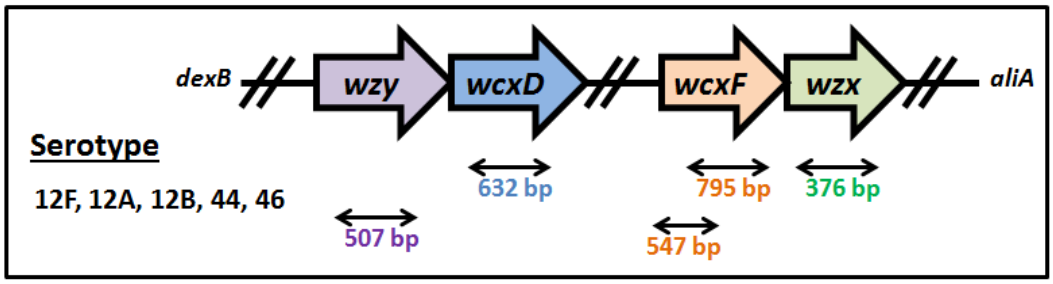


Figure 22. Diagram illustrating the repeating unit of *S. pneumoniae* serotypes 12F and 12A. Abbreviations: acetate (Ac); galactose (Gal); N-acetlyfucosamine (FucNAc); N-acetlygalactosamine (GalNAc); glucose (Glc); glycerol (Gro); N-acetlymanosamine (ManNAc); pyranose (*p*).

A)



B)

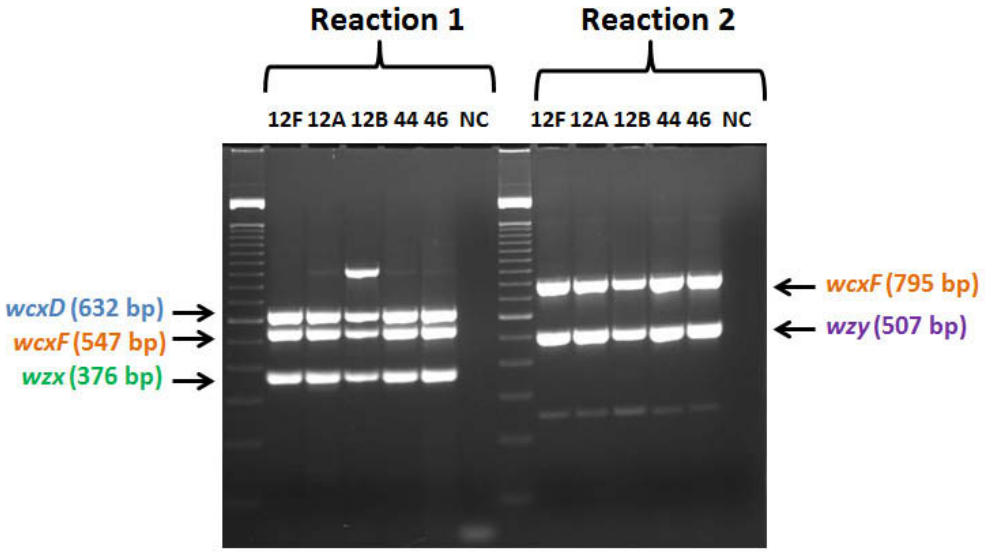


Figure 23. PCR amplification plan and results for serotypes 12F, 12A, 12B, 44 and 46. A) Diagram illustrating the desired PCR targets. B) Representative PCR reactions for each serotypes for PCR reactions 1 and 2. The negative control (NC) for each multiple PCR included. Amplicon sizes were compared to a 100 bp DNA ladder (Invitrogen).

Table 21. Summary of the PCR results for *S. pneumoniae* serotypes 12F, 12A, 12B, 44 and 46.

Target	Serotype					Other *S. pneumoniae* serotypes	Other streptococci
	12F (n=13)	12A (n=8)	12B (n=7)	44 (n=2)	46 (n=3)		
wzx	100% (13/13)	100% (8/8)	100% (7/7)	100% (2/2)	100% (3/3)	0% (0/87)	0% (0/32)
wzy	100% (13/13)	100% (8/8)	100% (7/7)	100% (2/2)	100% (3/3)	0% (0/87)	0% (0/32)
wcxD	100% (13/13)	100% (8/8)	100% (7/7)	100% (2/2)	100% (3/3)	0% (0/87)	0% (0/32)
wcxF	100% (13/13)	100% (8/8)	100% (7/7)	100% (2/2)	100% (3/3)	0% (0/87)	0% (0/32)

Table 22. Summary of sequencing results for *S. pneumoniae* serotypes 12F, 12A, 12B, 44, and 46.

Gene	Position	Reference sequence					Reference strains					Isolates											
												12A					12B		12F			44	46
		12A	12B	12F	44	46	12A	12B	12F	44	46	n=1	n=1	n=1	n=3	n=1	n=1	n=5	n=6	n=5	n=1	n=1	n=2
wcxD	781	TTG	TTG	TTG	TTG	ATG	TTG	TTG	TTG	TTG	ATG	TTG	ATG	ATG	ATG	TTG	TTG	TTG	TTG	TTG	TTG	TTG	ATG
	793	TAT	TAT	TAT	TAT	CAT	TAT	TAT	TAT	TAT	CAT	TAT	CAT	CAT	CAT	CAT	TAT	TAT	TAT	TAT	TAT	TAT	CAT
	805	TCA	TCA	TCA	TCA	CCA	TCA	TCA	TCA	TCA	CCA	TCA	CCA	CCA	CCA	TCA	TCA	TCA	TCA	TCA	TCA	TCA	CCA
	809	ATG	ATG	ATG	ATG	ACT	ATG	ATG	ATG	ATG	ACT	ATG	ACT	ACT	ACT	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ACT
	812	GCA	GTA	GTA	GTA	GTA	GCA	GTA	GTA	GTA	GTA	GTA	GTA	GTA	GTA	GTA	GTA	GTA	GTA	GTA	GTA	GTA	GTA
	845	GCT	GCT	GCT	GCT	GTT	GCT	GCT	GCT	GCT	GTT	GCT	GTT	GTT	GTT	GTT	GCT	GCT	GCT	GCT	GCT	GCT	GTT
wcxF	256	GCC	GCC	GCC	GCC	ACC	GCC	GCC	GCC	GCC	ACC	GC	ACC	GC	GCC	GCC	GCC	GCC	GCC	GCC	GAC	GCC	ACC
	512	ACG	ACG	ATG	ACG	ACG	ACG	ACG	ATG	ACG	ACG	ACG	ACG	ACG	ACG	ATG	ACG	ATG	ATG	ACG	ACG	ACG	ACG
	560	CTT	CTT	CTT	CCT	CTT	CTT	CTT	CTT	CCT	CTT	CTT	CTT	CTT	CTT	CTT	CTT	CTT	CTT	CTT	CTT	CCT	CTT
	703	ATA	CTA	CTA	CTA	CTA	ATA	CTA	ACA	CTA	CTA	CTA	CTA	CTA	ATA	CTA	CTA	CTA	CTA	CTA	CTA	CTA	CTA
	787	CTA	CTA	CTA	ATA	CTA	CTA	CTA	AGA	ATA	CTA	CTA	CTA	CTA	CTA	CTA	CTA	CTA	CTA	CTA	CTA	ATA	CTA
	889	GTT	GTT	GTT	GTT	ATT	GTT	GTT	GTT	GTT	ATT	GTT	ATT	GTT	GTT	GTT	GTT	GTT	GTT	GTT	GTT	GTT	ATT
	916	GGT	GGT	GGT	GCT	AGT	GGT	GGT	GGT	GCT	AGT	GGT	AGT	GGT	GGT	GGT	GGT	GGT	GGT	GGT	GGT	GCT	AGT
	1120	GCC	GCC	GCC	ACC	GCC	GCC	GCC	GCC	ACC	GCC	GCC	GCC	GCC	GCC	GCC	GCC	GCC	GCC	GCC	GCC	ACC	GCC
wzy	251	ATT	ACT	ACT	ACT	ACT	ATT	ACT	ACT	ACT	ACT	ACT	ACT	ACT	ACT	AC	ACT	ACT	ACT	ACT	ACT	ACT	ACT

The green corresponds to expected results, whereas red signifies discordant results.

Table 23. Discordant results between Quellung and sequenced-based serotyping for *S. pneumoniae* serotypes 12F, 12A, 12B, 44, and 46.

Serotype	Strains	cmPCR result	Sequencing result	Repeat Quellung result	Result resolution
12A	CDC 3762	12F/A/B/44/46	12B	pending	pending
12A	CDC 1706 CDC 4170 CDC 4233 CDC 4508 TIBDN 37347 TIBDN 40719	12F/A/B/44/46	12A/46	pending	pending
12B	TIBDN 40454 CSBN 29419 CSBN 27651 CSBN 29423 CSBN 40453	12F/A/B/44/46	12F	pending	pending
12F	NML SC14-0150 NML SC14-1506 NML SC14-3562 CDC 4438 CDC 5617 CDC 6566	12F/A/B/44/46	12B	pending	pending

4.3.7. PCR and Sequencing-Based Discrimination of Serotypes 15B and 15C

Following the detection of serotype 15B/C using the 496 bp segment of the *wzy* gene in the modified CDC cmPCR methods (Lang *et al.,* 2017), the discrimination of serotypes 15B from 15C relied on differences in tandem repeats of "TA" near position 413 of *wciZ* (Bentley *et al*., 2006). The *wciZ* encodes for an O-acetyltransferase, and additional TA insertions in the *wciZ* gene of serotype 15C leads to a frameshift. The resulting loss of function of the WciZ acetyltranferase leads to differences in acetylation in serotype 15C (Figure 24) (Kapatai *et al*., 2016). This study design primers to amplify a 606 bp fragment of *wciZ* for the differentiation of serotypes 15B and 15C, and primers

targeting the 496 bp segment of *wzy* was used as an internal control (Figure 25A). When *wciZ* and *wzy* monoplex PCR reactions were compared to a duplex PCR containing both targets, no differences were observed in band intensity (Figure 25B). The reference strains for *S. pneumoniae* serotypes 15B and 15C (Table 3) displayed expected amplicon sizes and sequences (Figure 25B). In the reproducibility analysis, both *wzy* and *wciZ* did not show any cross-reactions with 32 other streptococci, or the other 90 *S. pneumoniae* serotypes (Table 24). No discordant results were observed with the various serotype 15B and 15C strains of *S. pneumoniae*. The 11 serotype 15B, and 19 serotypes 15C strains, displayed the expected PCR results were consistent with expected results (Table 24). During sequence analysis, each serotype 15C displayed a frameshift insertion in *wciZ*; however, in contrast to the typical TA insertion for most (n=15) of the 15C serotypes, one strain (CDC 2015) had a "TATA" insertion at the same position, three strains had an "A" insertion (NML SC15-2294, NSHA 25-60, and CIRN BC05-7070).

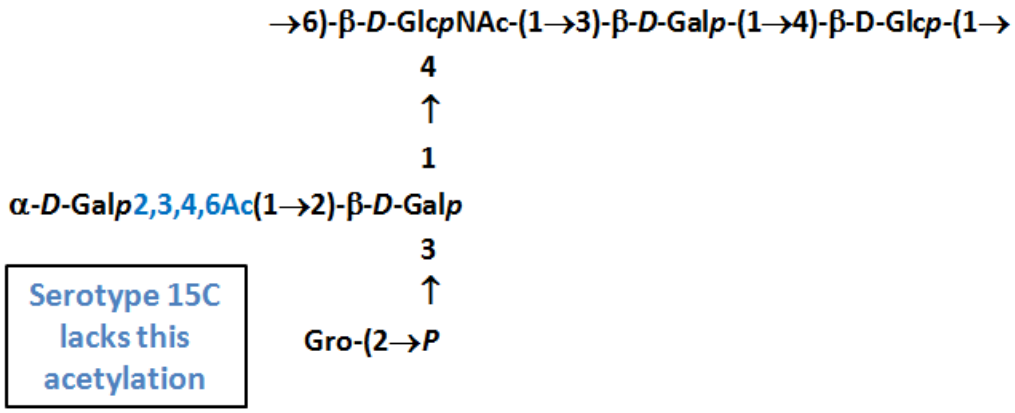
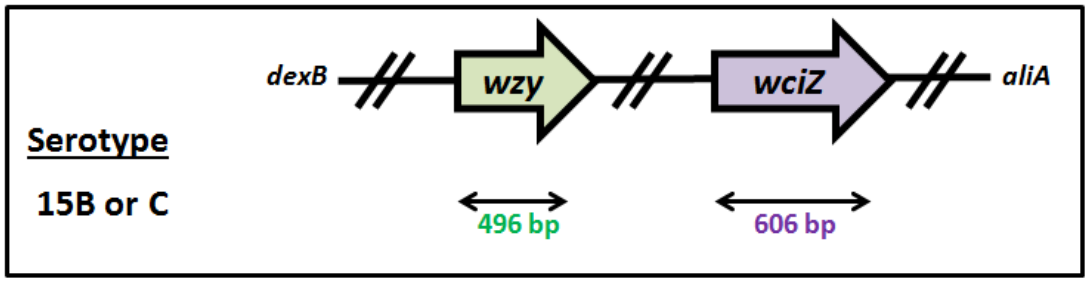


Figure 24. Rationale for the targets used for PCR and sequenced-based discrimination of *S. pneumoniae* serotypes 15B and 15C. Compared to serotype 15B, differences in the length of the TA tandem repeat region of the *wciZ* gene in serotype 15C lead to loss of function the encoded O-acetyltransferase, and results in differences in acetylation. Abbreviations: acetate (Ac); galactose (Gal); N-acetylgalactosamine (GalNAc); glucose (Glc); phosphate (*P*); pyranose (*p*).

A)



B)

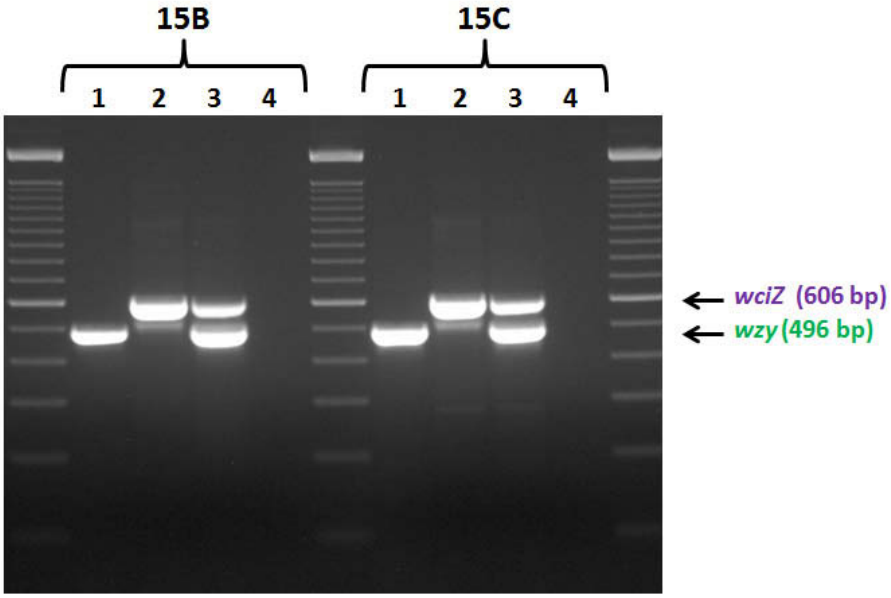


Figure 25. PCR amplification plan and results for serotypes 15B and 15C. A) Diagram illustrating the desired PCR targets. B) Representative PCR reactions for serotypes 15B and 15C. Monoplex PCR reactions for *wzy* (lane 1); *wciZ* (lane 2); duplex PCR (lane 3); and the negative control for duplex PCR (lane 4). Amplicon sizes were compared to a 100 bp DNA ladder (Invitrogen).

Table 24. Summary of the PCR and serotyping results for *S. pneumoniae* serotypes 15B and 15C.

Target	Serotype		Other *S. pneumoniae* serotypes	Other streptococci
	15B (n=11)	15C (n=19)		
wzy	100% (11/11)	100% (19/19)	0% (0/90)	0% (0/32)
wciZ	100% (11/11)	100% (19/19)	0% (0/90)	0% (0/32)
Frameshift near position 413 of *wciZ*	0% (0/11)	100% (19/19)	N/A	N/A

4.3.8. PCR and Sequencing-Based Discrimination of Serotypes 18C, 18F, 18A, and 18B

Following the detection of serotype 18C/F/A/B using the 573 bp segment of the *wzy* gene in the modified CDC cmPCR methods (Lang *et al.,* 2017), a PCR strategy was developed to discriminate serotypes 18F and 18A from serotypes 18B and 18C, and the later two serotypes could be discriminate by sequencing. The rationale behind this approach is based on the capsule structure differences in acetylation (Figure 26). The identification of serotype 18F is relies on the fact that it has an extra acetyltransferase gene (*wcxM*), and serotype 18A lacks the acetyltransferase gene *wciX* (Kapatai *et al*., 2016; Mavroidi *et al*., 2007). A PCR reaction targeting a 345 bp region of *wcxM* was used for the detection of serotype 18F, and a 155 bp DNA sequence spanning the *wzy* and *wciY* genes was used for serotype 18A detection (Figure 27A). Differentiation of serotypes 18B and 18C relied on PCR amplification and sequencing of a 678 bp segment spanning *wzx* and *wciX*, where a G to T substitution at position 168 in *wciX* leads to early stop codon in serotype 18B (Figure 27A) (Kapatai *et al*., 2016; Mavroidi *et al*., 2007). For the PCR, no differences were noted between monoplex reactions for each target and the multiplex PCR reaction. The reference strains for *S. pneumoniae* serotypes 18C, 18F, 18A, and 18B (Table 3) all displayed expected amplicon sizes (Figure 27B), as did the sequence analysis for serotypes 18B and 18C. In the reproducibility analysis, no targets showed cross-reactions with 32 other streptococci, or the other 88 *S. pneumoniae*

serotypes (Table 25). Concordance with the Quellung serotyping result was noted for 11/11 serotype 18C strains, 6/7 serotype 18F strains, 10/12 serotype 18A strains, and 8/22 serotype 18B (Table 25). The discordant results are summarized in Table 26 and are currently under investigation.

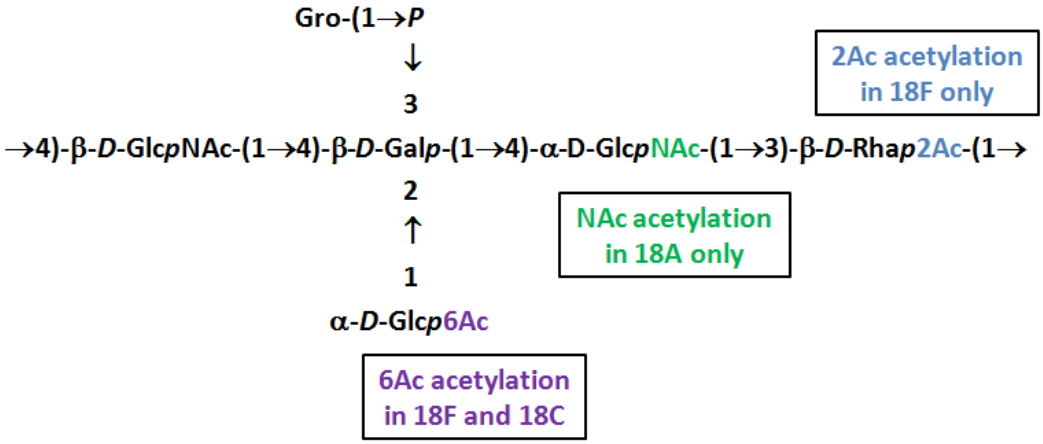
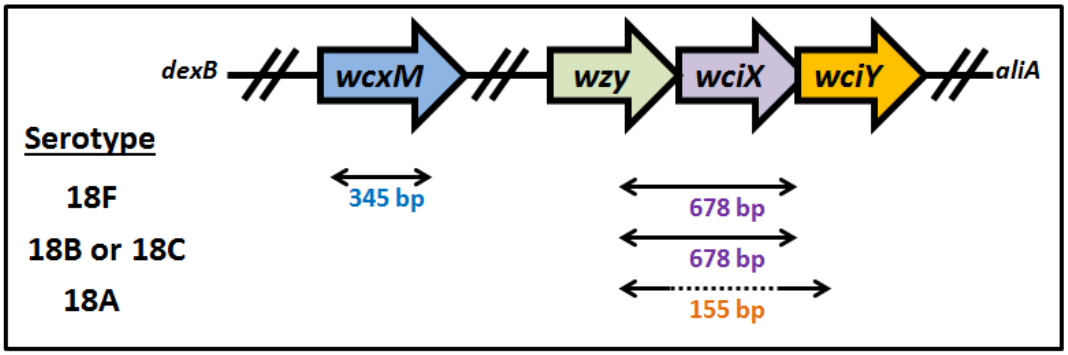


Figure 26. Diagram illustrating the repeating unit of *S. pneumoniae* serogroup 18. For each serotype (C, 18F, 18A, and 18B), differences in acetylation can be observed. Abbreviations: acetate (Ac); galactose (Gal); glucose (Glc); N-acetlyglucosamine (GlcNAc); glycerol (Gro); pyranose (*p*); rhamnose (Rha).

A)



B)

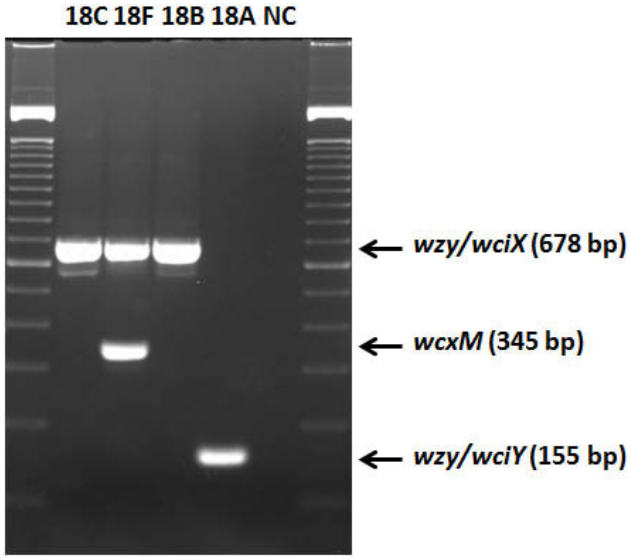


Figure 27. Strategy for PCR and sequence-based discrimination for serogroup 18. A) Diagram illustrating the desired PCR targets. B) Representative PCR reactions for serotypes 18C, 18F, 18A, and 18B. The negative control (NC) is also shown. Amplicon sizes were compared to a 100 bp DNA ladder (Invitrogen).

Table 25. Summary of the PCR and serotyping results for *S. pneumoniae* serotypes within serogroup 18.

Target	Serotype				Other *S. pneumoniae* serotypes	Other streptococci
	18C (n=11)	18F (n=7)	18A (n=12)	18B (n=22)		
wzy/wciY	0% (0/11)	14.3% (1/7)	83.3% (10/12)	0% (0/22)	0% (0/88)	0% (0/32)
wcxM	0% (0/11)	85.7% (6/7)	0% (0/12)	0% (0/22)	0% (0/88)	0% (0/32)
wciX	100% (11/11)	85.7% (6/7)	16.7% (2/12)	100% (22/22)	0% (0/88)	0% (0/32)
G at 168 in *wciX*	100% (11/11)	N/A	N/A	63.6% (14/22)	N/A	N/A
T at 168 in *wciX*	0% (0/11)	N/A	N/A	36.4% (8/22)	N/A	N/A

Table 26. Discordant results between Quellung and sequenced-based serotyping for *S. pneumoniae* serogroup 18.

Serotype	Strains	cmPCR result	Sequencing result	Repeat Quellung result	Result resolution
18F	CDC 4311	18C/F/B/A	18A	pending	pending
18A	CSBN 30969 CSBN 30377	18C/F/B/A	18C	pending	pending
18B	NML SC11-3743 NML SC12-0474 NML SC13-2464 CDC 269 CDC 5122 CDC 3185 TIBDN 29122 TIBDN 34360 TIBDN 41217 TIBDN 41383 CSBN 21738 CSBN 28457 CSBN 15437 CSBN 28141	18C/F/B/A	18C	pending	pending

4.3.9. Sequence-Informed, PCR-Based Discrimination of Serotypes 22F and 22A

Following the detection of serotype 22F/A using the 643 bp *wcwV* target in the modified CDC cmPCR methods (Lang *et al.,* 2017), the discrimination of serotypes 22F and 22A was performed by PCR only. In contrast to the sequences found in NCBI for serotypes 22F and 22A (which differed only by one bp in the *wcwA* gene), comparative genomics studies performed in this study and by Kapatai *et al*., (2016), showed that the sequence of serotype 22A was identical to that found in NCBI; however, the sequences for serotype 22F differed markedly in genes *wcwA* and *wcwC*. Kapatai *et al*. (2016) suggested this was likely due to an incorrect submission for serotype 22F by Bentley *et al*. (2006). The *wcwA* gene encodes a glycosyltransferase, and *wcwC* an O-acetyltranferase. While differences in these genes might confer serotype specificity, only the capsular structure of serotype 22F has been determined (Figure 28). Since there was virtually no similarity between the *wcwA* and *wcwC* genes of serotypes 22F and 22A, this study designed a multiplex PCR using primers pairs targeting a 250 bp region of serotype 22F *wcwA* (*wcwA*$_{22F}$), a 148 bp region of serotype 22F *wcwC* (*wcwC*$_{22F}$), and a 399 bp region spanning the *wcwA* and *wcwC* genes of serotype 22A (*wcwA/C*$_{22A}$) (Figure 29A). The 643 bp *wcwV* genes used for serotype 22F/A detection in the CDC cmPCR reactions was also used as an internal control (Figure 29A). The reference strains for *S. pneumoniae* serotypes 22F and 22A (Table 3) displayed expected amplicon sizes and sequences (Figure 29B). When monoplex PCR reactions for each target were compared to a multiplex PCR, only a subtle difference in the internal control (*wcwV*) was noted. In the reproducibility analysis, none of the PCR targets showed cross-reactions with 32 other streptococci, or the other 90 *S. pneumoniae* serotypes (Table 27). All serotype 22F (20/20) and 22A (10/10) were accurately identified.

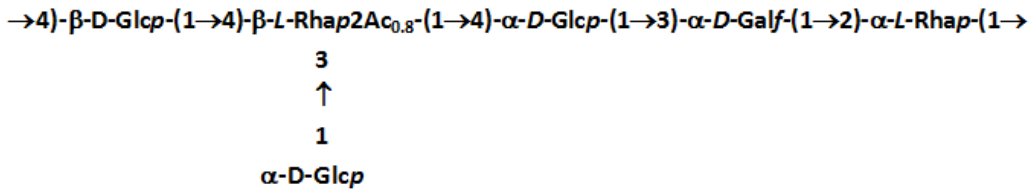
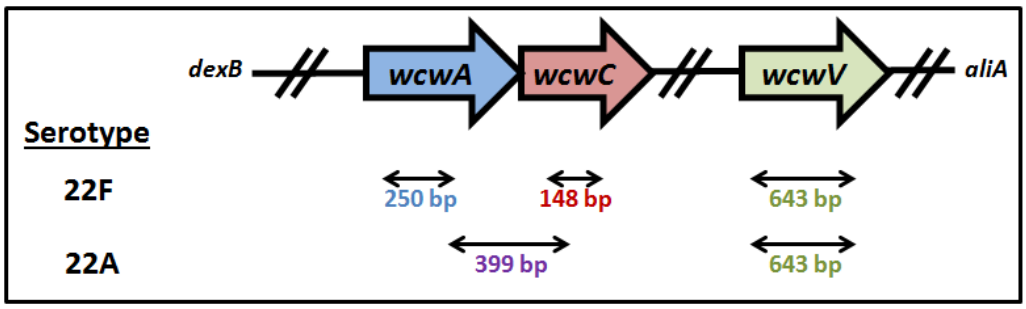


Figure 28. Diagram illustrating the repeating unit of *S. pneumoniae* serotype 22F. Abbreviations: acetate (Ac); furanose (*f*); glucose (Glc); pyranose (*p*); rhamnose (Rha).

A)



B)

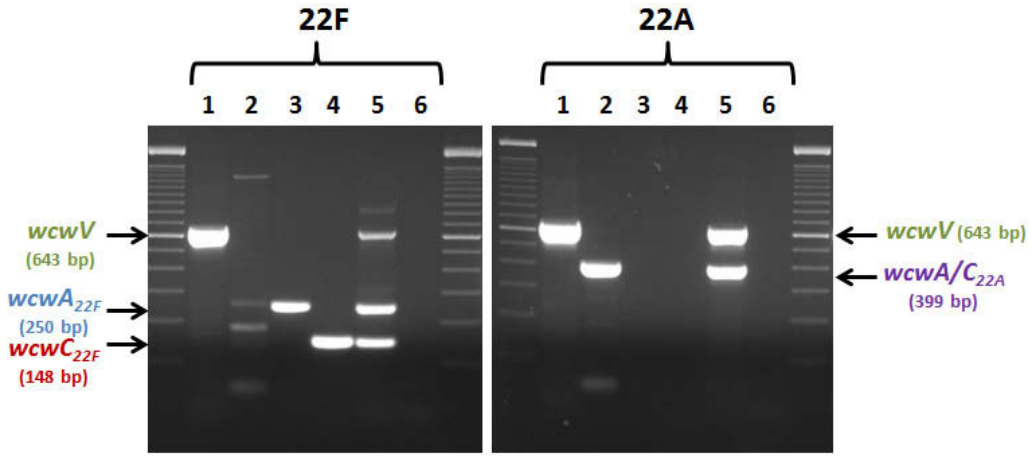


Figure 29. PCR amplification plan and results for serotypes 22F and 22A. A) Summary of the desired PCR targets. B) Representative PCR reactions for serotypes 22F and 22A. Monoplex PCR reactions for *wcwV* (lane 1); *wcwA/C*$_{22A}$ (lane 2); *wcwA*$_{22F}$ (lane 3); *wcwC*$_{22F}$ (lane 4); multiplex PCR for all targets (lane 5); and the negative control for the multiple PCR (lane 6). Amplicon sizes were compared to a 100 bp DNA ladder.

Table 27. Summary of the PCR results for *S. pneumoniae* serotypes 22F and 22A.

Target	Serotype 22F (n=20)	Serotype 22A (n=10)	Other *S. pneumoniae* serotypes	Other streptococci
wcwV	100% (20/20)	100% (10/10)	0% (0/88)	0% (0/32)
$wcwA_{22F}$	100% (20/20)	0% (0/10)	0% (0/90)	0% (0/32)
$wcwC_{22F}$	100% (20/20)	0% (0/10)	0% (0/90)	0% (0/32)
$wcwA/wcwC_{22A}$	0% (0/20)	100% (10/10)	0% (0/90)	0% (0/32)

4.3.10. PCR and Sequencing-Based Discrimination of Serotypes 33F, 33A, and 37

Following the detection of serotype 33F/A/37 using the 338 bp *wzy* target in the modified CDC cmPCR methods (Lang *et al.,* 2017), the discrimination of each serotypes was performed by a combination of PCR and sequencing. Since the *tts* gene is only present in serotype 37, primers were designed to target a 186 bp region of this gene. Tts is a transferase responsible for the synthase-dependent capsule synthesis, generating a unique capsule structure for serotype 37 (Figure 30). Of note, the *tts* gene is located outside the *cps* loci, but serotype 37 also has a complete (but non-functional) 33F *cps* locus (Bentley *et al*., 2006). For the detection and differentiation of serotypes 33F and 33A, primers were designed to target a 593 bp region of *wcjE*. WcjE is an O-acetyltranferase, and a frameshift insertion in position 433 of the *wcjE* gene of serotype 33F results in its inactivation, possibly explaining the differences in acetylation noted in the capsule structures of serotypes 33F and 33A (Figure 31A). When the *tts*, *wcjE*, and *wzy* multiplex PCR reactions was used, the reference strains for *S. pneumoniae* serotypes 33F, 33A and 37 (Table 3) displayed expected amplicon sizes and sequences, with one exception (Figure 31B). The *wcjE* gene in serotype 33F was absent. Since absence of this gene would have the same consequence as a frameshift insertion, the studies continued the reproducibility and specificity analyses (Table 28). In the reproducibility analysis, all targets were shown to be specific, with no cross-reactions with 32 other streptococci, or the other 89 *S. pneumoniae* serotypes (Table 28).

Accurate results were obtained for all 16 serotype 33F strains, the 6 serotype 33A strains, and the 10 serotype 37 strains. It should be noted that *wcjE* was absent in all 33F strains, and wild-type *wcjE* sequences were obtained for all serotype 33A strains (Table 28).

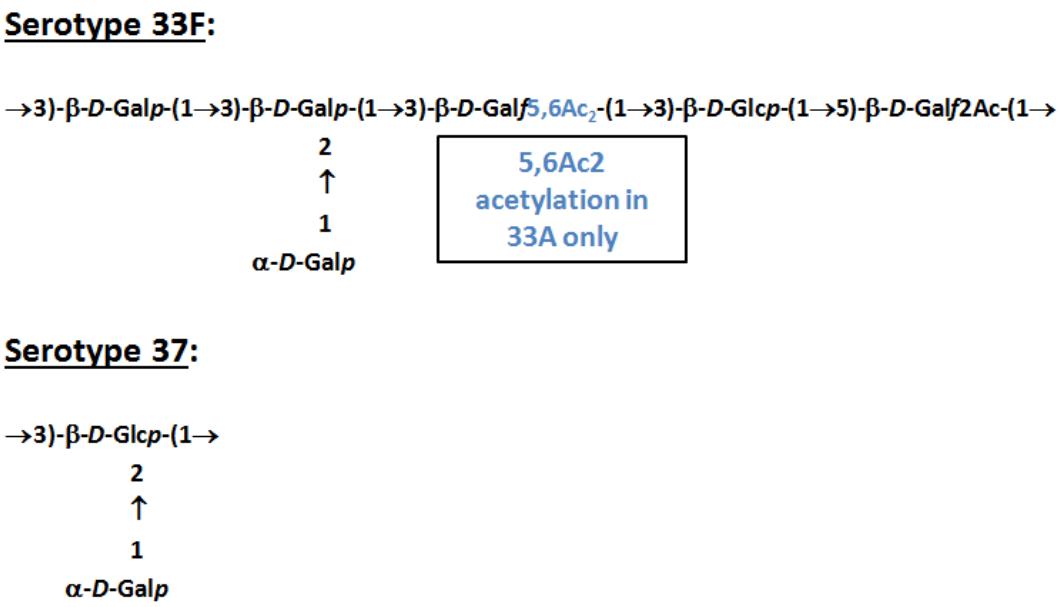
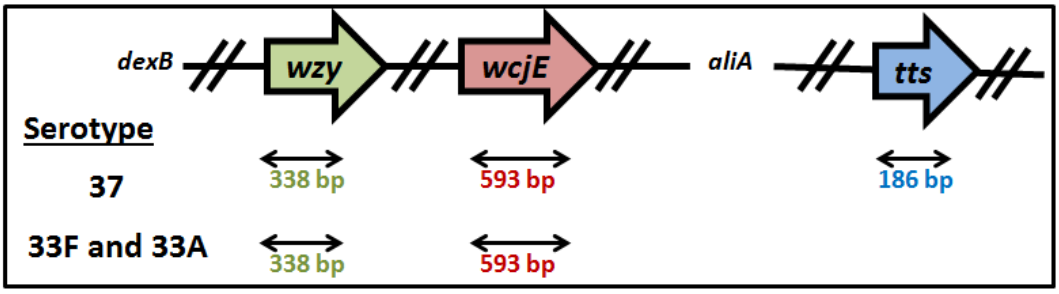


Figure 30. Diagram illustrating the repeating unit of *S. pneumoniae* serotypes 33F, 33A, and 37. Abbreviations: acetate (Ac); furanose (*f*); galactose (Gal); glucose (Glc); pyranose (*p*).

A)



B)

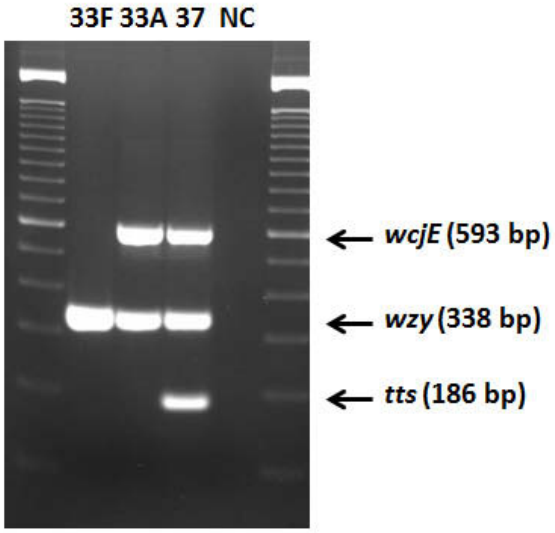


Figure 31. PCR amplification plan and results for serotypes 33F, 33A, and 37. A) Diagram illustrating the desired PCR targets. B) Representative PCR reactions for serotypes 33F, 33A, and 37. The negative control (NC) is also shown. Amplicon sizes were compared to a 100 bp DNA ladder (Invitrogen).

Table 28. Summary of the PCR results for *S. pneumoniae* serotypes 33F, 33A, and 37.

Target	Serotype			Other *S. pneumoniae* serotypes	Other streptococci
	33F (n=16)	33A (n=6)	37 (n=10)		
wzy	100% (16/16)	100% (6/6)	100% (10/10)	0% (0/89)	0% (0/32)
wcjE	0% (0/16)	100% (6/6)	100% (10/10)	0% (0/89)	0% (0/32)
tts	0% (0/16)	0% (0/6)	100% (10/10)	0% (0/89)	0% (0/32)

4.4. Summary of Concordant and Discordant Results for the PCR and Sequencing Approach to Serotype Discrimination for *S. pneumoniae*

Overall, A 78.6% (272/346) of the results were concordant with Quellung serotyping. More specifically, no discordant results were observed with the *S. pneumoniae* serotypes use to evaluate cmPCR groups 15B/C, 22F/A, and 33F/A/37. Only a limited number of discrepant results were observed for *S. pneumoniae* serotypes used for validation of cmPCR groups 6A/B/C/D, 7F/A, and 11A/D. The highest numbers of discrepant results were noted for serotypes within serogroups 9, 12, and 18.

It should also be noted that a higher proportion of discordant results were observed in strains serotyped by Quellung at Mount Sinai hospital for TIBDN [74% (17/23)] and CSBN [75% (18/24)] compared to those serotyped by the CDC [24% (18/76)] or by the NML [8% (18/221)]. Of note, the serotyping fro strains orgininating from CIRN, and NSHA were performed by the NML. While representatives of each serotype was not assessed from each laboratory (which could bias these proportions), retesting of all discordant results with Quellung serotyping is currently underway at the NML to help minimize technologists- or laboratory-specific errors. Following the retesting of the PCR group 7F/7A only, 10 discrepant results were resolved, increasing the overall concordance to 81.5% (282/346) (Table 29).

Table 29. Summary of concordant and discordant results for the PCR and sequencing approach to serotype discrimination for *S. pneumoniae.*

Serotype	No.	Concordant	Discordant before Quellung repeat at the NML	Concordant after Quellung repeat at the NML	Discordant after Quellung repeat at the NML
6A	8	8	0	N/A	N/A
6B	11	11	0	N/A	N/A
6C	20	20	0	N/A	N/A
6D	13	10	3	Pending	Pending
7F	22	22	0	N/A	N/A
7A	14	1	13	11	3
9V	11	5	6	Pending	Pending
9A	10	4	6	Pending	Pending
9N	17	17	0	N/A	N/A
9L	21	12	9	Pending	Pending
11A	18	18	0	N/A	N/A
11D	4	2	2	Pending	Pending
12F	13	7	6	Pending	Pending
12A	8	1	7	Pending	Pending
12B	7	2	5	Pending	Pending
44	2	2	0	N/A	N/A
46	3	3	0	N/A	N/A
15B	11	11	0	N/A	N/A
15C	19	19	0	N/A	N/A
18C	11	11	0	N/A	N/A
18F	7	6	1	Pending	Pending
18A	12	10	2	Pending	Pending
18B	22	8	14	Pending	Pending
22F	20	20	0	N/A	N/A
22A	10	10	0	N/A	N/A
33F	16	16	0	N/A	N/A
33A	6	6	0	N/A	N/A
37	10	10	0	N/A	N/A
Total	346	272	74	282	64

Chapter 5. Discussion

S. pneumoniae is a bacterium that poses a significant morbidity and mortality burden worldwide. Its polysaccharide capsule has proven to be a successful target for vaccination; however, extensive diversity in its capsule structures has made it challenging to develop a pan-pneumococcal vaccine. With the currently available polysaccharides or polysaccharide conjugate pneumococcal vaccines providing coverage of only a select number of serotypes, surveillance for serotype-specific disease and vaccine effectiveness studies requires ongoing and accurate methods for serotyping.

The Quellung reaction has long been the gold-standard for *S. pneumoniae* serotyping, but its requirement for live *S. pneumoniae*, and large panels of cross-reactive antisera, as well as the use of microscopy examination that can be subjective, are all limitations for large-scale epidemiological studies. In addition, isolates of *S. pneumoniae* must be actively producing a capsule in order to be serotyped by Quellung reaction (Park *et al.,* 2012; Salter *et al.,* 2012). The ability to produce capsule is dependent on environmental conditions, and can be lost with continuous subculture (Chao *et al.,* 2015; Shainheit *et al.,* 2015). For these reasons and with the subsequent availability of serotype-specific DNA sequences (Bentley *et al.,* 2006), interest in molecular methods such as PCR and sequencing has grown rapidly over recent years. Over the last decade it has been proven to be an acceptable method for serotype deduction in *S. pneumoniae*, and can be performed in any molecular laboratory able to perform PCR and agarose gel electrophoresis (Dias *et al.,* 2007; Jourdain *et al.,* 2011; Lang *et al.,* 2015 and 2017; Morais *et al.,* 2007; Pai *et al.,* 2006). On the other hand, cmPCR relies on serotype-specific genes within the *cps* loci, and is not be able to discriminate certain *S. pneumoniae* serotypes. This lack of serotype discrimination poses challenges for studies trying to assess the epidemiology of vaccine-preventable serotypes and impact of pneumococcal vaccines (CDC, 2014; Lang *et al*., 2015 and 2017; LeBlanc *et al*., 2017).

The limitations of cmPCR stem from the close similarity of the *cps* loci in some *S. pneumoniae* serotypes. While there was biological relevance for cmPCR using serotypes-specific gene targets within the *cps* loci, this study postulated that other molecular signatures outside the *cps* loci could be used to help discriminate *S. pneumoniae* serotypes. In serotype 37, for example, the *tts* gene, located outside of the *cps* locus, is solely responsible for capsule synthesis (Llull *et al.,* 1999; Mavroidi *et al.,* 2007). However, it is possible that other conserved genes specific to a given serotype could be present, even if it is unrelated to capsule biosynthesis. Recent studies have demonstrated that there is only a subset of genes shared among different serotypes in the core genome of *S. pneumoniae*, and finding unique serotype-specific genes seemed possible (Croucher *et al*., 2015; Kapatai *et al*., 2016). On the other hand, sequence variation within the *cps* loci are now known confer serotype specificity (Kapatai *et al*., 2016), and thus, a sequencing approach might also provide a suitable approach *S. pneumoniae* serotype discrimination.

With the possibility that genetic signatures contributing to serotype specificity could occur inside or outside the *cps* loci of *S. pneumoniae*, this study took two approaches to differentiate vaccine-preventable serotypes from others within cmPCR groups that lacked discrimination: 1) NGS and comparative genomics to identify novel PCR targets outside the *cps* loci; or 2) validation of sequence-based targets inside the *cps* loci. Overall, no PCR targets outside the *cps* loci were identified that enabled reproducible PCR-based discrimination of *S. pneumoniae* serotypes; however, the sequenced-based approach using targets within the *cps* loci showed promising results for the discrimination of *S. pneumoniae* serotypes.

5.1. Identification of Serotype-Specific Targets Outside the *cps* Loci of *S. pneumoniae*

The first objective was to identify novel PCR targets located outside the *cps* loci for discrimination of vaccine-preventable serotypes of *S. pneumoniae.* Following

identification of unique DNA sequences for each *S. pneumoniae* serotype using NGS and comparative genomics, four targets for each serotype were chosen for evaluation. Targets were classified and, if necessary, optimized to increase serotype specificity. While the bioinformatics approach was successful in identifying a list of serotype-specific DNA signatures that were used to design novel PCR targets, the evaluation of each selected target was unsuccessful. Most targets showed no amplification during PCR despite using high concentrations of DNA. To troubleshoot these issues, lower PCR annealing temperatures were used during PCR amplification, and various concentrations of $MgCl_2$ were used in the PCR reactions. This yielded no additional successes.

For *S. pneumoniae* serotypes where the novel PCR could be amplified in the reference strains, with amplicons of the expected sizes, the targets were developed into a multiplex reaction. For example, a multiplex PCR reaction was developed that could discriminate the reference strains of *S. pneumoniae* serotypes within the PCR group 12F/B/A/44/46 (Figure 11). Unfortunately, all of these targets were subsequently found to have at least one exclusion criterion. Some targets were inconsistently detected when tested with multiple isolates of a given serotype, and thus lacked reproducibility. Others targets were only detectable at high DNA concentrations (approximately 10^9 copies/ml), and were shown to be insensitive at lower DNA concentration (approximately 10^6 copies/ml). This would restrict their use to concentrated DNA obtained from culture isolates of *S. pneumoniae*, and perhaps preclude their use for clinical specimens. Finally, many targets chosen to discriminate serotypes with PCR groups 12F/12A/12B/44/46, 6A/6B/6C/6D, or 9V/9A, exhibited cross-reactions with *S. pneumoniae* serotypes within PCR groups, between other *S. pneumoniae* serotypes, or with other streptococci.

Overall, this study demonstrated the proof-of-principle that serotype-specific targets outside the *cps* loci are possible, but the lack or reproducibility or target cross-reactivity precluded it use for epidemiological studies. While none of the novel PCR

targets were acceptable for serotype differentiation, this approach might still be plausible with less restriction on cost and time, as our evaluations were generally limited to the selection of only four targets per serotype.

5.2. Identification of Serotype-Specific Genetic Signatures Within the *cps* Loci of *S. pneumoniae*

After little success with the novel PCR-based approach for serotype discrimination using target outside the *cps* loci, the study design was re-evaluated. While previous cmPCR techniques were not be able to discriminate all vaccine-preventable serotype of *S. pneumoniae* by targeting genes within the *cps* loci, this study hypothesized that unique genetic signatures like mutations may be used for serotype discrimination. Many studies have shown or alluded to differences in the *cps* loci as conferring serotype specificity (Bentley *et al*., 2006, Kapatai *et al*., 2016; Mavroidi *et al*., 2007). This prompted us to reconsider the use of unique genetic differences within the *cps* loci for *S. pneumoniae* serotype discrimination, using strains provided by various laboratories and reference facilities. At the same time this study was considering a sequence-based approach, Kapatai *et al.* (2016) published a bioinformatics pipeline for *S. pneumoniae* serotype deduction called PneumoCaT. The PneumoCaT NGS pipeline software is freely accessible online, and was demonstrated to have good concordance (over 99%) with Quellung serotyping. As such, this tool showed promise for molecular serotyping of *S. pneumoniae*. The tool contains sequence-specific signatures within the *cps* loci capable of differentiating 87 of the 92 serologically different serotypes, as well as two molecular types (6E and 23B1). This study tried to validate the genetic signatures conferring serotype specificity in PneumoCaT by performing Sanger sequencing following PCR amplification of regions flanking the mutations or genetic regions of interest in the *cps* loci. This study evaluated the genetic signatures for the 28 *S. pneumoniae* serotypes that fell into cmPCR groups, which included a vaccine-preventable serotype that could not be discriminated from closely related serotypes.

The subsections below will discuss the results for each serotype within these PCR groups.

5.2.1. PCR and Sequencing-Based Discrimination of Serotypes 6A, 6B, 6C, and 6D

The PCR and sequencing-based discrimination for serotypes 6A, 6B, 6C and 6D used *wciN*α to detect 6A/6B or *wciN*β to detect 6C/6D. Following sequencing of a second target (*wciP*), serotypes 6A and 6C isolates were expected to have a codon for serine at amino acid position 195, whereas serotypes 6B and 6D isolates were expected to have a codon for asparagine. Overall, this two-gene approach in the differentiation of serotypes within PCR group 6A/B/C/D worked well. All WciP sequences were accurate and most overall serotyping results was concordant with Quellung. Other groups have also targeted this region for the discrimination of different combinations of serotypes 6A, 6B, and 6C (Jacobs *et al.,* 2009; Jin *et al.,* 2009; Pai *et al.,* 2005). In 2006, Pai *et al.* targeted *WciP* to differentiate serotypes 6A and 6B using pyrosequencing. A cmPCR reaction using primers targeting *wciN*α or *wciN*β were previously used to differentiate serotypes 6A and 6C, respectively (Jacobs *et al.,* 2009). Jin *et al.,* (2009) modified both of these approaches to differentiate serotypes 6A, 6B, and 6C, prior to the discovery of serotype 6D.

Although the overall results for this PCR group looked promising, there were three discordant results for three isolates that were previously identified as serotype 6D by Quellung; two were likely false negatives (NML SC13-1651 and NML SC13-4981), and one typed as a 6B (NML SC12-0507). The two false negative serotype 6D isolates, were likely the result of low DNA concentrations, as the *lytA* Ct values were approximately 32. To further evaluate these isolates, it would be useful to repeat the PCR reactions for *wciN* and sequencing of *wciP*, as well as the cmPCR 6A/B/C/D. However, we were unfortunately unable to complete this step due to lack of additional template DNA. Additional DNA has been requested from the NML.

NML SC12-0507, was typed as a serotype 6B due to the presence of Asn195 in WciP and *wciN*α as opposed to *wciN*β that has typically been documented in serotype 6D. One possibility for the discordant result between Quellung serotyping and the molecular serotype result obtained in this study, would be the possibility of a false positive for factor 6d during Quellung serotyping or other mutations conferring a phenotype change (Table S1). A second possibility is that this isolate is a serogroup 6 variant, such as the recently reported serotype 6G (Oliver *et al.,* 2013). Oliver *et al.* (2013) demonstrated a variant version of *wciN*α, here called *wciN*α', capable of adding either glucose or galactose to the capsule repeat units. This bispecific-enzyme activity is accredited to a single amino acid change, A150T. When *wciN*α' is present in a serotype 6B *cps* locus, the enzyme creates a 6B/6D hybrid serotype, named 6G. Additionally, the repeat unit was shown to be expressed at a ratio of ~40:60 6B:6D repeat units (Oliver *et al.,* 2013). Therefore, if NML SC12-0507 is actually a 6G, it is possible that Quellung previously characterized it as a 6D, if 6D repeat units dominated, while also containing a *wciN*α that we detected with PCR. To confirm this hypothesis, it would be necessary to sequence *wciN*α to look for an A150T mutation. Finally, it is possible that this isolate represents a different serogroup 6 variant, as many have recently been reported (Burton *et al.,* 2016; Ko *et al.,* 2013; Oliver *et al.,* 2013; Park *et al.,* 2015; Sheppard *et al.,* 2010).

Similarly to the case for 6G, Oliver *et al.* (2013) also characterized the new variant, serotype 6F. Serotype 6F is also the product of a *wciN*α' mutant, but instead is present in a serotype 6A *cps* locus, and is therefore a hybrid of serotype 6A and 6C (Oliver *et al.,* 2013). A hybrid serotype of serotypes 6A and 6B was also discovered, and found to be the result of one of two different combinations of three mutations in *wciP* (Sheppard *et al.,* 2010). Another group subsequently mutated a serotype 6A isolate using one of those two combinations and generated the 6A/6B hybrid serotype, and proposed it be called serotype 6H (Park *et al.,* 2015). Finally, in 2011, Elberse *et al.* described isolates characterized by Quellung as serotype 6B, but differed by >5% in the

cps and proposed that they were a different serotype which they named serotype 6E, that was cross-reacting with 6B antisera. Another group has since shown that the capsule polysaccharide of these isolates is identical to that of serotype 6B and propose these isolates be considered a genetic variant of 6B instead, described as 6B with a class 2 *cps* locus (Burton *et al.,* 2016). It is clear that much variation with the *cps* locus of serogroup 6 is possible, and serial mutations and recombination events have been documented (Yun *et al.,* 2014).

Finally, a careful look at the raw data for Quellung serotyping may be warranted. Of the strains from serogroup 6 that were initially discordant during PneumoCaT validation, most were due to mis-reporting (Kapitai *et al*., 2016). Since serotype 6A reacts with factor sera 6b, serotype 6B with factor sera 6c, and serotype 6C with factor sera 6d, transcription errors could occur. While electronic data provided into PneumoCat could be reported automatically into a laboratory information system to avoid transcriptional errors, manual transcription like the analyses performed in this study should be done with caution.

It is unclear whether our strains with discrepant results represents the recently reported serogroup 6 variants, a new unreported variant, or is simply misclassified due to transcription errors or incorrect Quellung reactions. Further analyses should be conducted to evaluate any one of these possibilities.

5.2.2. PCR and Sequencing-Based Discrimination of Serotypes 7F and 7A

As previously noted, PCR and sequence-based discrimination of serotypes 7F and 7A relies on a frameshift thymine (T) insertion at position 587 of *wcwD* in serotype 7A (Kapatai *et al.,* 2016; Mavroidi *et al.,* 2007). This frameshift is thought to result in loss of function of WcwD glycosyltransferase, which leads to loss of a side branch in 7A repeat units. While the internal control (*wzy*) suggested all strains used were either 7F or 7A, and no cross-reactions were seen for either *wzy* or *wcwD,* there were unexpected

sequence results of *wcwD* in strains previously characterized as 7A by Quellung reaction. In contrast to strains of *S. pneumoniae* previously serotyped as 7F that were all were concordant with the expected *wcwD* sequences, only the results for reference strain of *S. pneumoniae* serotype 7A was concordant with Quellung serotyping.

The strains with discrepant serotype 7A results were re-tested using Quellung serotyping at the NML. Of, 13 strains initially identified as serotype 7A, 10 were resolved as serotype 7F. Interestingly, these strains had all previously been typed in one laboratory (TIBDN/CSBN isolates are all processed at Mount Sinai hospital). Upon careful review of the CSBN and TIBDN Quellung serotyping results, one of the 10 strains now resolved as 7F had an antisera reaction suggestive of serotype 7F (pool C+P+, serogroup 7+, and factors 7b+ and 7c-), in contrast to the expected result for serotype 7A (pool C+P+, serogroup 7+, and factors 7b+ and 7c+) (Table S1). As such, this result was clearly misinterpreted as serotype 7A. The remaining nine results now resolved as serotype 7F had Quellung reactions consistent with serotype 7F. Given the NML results, the factor 7c antisera result in these nine strains were likely false positive when tested in by CSBN or TIBDN (Table S1). It should also be noted that testing at TIBDN and CSBN was incomplete, as factors 7e or 7f were not done.

While 10 strains of serotype 7A were now resolved as 7F, an additional three strains previously identified as serotype 7A (by the NML) repeated the same. For these three isolates, it is possible that these strains also had false positive reactions for Quellung factor 7c; however, an alternative explanation could be that there are alternative mutations in *wcwD* or other genes that lead to a 7A phenotype. In fact, during the PneumoCaT validation, only 22.2% (2/9) of the serotype 7A isolates carried the frameshift insertion in *wcwD* sequence, and these two displayed a mixed profile at residue 587 (Kapatai *et al*., 2016). Even in the two where this mutation was observed, only 60-70% of the reads carried the insertion. This suggested that 77.8% (7/9) contained a wild-type *wcwD* sequence consistent with serotype 7F, and the 587

mutation may be prone to reversion. DNA from the strains with discrepant results in our study should be re-extracted and retested with NGS for the mutation at *wcwD* position 587, investigated for additional mutations in *wcwD*, as well as additional mutations in the *cps* loci.

5.2.3. The PCR and Sequencing-Based Discrimination of Serotypes 9V and 9A

The *cps* loci within serogroup 9 are highly similar; serotypes 9V and 9A both contain additional acetyltransferases (*wcjE* and *wcjD*) and the *wzy* target used in cmPCR is specific for 9V/A (Kapatai *et al.,* 2016). For differentiation between serotypes 9V and 9A, this study investigated a mutation in *wcjE*, which leads to the differences in O-acetylation in the capsule structure of serotype 9A (Bentley *et al.,* 2006; Calix *et al.,* 2011b and 2012b). The mutation targeted in this study was a frameshift deletion of guanine (G) at position 722 in *wcjE* in serotype 9A, and is the same target used by Kapatai *et al.* (2016). An internal control was also added to the PCR reaction. Initially, *wcjD* was used, but due to a significant reduction in band intensity upon duplexing with *wcjE*, an alternative internal control (*wzy*) was chosen (da Gloria Carvalho *et al.,* 2010). No band intensity differences in *wcjE* were noted with the *wzy* target in the duplex PCR.

When the *wzy*/*wcjE* duplex PCR reactions were assessed and the expected targets were amplified, no cross-reactions were seen with other *S. pneumoniae* serotypes or other streptococci. However, a total of 12 discordant results were observed between the previously obtained Quellung serotypes.

First, three discordant results each for strains previously serotyped as 9A and 9V by Quellung reaction did not amplify either the internal control or the *wcjE* target. Repeating Quellung would be helpful, as would checking the pneumococcal DNA concentration using quantitative *lytA* real-time PCR (Lang *et al*., 2015). If Quellung reactions repeat the same results, and the DNA concentration is sufficiently high, it

would be prudent to check sequence the entire *cps* loci to ensure there are no issues with the internal control, as it is used in the CDC cmPCR reactions (Lang *et al*., 2015).

In a second set of discordant results in this study, three *S. pneumoniae* isolates previously serotyped as 9A by Quellung reaction had a positive internal control but *wcjE* was not detected. Similarly, the internal control was positive and *wcjE* was not detected for three isolates previously serotyped as 9V. There are at least three possibilities for these discordant results. One possibility is a false negative result for Quellung factor 9g for isolates previously typed as 9A, or a false positive result for Quellung factor 9g for isolates previously typed as 9V (Table S1). To help investigate this possibility, Quellung reactions should be repeated at the NML for all six of the isolates. It should be noted that during the PneumoCaT validation, there were initially 6 isolates of serotype 9A, but only one showed concordance between the molecular and Quellung serotype (Kapatai *et al*., 2016). Upon repeat testing with Quellung serotyping, four were resolved as serotype 9V (which were added to the other 45 concordant serotype 9V results). This data suggests that repeat testing using Quellung reaction can help resolve discordant serotypes.

Another possibility for the discrepant results is that other genetic differences in *wcjE* are responsible for the phenotypic changes (Figure 32). Calix *et al.* (2011b) proved that inactivation of *wcjE* of in a serotype 9V strain resulted in expression of a serotype 9A phenotype; and there were various mechanisms for *wcjE* inactivation were described in clinical isolates of *S. pneumoniae* serotype 9A. These included deletions, insertions, duplications, or point mutations (Figure 32). In some cases, some of these 9A isolates expressed small amounts of 9V epitopes (Calix *et al.,* 2011b). The possibility of one serotype expressing variable amounts of two different capsule repeat units, and the possibility of multiple different mutations contributing to a given serotype poses issues for both phenotypic and genotypic serotyping techniques. To look for large insertions in *wcjE*, such as a phage, the elongation time during PCR could be extended. However, it

would be prudent to sequence the entire *cps* loci for isolates where the internal control is present but *wcjE* fails to amplify.

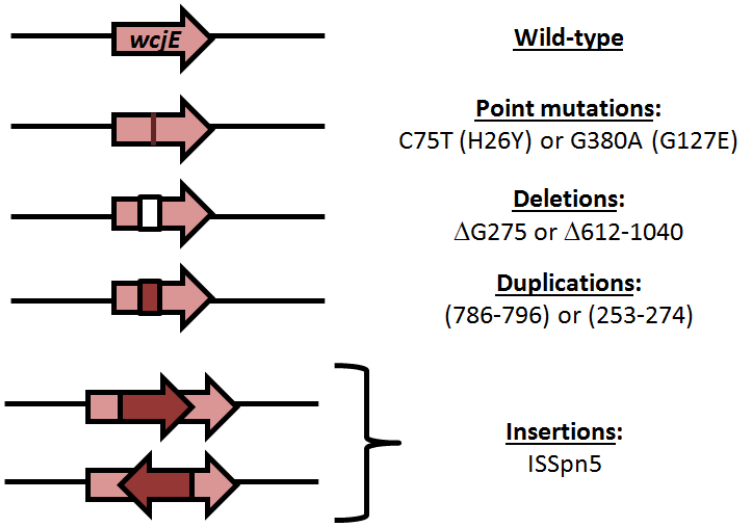


Figure 32. Possible mechanisms of *wcjE* inactivation leading to a serotype 9A phenotype. This diagram is inspired form data presented in Calix *et al*., 2011b.

5.2.4. The PCR and Sequencing-Based Discrimination of Serotypes 9N and 9L

The capsules of serotypes 9N and 9L differ only in the presence of glucose or galactose present in the capsule repeating unit, respectively (Kapatai *et al*., 2016; Mavroidi *et al*., 2007). As this capsular change is thought to occur through the activity of the glycosyltransferases WcjA and WcjB, molecular discrimination of serotype 9N and 9L relies on SNPs in two glycosyltransferase genes, *wcjA* and *wcjB*, as well as in the *wzy* and *wchA* genes. While sequencing results were concordant for all 17 isolates of serotype 9N and 12 isolates of serotype 9L, there were 9 discordant results in isolates previously characterized as serotype 9L. However, in these 9 isolates with discordant results, all 10 differentiation mutations across the four gene targets were suggestive that these were all serotype 9N. This likely represents false negative results for Quellung factor 9e (Table S1). It should be noted that all discordant results were previously serotyped in one laboratory (TIBDN/CSBN). As such, Quellung reactions should be repeated at the NML.

In the Kapatai *et al*. (2016) study, concordant results were noted for 43 isolates of serotype 9N, but only one (of two) isolates of serotype 9L showed the expected SNP profile. The other was described as a novel variant in serogroup 9, with sequences that were a mixture between the expected results for serotypes 9N and 9L. This novel 9 serotype showed Quellung reactions consistent with serotype 9L and a weak reaction for factor sera 9 g, which is usually only positive for serotype 9V (Table S1). While no cross-reactivity was noted in the specificity analysis, the possibility of serogroup 9 variants suggests that it would be prudent to check each isolate of serotype 9N and 9L with Quellung and PCR/sequencing reactions for serotypes 9V and 9A, and vice versa.

5.2.5. The PCR and Sequencing-Based Discrimination of Serotypes 11A and 11D

The *cps loci* of serotypes within the serogroup 11 are very similar, except for a few differences in acetyltransferase genes and mutations within a gene encoding a glycosyltransferase. The acetyltransferase genes *wcwC* and *wcjE* present in serotypes 11A, 11D and 11F can be used for differentiation of these three from serotypes 11B and 11C, which possess the unique acetyltransferase gene, *wcwR* (Mavroidi *et al.,* 2007). The *cps* loci of 11A, 11D, and 11F can further be differentiated using sequencing of a glycosyltransferase gene*, wcrL*, since mutation at nucleotide position 334 are serotype-specific (Bentley *et al*., 2006; Kapatai *et al.,* 2016; Mavroidi *et al.,* 2007; Oliver *et al.,* 2013). Mutations in *wcrL* have been shown to donor sugar specificity for WcrL, which is Glc*p* in serotype 11A compared to Glc*p*NAc in serotypes 11F, 11B, and 11C. Site-directed mutagenesis of *wcrL* at position 334, changing codon AAT to ACT has been shown sufficient to convert a serotype 11A to 11D (Oliver *et al.,* 2013). In this study, a multiplex PCR was developed for amplification of either *wcwC*/*wcrL* or *wcwR*. Overall, all *wcrL* sequencing results for serotypes 11A and 11D were concordant with those previously obtained by Quellung serotyping, except two. Isolates TIBDN 21411 and CARA GZ-11D were previously identified as 11D by Quellung, but were identified as serotype 11A by *wcrL* sequencing. It should be noted that no isolates of serotype 11D had be previously evaluated by PneumoCaT, only 46 isolates of serotype 11A (Kapatai *et al*., 2016).

Several possibilities could explain the two discrepancies noted in this study. The discordant results could be due to false positive reactions for Quellung factor 11f as well as false negative factor 11c (Table S1) The WcrL glycosyltransferase in serotype 11D was shown to be bifunctional, and 75% of the capsular polysaccharide repeating unit in serotype 11D is identical to 11A, possibly explaining cross-reactions between these two serotypes (Oliver *et al.,* 2013). It is also possible that additional genetic difference can lead to differences between serotypes 11A and 11D. Serotype 11A has also been associated with two distinct *cps* alleles: 11A-1 and 11A-2 (Calix et al., 2013). 11A-1 corresponds to the *cps* allele recovered in the majority of the 11A strains reported to date, and is identical to the 11D *cps* locus described by Bentley *et al*. (2006) with the exception of the *wcrL* glycosyltransferase gene single nucleotide polymorphism (Calix *et al*., 2011b; Camilli *et al*., 2011; Hiller *et al.*, 2007); 11A-2 corresponds to the *cps* allele reported for serotype 11A by Bentley *et al*. (2006), but is only found in a minority of serotype 11A isolates (Calix *et al*., 2013). The differences in genetic lineages of serotype 11A would suggest that the single nucleotide polymorphism (SNP) in the *wcrL* gene may not be sufficiently discriminatory, as previously thought by Kapatai *et al*. (2016).

To further complicate the story within serogroup 11, a new serotype named 11E was identified among isolates previously characterized as serotype 11A (Calix *et al*., 2010). Serotype 11E is described as containing a serotype 11A *cps* locus with a mutated or disrupted *wcjE*, through distinct mutations leading to a truncated protein, which include transposable element insertions, nucleotide duplications, or missense mutations. These mutations lead to modifications of the capsular polysaccharide structure, which differs from serotype 11A by total or partial absence of acetylation at a defined position: the C_6 position of the β-Gal sugar in the capsular repeating unit (Camilli *et al*., 2014; Zartler et al., 2009). In some cases, the mutated WcjE O-acetyltransferase is partially functional, and the strains are designated 11A serovariants (11Av) (Calix *et al*., 2010). The current literature suggests that serotypes 11A and 11E

represent two extremes of an antigenic spectrum, and 11Av is an intermediate between the two. These three all expresses a variable degree of O-acetylation due to variation in WcjE activity (Camilli *et al*., 2014).

Given all these phenotypic and genetic variations, the Quellung serotyping should be repeated at the NML, additional genetic analyses should be undertaken such as sequencing of the entire *wcrL* and *wcjE* genes, or even the entire *cps* loci. To expand our knowledge of the spectrum of serotype 11 variants, assessment of the capsular sugar composition should also be considered (Camilli *et al*., 2014; Oliver *et al*., 2013).

5.2.6. PCR and Sequencing-Based Discrimination of Serotypes 12F, 12A, 12B, 44 and 46

As seen with serotypes 9N and 9L, serotypes within PCR group 12F/A/B/44/46 were discriminated using various SNPs in three genes: *wcxD, wcxF*, and *wzy*. While functions have been described for these genes, inferences of genetic signatures to capsule phenotypes are limited since only the structures of serotypes 12F and 12A are known (Geno *et al*., 2015; Kapatai *et al.,* 2016; Mavroidi *et al.,* 2007). In this study, the sequencing results were concordant for 1/8 isolates of serotype 12A; 2/7 isolates of serotype 12B, 7/13 isolates of serotype 12F, and all isolates of serotype 44 (n=2) and 46 (n=3). However, many discordant results were noted for serotype 12A (7/8), 5/7 isolates of serotype 12B, and 6/12 isolates of serotype 12F.

For the discordant results obtained for isolates previously serotyped as serotype 12A, there were several sequence profiles that were inconsistent with this result. First, a single isolate previously serotyped as 12A by Quellung had sequence results consistent with serotype 12B. It is possible that the isolate is actually a serotype 12B and the discrepancy is due to false negative reactions for sera factor 12b and 12e (Table S1). Repeating the Quellung reaction at the NML would be the first step to resolve this discrepant result.

Another six isolates previously serotyped as serotype 12A by Quellung produced sequencing results resembling characteristics of both serotypes 12A and 46. Whether these strains represent sequence variants that arose from recombination events or perhaps an evolutionary ancestor remains to be elucidated. It is also possible that the mutations published by Kapatai *et al.* (2016) are not specific for serotype 46. These results are not likely to be explained by Quellung error alone, as there are many differences in the Quellung reactions for serotypes 12A and 46. A Quellung positive for 12A would be the result of positive pool E and R, group 12, and factor sera 12b, 12c and 12e. For serotype 46, Quellung reactions would be the result of positive pool I and type 46. To investigate the discrepant results, Quellung reactions at the NML should still be performed, as well as sequencing the entire *cps* loci to look for recombination events or other genetic clues. It may also be interesting to determine the genetic lineage for these serotypes, by using techniques such as NGS or multilocus sequence typing (MLST) (Deng *et al.,* 2016). It should be noted that all these isolates did not follow the recommended sera algorithm for Quellung serotyping, as factor 12e was not used, which could lead to serotype misclassification.

Finally, five isolates previously serotyped as 12B had sequence results characteristic of serotype 12F, and six isolates previously serotyped as 12F had sequence results characteristic of 12B. From the sequence-based discrimination described in Kapatai *et al*., (2016) that was also used in this study, only one codon at nucleotide position 512 of *wcxF* differentiates these two serotypes: the ACG codon of serotype 12B represents a threonine at amino acid position 170, whereas the ATG encodes methionine. During Quellung serotyping, both serotypes 12B and 12F are positive for factor sera 12b, but differences between the two comes from positive or negative results for factor 12c and 12e, respectively (Table S1). In the Kapatai *et al*. (2016) study, 24 of 69 serogroup 12 results were shown to be either a 12B to 12F or 12F to 12B discordance. Since all three serogroup 12 factor sera (12b, 12c, and 12e) are required for identification of serotype 12B, it is possible that auto-agglutination could

lead to incorrect scoring. In addition, weak reactions may be missed, and the serotype may be misclassified as 12F. While it is possible that each of these isolates had false reactions, it is also possible that the single mutation used for serotype discrimination is not specific. Quellung serotyping should be confirmed by the NML, and perhaps, the *cps* loci for serotypes 12B and 12F should also be investigated for additional mutations. Finally, additional isolates of serotypes 44 and 46 should also be tested as none were used in the PneumoCaT validation (Kapatai *et al*., 2016).

5.2.7. The PCR and Sequencing-Based Discrimination of Serotypes 15B and 15C

The capsules of serotypes 15B and 15C differ only in the presence or absence of O-acetylation (Jansson *et al.,* 1987). The genetic basis for this capsular difference was determined to be the result of a variable number of TA repeats in serotype 15C compared to 15B around position 413 of *wciZ* (van Selm *et al.,* 2003). Serotype 15B has eight TA repeats; serotype 15C have seven or nine TA repeats, leading to a frameshift that disrupts the open reading frame (ORF) of *wciZ* and results in a premature stop in translation in the WciZ O-acetyltransferase (Bentley *et al.,* 2006; van Selm *et al.,* 2003). The PCR and sequencing-based discrimination for serotypes 15B and 15C in this study took advantage of this genetic difference. Upon sequencing of *wciZ*, the number of TA repeats for serotype 15B was consistent with those seen in other studies (Kapatai *et al.,* 2016; van Selm *et al.,* 2003). In contrast to the expected TA insertion seen in 15 isolates of the 15C serotypes, one strain had a "TATA" insertion at the same position, and three strains simply had an "A" insertion. All these mutations would lead to a frameshift, and would thus likely have a similar detrimental effect on WciZ-mediated acetylation.

The sequence variations in the TA-rich region in *wciZ* are thought to explain the reversible serotype switching previously documented between serotypes 15B and 15C (van Selm *et al.,* 2003; Venkateswaran *et al.,* 1983). Short direct repeats can allow for pausing and dissociation of DNA polymerase during DNA replication, which could results in slipped-strand mispairing, which eventually translates into a frameshift and protein

truncation (Viguera *et al.,* 2001). Serotype switching from 15B to 15C may allow a mechanism for immune evasion, as O-acetyl groups have been documented to increase immunogenicity of bacterial sugars (Berry *et al.,* 2002). It is important to note that serotype switching may also interfere or be responsible for specificity in serologic typing techniques such as Quellung reaction. Although this study did not encounter discrepant results between sequencing results and serotype 15B and 15C determined by Quellung reaction, Kapatai *et al.* (2016) noted numerous some discrepant results were attributed to errors in record keeping. As noted for serogroup 6, the nomenclature of the factor sera can lead to confusion, as it can be very similar to the nomenclature for serotypes. For example, serotype 15B reacts with 15c factor sera, and Kapatai *et al*. (2016) reported several discordant cases where serotype 15B isolates had been originally recorded as 15C due to positive reaction with factor 15c.

Another interesting observation by Kapatai *et al*. (2016) was the presence of mixed 15B/C serotypes. In fact, during the PneumoCaT validation, there were 42 serotype 15B reported, and 26 serotype 15B/C, but only one serotype 15C (with a different frameshift mutation in *wciZ* which was not described) (Kapatai *et al*., 2016). These authors believed the isolates with discrepant results for serotypes 15B or 15C were due to the dominance of one serotype repeat unit over the other. It would be interesting to verify whether these mixed 15B/C serotypes could also be detected using NGS on our serotype 15C isolates, or whether these were observations were strain-specific. However, with the possibility of reversible switching between the 15B and 15C serotypes, it may not be possible to clearly distinguish between these two serotypes in the laboratory. In many publications, failure to discriminate serotype 15B from 15C is "solved" by naming these isolates serotype 15B/C (Dube *et al*., 2015; Kapatai *et al*., 2016; Laufer *et al*., 2010). This answer does not provide a solution for discrimination of the vaccine-preventable serotype 15B, and supports looking for additional mutations in *wciZ* or within other genes in the *cps* loci of serotypes 15B and 15C to help resolve isolates that don't harbour the typical *wciZ* genotype/capsule phenotype linkage.

5.2.8. The PCR and Sequencing-Based Discrimination of Serotypes 18C, 18F, 18A and 18B

The PCR and sequencing-based discrimination of serotype within serogroup 18 relies on differences in three acetyltransferase genes. Serotypes 18F and 18A can be discerned by PCR as the *cps* locus of serotype 18F contains an additional acetyltransferase gene (*wcxM*), and serotype 18A lacks the acetyltranferase gene *wciX* (Kapatai *et al.,* 2016; Mavroidi *et al.,* 2007). Differentiation of serotypes 18C and 18B required sequencing. More specifically, a G to T substitution at nucleotide position 168 in *wciX* leads to an early stop codon in serotype 18B (Kapatai *et al.,* 2016; Mavroidi *et al.,* 2007). This study designed a multiplex PCR with targets for serotype 18F (*wcxM*), for serotype 18A (the genes flanking *wciX: wzy* and *wciY*), and for serotype 18B and 18C (*wzy*/*wciX*). While the references strains and a number of other strains displayed expected results, many discordant results were noted with previous Quellung reaction results including one strain of serotype 18F, two serotype 18A, and 14 strains of serotype 18B.

First, the CDC strain 4311 was previously characterized as serotype 18F, but this is the only member of this serotype that was discordant. In contrast to the expected results, *wcxM* was absent in this strain and the *wzy*/*wciY* PCR was positive (suggesting *wciX* was absent). These two genetic signatures are consistent with serotype 18A, not 18F. During Quellung reactions, serotypes 18F and 18A share positive reactions for pool A and Q, as well as group 18; however, the factor sera reactions are opposite (Table S1). For serotype 18F, factors 18c, 18e, and 18f are positive and factor 18d is negative. For serotype 18A, factors 18C, 18e, and 18f are negative and factor 18d is positive. If the CDC strain 4311 is truly a serotype 18A as suggested by the genetic profile, this would suggest there was either a transcription error during recording of the results, or there were three false positive Quellung reactions (factors 18c, 18e, and 18f) and one false negative reaction (factor 18d). Quellung serotyping should be repeated at the NML, as

well as the PCR reactions from this study. If the results repeat the same, the entire *cps* locus should be sequenced.

The second set of discrepant results was CBSN strains 30969 and 30377, which were previously serotyped as 18A by Quellung reaction. Genetically, these were consistent with serotype 18C. Neither of these isolates had *wcxM* and both had the *wciX* gene; therefore, were not consistent with serotype 18F or 18A. Additionally, there was no premature stop in *wciX*, suggesting these strains were serotype 18C, not 18B. Upon careful review of the CSBN Quellung reaction results, antisera pools A and Q were positive, as was serogroup 18; however, factors 18c, 18e, and 18f were negative. While this is consistent with serotype 18B, factor 18d was not tested (which should have been positive for 18B). As none of factor serum was positive, one could argue that the serotypes for these two strains were not confirmed. For serotype 18C (as deduced genetically), factors 18c and 18e should have been positive. It is possible that two false negatives occurred (for factors 18c and 18e) during Quellung serotyping at CSBN. Both strains should be retested by Quellung reaction at the NML, as well retested by PCR and sequencing.

The third set of discrepant results was by far the most predominant. Fourteen strains previously characterized as serotype 18B by Quellung reaction were deduced as serotype 18C by molecular typing using PCR and sequencing. While both these serotypes are highly similar in phenotypic and genetic characteristics, it is not surprising to see this group of discordant results being predominant. With no *wcxM*, and *wciX* present with no premature stop, all fourteen strains were genetically similar to serotype 18C, not 18B. For the Quellung reaction, only a positive factor serum 18c differentiates serotype 18C from 18B. It is possible that false negative reactions for factor 18c occurred for all 14 strains, but this would need to be confirmed by retesting Quellung at the NML. A careful re-analysis of *wciX* mutations (or other mutations) should also be considered.

It should be noted that during the PneumoCaT validation, there three of the persistent discordant results involving isolates previously characterized as serotype 18C by Quellung: two were assigned as serotype 18B by PneumoCaT, and one as serotype 18A (Kapatai *et al*., 2016). Following retesting by both molecular and serological methods for serotyping, the two 18B/18C discordant results persisted, whereas the third serotype 18A isolate was reclassified as serotyped 18B (which was concordant with the NGS sequence results). These investigators reported that the 18B/C discordant results are not unexpected, as only factor sera 18c differentiates serotype 18B from 18C. They also suggested that a weak reaction or auto-agglutination could lead to these false predictions. This possibility could apply to our 18B/C discordant results. For the 18B/A discrepant result, it was suggested to look at the sequences of the UDP-pyranase mutases gene (*glf*), which could help discriminate these serotypes (Bentley *et al*., 2006; Kapatai *et al*., 2016). This target might also help explain discordant results between phenotype and the genotypic markers assessed in this study.

5.2.9. Sequence Informed, PCR-Based discrimination of Serotypes 22F and 22A

The *cps* loci of serotypes 22F and 22A described by Bentley *et al*. (2006), and the sequences submitted to NCBI by these investigators, differs only one nucleotide in the *wcwA* gene. This study had originally designed primers targeting a 399 bp region flanking this mutation, which spanned the glycosyl-transferase gene *wcwA* and the O-acetyltransferase gene *wcwC.* Interestingly, this target was only amplified in isolates of serotype 22A. At the time, the reason for the negative PCR amplification with isolates of serotype 22F was not understood. In a recent comparative genomics study by Kapatai *et al*., (2016), the explanation was revealed. All of their serotype 22A sequences were identical to that found in NCBI; however, the sequences for serotype 22F differed markedly in genes *wcwA* and *wcwC*. Kapatai *et al*. (2016) suggested this was likely due to an incorrect submission for serotype 22F sequence by Bentley *et al*. (2006). Using a multiplex reaction with two targets ($wcwA_{22F}$ and $wcwC_{22F}$) designed based on

sequences for the serotype 22F strains described by Kapatai *et al*. (2016), along with the 399 bp target specific for 22A (*wcwA/C*$_{22F}$) and an internal control (*wcwV)* (Lang *et al.,* 2017), all 20 strains of serotypes of 22F and 10 strains of serotype 22A were accurately discriminated, and showed no cross-reactions with 32 other streptococci or the other 90 *S. pneumoniae* serotypes. It was surprising that it took 10 years to note this error in NCBI. Future direction could include experiments to prove the *wcwA* and *wcwC* are responsible for the phenotype differences between serotypes 22F and 22A. To date, it is unknown how these genes affect the capsule of these serotypes, as the structure for 22A has yet to be determined.

5.2.10. PCR and Sequencing-Based Discrimination of Serotypes 33F, 33A and 37

The *cps* loci of serotypes 33F and 33A differ by a thymine (T) insertion at position 433 of the *wcjE* gene in serotype 33F, resulting in a frameshift and inactivation of the O-acetyltransferase (Kapatai *et al.,* 2016). Loss of function of the WcjE likely explains the difference in acetylation noted between the capsules of serotypes 33F and 33A (Spencer *et al.,* 2017). Serotype 37, on the other hand, has a complete but non-functional 33F *cps* locus (Bentley *et al.,* 2006). While the 33F *cps* loci genes are present (and can be amplified by PCR), its capsule is generated solely by the *tts* gene, which is located outside the *cps* locus and encodes a synthase-dependent transferase (Llull *et al.,* 1999). In a triplex PCR reaction containing primer targeting *wcjE*, *tts*, and an internal control (*wzy*), no cross-reactions or discordant results were observed between the serotypes determined by Quellung reaction. However, unlike serotype 33A, the *wcjE* sequence of all 33F isolates were not amplified with the *wcjE* primer pair. This result was surprising since only a point mutation in *wcjE* was expected to differentiate serotypes 33A and 33F (Bentley *et al*., 2006; Kapatai *et al.,* 2016; Mavroidi *et al*., 2007). While this study was able to accurately discriminate serotypes 33F, 33A, and 37 using PCR alone, sequencing was also performed in the *wcjE* amplicon for serotype 33A strains. All were consistent with the expected wild-type sequences described previously (Bentley *et al*., 2006). Further analyses should investigate why the serotype 33F strains did not amplify the

wcjE gene. Sequencing of *wcjE* gene by designing primer flanking this gene, or sequencing the entire *cps* loci would be useful. It should be noted that loss of function of WcjE, whether through absence of *wcjE*, or any inactivating mutations should lead to the same phenotype, as WcjE acetylation is only present in serotype 33A. On the other hand, serotype 33A has two membrane-bound O-acetyltransferase genes, *wciG* and *wcjE* (Spencer *et al*., 2017). *In vitro* generation of *wciG*-deficient variants in 33A and 33F, designated 33X1 (Δ*wciG*) and 33X2 (Δ*wciG*Δ*wcjE*), showed that WciG-mediated (not WcjE-mediated) O-acetylation is important for producing protective capsules in 33A (Spencer *et al*., 2017). While no *wciG*-deficient variants have been reported in clinical isolates to date, it would be interesting to sequence *wciG* in serotypes 33A and 33F to help understand acetylation and immune escape. During the PneumoCaT validation, 48 and 22 results were concordant for serotypes 33F and 37, respectively, but no serotype 33A were assessed (Kapatai *et al*., 2016). Initially, five serotype 33A were evaluated, but upon retesting three were resolved a serotype 33F, and two remained non-typeable. Failure to validate any serotype 33A strains questions the validity of the *wcjE*-based sequencing approach; however, our study showed concordant results for six clinical isolates. In addition, the Kapatai *et al*. (2016) did not comment on other possible mutation in *wcjE* for serotype 33F, nor its absence in any of the 48 serotype 33F strains evaluated.

5.3. Overall Discussion

Quellung reactions have long been the gold standard for *S. pneumoniae* serotyping, but in recent years, its accuracy and continued use for surveillance has been questioned (Kapatai *et al*., 2016; Siira *et al.,* 2012). The numerous limitations of Quellung reaction have prompted researchers to develop new technologies for *S. pneumonaie* serotyping, and facilitate large-scale epidemiological studies. Recent advances in serological methods have focused on moving away from the more labour intensive Quellung reaction and agglutination reactions, to technologies more amenable to automation like EIA and multiplexed serologic assays coupled to flow cytometry.

Molecular methods have also evolved to facilitate testing. Additionally they have tried to provide more accurate results by identifying which molecular signature leads to capsular switching events (Calix *et al.,* 2011a and 2012a; Kapatai *et al.,* 2016; Oliver *et al.,* 2013; Spencer *et al*., 2017; van Selm *et al.,* 2003). This study has taken cmPCR and sequencing one step closer to being applied in any molecular laboratory, by tackling the lack of serotype discrimination in current cmPCR reactions. In contrast to our initial strategy to target sequences outside the *cps* loci (which showed cross reactivity and poor reproducibility among serotypes), this study showed promising avenues of research for the molecular detection and discrimination of *S. pneumoniae* serotypes by targeting sequences within the *cps* loci.

Using molecular methods that can detect specific genetic signatures that are linked to defined capsular changes is crucial to define serotypes and understand the epidemiology of *S. pneumoniae*. Interestingly, this study noted a correlation between the highest number of discordant results between Quellung and sequencing for serotypes within PCR groups that required a large number of mutations to differentiate them. For example, serotypes within the PCR group 9N/L and 12F/A/B/44/46 had high numbers of discordant results, and each used at least four genetic targets for serotype discrimination. This data highlights that a PCR and sequencing based approach focusing on many mutations may be impractical, and that certain mutations may not be correlated to a given serotype. While identifying which gene mutation(s) leads to capsule difference is crucial to define a serotype, not all capsule structures are known (Geno *et al*. 2015). Even if assumptions are made based on genetic differences between serotypes, data supporting the genetic-phenotype link have only been studied in a limited number of isolates (Kapatai *et al.,* 2016). This study showed that some mutations targeted by PneumoCaT might not be accurate, or sufficient. In fact, it is important that the methods account for all possible genetic variants of a given serotype. In turn, each genetic variant should be confirmed by genetic analyses to be responsible for a given capsule structure change. For example, serotype 9A has been shown to arise from

serotype 9V through various different mutations (Calix *et al.,* 2011b), phenotype reversions have been noted between serotypes 15B and 15C (Kapatai *et al*., 2016), and new variants have emerged in serogroups 6, 9, and 11 (Burton *et al.,* 2016; Calix *et al.,* 2010, 2011b and 2013; Camilli *et al*., 2014; Ko *et al.,* 2013; Oliver *et al.,* 2013; Park *et al.,* 2015; Sheppard *et al.,* 2010). PneumoCaT would only capture a small subset of these variants, and these variants might explain some of the discrepant results noted in this study. Additional targets may be required. Regardless of the molecular method used for *S. pneumoniae* serotype deduction, it will be important to continuously monitor and update genetic variants signatures that arise from human immune selective pressures or from interactions with other microorganisms like the viridans group streptococci in anatomical sites such as the human nasopharynx (Croucher *et al.,* 2011).

One advantage to using PCR in combination with Sanger sequencing as opposed to NGS, is the potential to apply it directly to clinical specimens such as nasopharyngeal swabs. While the approach would require validation, it is important to highlight that this study did not observe any cross-reactions between PCR targets and other *Streptococcus* species. Similar successes were demonstrated for the PCR-based screening methods for *S. pneumoniae* (*lytA* and *cpsA* real-time PCR) and serotype deduction using cmPCR (Da Gloria Carvalho *et al*. 2010; Gillis *et al*., 2017; Lang *et al*., 2015 and 2017). As such, an algorithm could be evaluated using *lytA*/*cpsA* real-time PCR for *S. pneumoniae* detection in clinical specimens, following by serotype deduction using cmPCR, and serotype discrimination using PCR and Sanger sequencing when needed. Validation would be crucial as commensals in the oro- and nasopharynx are known to exchange genetic material with *S. pneumoniae*, and this plays an important role in their evolution (Coffey *et al*., 1993; Dowson *et al*., 1993; Kilian *et al*., 2014; Sauerbier *et al*., 2012). While it was previously assumed that a hallmark of *S. pneumoniae* virulence compared to other streptococci was its capsule, some closely related streptococci such as *S. mitis* and *S. oralis* found in the oro- and nasopharynx have been shown to contain capsule genes with high identity to those in *S. pneumoniae* (Sorsensen *et al*., 2016). For example, one

isolate of *S. mitis* was shown to contain an entire serotype 19F *cps* locus (Sorensen *et al*., 2016). The possibility of capsule genes in other organisms found in the same anatomical niches as *S. pneumoniae* could pose a challenge for software like PneumoCaT, which compares the *cps* genes sequences to a reference genome. Using a defined detection and serotyping algorithm for *S. pneumoniae* as proposed above could avert these challenges. In this study (as for PneumoCaT), the method validations began with cultured isolates of *S. pneumoniae*, but once the discrepant results are resolved, could be applied directly on clinical specimens for colonization studies (naso- or oropharyngeal swabs) or in defined cases with pneumococcal disease (i.e. respiratory secretions, blood, cerebral spinal fluids, or pleural fluids). For clarity, the intent of the molecular tools for *S. pneumoniae* serotyping would not be for clinical diagnosis but an aid to support pneumococcal disease surveillance and serotype epidemiology, which could change with implementation of new pneumococcal vaccines or vaccine recommendations (LeBlanc *et al*., 2017).

Capsular switching (or serotype switching) is the result of homologous recombination that partially or completely replaces a homologous region of the *cps* locus from another *S. pneumoniae* serotype or another *Streptococcus* species (Croucher *et al*., 2013; Salter *et al*., 2012). While reports of capsular switching predate pneumococcal vaccines and the use of antibiotics (Wyres *et al*., 2013), the primary concern regarding capsular switching is a vaccine-type (VT) to non-vaccine-type (NVT) conversion, which would result in vaccine failure (Andam and Hanage, 2015). Significantly higher rates of capsular switching seem to occur between closely related serotypes within a serogroup (Croucher *et al*., 2015). For serotypes applicable to this study, capsular switches have been reported for between serotypes 9V and 9A, 11A and 11D, 15B and 15C, 18B and 18C, and 33F and 33A (Calix *et al*., 2011; Croucher *et al*., 2015; Kapatai *et al*., 2016; Mavroidi *et al*., 2007; Oliver *et al*., 2013; van Selm *et al*., 2003). It has been proposed that capsule switching may allow some serotypes to escape the immune system. For example, serotype switching in serogroups 9 and 11 involves

mutations in *wcjE* (Calix *et al*., 2011 and 2013), which leads to a loss of O-acetate groups that have been shown to be epitopes for neutralizing antibodies (Berry et al., 2012). Monitoring the evolutionary trends in *S. pneumoniae* serotype switching will be critical for implementing successful novel vaccines. Although the successes of PCV13 have demonstrated thus far, it is important to keep in mind that the remaining NVT serotypes may provide a pool of new antigens from which capsule switching and new serotype emergence may result. Detection and serotyping methods for *S. pneumoniae* need to account for these possible variations.

It should be noted that a major limitation of this study (like the validation of PneumoCaT) is the low numbers for rare *S. pneumonaie* serotypes such as 7A, 11D, 44, and 46. As these serotypes have not been reported in any recent literature, it will be difficult to completely ascertain the robustness of molecular methods for discrimination of these serotypes. Even reference laboratories like the NML, CDC, and SSI were only able to provide a limited number of certain serotypes. New serotyping technologies should be able to discriminate all and any that could emerge through serotype replacement following new pneumococcal vaccines or vaccine recommendations (LeBlanc *et al*., 2017). To help validate new detection or serotyping methods for *S. pneumoniae*, there would be value in having a world biorepository for characterized *S. pneumoniae* isolates with well defined genotypic-phenotypic traits, as initiated in 2008 by the CDC Global Pneumococcal Strain Bank for serotypes characterized by Quellung reaction (https://www.cdc.gov/streplab/global-pneumo-strain-bank.html). While additional experiments are needed to resolve discrepant results noted in this study, the methods and approach could pave an avenue of research for *S. pneumoniae* strain characterization that would add to traditional Quellung serotyping.

5.4. Conclusions

Accurate detection and discrimination of *S. pneumoniae* serotypes is crucial to help characterize the epidemiology of *S. pneumoniae* and the impact of pneumococcal vaccines. Unlike the initial approach based on finding novel targets outside the *cps* loci, PCR and sequencing shows promise for the discrimination of *S. pneumoniae* serotypes. With the PCR and sequence-based strategy, no discordant results were observed with *S. pneumoniae* serotypes within cmPCR groups 15B/C, 22F/A, and 33F/A/37, which is a significant advance for the molecular surveillance of *S. pneumoniae.* While a limited number of discrepant remain to be resolved for other serotypes, the results of this study showed that PCR cmPCR coupled to sequence-based serotyping is of value for pneumococcal epidemiology, and may one day enable accurate molecular serotyping results without the need for sophisticated bioinformatics analyses. This could provide a simple solution for any research molecular laboratory, and could avoid the challenges noted with traditional Quellung serotyping.

Bibliography

Aanensen, D. M., Mavroidi, A., Bentley, S. D., Reeves, P. R. and Spratt, B. G. 2007. Predicted functions and linkage specificities of the products of the *Streptococcus pneumoniae* capsular biosynthetic loci. J Bacteriol **189(21):** 7856-7876.

AlonsoDeVelasco, E., Verheul, A. F., Verhoef, J., and Snippe, H. 1995. *Streptococcus pneumoniae*: virulence factors, pathogenesis, and vaccines. Microbiol Rev **59(4):** 591-603.

Andrews, S. 2014. FastQC: a quality control tool for high throughput sequence data. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. [Last accessed June 1 2017].

Austrian, R. 1976. The quellung reaction, a neglected microbiologic technique. Mt Sinai J. Med. **43:** 699–709.

Austrian, R. 1981a. Pneumococcus: the first one hundred years. Rev Infect Dis **3:** 183-189.

Austrian, R. 1981b. Some observations on the pneumococcus and on the current status of pneumococcal disease and its prevention. Rev Infect Dis **3**(suppl): SI-17.

Avery, O. T., MacLeod, C. M. and McCarty, M. 1944. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types - Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III. J Exper Med **79(2):** 137-158.

Batt, S. L., Charalambous, B. M., McHugh, T. D., Martin, S. and Gillespie, S. H. 2005. Novel PCR-restriction fragment length polymorphism method for determining serotypes or serogroups of *Streptococcus pneumoniae* isolates. J Clin Microbiol **43:** 2656-2661.

Beall, B. W., Gertz, R. E., Hulkower, R. L., Whitney, C. G., Moore, M. R. and Brueggemann, A. B. 2011. Shifting Genetic Structure of Invasive Serotype 19A Pneumococci in the United States. JID **203:** 1360-1368.

Beckler, E. and MacLeod, P. 1934. The Neufeld Method of Pneumococcus Type Determination as Carried Out in a Public Health Laboratory: A Study of 760 Typings. J Clin Invest **13(6):** 901-907.

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

Berry, D. S., Lynn, F., Lee, C. H., Frasch, C. E. and Bash, M. C. 2002. Effect of O acetylation of *Neisseria meningitidis* serogroup A capsular polysaccharide on development of functional immune responses. Infect Immun **70:** 3707–3713.

Bettinger, J. A., Scheifele, D. A., Kellner, J. D., Halperin, S. A., Vaudry, W., Law, B. and Tyrrell, G. 2010. The effect of routine vaccination on invasive pneumococcal infections in Canadian children, Immunization Monitoring Program, Active 2000-2007. Vaccine **28:** 2130-2136.

Bogaert, D., De Groot, R. and Hermans, P. W. M. 2004a. *Streptococcus pneumoniae* colonization: the key to pneumococcal disease. Lancet Infect Dis **4:** 144-154.

Bogaert, D., Germans, P. W., Adrian, P. V., Rumke, H. C. and de Groot, R. 2004b. Pneumococcal vaccines: an update on current strategies. Vaccine **22 (17-18):** 2209-2220.

Bonten, M. J. M., Huijts, M., Bolkenbaas, C., Webber, S., Patterson, S., Gault, S., van Werkhoven, C. H., van Deursen, A. M. M., Sanders, E. A. M., Vergeij, T. J. M., Patton, M., McDonough, A., Moradoghli-Haftvani, A., Smith, H., Mellelieu, T., Pride, M. W., Crowther, G., Schmolele-Thoma, B., Scott, D. A., Jansen, K. U., Lobatto, R., Oosterman, B., Visser, N., Caspers, E., Smorenburg, A., Emini, E. A., Gruber, W. C. and Grobbee, D. E. 2015. Polysaccharide conjugate vaccine against pneumococcal pneumonia in adults. NEJM. **372:** 1114-1125.

Bronsdon, M. A., O'Brien, K. L., Facklam, R. R., Whitney, C. G., Schwartz, B. and Carlone, G. M. 2004. Immunoblot method to detect *Streptococcus pneumoniae* and identify multiple serotypes from nasopharyngeal secretions. J Clin Microbiol **42:** 1596-1600.

Brueggemann, A. B., Griffiths, D. T., Meats, E., Peto, T., Crook, D. W. and Spratt, B. G. 2003. Clonal relationships between invasive and carriage *Streptococcus pneumoniae* and serotype- and clone-specific differences in invasive disease potential. J Infect Dis **187(9):** 1424-1432.

Bryant, K., Block, S. L., Baker, S. A., Gruber, W. C. and Scott, D. A. 2010. Safety and immunogenicity of a 13-valent pneumococcal conjugate vaccine. Pediatrics **125(5):** 866-75.

Burton, R. L., Geno, A., Saad, J. S. and Nahm, M. H. 2016. Pneumococcus with the "6E" *cps* locus produces serotype 6B capsular polysaccharide. J Clin Microbiol **54(4):** 967-971.

Butler, J. C., Tuomanen, E. I., Mitchell, T. J., Morrison, D. and Spratt, B. G. 2004. Epidemiology of pneumococcal disease. In: The pneumococcus. American society for microbiology, Washington DC, USA 148-168.

Cai, P., Moran, J., Pavliak, V., Deng, C., Khoury, N., Marcq, O. and Ruppen, M. E. 2012. NMR structural analysis of the capsular polysaccharide from *Streptococcus pneumoniae* serotype 6C. Carbohydr Res **351:** 98–107.

Calix, J. J., Brady, A. M., Du, V. Y., Saad, J. S. and Nahm, M. H. 2013 Spectrum of pneumococcal serotype 11A variants results from incomplete loss of capsule O-acetylation. J Clin Microbiol **52:** 758–765.
Calix, J. J. and Nahm, M. H. 2010 A new pneumococcal serotype, 11E, has a variably inactivated *wcjE* gene. J Infect Dis **202:** 29–38.

Calix, J. J., Nahm, M. H. and Zartler, E. R. 2011a. Elucidation of structural and antigenic properties of pneumococcal serotype 11A, 11B, 11C, and 11F polysaccharide capsules. J Bacteriol **193:** 5271–5278.

Calix, J. J., Oliver, M. B., Sherwood, L. K., Beall, B. W., Hollingshead, S. K. and Nahm, M. H. 2011b. *Streptococcus pneumoniae* serotype 9A isolates contain diverse mutations to *wcjE* that result in variable expression of serotype 9V-specific epitope. J Infect Dis **204:** 1585-1595.

Calix, J. J., Porambo, R. J., Brady, A. M., Larson, T. R., Yother, J., Abeygunwardana, C. and Nahm, N. H. 2012a. Biochemical, genetic, and serological characterization of two capsule subtypes among *Streptococcus pneumoniae* serotype 20 strains: discovery of a new pneumococcal serotype. J Biol Chem **287(33):** 27885-27894.

Calix, J. J., Saad, J. S., Brady, A. M. and Nahm, M. H. 2012b. Structural characterization of *Streptococcus pneumoniae* serotype 9A capsule polysaccharide reveals role of glycosyl 6-*O*-acetyltransferase *wcjE* in serotype 9V capsule biosynthesis and immunogenicity. J Biol Chem **287 (17):** 13996-14003.

Camilli, R., Bonnal, R. J., Del Grosso, M., Iacono, M. and Corti, G. 2011. Complete genome sequence of a serotype 11A, ST62 *Streptococcus pneumoniae* invasive isolate. BMC Microbiol **11**: 25.

Centers for Disease Control and Prevention (CDC). 2005. Direct and indirect effects of routine vaccination of children with 7-valent pneumococcal conjugate vaccine on incidence of invasive pneumococcal disease--United States, 1998-2003. MMWR Morb Mortal Wkly Rep. 2005. **54:** 893–897.

Centers for Disease Control and Prevention (CDC). 2013. Active Bacterial Core Surveillance Report, Emerging Infections Program Network, *Streptococcus pneumoniae*, 2013. Available online: https://www.cdc.gov/abcs/reports findings/survreports/spneu13.pdf. [Last accessed June 1 2017]

Centers for Disease Control and Prevention (CDC). 2014. Comparison of a Real-Time Multiplex PCR and Sequetyping Assay for Pneumococcal Serotyping. Available online: https://www.cdc.gov/streplab/protocols.html. [Last accessed June 1 2017]

Centers for Disease Control and Prevention (CDC). 2015b. The Pink Book: Course Textbook - 13th Edition. Available online: https://www.cdc.gov/vaccines/pubs/pinkbook/index.html. [Last accessed June 1 2017]

Centers for Disease Control and Prevention (CDC). 2015a. Pneumococcal disease: Types of infections. Available online: https://www.cdc.gov/pneumococcal/about/infection-types.html. [Last accessed June 1 2017]

Centers for Disease Control and Prevention (CDC). 2016. Surveillance and reporting. Available online: https://www.cdc.gov/pneumococcal/surveillance.html. [Last accessed June 1 2017]

Chaffin, D. O., Beres, S., Yim, H. and Rubens, C. 2000. The serotype of type Ia and III group B streptococci is determined by the polymerase gene within the polycistronic capsule operon. J Bacteriol **182:** 4466-4477.

Chao, Y., Marks, L. R., Pettigrew, M. M. and Hakansson, A. P. 2015. *Streptococcus pneumoniae* biofilm formation and dispersion during colonization and disease. Front Cell Infect Microbiol **4:**194.

Congeni, B. L., Igel, H. J. and Platt, M. S. 1984. Evaluation of a latex particle agglutination kit in pneumococcal disease. Pediatr Infect Dis **5:** 417-419.

Croucher, N. J., Coupland, P. G., Stevenson, A. E., Callendrello, A., Bentley, S. D. and Hanage, W. P. 2014. Diversification of bacterial genome content through distinct mechanisms over different timescales. Nature Commun. **5:** 5471.

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

da Gloria Carvalho, M., Pimenta, F. C., Jackson, D., Roundtree, A., Ahmad, Y., Millar, E. V., O'Brien, K. L., Whitney, C. G., Cohen, A. L. and Beall, B. W. 2010. Revisiting pneumococcal carriage by use of broth enrichment and PCR techniques for enhanced detection of carriage and serotypes. J Clin Microbiol **48(5):** 1611-1618.

Deibel, R. H. and Seeley, H. W. Jr. 1974. Family II: *Streptococcuceae*. Fam, p 490-517. *In* Buchanan, R. E. and Gibbons, N. E. (ed.) Bergy's manual of determinative bacteriology, **8th ed**. Williams & Wilkins. Baltimore, MD.

Deng, X., Peirano, G., Schillberg, E., Mazzulli, T., Gray-Owen, S. D., Wylie, J. L., Robinson, D. A., Mahmud, S. M. and Pillai, D. R. 2016. Whole-Genome Sequencing Reveals the Origin and Rapid Evolution of an Emerging Outbreak Strain of *Streptococcus pneumoniae* 12F. Clin Infect Dis **62(9):** 1126-1132.

Dhoubhadel, B. G., Yasunami, M., Yoshida, L. M., Thi, H. A., Thi, T. H., Thi, T. A., Watanabe, K., Suzuki, M., Morimoto, K., Dang, D. A. and Ariyoshi, K. A novel high-throughput method for molecular serotyping and serotype-specific quantification of *Streptococcus pneumoniae* using a nanofluidic real-time PCR system. 2014. J Med Microbiol. **63(Pt 4)**: 528-539.

Dias, C. A., Teixeira, L. M., Carvalho, M. D. G. and Beall, B. 2007. Sequential multiplex PCR for determining capsular serotypes of pneumococci recovered from Brazilian children. J Med Microbiol **56:**1185-1188.

Dube, F. S., van Mens, S. P., Robberts, L., Wolter, N., Nicol, P., Mafofo, J., Africa, S., Zar, H. J. and Nicol, M. P. 2015. Comparison of a Real-Time Multiplex PCR and Sequetyping Assay for Pneumococcal Serotyping. PLoS One **10(9):** e0137349.

Drijkoningen, J. J. C. and Rohde, G. G. U. 2014. Pneumococcal infection in adults: burden of disease. Clin Microbiol Infect **20:** 45-51.

Eddy, B. 1944. Cross-reactions between the several pneumococcic types and their significance in the preparation of polyvalent antiserum. Public Health Rep **59:** 485–499.

Elberse, K., Witteveen, S., van der Heide, H., van de Pol, I., Schot, C., van der Ende, A., Berbers, G. and Schouls, L. 2011. Sequence diversity within the capsular genes of *Streptococcus pneumoniae* serotype 6 and 19. **PLoS One 6(9):** e25018.

Fenoll, A., Jado, I., Vicioso, D. and Casal, J. 1997. Dot blot assay for the serotyping of pneumococci. J Clin Microbiol **35:** 764-766.

Fraenkel, A. 1886. Weitere Beitrage zur Lehre von den Mikrococcen der genuinen fibrinosen Pneumonie. Zeitschrift filr Klinische Medicin **11:** 437-58.

Garcia, E., Arrecubieta, C., Munoz, R., Mollerach, M. and Lopez, R. 1997. A functional analysis of *Streptococcus pneumoniae* genes involved in the synthesis of type 1 and type 3 capsular polysaccharides. Microb Drug Resis **3:** 73-88.

Geno, K.A., Gilbert, G.L., Song, J.Y., Skovsted, I.C., Klungman, K.P., Jones, C., Konradsen, H.B. and Nahm, M. H. 2015. Pneumococcal capsules and their types: past, present and future. Clin Microbiol Rev **28(3)**: 871-896.

Gillis, H. D., Lang, A. L. S., ElSherif, M., Martin, I., Hatchette, T. F., McNeil, S. A., LeBlanc, J. J. 2017. Assessing the diagnostic accuracy of PCR-based detection of *Streptococcus pneumoniae* from nasopharyngeal swabs collected for viral studies in Canadian adults hospitalized with community-acquired pneumonia: a Serious Outcomes Surveillance (SOS) Network of the Canadian Immunization Research Network (CIRN) study. BMJ Open. (bmjopen-2016-015008; accepted for publication).

Grabenstein, J. D. and Klugman, K. P. 2012. A century of pneumococcal vaccination research in humans. Clin Microbiol Infect **18**(Suppl 5): S15–S24.

Gram, C. 1884. Ueber die isolierte F&bung der Schizomyceten in Schnittund Trockenpraparaten. Fortschr Med **2:** 185-189.

Griffith, F. 1928. The significance of pneumococcal types. J Hyg (Lond) **27:** 113-159.

Gruber, W. C. and Grobbee, D. E. 2015. Polysaccharide conjugate vaccine against pneumococcal pneumonia in adults. NEJM **372(12):** 1114-1125.

Guidolin, A., Morona, J. K., Morona, R., Hansman, D. and Paton, J.C. 1994. Nucleotide sequence analysis of genes essential for capsular polysaccharide biosynthesis in *Streptococcus pneumoniae* type 19F. Infect Immun **62:** 5384-5396.

Heidelberger, M. and Avery, O. T. 1923. The soluble specific substance of pneumococcus. J Exp Med **38:** 73-79.

Heilmann, C. 1990. Human B and T lymphocyte responses to vaccination with pneumococcal polysaccharides. APMIS Suppl **15:** 1-23.

Hendley, J. O., Sande, M. A., Stewart, P. M. and Gwaltney, J. M. jr. 1975. Spread of *Streptococcus pneumoniae* in families. 1. Carriage rates and distribution of types. J Infect Dis **132:** 55-61.

Hicks, L. A., Harrison, L. A., Flannery, B., Hadler, J. L., Schaffner, W., Craig, A. S., Jackson, D., Thomas, A., Beall, B., Lynfield, R., Reingold, A., Farley, M. M. and Whitney, C. G. 2007. Incidence of pneumococcal disease due to non-pneumococcal conjugate vaccine (PCV7) serotypes in the United States during the era of widespread PCV7 vaccination, 1998-2004. J Infect Dis **196(9):** 1346-1354.

Hiller, N. L., Janto, B., Hogg, J. S., Boissy, R., Yu, S., Powell, E., Keefe, R., Ehrlich, N. E., Shen, K., Hayes, J., Barbadora, K., Klimke, W., Dernovoy, S., Tatusova, T., Parkhill, J., Bentley, S. D., Post, C., Ehrlich, G. D. and Hu, F. Z. (2007) Comparative genomic analyses of seventeen Streptococcus pneumoniae strains: insights into the pneumococcal supragenome. J Bacteriol **189:** 8186–8195.

Huang, S. S., Johnson, K. M., Ray, G. T., Wroe, P., Lieu, T. A., Moore, M. R., Zell, E. R., Linder, J. A., Grijalva, C. G., Metly, J. P. and Finkelstein, J. A. 2011. Healthcare utilization and cost of pneumococcal disease in the United States. Vaccine **29(18):** 3398-3412.

Huijts, S. M., Pride, M. W., Vos, J. M. Jansen, K. U., Webber, C., Gruber, W., Boersma, W. G., Snijders, D., Kluytmans, J. A., van der Lee, I., Kuipers, B. A., van der Ende, A. and Bonten, M. J. 2013. Diagnostic accuracy of a serotype-specific antigen test in community-acquired pneumonia. Eur Resp J **42:** 1283-1290.

Iannelli, F., Pearce, B. J. and Pozzi, G. 1999. The type 2 capsule locus of *Streptococcus pneumoniae.* J Bacteriol **181:** 2652-2654.

Ip, M., Liyanapathirana, V., Ang, I., Fung, K. S., Ng, T. K., Zhou, H. and Tsang, D. N. 2014. Direct detection and prediction of all pneumococcal serogroups by target enrichment-based next-generation sequencing. J Clin Microbiol **52(12):** 4244-4252.

Jacobs, M. R., Bajaksouzian, S., Bonomo, R. A., Good, C. E., Windau, A. R., Hujer, A. M., Massire, C., Melton, R., Blyn, L. B., Ecker, D. J. and Sampath, R. 2009. Occurrence, distribution, and origins of *Streptococcus pneumoniae* serotype 6C, a recently recognized serotype. J Clin Microbiol **47(1):** 64-72.

Jain, S., Self, W. H., Wunderink, R. G., Fakhran, S., Nalk, R., Bramley, A. M., Reed, C., Grijalva, C. G., Anderson, E. J., Courtney, D. M., Chappell, J. D., Qi, C., Hart, E. M., Carroll, F., Trabue, C., Donnelly, H. K., Williams, D. J., Zhu, Y., Arnold, S. R., Ampofo, K., Waterer, G. W., Levine, M., Lindstrom, S., Winchell, J. M., Katz, J. M., Erdman, D., Schneider, E., Hicks, L. A., McCullers, J. A., Pavia, A. T., Edwards, K. M. and Finelli, L. for the CDC EPIC Study Team. 2015. Community-acquired pneumonia requiring hospitalization among U. S. adults. NEJM **373(5):** 415-427.

Janeway, C. A., Jr, Travers, P., Walport, M. and Shlomchik, M. J. 2001. Immunobiology: the immune system in health and disease, **5th ed**. Garland Science, New York, NY.

Jansson, P. E., Lindberg, B., Lindquist, U. and Ljungberg, **J.** 1987. Structural studies of the capsular polysaccharide from *Streptococcus pneumoniae* types 15B and 15C. Carbohydr Res **162:** 111–116.

Jauneikatie, E., Tocheva, A. S., Jefferies, J. M. C., Gladstone, R. A., Faust, S. N., Christodoulides, M., Hibberd, M. L. and Clarke, S. C. 2015. Current methods for capsular typing of *Streptococcus pneumoniae*. J Microbiol Methods **13:** 41-49.

Jin, P., Xiao, M., Kong, F., Oftadeh, S., Zhou, F., Liu, C. and Gilbert, G. L. 2009. Simple, accurate, serotype-specific PCR assay to differentiate *Streptococcus pneumoniae* serotypes 6A, 6B and 6C. J Clin Microbiol **47(8):** 2470-2474.

Jourdain, S., Drèze, P. A., Vandeven, J., Verhaegen, J., Van Melderen, L. and Smeesters, P. R. 2011. Sequential multiplex PCR assay for determining capsular serotypes of colonizing *S. pneumoniae*. BMC Infect Dis **11:** 100.

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

Kamerling, J. P. 2000. Pneumococcal polysaccharides: a chemical view. *In* Tomasz A (ed.), *Streptococcus pneumoniae* molecular biology and mechanisms of disease. Mary Ann Liebert, Inc. Larchmont, NY.

Kapatai, G., Sheppard, C. L., Al-Shahib, A., Litt, D. J., Underwood, A. P., Harrison, T. G. and Dry, K. F. 2016. Whole genome sequencing of *Streptococcus pneumoniae:* development, evaluation and verification of targets for serogroup and serotype prediction using an automated pipeline. Peer J **4:** e2477.

Kauffman, F., Morch, E. and Schmith, K. 1940. On the serology of the pneumococcus-group. J Immunol **239:** 397-426.

Kellner, J. D., Vanderkooi, O. G., MacDonald, J., Church, D. L., Tyrrell, G. J. and Scheifele, D. W. 2009. Changing epidemiology of invasive pneumococcal disease in Canada, 1998-2007: Update from the Calgary-area *Streptococcus pneumoniae* research (CASPER) study. CID **49:** 205-212.

Kieninger, D., Kueper, K. and Steul, K. 2008. Safety and immunologic non-inferiority of 13-valent pneumococcal conjugate vaccine compared to 7-valent pneumococcal conjugate vaccine given as a 4-dose series in healthy infants and toddlers. World Vaccine Congress; 2008 Apr 21; Arlington, Virginia.

Kilian, M., Riley, D. R., Jensen, A., Bruggemann, H. and Tettelin, H. 2014. Parallel evolution of *Streptococcus pmeumoniae* and *Streptococcus mitis* to pathogenic and mutualistic lifestyles. mBio **5(4):** e14090-14.

Kim, J. O. and Weisner, J. N. 1998. Association of intrastrain phase variation in quantity of capsular polysaccharide and teichoic acid with the virulence of *Streptococcus pneumoniae.* J Infect Dis **177:** 368-377.

Kim, Y. K., LaFon, D. and Nahm, M. H. 2016. Indirect Effects of Pneumococcal Conjugate Vaccines in National Immunization Programs for Children on Adult Pneumococcal Disease. Infect Chemother: **48(4):** 257-266

Kirkman, J. B. Jr., Fischer, J. and Pagano, J. S. 1970. A microtiter plate technique for the agglutination typing of *Diplococcus pneumoniae.* J Infect Dis **121:** 217-221.

Ko, K. S., Baek, J. Y. and Song, J. 2013. Capsular gene sequences and genotypes of "serotype 6E" *Streptococcus pneumoniae* isolates. J Clin Microbiol **51(10):** 3395-3399.

Kolkman, M. A., van der Zeijst, B. A. and Nuikten, P.J. 1998. Diversity of capsular polysaccharide synthesis gene clusters in *Streptococcus pneumoniae*. J Biochem **123:** 937-945.

Kostyukova, N. N., Volkova, M. O., Ivanova, V. V. and Kvetnaya, A. S. 1995. A study of pathogenic factors of *Streptococcus pneumoniae* strains causing meningitis. FEMS Immunol Med Microbiol **10:** 133-137.

Kraicer-Melamed, H., O'Donnell, S. and Quach, C. 2016. The effectiveness of pneumococcal polysaccharide vaccine 23 (PPV23) in the general population of 50 years of age and older: A systemic review and meta-analysis. Vaccine **34:** 1540-1550.

Lafong, A. C. and Crothers, E. 1988. Simple latex agglutination method for typing pneumococci. J Clin Pathol **41:** 230–231.

Lang, A. L. S., McNeil, S. A., Hatchette, T. F., Elsherif, M., Martin, I. and LeBlanc, J. J. 2015. Detection and prediction of *Streptococcus pneumoniae* serotypes directly from nasopharyngeal swabs using PCR. J Med Microbiol **64(8):** 836-844.

Lang, A. L. S., Gillis, H. D., Elsherif, M., Martin, I., Hatchette, T. F., McNeil, S.A. and LeBlanc, J. 2017. Refining PCR-based serotyping of vaccine-preventable *Streptococcus pneumoniae*. J Epi Research **3(1):** 28-36.

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

Lankinen, K. S., Rintamaki, S., Syrjanen, R., Kipli, T., Ruutu, P. and Leinonen, M. 2004. Type-specific enzyme Immunoassay for detection of pneumococcal capsular polysaccharide antigens in nasopharyngeal specimens. J Microbiol Methods **56:** 193-199.

Laufer, A. S., Thomas, J. C., Figueira, M., Gent, J. F., Pelton, S. I., Pettigrew, M. M. 2010. Capacity of serotype 19A and 15B/C *Streptococcus pneumoniae* isolates for experimental otitis media: implications for the conjugate vaccine. Vaccine **28:** 2450-2457

Leal, J., Vanderkooi, O. G., Church, D. L., MacDonald, J., Tyrrell, G. J. and Kellner, J. D. 2012. Eradication of invasive pneumococcal disease due to the seven-valent pneumococcal conjugate vaccine serotypes in Calgary, Alberta. Ped infect dis j **31 (9):** 169-175.

LeBlanc, J. J., ElSherif, M., Ye, L., MacKinnon-Cameron, D., Li, L., Ambrose, A., Hatchette, T. F., Lang, A. L., Gillis, H., Martin, I., Andrew, M. K., Boivin, G., Bowie, W., Green, K., Johnstone, J., Loeb, M., McCarthy, A., McGeer, A., Moraca, S., Semret, M., Stiver, G., Trottier, S., Valiquette, L., Webster, D. and McNeil SA, on behalf of the Serious Outcomes Surveillance (SOS) Network of the Canadian Immunization Research Network (CIRN). 2017. Vaccine **35**: 3647-3654.

Leeming, J. P., Cartwright, K., Morris, R., Martin, S. A. and Smith, M. D. 2005. Diagnosis of invasive pneumococcal infection by serotype-specific urinary antigen detection. J Clin Microbiol **43:** 4972-4976.

Leung, M., Bryson, K., Freystatter, K., Pichon, B., Edwards, G., Charalambous, B. M. and Gillespie, S. H. 2012. Sequetyping: Serotyping *Streptococcus pneumoniae* by a single PCR-sequencing strategy. J Clin Microbiol **50(7):** 2419-2427.

Lewis, A. L., Nizet, V. and Varki, A. 2004. Discovery and characterization of sialic acid O-acetylation in group B Streptococcus. Proc Natl Acad Sci USA **101:** 11123–11128.

Li, J., Li, J. W., Feng, Z., Wang, J., An, H., Liu, Y., Wang, Y., Wang, K., Zhang, X., Miao, Z., Liang, W., Sebra, R., Wang, G., Wang, W. C. and Zhang, J. R. 2016. Epigenetic Switch Driven by DNA Inversions Dictates Phase Variation in *Streptococcus pneumoniae*. PLoS Pathog **12(7):** e1005762.

Lin, T. Y., Shah, N. K., Brooks, D. and Garcia, C. S. 2010. Summary of invasive pneumococcal disease burden among children in the Asia-Pacific region. Vaccine **28(48):** 7589-7605.

Llull, D., Munoz, R., Lopez, R., and Ernesto Garcia. 1999. A single gene (*tts*) located outside the *cap* locus directs the formation of *Streptococcus pneumoniae* type 37 capsule polysachharide: type 37 pneumococci are natural, genetically binary strains. J Exp Med **190 (2):** 241-251.

Liyanapathirana, V., Ang, I. , Tsang, D., Fung, K., Ng, T. K., Zhou, H. and Ip, M. 2014. Application of a target enrichment-based next-generation sequencing protocol for identification and sequence based prediction of pneumococcal serotypes. BMC Microbiol **14(60)**; doi 10.1186/1471-2180-14-60.

Loharikar, A., Dumolard, L., Chu, S., Hyde, T., Goodman, T. and Mantel, C. 2016. Status of New Vaccine Introduction — Worldwide, September 2016. MMWR Morb Mortal Wkly Rep **65:** 1136–1140.

Lund, E. 1970. On the nomenclature of the pneumococcal types. Int J Syst Bacteriol **20:** 321-323.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D. W., Yiu, S. M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T. W. and Wang, J. 2012**.** SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience **1(1):** 18.

MacLeod, C. M. and Krauss, M. R. 1950. Relation of virulence of pneumococcal strains for mice to the quantity of capsular polysaccharide formed in vitro. J Exp Med **92:** 1-9.

Magee, A. D. and Yother, J. 2001. Requirement for capsule in colonization by *Streptococcus pneumoniae.* Infect Immun **69:** 3755-3761.

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

Massire, C., Gertz, R. E. Jr., Svoboda, P., Levert, K., Reed, M. S., Pohl, J., Kreft, R., Li, F., White, N., Ranken, R., Blyn, L. B., Ecker, D. J., Sampath, R. and Beall, B. 2012. Concurrent serotyping and genotyping of pneumococci by use of PCR and electrospray ionization mass spectrometry. J Clin Microbiol **50:** 2018-2025.

Mavroidi, A., Aanensen, D. M., Godoy, D., Skovsted, I. C., Kaltoft, M. S., Reeves, P. R., Bentley, S. D. and Spratt, B. G. 2007. Genetic relatedness of the *Streptococcus pneumoniae* capsular biosynthetic loci. J Bacteriol **189:** 7841-7855.

McCool, T. L., Cate, T. R., Moy, G and Weiser, J. N. 2002. The immune response to pneumococcal proteins during experimental human carriage. J Exp Med **195:** 359-365.

McNeil, S. A., Qizilbash, N., Ye, J., Gray, S., Giovanni, Z., Munson, S., Dartois, N. and Laferriere, C. 2016. A retrospective study of the clinical burden of hospitalized all-cause and pneumococcal pneumonia in Canada. Can Respr J **pii:** 17154.

McNeil, S., Shinde, V., Andrew, M., Hatchette, T., LeBlanc, J., Ambrose, A., Boivin, G., Bowie, W., Diaz-Mitoma, F., Elsherif, M., Green, K., Haguinet, F., Halperin, S., Ibarguchi, B., Katz, K., Langley, J., Lagace-Wiens, P., Light, B., Loeb, M., McElhaney, J., Mackinnon-Cameron, D., McCarthy, A., Poirier, M., Powis, J., Richardson, D., Semret, M., Smith, S., Smyth, D., Stiver, G., Trottier, S., Valiquette, L., Webster, D., Ye, L. and McGeer, A for Public Health Agency of Canada/Canadian Institutes of Health Research Influenza Research Network (PCIRN) Serious Outcomes Surveillance Network; Toronto Invasive Bacterial Diseases Network (TIBDN). 2014. Interim estimates of 2013/14 influenza clinical severity and vaccine effectiveness in the prevention of laboratory-confirmed influenza-related hospitalisation, Canada, February 2014. EuroSurveillance **19:** pii 20729.

Moore, M. R., Gertz, R. E., Jr, Woodbury, R. L., Barkocy-Gallagher, G. A., Schaffner, W., Lexau, C., Gershman, K., Reingold, A., Farley, M., Harrison, L. H., Hadler, J. L., Bennett, N. M., Thomas, A. R., McGee, L., Pilishvili, T., Bruggemann, A. B., Whitney, C. G., Jorgensen, J. H. and Beall, B. 2008. Population snapshot of emergent *Streptococcus pneumoniae* serotype 19A in the United States, 2005. J Infect Dis **197:** 1016-1027.

Moore, M. R., Link-Gelles, R., Schaffner, W., Lynfield, R., Lexau, C., Bennett, N. M., Petit, S., Zansky, S. M., Harrison, L. H., Reingold, A., Miller, L., Scherzinger, K., Thomas, A., Farley, M. M., Zell, E. R., Taylor, T Jr.. H., Pondo, T, Rodgers, L., McGee, L., Beall, B., Jorgensen, J. H. and Whitney, C. G. 2015. Effect of use of 13-valent pneumococcal conjugate vaccine in children on invasive pneumococcal disease in children and adults in the USA: analysis of multisite, population-based surveillance. Lancet Infect Dis **15:** 301-309.

Morais, L., da Gloria Carvalho, M., Roca, A., Flannery, B., Mandomando, I., Soriano-Gabarró, M., Sigauque, B., Alonso, P. & Beall, B. 2007. Sequential multiplex PCR for identifying pneumococcal capsular serotypes from south-Saharan African clinical isolates. J Med Microbiol **56**: 1181-1184.

Morganroth, J. and Levy, R. 1911. Chemotherapie der Pneumokokkeninfektion. Berliner Klinische Wochenschrift **48:** 1560-1583.

Morrow, A., De Wals, P., Petit, G., Guay, M. and Erickson, L. J. 2007. The burden of pneumococcal disease in the Canadian population before routine use of pneumococcal conjugate vaccine. Can J infect dis and med microbial **18(2):** 121-127.

Moscoco, M. and Garcia, E. 2009. Transcriptional regulation of capsular polysaccharide biosynthesis locus of *Streptococcus pneumoniae:* A bioinformatics analysis. DNA research **16:** 177-186.

Musher, D. M. 1992. Infections caused by *Streptococcus pneumoniae*: clinical spectrum, pathogenesis, immunity and treatment. Clin Infect Dis **14:** 801-807.

National Advisory Committee on Immunization (NACI). 2002. Statement on recommended use of pneumococcal conjugate vaccine. Canada Communicable Disease Report **28.** Available online: http://www.collectionscanada.gc.ca/webarchives/20071116032814/http://www.phac-aspc.gc.ca/publicat/ccdr-rmtc/02vol28/28sup/acs2.html. [Last accessed June 1 2017]

National Advisory Committee on Immunization (NACI). 2010. Update on the use of conjugate pneumococcal vaccines in childhood. Canada Communicable Disease Report **36.** Available online: http://www.phac-aspc.gc.ca/publicat/ccdr-rmtc/10vol36/acs-12/index-eng.php. [Last accessed March 2017].

National Advisory Committee on Immunization (NACI). 2015. Re-Immunization with Polysaccharide 23-Valent Pneumococcal Vaccine (Pneu-P-23). Available online: https://www.canada.ca/en/public-health/services/publications/healthy-living/re-immunization-with-polysaccharide-23-valent-pneumococcal-vaccine-pneu-p-23.html. [Last accessed June 1 2017].

National Advisory Committee on Immunization (NACI). 2016. Update on the use of 13-valent pneumococcal conjugate vaccine (PNEU-C-13) in addition to 23-valent pneumococcal polysaccharide vaccine (PNEU-P-23) in immunocompetent adults 65 years of age and older – Interim Recommendation. Available online: https://www.canada.ca/en/public-health/services/publications/healthy-living/update-use-of-13-valent-pneumococcal-conjugate-vaccine-pneu-c-13-in-addition-to-23-valent-pneumococcal-polysaccharide-vaccine-pneu-p-23-immunocompetent-adults-65-years-and-older-interim-recommendation.html. [Last accessed June 1 2017].

Neufeld, F. 1900. Über eine spezifische bakteriolytische Wirkung der Galle. Zeitschrift fur Hygiene Infektionskrankheiten. **34**: 454-64.

Neufeld, F. 1902. Über die Agglutination der Pneumokokken und über die Theorieen der Agglutination. Zeitschrift fur Hygiene Infektionskrankheiten. **40:** 54-72.

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

Nuorti, J. P. and Whitney, C. G. 2010. 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 **59**: 1–18.

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

Oliver, M. B., van der Linden, M. P. G., Kunztel, S. A., Saad, J. S. and Nahm, M. H. 2013. Discovery of *Streptococcus pneumoniae* serotype 7 variants with glycosyltransferases synthesizing two different repeat units. J Biol Chem **288(36):** 25976-259865.

Orsi, A., Ansaldi, F., Trucchi, C., Rosselli, R. and Icardi G. 2016. Pneumococcus and the Elderly in Italy: A Summary of Available Evidence Regarding Carriage, Clinical Burden of Lower Respiratory Tract Infections and On-Field Effectiveness of PCV13 Vaccination. Int J Mol Sci **17(7)** pii: E1140.

Pai, R., Gertz, R. E. & Beall, B. 2006. Sequential multiplex PCR approach for determining capsular serotypes of *Streptococcus pneumoniae* isolates. J Clin Microbiol **44**: 124-131.

Pai, R., Limor, J. and Beall, B. 2005. Use of pyrosequencing to differentiate *Streptococcus pneumoniae* serotypes 6A and 6B. J Clin Microbiol **43(9):** 4820-4822.

Park, I. H., Geno, A., Yu, J., Oliver, M. B., Kim, K. and Nahm, M. H. 2015. Genetic, biochemical and serological characterization of new pneumococcal serotype, 6H, and generation of pneumococcal strain producing three different capsular repeat unit. Clin Vac Immunol **22(3):** 313-318.

Park, I. H., Kim, K. H., Andrade, A. L., Briles, D. E., McDaniel, L. S. and Nahm, M. H. 2012. Nontypeable pneumococci can be divided into multiple cps types, including one type expressing the novel gene *pspK*. MBio. **3(3):** pii: e00035-12.

Pasteur, L. 1881. Note sur la maladie novella provoquee par la salive d'un efant mort de la rage. Bull Acad Med (Paris) **10:** 94-103.

Perkins, M. D., Mirrett, S. and Reller, L. B. 1995. Rapid bacterial antigen detection is not clinically useful. J Clin Microbiol **33(6):** 1486-1491.

Pichichero, M. E. 2013. Protein carriers of conjugate vaccines: Characteristics, development, and clinical trials. Hum Vaccin Immunother **9(12):** 2505–2523.

Pilishvili, T., Lexau, C., Farley, M. M., Hadler, J., Harrison, L. H., Bennett, N. M., Reingold, A., Thomas, A., Schaffner, W., Craig, A. S., Smith, P. J., Beall, B. W., Whitney, C. G. and Moore, M. R. for the Active Bacterial Core Surveillance/Emerging Infections Program Network. 2010. Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. J Infect Dis **201**: 32-41.

Pillai, D. R., Shahinas, D., Buzina, A., Pollock, R. A., Lau, R., Khairnar, K., Wong, A., Farrell, D. J., Green, K., McGeer, A. and Low, D. E. 2009. Genome-wide dissection of globally emergent multi-drug resistant serotype 19A *Streptococcus pneumoniae*. BMC Genomics **10:** 642.

Pimenta, F. C., Roundtree, A., Soysal, A., Bakir, M., du Plessis, M., Wolter, N., von Gottberg, A., McGee, L., Carvalho, M. d. G. and Beall, B. 2013. Sequential triplex real-time PCR assay for detecting 21 pneumococcal capsular serotypes that account for a high global disease burden. J Clin Microbiol **51:** 647-652.

Porter, D. B., Ortika, B. D. and Satzke, C. 2014. Capsular serotyping of *Streptococcus pneumoniae* by latex agglutination. J Vis Exp **91:** e51747.

Pride, M. W., Huijts, S. M., Wu, K., **Souza, V., Passador, S., Tinder, C., Song, E., Elfassy, A., McNeil, L., Menton, R., French, R., Callahan, J., Webber, C., Grfuber, W. C., Bonten, M. J. and Jansen, K. U.** 2012. Validation of an immunodiagnostic assay for detection of 13 *Streptococcus pneumoniae* serotype-specific polysaccharides in human urine. Clin Vaccine Immunol **19:** 1131-1141.

Public Health Agency of Canada (PHAC). 2012. *Streptococcus pneumoniae*: Pathogen safety data sheet - Infectious substances. Available online: http://www.phac-aspc.gc.ca/lab-bio/res/psds-ftss/streptococcus-pneumoniae-eng.php#footnote4. [Last accessed June 1 2017].

Public Health Agency of Canada (PHAC). 2015. Streptococcus and STI unit, National Microbiology Laboratory. National laboratory surveillance of invasive streptococcal disease in Canada, annual summary 2014. Available online: https://www.canada.ca/en/public-health/services/publications/drugs-health-products/national-laboratory-surveillance-invasive-streptococcal-disease-canada-annual-summary-2014.html [Last accessed June 1 2017]

Public Health Agency of Canada (PHAC). 2016. Canadian immunization guide part 4: Active vaccines: Pneumococcal vaccine. Available online: https://www.canada.ca/en/public-health/services/publications/healthy-living/canadian-immunization-guide-part-4-active-vaccines.html. [Last accessed June 1 2017].

Public Health Agency of Canada (PHAC). 2017. Streptococcus and STI unit, National Microbiology Laboratory. National laboratory surveillance of invasive streptococcal disease in Canada, annual summary 2015. Ahead of print.

Rajam, G., Carlone, G. M. and Romero-Steiner, S. 2007. Functional antibodies to the O-acetylated pneumococcal serotype 15B capsular polysaccharide have low cross-reactivities with serotype 15C. Clin Vaccine Immunol **14:** 1223–1227.

Reeves, R. E. and Goebel, W. F. 1941. Chemoimmunological studies on the soluble specific substance of pneumococcus. V. The structure of the type III polysaccharide. J Biol Chem **139**: 511–519.

Regev-Yochay, G., Rav, M., Dagan, R., Porat, N., Shainberg, B., Pinco, E., Keller, N. and Rubinstein, E. 2004. Nasopharyngeal carriage of *Streptococcus pneumoniae* by adults and children in community and family settings. Clin Infect Dis **38:** 632-639.

Ritcher, S. S., Heilmann, K. P., Dohrn, C. L., Riahi, F., Beckmann, S. E., Doern and G. V. 2008. Accuracy of phenotypic methods for identification of *Streptococcus pneumoniae isolates* included in surveillance programs. J Clin Microbiol **46:** 2184-2188.

Ritcher, S. S., Heilmann, K. P., Dohrn, C. L., Riahi, F., Diekema, D. J. and Doern, G. V. 2013. Pneumococcal serotypes before and after introduction of conjugate vaccines, United States, 1999-2011. Emerg Infect Dis **19(7):** 1074-1083.

Robbins, J. B., Austrian, R., Lee, C. J., Rastogi, S. C., Schiffman, G., Henrichsen, J., Makela, P. H., Broome, C. V., Facklam, R. R., Tiesjema, R. H., and Parke, J. C. Jr. 1983. Considerations for formulating the second-generation pneumococcal capsular polysaccharide vaccine with emphasis on the crossreactive types within groups. J Infect Dis **148**: 1136–1159.

Rudnick, W., Zhong, L., Shigayeva, A., Low, D. E., Green, K., Plevneshi, A., Devlin, R., Downey, J., Katz, K., Kitai, I., Krajden, S., Ostrowska, K., Richardson, D., Richardson, S., Sarabia, A., Silverman, M., Simor, A. E., Tyrrell, G. and McGeer, A. 2013. Pneumococcal vaccination programs and the burden of invasive pneumococcal disease in Ontario, Canada, 1995-2011. Vaccine **31:** 5863-5871.

Sabin, A. B. 1933. Immediate Pneumococcus Typing Directly from Sputum by the Neufeld Reaction. J Amer Med Ass **100(20):** 1584–1586.

Saha, S. K., Darmstadt, G. L., Baqui, A. H., Hossain, B., Islam, M., Foster, D., Al-Emran, H., Naheed, A., Arifeen, S. E., Luby, S. P., Satosham, M. and Crook, D. 2008. Identification of serotype in culture negative pneumococcal meningitis using sequential multiplex PCR: implementation for surveillance and vaccine design. PLoS One **3**: e3576.

Salter, S. J., Hinds, J., Gould, K. A., Lambertsen, L., Hanage, W. P., Antonio, M., Turner, P., Hermans, P. W., Bootsma, H. J., O'Brien, K. L. and Bentley, S. D. 2012. Variation at the capsule locus, cps, of mistyped and non-typable *Streptococcus pneumoniae* isolates. Microbiol **158(6)**: 1560-1569.

Schaffner, A., Michel-Harder, C. and Yeginsoy, S. 1991. Detection of capsular polysaccharide in serum for the diagnosis of pneumococcal pneumonia: clinical and experimental evalution. J Infect Dis **163:** 1094-1102.

Selva, L., del Amo, E., Brotons, P. and Munoz-Almagro, C. 2012. Rapid and easy identification of capsular serotypes of *Streptococcus pneumoniae* by use of fragment analysis by automated fluorescence-based capillary electrophoresis. J Clin Microbiol **50:** 3451-3457.

Shainheit, M. G., Valentino, M. D., Gilmore, M. S. and Camilli, A. 2015. Mutations in pneumococcal *cpsE* generated via *in vitro* serial passaging reveal a potential mechanism of reduced encapsulation utilized by a conjunctival isolate. J Bacteriol **197(10):** 1781-1791.

Sheppard, C. L., Pichon, B., George, R. C. and Hall, L. M. C. 2010. *Streptococcus pneumoniae* isolates expressing a capsule with epitopes of both serotypes 6A and 6B. Clin Vacc Immunol **17(11):** 1820-1822.

Siira, L., Kaijalainen, T., Lambertsen, L., Nahm, M. H., Toropainen, M. and Virolainen, A. 2012. From Quellung to multiplex PCR, and back when needed, in pneumococcal serotyping. J Clin Microbiol **50(8):** 2727-2731.

Simell, B., Auranen, K., Kayhty, H., Goldblatt, D., Dagen, R. and O'Brien, K. L. 2012. The fundamental link between pneumococcal carriage and disease. Expert Rev Vaccines **11(7):** 841-855.

Skovsted, I. C. 2015. *Streptococcus pneumoniae*, p 1-48. *In* Everdal, P. L. *et al.* (ed.) Textbook in diagnosis, serotyping, virulence factors and enzyme-linked immunosorbant assay (ELISA) for measuring pneumococcal antibodies, **2nd ed.** Statens Serum Institut, Denmark.

Slotved, H. C., Kaltoft, M., Skovsted, I. C., Kerrn, M. B. and Espersen F. 2004. Simple, rapid latex agglutination test for serotyping of pneumococci (Pneumotest-Latex). J Clin Microbiol **42:** 2518–2522.

Smart, L. E. and Henrichsen, J. 1986. An alternative approach to typing of *Streptococcus pneumoniae* strains by coagulation. Acta Pathol Microbiol Immunol. **94:** 409-413.

Song, J. Y., Nahm, M. H. and Moseley, M. A. 2012. Clinical implications of pneumococcal serotypes: Invasive disease potential, clinical presentations, and antibiotic resistance. J Korean Med Sci **28:** 4-15.

Sorensen, U.B. 1993. Typing of pneumococci by using 12 pooled antisera. J Clin Microbiol **31**: 2097–2100.

Sorensen, U. B. S., Yao, K., Yang, Y., Tettelin, H. and Kilian, M. 2016. Capsular polysaccharide expression in commensal *Streptococcus* species: genetic and antigenic similarities to *Streptococcus pneumoniae*. mBio **7(6):** e01844-16.

Sorensen, U. B. S., Henrichsen, J., Chem, H. C. and Szu, S. C. 1990. Covalent linkage between the capsular polysaccharide and the cell wall peptidoglycan of *Streptococcus pneumoniae* revealed by immunochemical methods. Microb. Pathog **8:** 325-334.

Spellerberg, B. and Brandt, C. 2015. *Streptococcus,* p 383-394. *In* Dumler (ed), Manual of Clinical Micrbiology (MCM), **11th ed, vol 1**. Herndon, VA.

Spencer, B. L., Saad, J. S., Shenoy, A. T., Orihuela, C. J. and Nahm, M. H. 2017. Position of O-acetylation within the capsular repeat unit impacts the biological properties of pneumococcal serotypes 33A and 33F. Infect Immunol doi: 10.1128/IAI.00132-17.

Stein, K. E. 1992. Thymus-independent and thymus-dependent responses to polysaccharide antigens. J Infect Dis **165**(Suppl 1):S49–S52.

Sternberg, G. M. 1881. A fatal form of septicaemia in the rabbit, produced by the subcutaneous injection of human saliva. Annual Reports of the National Board of Health **3**: 87-108.

Syrjanen, R. K., Auranen, K. J., Leino, T. M., Kilpi, T. M. and Makela, P. H. 2005. Pneumococcal acute otitis media in relation to pneumococcal nasopharyngeal carriage. Pediatr Infect Dis J **24(9):** 801-806.

Tomcyzk, S., Bennett, N. M., Stoecker, C., Gierke, R., Moore, M. R., Whitney, C. G., Hadler, S. and Pilishvili, T; Centre for Disease Control and Prevention (CDC). 2014. Use of 13-valent pneumococcal conjugate vaccine and 23-valent pneumococcal polysaccharide vaccine among adults aged ≥ 65 years: Recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Morb Mortal Wkly Rep **63:** 822-825.

Tomita, Y., Okamoto, A., Yamata, K., Yagi, T., Hasegawa, Y. and Ohta, M. 2011. A new microarray system to detect *Streptococcus pneumoniae* serotypes. J Biomed Research International **11**; doi 10.1155/2011/352736.

Trzciński, K., Bogaert, D., Wyllie, A., Chu M. L. J. N., van der Ende, A., Bruin, J. P., van den Dobbelsteen, G., Veenhoven, R. H. and Sanders, E. A. M. 2013. Superiority of trans-oral over trans-nasal sampling in detecting *Streptococcus pneumoniae* colonization in adults. PLoS One **8(3):** e60520.

Tyrrell, G. J., Lovgren, M., Chui, N., Minion, J., Garg, S., Kellner, J. D. and Marrie, T. J. 2009. Serotypes and antimicrobial susceptibilities of invasive *Streptococcus pneumoniae* pre- and post-seven valent pneumococcal conjugate vaccine introduction in Alberta, Canada 2000-2006. Vaccine **27:** 3553-3560.

Vammen, B. 1939. Serological variants of pneumococcus types 9 and 10. J Immunol **37:** 359–365.

Van Selm, S., van Cann, L. M., Kolkman, M. A. B., van der Zeijst, B. A. M. and van Putten, J. P. M. 2003. Genetic basis for the structural difference between *Streptococcus pneumoniae* serotype 15B and 15C capsular polysaccharides. Infect and Immun **71(11):** 6192-6198.

Venkateswaran, P. S., Stanton, N. and Austrian, R. 1983. Type variation of strains of *Streptococcus pneumoniae* in capsular serogroup 15. J Infect Dis **147:** 1041–1054.

Viguera, E., Canceill, D. and Ehrlich, S. D. 2001. Replication slippage involves DNA polymerase pausing and dissociation. EMBO J **20:** 2587–2595.

Wang, Q., Wang, M., Kong, F., Gilbert, G. L., Cao, B., Wang, L. and Feng, L. 2007. Development of a DNA microarray to identify the *Streptococcus pneumoniae* serotypes contained in the 23-valent pneumococcal polysaccharide vaccine and closely related serotypes. J Microbiol Methods **68:** 128-136.

Watson, D. A., Musher, D. M., Jacobson, J. W. and Verhoef, J. 1993. A brief history of the pneumococcus in biomedical research: A panoply of scientific discovery. Clin Infect Dis **17:** 913-924.

Weiser, J. N., Austrian, R., Sreenivasan, P. K. and Masure, H. R. 1994. Phase variation in pneumococcal opacity: relationship between colonial morphology and nasopharyngeal colonization. Infect Immun **62:** 2582-2589.

World Health Organization (WHO). 2012. Estimated Hib and pneumococcal deaths for children under 5 years of age, 2008. Available online: http://www.who.int/immunization/monitoring_surveillance/burden/estimates/Pneumo_hib/en/. [Last accessed June 1 2017]

Winkelstein, J. A. 1981. The role of complement in the host's defense against *Streptococcus pneumoniae.* Rev Infect Dis **3:** 289-298.

Winslow, C. E. A., Broadhurst, J., Buchanan, R. E., Krumwiede, C. Jr., Rogers, L. A. and Smith, G. H. 1920. The families and genera of the bacteria: final report of the committee of the Society of American Bacteriologists on characterization and classification of bacterial types. J Bacterial **5 (3):** 191-229.

Wyllie, A. L., Rümke, L. W., Arp, K., Bosch, A. A., Bruin, J. P., Rots, N.Y., Wijmenga-Monsuur, A. J., Sanders, E. A. and Trzciński, K. 2016. Molecular surveillance on *Streptococcus pneumoniae* carriage in non-elderly adults; little evidence for pneumococcal circulation independent from the reservoir in children. Sci Rep 2016 **6**: 34888.

Yother, J. 2011. Capsules of *Streptococcus pneumoniae* and other bacteria: paradigms for polysaccharide biosynthesis and regulation. Annu Rev Microbiol **65:** 563-581.

Yun, K. W., Cho, E. Y., Choi, E. H. and Lee, H. J. 2014. Capsular polysaccharide gene diversity of pneumococcal serotypes 6A, 6B, 6C, and 6D. Int J Med Microbiol **(8):** 1109-1117.

Zartler, E. R., Porambo, R. J. , Anderson, C. L., Chen, L. H., Yu, J., and Nahm, M. H. 2009. Structure of the capsular polysaccharide of pneumococcal serotype 11A reveals a novel acetylglycerol that is the structural basis for 11A subtypes. J Biol Chem **284:** 7318–7329.

Zerbino, D. R. and Birney, E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome research **18**: 821-829.

Zhou, F., Kong, F., Tong, Z. and Gilbert, G. L. 2007. Identification of less-common *Streptococcus pneumoniae* serotypes by multiplex PCR-based reverse line blot hybridization assay. J Clin Microbiol **45:** 3411

Appendix

Table S1. Antisera required for serotyping *S. pneumoniae* using the Quellung reaction.

Serotype	Quellung antisera				Relevant to this study
	Pool	Type	Group	Factor	
1	A, P	1	N/A	N/A	No
2	A, T	2	N/A	N/A	No
3	B,R	3	N/A	N/A	No
4	A,R	4	N/A	N/A	No
5	A,S	5	N/A	N/A	No
6A	B, Q	N/A	6	6b+ 6c- 6d-	Yes
6B				6b- 6c+ 6d-	Yes
6C				6b- 6c- 6d+	Yes
6D				6b- 6c+ 6d+	Yes
7F	C, P	N/A	7	7b+ 7c- 7e- 7f-	Yes
7A				7b(+) 7c+ 7e- 7f-	Yes
7B				7b- 7c- 7e+ 7f-	Yes
7C				7b- 7c- 7e- 7f+	Yes
9A	D,R	N/A	9	9b- 9d+ 9e- 9g-	Yes
9V				9b- 9d+ 9e- 9g+	Yes
9N				9b+ 9d- 9e+ 9g-	Yes
9L				9b+ 9d- 9e- 9g-	Yes
10F	E, S	N/A	10	10b+ 10d- 10f-	No
10A				10b- 10d+ 10f-	No
10B				10b+ 10d+ 10f-	No
10C				10b+ 10d- 10f+	No
11F	D, T	N/A	11	11b+ 11c- 11f- 11g+	No
11A				11b- 11c+ 11f- 11g-	Yes
11B				11b+ 11c- 11f+ 11g+	No
11C				11b+ 11c+ 11f+ 11g-	No
11D				11b- 11c- 11f+ 11g-	Yes
12F	E, R	N/A	12	12b+ 12c- 12e-	Yes
12A				12b- 12c+ 12e-	Yes
12B				12b+ 12c+ 12e+	Yes
13	H	13	N/A	N/A	No
14	H, P	14	N/A	N/A	No

Serotype	Quellung antisera				Relevant to this study
	Pool	Type	Group	Factor	
15F	H, S	N/A	15	15b+ 15c+ 15e- 15h-	No
15A				15b- 15c+ 15e- 15h-	No
15B				15b+ 15c- 15e+ 15h+	Yes
15C				15b- 15c- 15e+ 15h-	Yes
16F	D	N/A	16	16b+ 16c-	No
16A				16b- 16c+	No
17F	F, S	N/A	17	17b+ 17c-	No
17A				17b- 17c+	No
18F	A, Q	N/A	18	18c+ 18d- 18e+ 18f+	Yes
18A				18c- 18d+ 18e- 18f-	Yes
18B				18c- 18d- 18e+ 18f-	Yes
18C				18c+ 18d- 18e+ 18f-	Yes
19F	B, P	N/A	19	19b+ 19c- 19f- 7h-	No
19A				19b- 19c+ 19f- 7h-	No
19B				19b- 19c- 19f- 7h+	No
19C				19b- 19c- 19f+ 7h+	No
20	C, T	20	N/A	N/A	No
21	E	21	N/A	N/A	No
22F	F, T	N/A	22	22b+ 22c-	Yes
22A				22b- 22c+	Yes
23F	H, Q	N/A	23	23b+ 23c- 23d-	No
23A				23b- 23c+ 23d-	No
23B				23b- 23c- 23d+	No
24F	C	N/A	24	24c- 24d+ 24e-	No
24A				24c+ 24d+ 24e-	No
24B				24c- 24d- 24e+	No
25F	I	N/A	25	25b+ 25c-	No
25A				25b- 25c+	No
27	F	27	N/A	N/A	No
28F	H	N/A	28	28b+ 28c-	No
28A				28b- 28c+	No
29	G	29	N/A	N/A	No
31	C	31	N/A	N/A	No

Serotype	Quellung antisera				Relevant to this study
	Pool	Type	Group	Factor	
32F	F	N/A	32	32a+ 32b-	No
32A				32a+ 32b+	No
33F	E, T	N/A	33	33b+ 33e- 33f- 6a- 20b-	Yes
33A				33b+ 33e- 33f- 6a- 20b+	Yes
33B				33b- 33e- 33f+ 6a- 20b-	No
33C				33b- 33e+ 33f(+) 6a- 20b-	No
33D				33b- 33e- 33f+ 6a+ 20b-	No
34	G	34	N/A	N/A	No
35F	G	N/A	35	35a+ 35b+ 35c- 29b- 42a-	No
35A				35a+ 35b- 35c+ 29b- 42a-	No
35B				35a+ 35b- 35c+ 29b+ 42a-	No
35C				35a+ 35b- 35c+ 29b- 42a+	No
36	D	36	N/A	N/A	No
37	D	37	N/A	N/A	Yes
38	I	38	N/A	N/A	No
39	E	39	N/A	N/A	No
40	C	40	N/A	N/A	No
41F	F	N/A	41	41a+ 41b+	No
41A				41a+ 41b-	No
42	G	42	N/A	N/A	No
43	I	43	N/A	N/A	No
44	I	44	N/A	N/A	No
45	I	45	N/A	N/A	No
46	I	46	N/A	N/A	No
47F	G	N/A	47	47a+ 43b-	No
47A				47a+ 43b+	No
48	I	48	N/A	N/A	No