Molecular Characterization of Twelve Predominant *Streptococcus pneumoniae*

Serotypes Causing Invasive Infections in Canada from 2011 to 2015

by

Alyssa Rose Golden

A Thesis Submitted to the Faculty of Graduate Studies of

The University of Manitoba

in Partial Fulfillment of the Requirements of the Degree of

Doctor of Philosophy

Department of Medical Microbiology and Infectious Diseases

Rady Faculty of Health Sciences

University of Manitoba

Winnipeg, MB, Canada

ABSTRACT

Streptococcus pneumoniae is a highly diverse and commonly antimicrobial-resistant organism capable of causing invasive disease. The purpose of this thesis was to thoroughly characterize twelve predominant serotypes of *S. pneumoniae* causing invasive infections in Canada from 2011 to 2015. These serotypes included 3, 6C, 7F, 8, 9N, 11A, 12F, 15A, 19A, 22F, 33F and 35B. We hypothesized that these common serotypes would frequently be antimicrobial resistant and/or multidrug resistant (MDR), genetically similar within the serotype and virulent due to the presence of pneumococcal pilus genes.

Invasive *S. pneumoniae* isolates were obtained from the SAVE study (*S. pneumoniae* Serotyping and Antimicrobial Susceptibility: Assessment for Vaccine Efficacy in Canada) through a partnership between the Canadian Antimicrobial Resistance Alliance and the National Microbiology Laboratory. All strains were tested for antimicrobial susceptibilities and pneumococcal pilus presence. A subset of the above-listed twelve serotypes were characterized using PFGE, MLST and WGS. WGS analyses included phylogeny, recombination prediction and resistance and virulence gene extraction. Capsular switch variants were identified using a penicillin-binding protein typing scheme.

The hypotheses of this study were found to be closely linked based on the serotype in question. Serotypes 6C, 15A, 19A and 33F were found to be highly diverse, demonstrating numerous sequence types within the serotype and the most significant amounts of predicted recombination. These four serotypes accounted for most of the MDR *S. pneumoniae* identified in this thesis; in contrast, serotypes such as 7F and 22F were highly clonal and uncommonly resistant to antimicrobials. The presence of pilus genes was also associated with clonal type, but variably associated with MDR and virulence. PI-1 was identified in a subset of serotype 6C and

in antimicrobial susceptible 15A isolates. PI-2 genes were present in all serotype 7F and two-thirds of 11A isolates. The dual genotype was only present in extensively drug resistant serotype 19A.

Recombination had a significant impact on the diversity of the Canadian pneumococcal population over the course of this project, as was evident by the numerous capsular switch variants identified by this study. Capsule switching led to vaccine escape and increased diversity through acquisition of resistance determinants or MDR phenotypes not normally associated with certain serotypes. WGS was highly discriminatory and provided a wealth of information not normally accessible using traditional sub-typing methods. This powerful tool will become increasingly important in the surveillance of pathogens in Canada, particularly for elucidating genetic differences within the diverse pneumococcal population.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

5%-SBA	Trypticase soy agar with 5% sheep blood
%I	% Intermediate
%NS	% Nonsusceptible
%R	% Resistant
%S	% Susceptible
ABCs	Active Bacterial Core Surveillance
AMC	Amoxicillin/clavulanic acid
ATCC	American Type Culture Collection
BESST	Baseline Epidemiology of *Streptococcus pneumoniae* Serotypes in Canada prior to the Introduction of the 13-valent Pneumococcal Vaccine
BLAST	Basic Local Alignment Search Tool
bp	Basepair
CARA	Canadian Antimicrobial Resistance Alliance
CAP	Community-acquired pneumonia
CANWARD	Canadian Ward Surveillance Study
CC	Clonal cluster
CDC	Centers for Disease Control and Prevention
CFU	Colony-forming unit
cgMLST	Core genome multi-locus sequence typing
CHEF	Contour-clamped homogeneous electric field
CHL	Chloramphenicol
CLD	Clindamycin
CLR	Clarithromycin
CLSI	Clinical and Laboratory Standards Institute
cm	Centimetre
CPHLN	Canadian Public Health Laboratory Network
CSF	Cerebrospinal fluid
CSP	Competence-stimulating peptide
dH$_2$O	Distilled H$_2$O
DLV	Double-locus variant
dNTP	Deoxynucleotide triphosphates
DNA	Deoxyribonucleic acid
DOX	Doxycycline
dsDNA	Double-stranded DNA
EDTA	Ethylenediaminetetraacetic acid
EMBOSS	European Molecular Biology Open Software Suite
EUCAST	European Committee on Antimicrobial Susceptibility Testing
g	Gram
GDR	Gender
GPS	Global Pneumococcal Sequencing Project
Hib	*Haemophilus influenzae* type b
Ig	Immunoglobulin
IPD	Invasive pneumococcal disease
IS	Insertion sequence

iv	Intravenous
Kb	Kilobase
LEV	Levofloxacin
Mb	Megabase
MDR	Multidrug resistant
mg	Milligram
MIC	Minimum inhibitory concentration
mL	Millilitre
MLST	Multi-locus sequence typing
mm	Millimetre
mM	Millimolar
NCBI	National Center for Biotechnology Information
ng	Nanogram
nm	Nanometre
PBP	Penicillin-binding protein
PCR	Polymerase chain reaction
PCV	Pneumococcal conjugate vaccine
PEN	Penicillin
PFGE	Pulsed-field gel electrophoresis
pH	Power of hydrogen
PHAC-NML	Public Health Agency of Canada – National Microbiology Laboratory
PI-	Pilus-encoding island
PMEN	Pneumococcal Molecular Epidemiology Network
PPV	Pneumococcal polysaccharide vaccine
PROV	Province
QRDR	Quinolone resistance-determining region
rRNA	Ribosomal-ribonucleic acid
SAVE	*Streptococcus pneumoniae* Serotyping and Antimicrobial Susceptibility: Assessment for Vaccine Efficacy in Canada
SLV	Single-locus variant
SNV	Single nucleotide variant
SPN	*Streptococcus pneumoniae*
SRST	Short-read sequence typing
ssDNA	Single-stranded DNA
ST	Sequence type
SXT	Trimethoprim-sulfamethoxazole
TAE	Tris-acetate-EDTA
TBE	Tris-borate-EDTA
TE	Tris-EDTA
TZP	Piperacillin/tazobactam
UV	Ultraviolet
V	Volt
wgMLST	Whole genome multi-locus sequence typing
WGS	Whole genome sequencing
WHO	World Health Organization
XDR	Extensively drug resistant

YR	Year
μg	Microgram
μL	Microlitre

1. INTRODUCTION

1.1 *Streptococcus pneumoniae*

1.1.1 Colony Morphology and Identification

Streptococci are a group of Gram-positive, catalase-negative, facultative anaerobic bacteria that are usually less than 2 μm in diameter (1, 2). *S. pneumoniae*, a member of the viridans group of streptococci, tends to grow as lancet-shaped diplococci or as short chains of variable length in liquid. Isolates of *S. pneumoniae* are differentiated from other viridans group streptococci by their bile (deoxycholate) solubility and sensitivity to optochin (1, 2). *S. pneumoniae* is grown on media enriched with blood, and is optimally incubated at 35 - 37°C in a 5% CO_2 environment (1). When grown on blood agar, *S. pneumoniae* demonstrates α-hemolysis, a partial degradation of the blood cells identifiable by a green discoloration surrounding the colonies. This degradation is due to the pneumococcal toxin pneumolysin, an enzyme demonstrating hemolytic activity due to its ability to form pores in cholesterol-rich membranes (2). In contrast to other viridans group streptococci, strains of *S. pneumoniae* produce variable amounts of capsular polysaccharide that may cause the colonies to have a wet, or even mucoid, appearance. Many colonies also display a navel-like depression, which is caused by the lytic enzyme autolysin (1).

1.1.2 *S. pneumoniae* Genome and Natural Transformation

S. pneumoniae has a covalently closed, circular genome of approximately 2.1 million basepairs (bp) of DNA, with a low guanine-cytosine content of 34 – 46% (1, 3). The pneumococcal genome demonstrates a high level of plasticity, with strains varying by up to 10%. Strains may possess conjugative transposons that confer antimicrobial resistance to classes such as macrolides and tetracyclines, as well as an unusually large number of insertion sequences that

can account for up to 5% of the genome (3, 4). Additionally, the genome contains numerous copies of direct-repeat DNA elements. These repeats provide recombination hotspots and help account for the genetic variability between strains (3).

Natural transformation is the ability of a bacterial cell to take up exogenous DNA and utilize homologous recombination to incorporate this DNA into its genome (4). Since the original discovery of natural transformation in *S. pneumoniae* by Frederick Griffith in 1928 (5), and the subsequent use of his methods to determine the hereditary nature of DNA in 1944 (6), *S. pneumoniae* has been a model organism for the study of natural transformation (4). Transformation requires that bacterial cells be in a state of competence, which simply means that the appropriate signals have been received to turn on transient synthesis of the numerous proteins necessary in the uptake and recombination of exogenous DNA (4, 7, 8).

The transformation process in *S. pneumoniae* can be divided into three different steps (4). First, the appropriate internal and external signals must be received to turn on competence. ComABCDE create a quorum-sensing-like system that functions as a biological switch to incorporate these signals. Production of the competence-stimulating peptide (CSP) triggers a series of steps that result in the phosphorylation of response regulator ComE (ComE-P) and the secretion of CSP to the extracellular environment. The ComABCDE switch functions as an autocatalytic loop. A build-up of CSP in the extracellular environment causes increased ComE-P binding to the promotor region of the ComABCDE operons, which activates the loop; eventually this causes a rapid increase of ComE-P that binds to the promoter region of early competence genes, thus stimulating competence (4, 7). The extracellular CSP also serves to stimulate competence in the entire population of cells (4). Only certain conditions allow CSP to accumulate enough to turn on competence; these include stressors such as antibiotic or vaccine

use, or the need to repair damaged DNA (4, 9). The second step in pneumococcal transformation involves the expression of about 20 early and late competence genes. ComE-P regulates the induction of early competence genes, including ComX, which controls the expression of the late competence genes (4, 7).

Once ComX has activated the late competence genes, the third part of the pneumococcal transformation process can occur, which is the actual uptake and recombination of DNA. The primary receptor for exogenous DNA on the surface of pneumococcal cells is a type IV-like pilus (4, 8). The non-transforming strand of DNA is degraded by EndA endonuclease, while the complementary strand (ssDNA) is transported into the cell in a 3' to 5' direction. The rate of uptake of DNA in this manner is controlled by the ComFA helicase (4). ssDNA is coated with single-stranded binding proteins to prevent degradation by endogenous nucleases. If the internalized DNA shares sufficient homology with the recipient genome, it will undergo homologous recombination with the assistance of DprA and RecA proteins (4, 8). DprA is also implicated in the gradual termination of the competent state by sequestering ComE-P, thus preventing ComE-driven transcription of competence genes (4).

Interestingly, *S. pneumoniae* has developed mechanisms to ensure it has access to exogenous DNA. One of these mechanisms is fratricide, where competent *S. pneumoniae* cells lyse non-competent "sibling" cells to allow access to homologous donor DNA (4, 7). This is achieved during late competence by the secretion of a murein hydrolase (CbpD) that lyses non-competent pneumococci and closely-related species. Competent cells are protected from lysis by the production of early "immunity" protein ComM (4). This system is restricted to closely related strains, and is thought to both maintain genomic integrity and provide allelic diversity that could enhance the fitness of the community (7).

1.1.3 Pathogenesis and Respiratory Infection

The main ecological niche of *S. pneumoniae* is the nasopharynx (10–12). It colonizes this area asymptomatically as part of the normal flora in healthy individuals (13, 14). The rate of carriage of *S. pneumoniae* is highest in children, peaking at approximately 2-3 years of age and decreasing with age to <10% carriage in adults (10). *S. pneumoniae* is not generally highly contagious; it is transmitted by direct contact with contaminated respiratory droplets or secretions (3, 10, 11). Because it is carried in the respiratory tract, carriage is an important factor in horizontal transmission to other individuals (13); as children have the highest carriage rates, they are an important reservoir for contiguous spread of disease, and the main source of *S. pneumoniae* transmission to adults (11, 12). Incidences of *S. pneumoniae* outbreaks are often seen in closed populations and areas of crowding, such as hospitals, day cares, orphanages, prisons, shelters, military bases, long term care facilities and slums (3, 11).

S. pneumoniae enters the body through the nasal cavity and attaches to the nasopharyngeal epithelium (10). Aside from colonizing this region, *S. pneumoniae* can cause disease by contiguous spread to the sinuses (sinusitis) or middle ear (acute otitis media), or by aspiration into the lungs to cause pneumococcal pneumonia (10, 14). Otitis media and sinusitis are considered to be mild respiratory diseases (10, 15), however pneumococcal pneumonia is a significant cause of morbidity and mortality. The Centers for Disease Control and Prevention (CDC) estimates that *S. pneumoniae* causes up to 36% of community-acquired pneumonia (CAP) in adults, making it the leading cause of CAP. Approximately 400,000 adults are hospitalized with pneumococcal pneumonia each year in the United States (16). Rates of pneumococcal pneumonia in children vary considerably depending on location, however the World Health Organization (WHO) estimates that 8.7-52% of pneumococcal pneumonia cases

occur in infants less than six months of age (15).

1.1.4 Invasive Pneumococcal Disease (IPD)

Individuals who are infected by or carry *S. pneumoniae* are at risk to develop IPD (11). In particular, children less than two years of age, adults over 65 years and immunocompromised individuals are at highest risk for development of invasive disease (11, 12, 16). Invasive disease is defined as the invasion of normally sterile body sites, including the blood stream and sites secondary to the blood stream (15); pneumococcus accomplishes this by penetrating the mucosal barrier of the lower respiratory tract (10). The incidence of IPD varies substantially depending on the country; the CDC reported in 2014 that 90% of IPD in the United States was in adults (17), whereas the WHO estimates that on average, 75% of IPD worldwide is in children less than two years of age (15).

Pneumococcal bacteremia is a significant source of morbidity and mortality in both adults and children. Bacteremia is a common complication of pneumococcal pneumonia, occurring in approximately 25-30% of adult pneumonia cases and accounting for 12-16% of IPD in children (16). Bacteremia also occurs without the presence of pneumonia, and has a mortality rate as high as 60% in elderly patients. In children, bacteremia without an identifiable focus of infection (occult bacteremia) is the most common form of IPD in children less than two, accounting for about 70% of IPD in this age group (16). Another form of IPD is pneumococcal meningitis, which has been the most common form of meningitis in children since the widespread use of the *Haemophilus influenzae* Type b (Hib) vaccine caused a decline in Hib invasive disease (16). *S. pneumoniae* is estimated to cause 50% of bacterial meningitis cases in the United States, with the vast majority of these cases being in children less than two years of age (15, 16). The CDC estimates that the case-fatality rate of pneumococcal meningitis is 8% in children and 22% in

adults, however those who survive commonly suffer from long-term neurological sequelae (15, 16). In addition to these more common manifestations of IPD, *S. pneumoniae* can also be isolated from pleural, peritoneal, pericardial, bone or joint fluid (13, 14).

1.2 Treatment of IPD

In the 1940s, the discovery and mass production of penicillin replaced sulfonamides for treatment of pneumococcal pneumonia; it became the treatment of choice for its efficacy, limited toxicity in patients and uniform susceptibility (18). However, reduced susceptibility to penicillin was identified in clinical isolates in 1965 (19), with full resistance discovered shortly thereafter in 1967 (20). Since this time, there has been increasing resistance to many antimicrobial classes; the selective pressure of antibiotic use drives adaptation in the nasopharynx, where a large and genetically diverse gene pool of streptococcal species is available and can transfer resistance determinants to generate nonsusceptible *S. pneumoniae* (4, 21). Multidrug resistant pneumococci (MDR, defined as resistance to three or more antimicrobial classes) were first identified in pediatric isolates in 1977 (22), and have since become a significant global concern. The Pneumococcal Molecular Epidemiology Network (PMEN), established in 1997, noted that the spread of MDR *S. pneumoniae* was due to independent propagation of a few highly resistant clones (23). These MDR isolates are of particular concern as there are decreased treatment options available for serious invasive infections. In 2013, the CDC published a list of organisms considered to be antibiotic resistance threats in the United States. Drug-resistant *S. pneumoniae* was classified as a serious threat, requiring "prompt and sustained action to ensure the problem does not grow" (24).

Acute bacterial infections are treated empirically with broad-spectrum antimicrobial agents, considering common etiologic pathogens, probability of a certain pathogen and local or

regional resistance trends (21). β-lactam antibiotics are the most widely used and effective antibiotics against pneumococcal infection (21), however due to the high rates of penicillin resistance and MDR, it is crucial to perform antimicrobial susceptibility testing to ensure the right agent is chosen to treat resistant strains.

Empiric therapy for suspected bacterial meningitis is vancomycin plus a third-generation cephalosporin (such as ceftriaxone or cefotaxime) (25). Once susceptibility results are obtained, treatment can be adjusted. If the strain is susceptible to penicillin (MIC ≤0.06 μg/mL), empiric therapy is discontinued in favour of high dose penicillin or ampicillin (3, 25–27). For strains that are resistant to penicillin (MIC ≥0.12 μg/mL), but susceptible to third-generation cephalosporins, vancomycin is discontinued. Lastly, if the strain is resistant to penicillin and nonsusceptible to third-generation cephalosporins, empiric treatment is continued (26, 27). Dexamethasone, a corticosteroid with anti-inflammatory properties, may be included as part of the treatment to help prevent the development of neurological sequelae (3, 27). Cephalosporins are also used for the treatment of bacteremic pneumococcal pneumonia, often in tandem with a macrolide. This combination has been shown to decrease mortality associated with bacteremic pneumococcal pneumonia in adults (21).

1.3 Resistance to Common Therapy

1.3.1 Penicillin and β-Lactams

β-lactam antimicrobials are cell-wall active agents that function by irreversibly binding to the transpeptidase domain of penicillin-binding proteins (PBPs) to prevent peptidoglycan cross-linking (21). Reduced susceptibility and resistance to this class occurs in a stepwise manner, with mutations building within three different genes (*pbp1A*, *pbp2B* and *pbp2X*) to create proteins with decreased affinity for the antimicrobial. Although numerous combinations of transpeptidase

domains exist within the pneumococcal population, certain mutations are commonly associated with certain levels of resistance (21). Alterations within PBP2X confer low-level resistance or reduced susceptibility to most β-lactams, including cephalosporins; conversely, PBP2B lacks the ability to bind cephalosporins and confers only low-level penicillin resistance or reduced susceptibility. Studies have determined that high-level β-lactam resistance requires PBP1A alterations in combination with low affinity PBP2B and PBP2X (21, 28). Studies conducted in the early 2010s in Canada and the United States both noted penicillin resistance to be approximately 14% in isolates from all sources (29, 30). However, the percent nonsusceptibility in the United States was much higher than in Canada, at 35.0% and 19.2% penicillin-nonsusceptible, respectively, highlighting the need for routine susceptibility testing due to regional differences in β-lactam resistance (29, 30).

1.3.2 Macrolides, Lincosamides and Tetracyclines

Macrolide-resistant pneumococci are common in Canada and are growingly more frequently isolated than penicillin-resistant pneumococci in parts of the world, including the United States (21, 29). A recent Canadian study reported almost 13% of blood isolates and over 25% of respiratory isolates as nonsusceptible to macrolide agents (30). Resistance to this antimicrobial class is mediated by two mechanisms, efflux and ribosomal modification. Efflux is facilitated by the presence of *mef* class genes, particularly *mefA* and *mefE*, and is associated with low-level resistance. Methylation of 23S rRNA is mediated by *ermB*, resulting in high-level resistance to macrolides as well as resistance to lincosamides; this is known as the MLS$_B$ phenotype (21). Many isolates possessing both *mefA* and *ermB* have been isolated in Canada (31, 32); this dual genotype has recently been attributed to the presence of the *Tn*2010 element (33). *Tn*2010 is a derivative of *Tn*916 transposons, a family known for carrying *tetM*. For this reason,

ermB-mediated macrolide resistance and tetracycline resistance are frequently identified together. *TetM* functions by protecting the 30S ribosome from antimicrobial binding; *tetO* functions similarly, however this gene has rarely been reported in pneumococci (21). Canadian studies have reported low levels of tetracycline resistance, with <5% of blood isolates and <7% of isolates from all sources demonstrating nonsusceptibility to doxycycline (30, 34); in contrast, a recent American study identified 19.9% of *S. pneumoniae* isolates to be tetracycline-resistant (29).

1.3.3 Fluoroquinolones

Resistance to fluoroquinolones in *S. pneumoniae* is primarily due to point mutations in the genes of type II topoisomerase enzymes, particularly DNA topoisomerase IV (*parC*) and DNA gyrase (*gyrA*). Alterations to the quinolone resistance-determining regions (QRDR, the domain bound to DNA during enzymatic action) of these genes prevent fluoroquinolone binding, thus allowing DNA replication to continue (21, 35). Interestingly, in *S. pneumoniae* different fluoroquinolones have preferential targets for antibiotic action; ciprofloxacin and levofloxacin preferentially target *parC*, while moxifloxacin preferentially binds *gyrA* (21, 35). The most commonly noted alterations to these genes that confer resistance to fluoroquinolones are both serine point mutations, at residue 79 of *parC* and residue 81 of *gyrA* (36). Although antibiotic efflux has been studied as a cause of low-level fluoroquinolone resistance, this mechanism is not well characterized (21). Resistance to fluoroquinolones is generally low due to the high fitness cost associated with point mutations in DNA replication genes (21); in Canada and the United States, the rate of resistance has been reported at <1% (29, 30, 37).

1.3.4 Folic Acid Pathway Inhibitors

Trimethoprim and sulfamethoxazole are two synergistic agents administered in combination that inhibit successive steps in the pathway to synthesize folate. *S. pneumoniae* develops resistance to each component through chromosomal mutations in different genes. Resistance to trimethoprim develops due to a single amino acid substitution (Ile100Leu) in *folA*, the gene for dihydrofolate reductase (21). Sulfamethoxazole resistance is caused by localized insertions of one or two codons between Arg58 and Ile66 of *folP* (encoding dihydropteroate synthase) (38). Presence of one of these alterations individually has been demonstrated to confer intermediate resistance to trimethoprim-sulfamethoxazole, while both alterations are required for full resistance (21). Studies have noted nonsusceptibility to trimethoprim-sulfamethoxazole to be much higher in the United States than Canada, with 28.8% and 14.1% of all isolates demonstrating nonsusceptibility, respectively (29, 30).

1.3.5 Vancomycin, Linezolid and Other Options

Vancomycin is considered an antimicrobial of last resort and is a crucial component of the empiric treatment regimen for bacterial meningitis. To date, there have been no reports of vancomycin resistance in *S. pneumoniae*; however, several studies have identified vancomycin-tolerant clinical isolates. Although the full mechanism is unclear, tolerance seems to include a combination of defective autolysin production and alterations within genes associated with cell death (*vncS*, *vex2* and *pep27*) (39–41).

Linezolid-nonsusceptible pneumococci are rare, accounting for <1% of isolates tested over a 12-year period in the United States (42). Reduced susceptibility has been associated with ribosomal protein modification, including *cfr*-mediated methylation of the 23S rRNA, mutations in the 23S rRNA and alterations or deletions in *rplD* (21, 42).

Chloramphenicol resistance in *S. pneumoniae* occurs through enzymatic inactivation of the antimicrobial by production of a chloramphenicol acetyltransferase. This enzyme is encoded by the *cat* gene and is carried on conjugative transposons. Chloramphenicol is rarely used to treat pneumococcal infection due to severe side effects (21).

1.4 Virulence in *S. pneumoniae*

1.4.1 Pneumococcal Capsule

The pneumococcal polysaccharide capsule has long been agreed upon as the most important virulence factor possessed by *S. pneumoniae* (43). There is a large amount of diversity amongst pneumococcal capsules, including capsule thickness (ranging between 200 and 400 nm) and composition (44). Extensive study of the capsule has led to the identification of 97 distinct capsular types known as serotypes. Although many serotypes arose from point mutations and serotype divergence, a large proportion of the genetic diversity is due to the ease at which *S. pneumoniae* takes up DNA from its surroundings. Each serotype has the ability to elicit type-specific protective immunity, however there is occasionally cross-reactivity for serotypes within the same serogroup (43). This characteristic is the basis for phenotypic typing of *S. pneumoniae* isolates by the Quellung reaction, which has been in use since the early 1900s (45). Typing antisera are added to a prepared slide and viewed with phase microscopy; cells becoming visible due to capsule swelling and agglutination is considered a positive reaction. Many different specificity levels of antisera are utilized to encompass the incredible diversity of *S. pneumoniae* capsules.

1.4.1.1 Capsule Function and Association with Virulence

Possession of a capsule is essential for the survival of *S. pneumoniae* in blood, and is associated with its ability to cause invasive disease (46). Presence of the capsule reduces

entrapment of the bacterium in mucus, an innate host defense mechanism, facilitating easier invasion. Capsular polysaccharide is highly negatively charged which facilitates repulsion from mucopolysaccharides, which contain large amounts of positively charged sialic acid (44). The pneumococcal capsule also functions to impair complement and phagocytosis. Capsular polysaccharide blocks the deposition of complement factor C3b on the cell surface (47); if any does manage to attach to the cell surface, the capsule limits the recognition of C3b by other complement factors (48). It also functions to prevent the activity of opsonins, which limits the extent to which cells are phagocytosed (47).

S. pneumoniae is both a pathogenic and commensal organism; because of this, it has adaptive mechanisms to survive in both environments. As the capsule functions largely as a protective mechanism, maximal expression of capsule polysaccharide is necessary for systemic virulence. Conversely, less capsule polysaccharide is required for adequate adherence and survival in the nasopharynx (44, 48). This relationship has been verified by *in vitro* studies which have shown that strains that produce more capsule *in vitro* are more virulent *in vivo* (47). However, many strains that frequently colonize are also significant sources of invasive disease. This is possible due to phase variation, which involves the spontaneous transition of the bacterium from a "transparent" to "opaque" morphology. The transparent phase produces a thinner, less abundant capsule, and is selected for when increased adherence to epithelial cells is necessary to promote colonization. The opaque morphology is selected for during the transition between the mucosal surface and the bloodstream. This morphology contains a thicker capsule, which makes the pathogen more resistant to opsonophagocytosis and therefore more virulent and better suited to cause invasive disease (43, 44). Phase variation also helps to regulate the expression of capsule, which is a significant metabolic burden to produce (43). Additionally, a

recent study determined that capsule size is correlated to frequency of recombination, with larger capsules being associated with more recombination events (49).

1.4.1.2 Capsule Locus

The genetic locus that contains the genes for capsule production is called *cps*. This single transcriptional region is, on average, approximately 20.7kb in length and can be found between the genes *dexB* and *aliA*, which do not participate in capsule synthesis (46). The first four genes in the cps locus are *cpsABCD*, which are widely conserved sequences in almost all *S. pneumoniae* genomes (43). The four proteins encoded by these genes function to regulate capsule expression level, including processing, export and attachment (46, 50). Most of the potential coding sequences in the central part of the locus encode functional transferases, while the 3'-end genes encode sugar precursors (44, 46, 48). These downstream genes are serotype specific, and are often conserved within a serogroup. The functional promoter for the *cps* locus is located directly upstream of *cpsA*, and the unit is transcribed as an operon (48).

There are two pathways by which capsular polysaccharides are synthesized. The first is the Wzy-dependent pathway, so named due to its reliance on the *cps*-encoded Wzy polymerase (44). This method is utilized by all but two pneumococcal serotypes, resulting in complicated polymers that contain multiple types of sugar units, glycosidic linkages and branching (48). Briefly, sugar repeat units are synthesized by glycosyltransferases on the inner membrane of the cell; units are then flipped to the outer membrane so they may be linked into the full polysaccharide chain by Wzy polymerase (48). Serotypes 3 and 37 are the only two serotypes currently known to use the synthase pathway. This pathway is much simpler than the Wzy-dependent pathway in that the repeat units are single sugars or alternations between two sugars (48, 50). Synthesis by this mechanism is highly processive, with a single enzyme transferring

sugars directly to a lipid acceptor to elongate the polysaccharide (43, 46).

1.4.2 Pneumococcal Pili

The pilus is a structure that is most commonly associated with Gram-negative bacteria, however, pili have recently been discovered in Gram-positive pathogens. The Gram-positive pilus-like structure was first seen in *Corynebacterium renale*, and was soon after found to be present in many *Streptococcus* species (51). A subset of *S. pneumoniae* isolates have long, rod-like pili that mediate adherence to epithelial cells and interact with cellular matrix proteins (51, 52). Because the polysaccharide capsule is often abundant and prevents colonization, many *S. pneumoniae* serotypes must rely on other structures to mediate adherence; studies have shown that serotypes with pili, composed of covalently linked pilin subunits, are able to outcompete pilus-deficient strains in mouse models of colonization (52, 53).

Analysis of the genomic organization of invasive streptococci suggest that the genes necessary to express pili were acquired by horizontal transfer of a pathogenicity island (51). There are two different genetic regions that encode pilus proteins in *S. pneumoniae*. The first, and most well characterized, is the 14kb *rlrA* pathogenicity island that encodes pilus-type 1 (PI-1). This island encodes a transcriptional regulator, three pilin subunits (RrgABC), and three sortase enzymes (SrtBCD) (52). A second, more recently discovered pilus-encoding region is found between the genes *pepT* and *hemH*. This 6.6kb island contains two genes encoding sortase proteins, two pilin subunits (PitAB) and a signal peptidase (SipA) which make up pilus-type 2 (PI-2) (52, 54). Pili were originally only identified in a small proportion of *S. pneumoniae* isolates. An early study by Bagnoli *et al*. collected a total of 305 *S. pneumoniae* isolates and reported that 31.5% and 16.4% of isolates contained PI-1 and PI-2, respectively (52).

1.4.3 Other Virulence Factors of Interest

In addition to the pneumococcal capsule and pili, *S. pneumoniae* isolates possess many virulence factors that participate in both colonization and active infection. One of the most well studied groups of pneumococcal virulence factors are the surface proteins. Like the capsule, surface proteins are essential for immune evasion, with many studies noting decreased virulence in surface protein knockout mutants (44). Pneumococcal surface protein A (PspA) binds C-reactive protein to block complement deposition from the classical pathway, thus preventing opsonization (55). This protein is ubiquitous on the cell-surface of pneumococci, however it is associated with a high degree of variability between strains and clones. Based on the sequence of the N-terminal region, PspA can be grouped into three families and six subsequent clades (44). Pneumococcal surface protein C (PspC) is similar in structure to PspA. It is also able to inhibit complement deposition by binding factor H of the alternative pathway (55, 56). Importantly, PspC also interacts with polymeric immunoglobulin (Ig) receptors and secretory IgA. Studies have determined that this interaction may be critical to the translocation of *S. pneumoniae* from the nasopharynx to normally sterile tissues (56).

A second group of key virulence factors are the neuraminidases (or sialidases), proteins which are crucial to pneumococcal colonization. *S. pneumoniae* is known to produce at least three neuraminidases, two of which are well characterized (57). Neuraminidase A (NanA) is present in all pneumococci while Neuraminidase B (NanB) is present in most *S. pneumoniae* strains (44). Both proteins function to cleave terminal sialic acid residues from various substrates to expose host cell surface receptors for adherence, thus promoting colonization (57–59). *In vivo*, these proteins likely have different functions, as suggested by the unique pH optima possessed by each protein and the surface anchorage domain possessed only by NanA (44, 57, 59). Recent

studies have noted that *nanB* deficient mutants demonstrated reduced NanA activity, suggesting that NanB may possess some role in regulating the expression of *nanA* (57). Neuraminidase C is currently not well characterized and there is no direct evidence of a biological role, however it may play a tissue-specific role in meningitis (44, 58).

1.5 Pneumococcal Vaccines and their Implications

As previously described, *S. pneumoniae* is a significant cause of morbidity and mortality worldwide. The first true vaccine against pneumococcal disease was a whole cell preparation developed in 1911. Though the discovery of penicillin provided excellent treatment against pneumococcal infection, the rapid development of resistance increased the desire to create new polysaccharide vaccine formulations in the mid- to late-1900s (60).

1.5.1 Pneumococcal Polysaccharide Vaccine

The 23-valent pneumococcal polysaccharide vaccine (PPV23) was licensed in 1983 and is still in use today (60). Included in this vaccine are serotypes 1, 2, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19F, 19A, 20, 22F, 23F and 33F. This vaccine is recommended by the Canadian National Advisory Committee on Immunization (NACI) for routine preventative use in adults over the age of 65 (61). Unfortunately, PPV23 is not suitable for use in children as it does not adequately stimulate the immune system; bacterial polysaccharides are T-cell-independent antigens which rely on immediate B-cell differentiation into antibody-producing cells. As T-cells are not contacted, B-cells cannot differentiate into memory cells that can rapidly respond and produce antibodies upon subsequent exposure to antigen (60).

1.5.2 Conjugate Vaccines

Early research demonstrated that coupling polysaccharides to a carrier protein greatly enhanced their immunogenicity by eliciting a T-cell-dependent immune response. The strongly immunogenic carrier protein is taken up by antigen presenting cells, degraded into peptides and presented to helper T-cells, which become activated. Simultaneously, polysaccharide molecules are recognized by specific B-cells, broken down and presented on the cell surface; activated T-cells recognize the complex on the B-cell surface, and are stimulated to differentiate into plasma cells that secrete polysaccharide specific antibodies. A subset of these B-cells mature into memory B-cells that are able to respond to subsequent exposure to antigen. Additionally, stimulation of mucosal immunity has the benefit of eradicating serotypes of interest from nasopharyngeal carriage (60). As of 2017, three pneumococcal conjugate vaccines (PCV) have been licensed for use and one is undergoing clinical trials. Most PCVs are coupled to cross-reactive material CRM_{197}, a non-toxic variant of diphtheria toxoid isolated from *Corynebacterium diphtheriae* (62); however, one formulation contains polysaccharides conjugated to one of *H. influenzae* protein D, diphtheria toxoid or tetanus toxoid (60).

1.5.2.1 PCV-7

The first PCV was licensed in the United States and Canada in 2000 and 2001, respectively. PCV-7 included the seven serotypes most commonly associated with IPD at the time (4, 6B, 9V, 14, 18C, 19F and 23F), which were responsible for over 80% of invasive infections in North American children (60, 63). PCV-7 was added to the routine immunization schedule gradually in Canada, reaching all provinces between 2002 and 2005 (63). Active vaccination with PCV-7 resulted in significant overall decreases in IPD incidence in children, particularly due to vaccine and vaccine-related serotypes (63, 64). Indirect effects were also

noted, particularly herd protection; as PCV-7 reduced asymptomatic carriage of vaccine serotypes in the nasopharynx, transmission from children to other susceptible individuals was subsequently reduced. Studies noted a decrease in prevalence of IPD in non-vaccinated age groups such as siblings, parents and grandparents (64).

1.5.2.2 Serotype Replacement and Capsular Switching

Although PCV-7 was successful at decreasing the overall prevalence of IPD, vaccine use resulted in a major shift in the distribution of serotypes causing IPD. The primary reason for this was the phenomenon known as serotype replacement. As the prevalence of IPD and carriage caused by vaccine serotypes decreased due to PCV-7 use, non-vaccine serotypes expanded to fill the nasopharyngeal niche vacated by PCV-7 serotypes. Non-vaccine serotypes that increased in prevalence in the post-PCV-7 era included 3, 22F, 7F, 5, 15C and most significantly 19A (63, 65).

The increase in non-vaccine serotypes in the post-PCV-7 era was assisted by capsular switching. As depicted in Figure 1.1, capsular switching involves the acquisition of an alternative capsule locus through recombination (66). Although capsular switching is a regular occurrence in the pneumococcal population, it is most problematic for vaccine effectiveness when the switch involves a vaccine to non-vaccine serotype swap (67). If the recipient strain is a vaccine serotype and it acquires donor DNA containing non-vaccine capsular genes, the resulting progeny strain will no longer possess a capsule included in the vaccine; this strain would be deemed a "vaccine escape recombinant". An early study by Brueggemann *et al*. characterized vaccine escape recombinants that emerged shortly after PCV-7 introduction in 2003, where virulent vaccine serotype 4 strains acquired the capsule of serotype 19A. This new clone was highly successful and rapidly expanded to help facilitate the increasing prevalence of serotype 19A (66).

Importantly, capsular switch recombination is a key factor in the transmission of penicillin resistance. As two of the three PBPs responsible for penicillin resistance reside on either side of the capsule locus, one (or both) of these genes could be transferred along with the capsule locus into the progeny strain (Figure 1.1).

FIGURE 1.1: Diagrammatic representation of pneumococcal capsular switching.
A segment of DNA containing the capsule locus obtained from a "donor" strain is recombined
into the "recipient" strain to generate "progeny". Putative recombination points for progeny are
indicated with arrows.
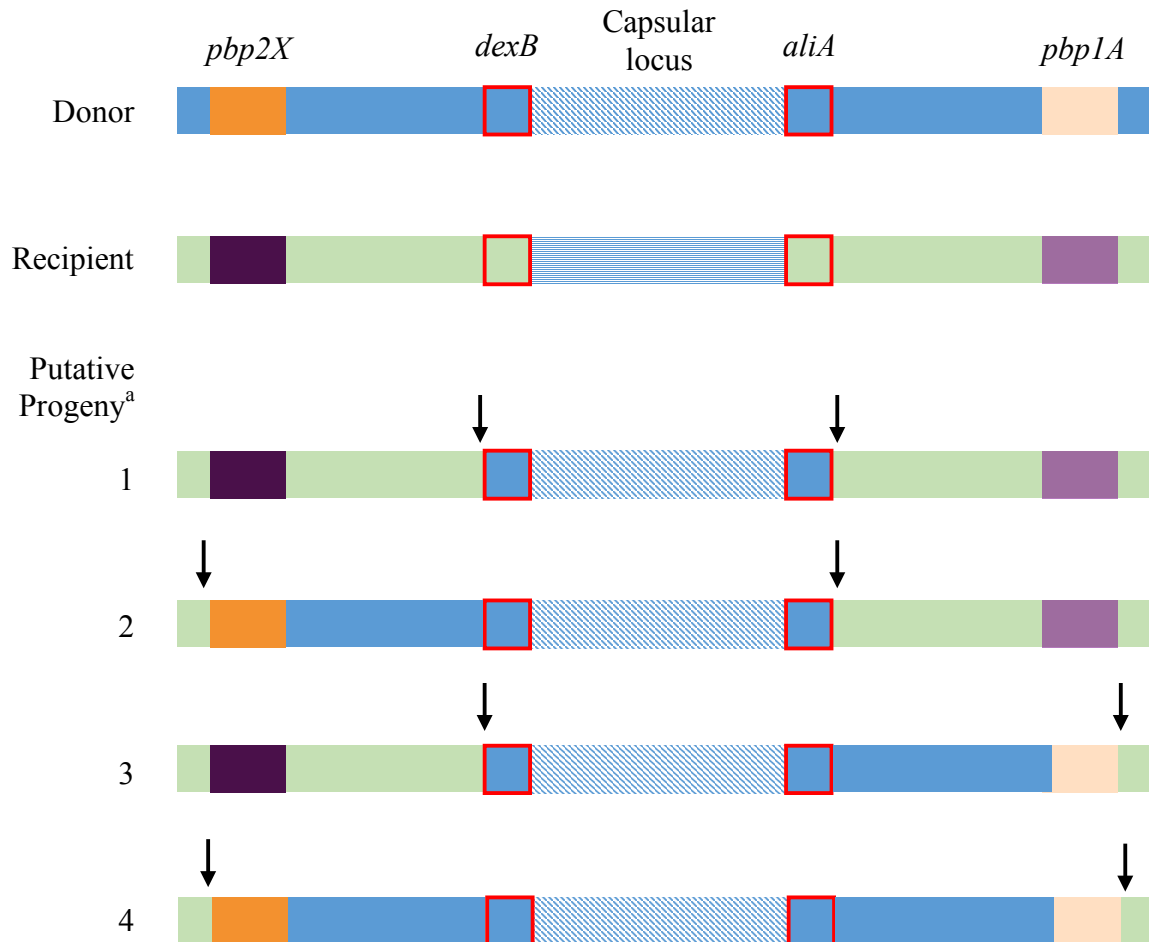
[a] 1 - transfer of capsule locus only; 2 – transfer of segment including *pbp2X* up to and including
the capsule locus; 3 – transfer of segment including capsule locus up to and including *pbp1A*; 4 –
transfer of entire segment from *pbp2X* to *pbp1A*.

1.5.2.3 PCV-13

In response to the serotype shift noted in the post-PCV-7 era, new conjugate vaccine formulations were developed to include those types that had increased in prevalence, including a 10-valent vaccine (PHiD-CV, composed of PCV-7 serotypes plus 1, 5 and 7F). PCV-13 (including PHiD-CV serotypes plus 3, 6A and 19A) was introduced in Canada in 2010 and replaced PCV-7 as the routine vaccine recommended by NACI (68). Guidelines have recently been updated to recommend the use of PCV-13 in vaccine-naïve adults over the age of 65 to protect against CAP and IPD (61). The increased coverage of pneumococcal serotypes provided by PCV-13 is desirable, however with continued use of PCV-13, it is expected that non-vaccine serotypes will again increase in prevalence as a cause of IPD.

1.5.3 Future Vaccines

Following the trend to increase the valency of PCVs with each new formulation, a 15-valent preparation (PCV-13 serotypes plus 22F and 33F) is currently entering Phase II clinical trials in the United States. Several trials in both infants and adults have indicated that the 15-valent formulation demonstrates an acceptable safety profile and induces serotype-specific antibodies and opsonophagocytic killing activity for all 15 serotypes (69, 70). To our knowledge, there is no timeline for introduction of this vaccine to the market.

All currently licensed pneumococcal vaccines are based on the generation of antibodies against capsular polysaccharide, meaning protection is serotype-based. The limited number of serotypes included, worldwide serotype variation and constant threat of serotype replacement or vaccine escape has led many to study the effects of protein-based vaccines as an alternative. Many pneumococcal proteins are universally expressed at high levels during pathogenesis, making them ideal targets in all strains, regardless of serotype. Several pneumococcal protein

vaccine candidates have already entered Phase I or II clinical trials, including PspA, PsaA, pneumolysin and Pht family proteins (71). Additional proteins that have demonstrated immunogenicity are PspC, NanA, PiaA and autolysin (71, 72).

1.6 Evolution of *S. pneumoniae* Genetic Characterization

S. pneumoniae demonstrates incredible genetic diversity resulting in different serotypes, antimicrobial resistance and virulence gene patterns. For this reason, it is incredibly important to genetically characterize pneumococcal isolates to determine the genetic basis of phenotypic traits and the relatedness of isolates within a population. Most frequently used characterization tools are considered "sub-typing" methods, as they provide a small fraction of the total information in a genome (73). Such methods include polymerase chain reaction (PCR) to determine the presence of genes and individual gene sequencing to identify mutations; to determine genetic relatedness, pulsed-field gel electrophoresis (PFGE) and multi-locus sequence typing (MLST) are often used. While MLST utilizes the sequences of seven housekeeping genes to assign an arbitrary sequence type (ST), PFGE is a DNA fingerprinting method that compares restriction digest patterns between isolates. The above listed methods were used successfully in tandem to assemble the PMEN database of global clones, which identifies persistent antimicrobial resistant and MDR clones of *S. pneumoniae* that are widely distributed both nationally and internationally (23). As noted by Klugman in 2002, genetic analysis of these successful international clones has been crucial in understanding the spread of antimicrobial resistance in *S. pneumoniae* (74); the frequency of recombination and overall diversity of the pneumococcal population necessitates that genetic analysis continue on an ongoing basis.

Although many laboratories still rely on sub-typing methods despite the limited amount of information they provide, whole genome sequencing (WGS) is growingly becoming the

22

method of choice to investigate pathogens. WGS is associated with rapidly decreasing costs, short completion times and unambiguous examination of the total genetic content of a strain at the single nucleotide level (73); the latter makes WGS particularly useful for studying *S. pneumoniae* due to its enormous genetic plasticity. Public health and clinical microbiology laboratories are beginning to implement WGS for identification, typing and outbreak investigation, as the universal and unbiased approach removes the need for multiple tests, as all information can be obtained in one assay (73, 75).

1.7 The SAVE Study

1.7.1 Premise, Isolate Collection, Confirmation and Serotyping

The SAVE Study (<u>S</u>. pneumoniae <u>S</u>erotyping and <u>A</u>ntimicrobial Susceptibility: <u>A</u>ssessment for <u>V</u>accine <u>E</u>fficacy in Canada) is an annual, ongoing study that began in 2011, following the introduction of PCV-13 in Canada. Invasive *S. pneumoniae* collected from sterile sites were forwarded from Canadian hospitals and provincial health laboratories (the Canadian Public Health Laboratory Network, CPHLN) to the Streptococcus and STI Unit at the Public Health Agency of Canada – National Microbiology Laboratory (PHAC-NML) for serotyping. Through a collaboration between the Canadian Antimicrobial Resistance Alliance (CARA) and PHAC-NML, isolates from participating CPHLN sites (as listed in Table 1.1) were forwarded to CARA, at the Health Sciences Centre, Winnipeg, MB, as frozen stocks. In total, 6,272 invasive isolates of *S. pneumoniae* collected as part of the SAVE study were forwarded to CARA for antimicrobial susceptibility testing. The annual numbers of isolates were: 1,379 isolates from 2011, 1,285 from 2012, 1,138 from 2013, 1,274 from 2014 and 1,196 from 2015.

TABLE 1.1: Participating CPHLN sites for the SAVE 2011-2015 study.

City, Province	CPHLN Site(s)	Investigator
Regina, Saskatchewan	Saskatchewan Disease Control Laboratory	Dr. G. Horsman
Winnipeg, Manitoba	Cadham Provincial Laboratory	Dr. P. Van Caeseele
Toronto, Ontario	Public Health Ontario	Dr. J. B. Gubbay
Ste-Anne-de-Bellevue, Québec	Laboratoire de santé publique du Québec	Dr. B. Lefebvre
St. John's, Newfoundland and Labrador	Newfoundland and Labrador Public Health Laboratory	Dr. G. Zahariadis

Various cities, New Brunswick [a]	Horizon Health Network
	  Zone 2 (Saint John)
	  Zone 7 (Miramichi)
	Vitalité Health Network
	  Zone 1A (Moncton)
	  Zone 4 (Edmundston)
	  Zone 5 (Campbellton)
	  Zone 6 (Bathurst)
		Dr. R. Gad

Halifax, Nova Scotia	Queen Elizabeth II Health Sciences Centre	Dr. D. Haldane
Charlottetown, Prince Edward Island	Health PEI	Dr. G. German

[a], New Brunswick does not have a central provincial health laboratory, thus isolates are sent from individual hospitals in eight health zones that belong to two networks. CARA receives isolates from six of these zones.

Prior to sending *S. pneumoniae* isolates to CARA, isolate confirmation and serotyping was performed by the Streptococcus and STI Unit at PHAC-NML. Isolates were confirmed as *S. pneumoniae* by colony morphology and α-hemolysis on Trypticase soy agar plates with 5% sheep blood (5%-SBA), as well as the results of Bile Solubility and Optochin Tests (1). Serotyping was performed using the Quellung reaction and standard protocols (45), with commercial pool, group, type and factor specific antisera from the Statens Serum Institut (Copenhagen, Denmark). Isolates that were nontypeable by the Quellung reaction were tested for the presence of the *cpsA* gene using the CDC's polymerase chain reaction methodology (http://www.cdc.gov/ncidod/biotech/strep/pcr.htm) and verified for species by *rpoB* sequence typing (76).

1.7.2 Distribution of Invasive *S. pneumoniae* in Canada

1.7.2.1 Serotypes

The annual serotype distribution of invasive pneumococcal isolates collected as part of the SAVE study is presented as Figure 1.2 and includes serotypes with n≥30. Serotypes for which less than 30 isolates were collected over the 2011-2015 study period included: 29 (n=28), 33A (n=28), 18C (n=26), NT (n=26), 9V (n=25), 14 (n=24), 6B (n=20), 23F (n=18), 1 (n=17), 5 (n=17), 28A (n=17), 13 (n=12), 37 (n=10), 24B (n=8), 11B (n=7), 6D (n=5), 9L (n=3), 10F (n=3), 24 (n=3), 27 (n=3), 22A (n=2), 25A (n=2), 35A (n=2) and one isolate each of serotypes 7A, 7B, 9A, 10B, 12A, 15F, 18A, 35C, 42 and 45.

FIGURE 1.2: Annual serotype distribution of invasive *S. pneumoniae* isolates (n≥30) collected by the SAVE study, 2011-2015. Serotypes are arranged in descending order of total isolates collected.
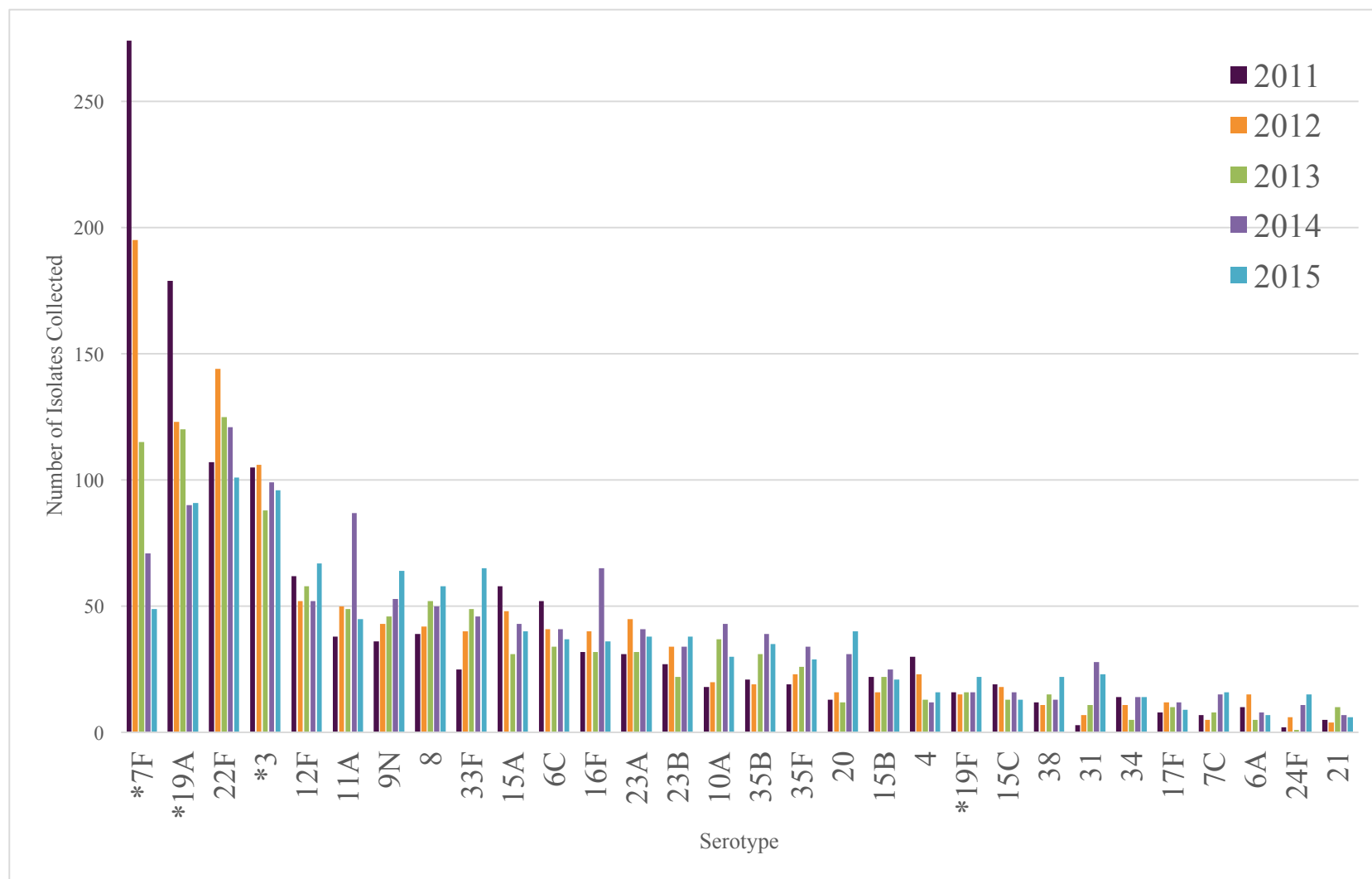
* PCV-13 vaccine serotype.

Overall, the ten most common serotypes collected were 7F (704, 11.2%), 19A (603, 9.6%), 22F (598, 9.5%), 3 (494, 7.9%), 12F (291, 4.6%), 11A (269, 4.3%), 9N (242, 3.9%), 8 (241, 3.8%), 33F (225, 3.6%) and 15A (220, 3.5%). However, the ten most common serotypes differed depending on the year, as depicted in Table 1.2. Over the five-year study period, statistically significant decreases in prevalence were observed for serotypes 7F and 19A, while increases were observed for serotypes 8, 9N, 33F and 35B (77). Additional trend data and *P*-values for the above mentioned increases/decreases can be found in publications specific to the SAVE study and in the 2014 annual report on invasive streptococcal disease published by the Streptococcus and STI Unit (77, 78).

TABLE 1.2: The ten most common serotypes collected annually by the SAVE study, 2011-2015.

	Year (n)				
	2011 (1,379)	2012 (1,285)	2013 (1,138)	2014 (1,274)	2015 (1,196)
Serotype (n, %)	7F* (274, 19.9)	7F* (195, 15.2)	22F (125, 11.0)	22F (121, 9.5)	22F (101, 8.4)
	19A* (179, 13.0)	22F (144, 11.2)	19A* (120, 10.5)	3* (99, 7.8)	3* (96, 8.0)
	22F (107, 7.8)	19A* (123, 9.6)	7F* (115, 10.1)	19A* (90, 7.1)	19A* (91, 7.6)
	3* (105, 7.6)	3* (106, 8.2)	3 * (88, 7.7)	11A (87, 6.8)	12F (67, 5.6)
	12F (62, 4.5)	12F (52, 4.0)	12F (58, 5.1)	7F* (71, 5.6)	33F (65, 5.4)
	15A (58, 4.2)	11A (50, 3.9)	8 (52, 4.6)	16F (65, 5.1)	9N (64, 5.4)
	6C (52, 3.8)	15A (48, 3.7)	11A (49, 4.3)	9N (53, 4.2)	8 (58, 4.8)
	8 (39, 2.8)	23A (45, 3.5)	33F (49, 4.3)	12F (52, 4.1)	7F* (49, 4.1)
	11A (38, 2.8)	9N (43, 3.3)	9N (46, 4.0)	8 (50, 3.9)	11A (45, 3.8)
	9N (36, 2.6)	8 (42, 3.3)	10A (37, 3.3)	33F (46, 3.6)	15A (40, 3.3)

* PCV-13 vaccine serotype.

1.7.2.2 Demographics

The regional distribution of invasive pneumococcal isolates collected as part of the SAVE study is presented as Figure 1.3. For the purposes of the SAVE study and all subsequent analysis presented in this thesis, regions were defined as Western (Manitoba and Saskatchewan), Central (Ontario and Quebec) and Eastern (New Brunswick, Nova Scotia, Prince Edward Island and Newfoundland and Labrador). True to the population density of Canada, the greatest proportion of isolates was collected from the Central region (66.3%, 4,161/6,272). Isolates from Western Canada accounted for 21.7% (1,358/6,272) of isolates collected, while the remaining 12.0% (753/6,272) were obtained from Eastern Canada. The Cochran-Armitage trend test did not identify significant differences in collection proportions from any region over the five-year study period ($P \geq 0.500$).

Age group and gender information associated with invasive pneumococcal isolates collected from the SAVE study is presented in Table 1.3. In general, the majority of isolates were collected from adults. Only 13.8% (865/6,272) of isolates were collected from individuals under the age of 18. Adults over the age of 65 comprised the largest age category, accounting for 37.6% (2,360/6,272) of isolates collected over the five-year period. Invasive pneumococcal isolates were consistently more commonly associated with male patients than female patients. Overall, isolates obtained from males accounted for 51.6% (3,238/6,272), while isolates from females made up only 44.8% (2,807/6,272). Age and gender information was not provided for 2.8% (174/6,272) and 3.6% (227/6,272) of isolates, respectively.

Invasive isolates included in this study were most commonly obtained from blood (91.1%, 5,713/6,272). Only 3.9% (245/6,272) of isolates were obtained from CSF, while the remaining 5.0% (314/6,272) were collected from other sterile body sites.

FIGURE 1.3: Geographic distribution by region[a] of invasive *S. pneumoniae* isolates collected by the SAVE study, 2011-2015.

NB/NLA/
NS/PEI: 12.0%
2011: 10.9 (150)
2012: 12.5 (161)
2013: 13.0 (148)
2014: 11.8 (150)
2015: 12.0 (144)

ON/QC: 66.3%
2011: 68.1 (939)
2012: 65.5 (842)
2013: 64.2 (731)
2014: 67.3 (857)
2015: 66.2 (792)

MB/SK: 21.7%
2011: 21.0 (290)
2012: 21.9 (282)
2013: 22.8 (259)
2014: 21.0 (267)
2015: 21.7 (260)

Central Canada
Eastern Canada
Western Canada
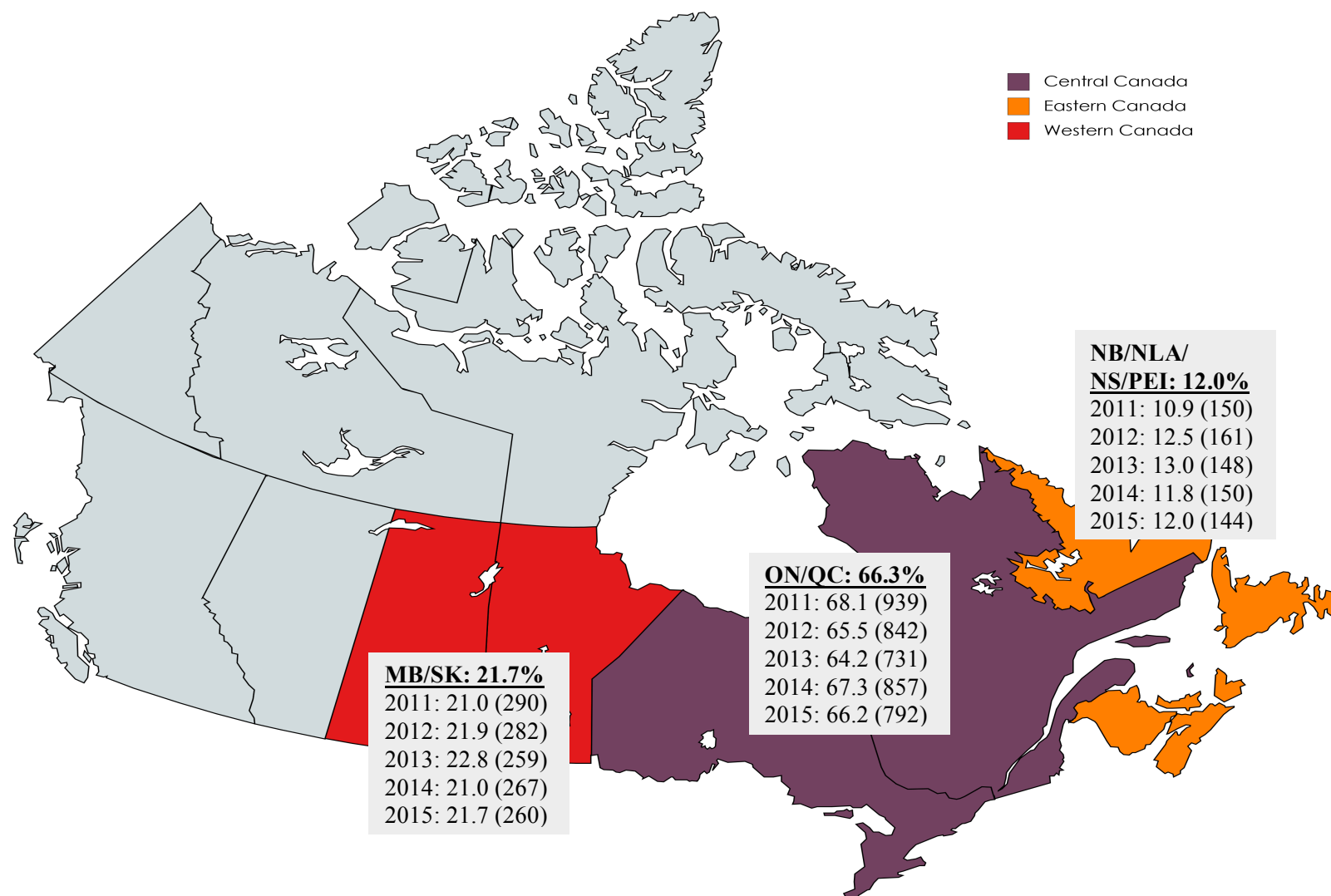
[a] Western Canada: Manitoba and Saskatchewan; Central Canada: Ontario and Quebec; Eastern Canada: New Brunswick, Newfoundland and Labrador, Nova Scotia and Prince Edward Island. Total N=6,272; 2011 n=1,379; 2012 n=1,285; 2013 n=1,138; 2014 n=1,274; 2015 n=1,196.

TABLE 1.3: Patient demographics associated with invasive *S. pneumoniae* collected by the SAVE study, 2011-2015.

Demographic Parameter	Year (%, n)					Total (n=6,272)
	2011 (n=1,379)	2012 (n=1,285)	2013 (n=1,138)	2014 (n=1,274)	2015 (n=1,196)	
Age (years)						
<1	5.1 (71)	2.9 (37)	3.1 (35)	2.4 (31)	1.8 (21)	3.1 (195)
1 - <2	4.6 (63)	2.4 (31)	3.0 (34)	4.2 (54)	2.5 (30)	3.4 (212)
2 - <6	5.5 (76)	5.1 (64)	3.2 (36)	3.8 (49)	2.8 (33)	4.1 (258)
6 - <18	4.6 (63)	3.8 (49)	2.6 (30)	2.0 (26)	2.7 (32)	3.2 (200)
18 - <50	20.2 (282)	22.1 (284)	18.5 (210)	18.1 (230)	19.4 (232)	19.7 (1238)
50 - <65	25.3 (349)	25.0 (321)	23.5 (267)	27.8 (354)	28.8 (344)	26.1 (1635)
≥65	34.2 (472)	35.1 (451)	37.6 (428)	40.1 (511)	41.6 (498)	37.6 (2360)
Not Provided	0.2 (3)	3.7 (48)	8.6 (98)	1.5 (19)	0.5 (6)	2.8 (174)
Gender						
Male	51.4 (709)	52.8 (679)	55.3 (629)	49.9 (636)	48.9 (585)	51.6 (3238)
Female	43.4 (599)	44.7 (575)	42.5 (484)	47.8 (609)	45.2 (540)	44.8 (2807)
Not Provided	5.2 (71)	2.4 (31)	2.2 (25)	2.3 (29)	5.9 (71)	3.6 (227)

2. SCIENTIFIC RATIONALE AND OBJECTIVES

S. pneumoniae is a significant source of morbidity and mortality worldwide. This organism can asymptomatically colonize the nasopharynx, as well as cause serious respiratory and invasive infections. Conjugate vaccine use in Canada has been a widespread success, resulting in significant overall decreases in IPD, particularly due to vaccine serotypes. However, after over a decade of conjugate vaccine use, the serotype distribution continues to shift dramatically in favour of types not included in vaccine formulations. This shift is primarily due to serotype replacement, where non-vaccine serotypes rise to occupy the nasopharyngeal niche vacated by vaccine types, and vaccine escape through recombination-mediated capsular switching events. Non-vaccine serotypes can be significant sources of antimicrobial resistance, and capsule switching is a particularly effective way of perpetuating the spread of penicillin and β-lactam resistance throughout the pneumococcal population.

The purpose of this thesis was to characterize twelve predominant *S. pneumoniae* serotypes causing invasive infections in Canada from 2011 to 2015, including three vaccine and nine non-vaccine types. The primary objective of this work was to utilize a variety of genotypic methods to characterize twelve prevalent serotypes (3, 6C, 7F, 8, 9N, 11A, 12F, 15A, 19A, 22F, 33F and 35B) and to ascertain genetic relationships within the serotype. Serotypes 6C and 35B were not part of the ten most common serotypes overall; however, 6C had previously been characterized as a member of the top ten and fell to eleventh after 2014. Serotype 35B was selected due to its increasing penicillin-nonsusceptibility and previously characterized piliation (79).

To address this objective, a cohort of isolates was selected from the SAVE study for analysis. The SAVE study began in 2011 to monitor invasive *S. pneumoniae* isolates in Canada

following the introduction of PCV-13. As *S. pneumoniae* is known for its diversity, isolates were characterized for genetic relatedness. The constantly evolving field of molecular characterization necessitated the use of three different methods: PFGE, MLST and WGS. WGS was also used to determine the extent of recombination within serotypes. Antimicrobial susceptibilities obtained using broth microdilution were analyzed to determine the rate of MDR in these serotypes, and resistance determinants were identified using WGS analysis. As *S. pneumoniae* is associated with carriage as well as invasive disease, several virulence genes pertaining to both colonization and infection were investigated using WGS and PCR. Due to time and cost constraints, WGS was only performed on a small subset of the twelve serotypes of interest.

A secondary objective was to identify relationships within the broader Canadian pneumococcal population collected by the SAVE study. To address this, a selection of "background" isolates were randomly selected from the large pool of SAVE isolates to include a diverse group of serotypes in the WGS analyses. Additionally, to obtain information on *S. pneumoniae* prior to the introduction of PCV-13 in Canada, a random selection of isolates was obtained from the BESST study (2007-2009) for comparison. These isolates were acceptable for comparison as they were chosen from a pool of isolates that matched the source, provinces of isolation and age group of isolates selected from SAVE. Isolates in this cohort were only investigated using WGS analyses described above. Inclusion of a breadth of isolates allowed for the identification of capsular switch variants that may pose a threat to vaccine or treatment efficacy.

3. HYPOTHESES

The hypotheses of this research evolved over the course of the study. Initial hypotheses focused on specific pneumococcal serotypes and antimicrobial resistance; however, the growing trend towards use of increasingly discriminatory methods of determining genetic relationships necessitated the expansion of the study to included additional methods and an expanded cohort of serotypes. The hypotheses of this study included:

(A) Predominant *S. pneumoniae* serotypes (3, 6C, 7F, 8, 9N, 11A, 12F, 15A, 19A, 22F, 33F and 35B) will frequently demonstrate antimicrobial resistance and/or MDR to key antimicrobial agents.

(B) Predominant *S. pneumoniae* serotypes will be genetically similar within the serotype.

(C) Predominant *S. pneumoniae* serotypes will be virulent due to the presence of PI-1 and PI-2 pilus-encoding islands.

(D) Capsular switch variants will contribute to the increasing prevalence of non-vaccine serotypes in Canada.

(E) WGS will provide increased power over PFGE and MLST for discriminating between closely related isolates or groups of isolates; specifically, the rank order of discriminatory power will be WGS >> MLST > PFGE.

4. MATERIALS AND METHODS

4.1 Antimicrobial Susceptibility Testing

4.1.1 Antimicrobial Preparation

Antimicrobial stock solutions were prepared by reconstituting laboratory grade powders in accordance with Clinical and Laboratory Standards Institute (CLSI) guidelines (80). The antimicrobial activity of each solution was confirmed using the broth macrodilution method described by CLSI, and the following quality control strains: *S. pneumoniae* ATCC® 49619, *Enterococcus faecalis* ATCC® 29212, *Escherichia coli* ATCC® 25922, *Staphylococcus aureus* ATCC® 29213 and *Pseudomonas aeruginosa* ATCC® 27853.

4.1.2 Broth Microdilution

Following two subcultures from frozen stock on 5%-SBA, antimicrobial susceptibilities were tested on all *S. pneumoniae* isolates collected from the SAVE study using the broth microdilution method detailed by CLSI M07-A10 (80). 96-well microtitre panels were custom-designed and prepared in-house using doubling antimicrobial dilutions in 100 μL of cation-adjusted Mueller-Hinton broth plus 4% lysed horse blood. Antimicrobials tested included: penicillin, amoxicillin-clavulanate, ceftaroline, ceftriaxone, cefuroxime, ertapenem, imipenem, meropenem, piperacillin-tazobactam, clarithromycin, clindamycin, doxycycline, tigecycline, chloramphenicol, trimethoprim-sulfamethoxazole, vancomycin, daptomycin, levofloxacin, moxifloxacin and linezolid. Bacterial suspensions were prepared in sterile water using the direct colony suspension method described by CLSI to achieve a turbidity equivalent to a 0.5 McFarland turbidity standard, or approximately 1-2 x 10^8 colony forming units (CFU) per mL. Suspensions were diluted in sterile water and inoculated into the microtitre panels at an approximate concentration of 5 x 10^5 CFU/mL. Plates were incubated at 35°C with 5% CO_2 for

20-24 hours.

4.1.3 MIC Interpretive Criteria

The MIC of an isolate is defined as the lowest concentration of an antimicrobial that completely inhibits visible growth. MIC interpretations were made as defined by CLSI M100-S26 (2016) breakpoints (81). MDR was defined as resistance to three or more classes of antimicrobials, while XDR was defined as resistance to penicillin plus resistance to at least four other classes of antimicrobials (31). Penicillin resistance was defined as an MIC ≥ 2 µg/mL.

4.2 Polymerase Chain Reaction (PCR) Detection of Pilus-Encoding Islands

All isolates demonstrating a serotype included in the PCV-13 and PCV-15 vaccines or one of the twelve serotypes of interest collected in Canada had DNA extracted for PCR. This included serotypes 1, 3, 4, 5, 6A, 6B, 6C, 7F, 8, 9N, 9V, 11A, 12F, 14, 15A, 18C, 19A, 19F, 22F, 23F, 33F and 35B.

4.2.1 Lysate Preparation

Bacterial isolates were subcultured from frozen stock on 5%-SBA. Once purity was confirmed, a very small loopful of bacteria was suspended in 100 µL of lysis solution containing 10 µL of 2M Tris-HCl (pH 7.6), 5 µL of Triton™ X-100 (Sigma-Aldrich, St. Louis, Missouri), 1 µL of 25 mg/mL Proteinase K (Sigma-Aldrich) and 84 µL of sterile distilled water (dH$_2$O). Bacterial lysis was achieved by incubating the cell suspensions at 60°C for 10 minutes followed by 94°C for 5 minutes. Lysates were stored at -20°C.

4.2.2 Amplification of PI-1 and PI-2 Regions

For each isolate, four separate PCR reactions (two per pilus type) were performed as previously described (54); primer sequences used to performed these reactions are listed in Table 4.1. The presence of PI-1 and PI-2 was detected by amplifying specific pilus genes *srt* and *sip*,

respectively, utilizing primer sets Rlr_srtC_F and Rlr_srtD_R (PI-1) and sipA_up and sipA_do (PI-2). To determine the absence of the two pili, flanking genes were amplified using primer sets Rlr_up_F and Rlr_do_R, and pepT_F and hemH_R for PI-1 and PI-2, respectively. If an isolate produced a PCR product for both reactions, this could indicate the presence of the PI in an alternate insertion site.

TABLE 4.1: Primers used to identify both the presence and the absence of pneumococcal pili PI-1 and PI-2.

Pilus	Reaction Primer	Primer Sequence	Size (bp)
PI-1	Presence		
	Rlr_srtC_F	5' GGGGAAGATTATGCGACCTT 3'	~600-700
	Rlr_srtD_R	5' GCTTGGCTCTGCACGGTGCC 3'	bp
	Absence		
	Rlr_up_F	5' CTTCCACGAAGTTCTTTCAATGG 3'	~700 bp
	Rlr_do_R	5' GTCTTAGAATATCATGGTTTACGTGC 3'	
PI-2	Presence		
	sipA_up	5' CTCTAGGAGGGATCTTCTTTATCATC 3'	~500 bp
	sipA_do	5' CTACAGCCGTGTTCGATTGTCC 3'	
	Absence		
	pepT_F	5' TAAGAAGCGGTCCAAGAGATTTGG 3'	~500 bp
	hemH_R	5' AATAATGGGGCTCCAAAATCAAGC 3'	

Each reaction was conducted in PCR tubes containing a total volume of 50 μL, including 10 μL of 5X Green GoTaq® Flexi Buffer, 5 μL of 25 mM MgCl$_2$, 1 μL of 10mM dNTP mixture and 0.25 μL of GoTaq® Flexi DNA Polymerase (Promega, Madison, Wisconsin), as well as 0.5 μL of each of the forward and reverse primers (Invitrogen), 30.75 μL of dH$_2$O and 2 μL of bacterial lysate. The positive control for both presence reactions was clinical isolate 12289 (possesses both pili), with clinical isolate 12464 (possesses neither pili) serving as the negative control. Conversely, 12464 was the positive control for the absence reactions, with 12289 as the negative control. Both isolates were obtained from the CANWARD study. Cycling parameters used on the GeneAmp® PCR System 9700 (Applied Biosystems, Foster City, California) were: initial denaturation at 94°C for 2 minutes, followed by 30 cycles of 94°C for 30 seconds (denaturation), 45°C (PI-1) or 55°C (PI-2) for 30 seconds (annealing) and 72°C for 45 seconds (extension). The PI-1 reactions often demonstrated non-specific binding; when this occurred repeatedly, the annealing temperature was increased to 50°C.

4.2.3 Agarose Gel Electrophoresis

The PCR products were visualized using agarose gel electrophoresis. 2% agarose gels were prepared by dissolving 2g of agarose (Invitrogen, Carlsbad, California) in 100 mL of 0.5X Tris-Borate-EDTA buffer (TBE, 45mM Tris-borate, 1mM EDTA [pH 8.3 ± 0.1]). This solution was brought to a rolling boil, and then allowed to cool at room temperature until reaching approximately 50°C. Upon reaching this temperature, 1 μL of ethidium bromide was added and swirled to ensure proper mixing. The gel was poured into a casting tray with combs (totaling 40 wells) and left to cool for approximately 30 minutes at room temperature.

Following cooling, gels were placed into an electrophoresis chamber filled with fresh 0.5X TBE buffer. Wells located at either end of a row were loaded with 5 μL of TrackIt™ 100 bp

DNA ladder (Invitrogen). All other wells were loaded with 15 μL of PCR product. Gels were run at 100V for 30 minutes and visualized using the AlphaImager HP (Alpha Innotech, Santa Clara, California).

4.3 Pulsed-Field Gel Electrophoresis (PFGE)

The first method utilized to determine the genetic relatedness of *S. pneumoniae* isolates was PFGE. The following methods were adapted from McEllistrem *et al*. and Louie *et al*. (82, 83). Ten randomly selected isolates per year of each of the eleven most common serotypes were included for analysis (50 of each serotype, 550 total isolates), as well as a sample of 35 serotype 35B isolates. Isolates were selected using a random number generator.

4.3.1 Preparation of Genomic DNA

Isolates were grown on 5%-SBA from frozen stock. Once purity was ensured, cultures were swabbed onto a full plate for confluent growth and incubated overnight. The mucoidy of serotype 3 isolates required two full plates of confluent growth to ensure adequate visualization of the PFGE pattern. Using the narrow edge of a sterile, disposable glass slide (Leica Microsystems, Wetzlar, Germany), bacterial growth was scraped into a pile in the centre of the plate. A sterile loop was used to suspend the colonies in 1 mL of cell suspension buffer (10 mM Tris-HCl [pH 7.2], 20 mM NaCl, 50 mM EDTA [pH 8.0]) to an optical density of 2.4-2.5 at 560nm. To cast plugs, 100 μL of bacterial suspension was added to 100 μL of 1.6% low-melt agarose (BioRad Laboratories, Hercules, California) made by boiling 0.16g of agarose in 10 mL 1X TE (10 mM Tris-HCl [pH 7.5], 1 mM EDTA [pH 7.5]) buffer. This mixture was pipetted gently to mix without forming bubbles and immediately dispensed into disposable plug molds (Bio-Rad Laboratories). Plugs were allowed to solidify for 10-15 minutes at room temperature, or 5 minutes at 4°C.

Once solidified, plugs were transferred to 1 mL of lysis solution (10 mM Tris-HCl [pH 7.2], 50 mM NaCl, 50 mM EDTA [pH 8.0], 0.5% Brij-58, 0.2% deoxycholate, 0.5% sarcosyl, 1 mg/mL lysozyme) and incubated for at least one hour in a 37°C water bath. Following incubation, the lysis buffer was removed and replaced with 1 mL of ESP solution (250 mM EDTA [pH 9.0], 1% sarcosyl, 200 µg/mL Proteinase K). Plugs were then incubated overnight in a 50°C water bath. After overnight incubation, plugs were rinsed once with 1 mL of 1X TE buffer. Rinse buffer was immediately removed, followed by at least four 30 minute washes with 1 mL 1X TE buffer at 37°C. Plugs were stored in 1 mL of fresh 1X TE buffer at 4°C for up to six months.

4.3.2 Restriction Endonuclease Digestion

Approximately one-third of each plug was cut off using a sterile scalpel and placed in a 0.6 mL tube. A restriction enzyme solution (150 µL per isolate) was prepared using 135 µL of sterile dH$_2$O, 15 µL of 10X NEBuffer 4 (New England Biolabs, Mississauga, ON) and 25U of *Sma*I (1.25 µL of 20000U/mL stock). Plug slices were covered with 150 µL of restriction enzyme solution and mixed gently to remove bubbles and ensure submersion. Plugs were incubated at room temperature for at least two hours.

4.3.3 Electrophoresis

To prepare the agarose gel, 2000 mL of fresh 0.5X TBE buffer was prepared by diluting 10X TBE in dH$_2$O. A 1% SeaKem® Gold agarose gel (Lonza, Basel, Switzerland) was prepared by boiling 1-1.5 g of agarose in 100-150 mL 0.5X TBE buffer, depending on the number of samples being run. After boiling and sufficient cooling, solidified agarose was removed from the top surface and the agarose was poured into a mold with combs placed 1-2 mm above the bottom surface. The gel was allowed to solidify for at least one hour.

Approximately 30 minutes prior to running the gel, the remaining fresh 0.5X TBE buffer was poured into the electrophoresis chamber. The buffer was allowed to cool to 9°C by turning on the power supply, the pump (set at 75-80) and the cooling module. After carefully removing the restriction enzyme solution, plug slices were melted at 65°C for 10 minutes. The comb was removed from the gel and 30 µL of each sample was loaded into each well, taking care to ensure no bubbles were created. Samples were allowed to solidify at room temperature for 5 minutes. During this time, a DNA size standard (Lambda ladder, Bio-Rad Laboratories), was loaded using a sterile scalpel and spatula to the first and last wells in each row. Restriction fragments were resolved in a contour-clamped homogenous electric field apparatus (CHEF DRIII; Bio-Rad Laboratories) with an initial switch time of 2 seconds, a final switch time of 30 seconds, voltage of 200 V (6 V/cm), an included angle of 120 and a run time of 18.5 hours. Once completed, the gel was stained in the dark for 50 minutes with SYBR® Green (Molecular Probes, Eugene, Oregon) in 1X TE buffer, followed by at least four rinses with dH_2O to destain. The gel was visualized under UV light using the AlphaImager HP and an image saved for future analysis.

4.3.4 Pattern Analysis

S. pneumoniae patterns were analyzed using BioNumerics v. 3.5 software (Applied Maths, Austin, Texas). Clusters were analyzed using the unweighted pair group method and percent similarity was calculated using the dice coefficient. Isolates were determined to be genetically related if they clustered with at least 80% similarity.

4.4 Multi-Locus Sequence Typing (MLST)

MLST was performed on the same 550 isolates included in the previously described PFGE analysis. In addition, 25 of the above-mentioned serotype 35B isolates were typed.

4.4.1 DNA Extraction

Bacterial isolates were cultured from frozen stock on 5%-SBA. Once purity was confirmed, a small loopful of bacteria was suspended in 100 μL of QuickExtract™ DNA Extraction Solution (Epicentre, Madison, Wisconsin). The solutions were vortexed vigorously and allowed to sit at room temperature for 5 minutes. DNA was stored at 4°C until needed and briefly centrifuged before use to pellet insoluble material.

4.4.2 Amplification of Seven Housekeeping Genes

The MLST scheme for *S. pneumoniae* involves the amplification of seven housekeeping genes as originally described by Enright and Spratt (84): *aroE* (shikimate dehydrogenase), *gdh* (glucose-6-phosphate dehydrogenase), *gki* (glucose kinase), *recP* (transketolase), *spi* (signal peptidase I), *xpt* (xanthine phosphoribosyltransferase) and *ddl* (D-alanine-D-alanine ligase). These genes were chosen as they were considered the most polymorphic within the pneumococcal genome, with maximal nucleotide sequence divergence (84). Primers utilized for this typing scheme are listed in Table 4.2, with only two being derived from the original MLST scheme (*gdh* and *gki*) (84). The remaining five genes (*aroE*, *recP*, *spi*, *xpt* and *ddl*) were amplified using modified primers suggested by the CDC (85). All PCR products were approximately 500bp in length.

TABLE 4.2: Primers used to amplify the seven housekeeping genes of the *S. pneumoniae* MLST scheme.

Gene Primer	Primer Sequence	Reference
aroE		
aroE-F	5' TCCTATTAAGCATTCTATTTCTCCCTTC 3'	(85)
aroE-R	5' ACAGGAGAGGATTGGCCATCCATGCCCACACTG 3'	
gdh		
gdh-F	5' ATGGACAAACCAGC(G/A/T/C)AG(C/T)TT 3'	(84)
gdh-R	5' GCTTGAGGTCCCAT(G/A)CT(G/A/T/C)CC 3'	
gki		
gki-F	5' GGCATTGGAATGGGATCACC 3'	(84)
gki-R	5' TCTCCCGCAGCTGACAC 3'	
recP		
recA-F	5' GAATGTGTGATTCAATAATCACCTCAAATAGAAGG 3'	(85)
recA-R	5' TGCTGTTTCGATAGCAGCATGGATGGCTTCC 3'	
spi		
spi-F	5' CGCTTAGAAAGGTAAGTTATGAATTT 3'	(85)
spi-R	5' GAAGAGGCTGAGATTGGTGATTCTCGGCC 3'	
xpt		
xpt-F	5' TTAACTTTTAGACTTTAGGAGGTCTTATG 3'	(85)
xpt-R	5' CGGCTGCTTGCGAGTGTTTTTCTTGAG 3'	
ddl		
ddl-F	5' TAAAATCACGACTAAGCGTGTTCTGG 3'	(85)
ddl-R	5' AAGTAGTGGGTACATAGACCACTGGG 3'	

Reactions were conducted in MicroAmp® Optical 96-well reaction plates (Applied

Biosystems), with each well containing a total volume of 35 μL. Each individual reaction

contained 17.5 μL of Bioline MyTaq™ HS Mix (FroggaBio Inc., Toronto, Ontario), 0.7 μL of

each of the forward and reverse primers (Invitrogen), 15.1 μL of dH$_2$O and 1 μL of bacterial

template. The *spi* reaction was determined to work more effectively with a smaller amount of

template, so for this reaction the amount of template in each well was reduced to 0.5 μL. Cycling

parameters used on the GeneAmp® PCR System 9700 (Applied Biosystems) were: initial

denaturation at 95°C for 5 minutes, followed by 35 cycles of 95°C for 15 seconds (denaturation),

55°C for 15 seconds (annealing) and 72°C for 15 seconds (extension).

4.4.3 Purification of Sequencing Products

The presence of PCR products was ensured prior to clean-up by running an agarose gel

using the E-Gel Precast Agarose Electrophoresis System and E-Gel® 96 Agarose Gels

(Invitrogen). Clean-up of PCR products was performed using the PCRClean DX™ purification

system (Aline Biosciences, Woburn, Massachusetts). Briefly, 63 μL of PCRClean DX™

magnetic beads were added to each well to bind PCR products greater than 120 bp. After a 5-

minute incubation at room temperature to ensure maximum recovery, a 96-well magnetic plate

was used to separate the magnetic bead-bound PCR products from the remaining solution. While

still on the magnetic plate, the supernatant was removed and the magnetic beads were washed

twice with 200 μL of 80% ethanol. After a 3-minute air dry to evaporate any residual ethanol, the

sample plate was removed from the magnetic plate and 30 μL of reagent grade water was added

to elute the PCR products. The beads were again separated from the eluate using the magnetic

plate, and the eluate was transferred to a clean plate for further analysis. Purified PCR products

were sequenced using the BigDye™ Terminator v3.1 Cycle Sequencing Kit (Applied

Biosystems). Sequencing was performed by the Genomics Core Facility at PHAC-NML using the 3730XL DNA Analyzer (Applied Biosystems) according to the manufacturer's instructions.

4.4.4 Sequence Analysis and Designation of Allelic Profiles

Sequence analysis was performed using Lasergene SeqMan v.7.0.0 (DNASTAR, Madison, Wisconsin). Forward and reverse reads were assembled to form contigs, which were scanned for quality; contigs with a poor percent match value were removed for repeat sequencing. Based on the primers used, each contig was trimmed to a specific length and start sequence, as described in Table 4.3. Trimmed sequences were uploaded to the *S. pneumoniae* MLST database (http://spneumoniae.mlst.net), where a unique allele number was assigned to each of the seven housekeeping genes. The site also took these seven allele numbers and assigned a sequence type (ST) to each isolate based on previously deposited data to the database. Isolates differing at one of seven alleles were considered single-locus variants (SLV), while those differing at two alleles were considered double-locus variants (DLV). During the course of this study, the MLST database for *S. pneumoniae* migrated from http://spneumoniae.mlst.net to http://pubmlst.org/spneumoniae. After the migration, it was no longer necessary to trim sequences prior to uploading. To identify commonly circulating clones, STs were compared to the PMEN database (http://www.sph.emory.edu/PMEN). The PMEN database currently recognizes 43 clones that have wide geographic distribution (isolated on at least two continents, in different locations) and are either a clone resistant to a key antimicrobial, or a susceptible clone that has wide clinical implications. Minimum spanning trees were generated using the freely available PhyloViZ 2.0 software (86).

TABLE 4.3: Start sequences and correct lengths necessary to submit the seven *S. pneumoniae* MLST housekeeping genes for allelic assignment by the database located at http://spneumoniae.mlst.net.

Gene	Start Sequence	Length Required for Submission
aroE	GAAGCGAGT	405 bp
gdh	AGAACAT	460 bp
gki	ACCCTTCAA	483 bp
recP	CTCAACCAAA	450 bp
spi	GTATCTTTT	474 bp
xpt	GGTGATAA	486 bp
ddl	GCTAAAAT	441 bp

4.5 Whole Genome Sequencing (WGS)

4.5.1 Isolate Selection

Due to time and cost constraints, a total of 192 isolates (two full 96-well plates) were selected for WGS. An initial 83 isolates from the SAVE study were specifically selected from the twelve serotypes of interest due to previous characterization (outlined in Sections 4.1, 4.2 and 4.4) that indicated MDR, novel MLST sequence types and/or the potential to be a capsular switch variant. To achieve broader coverage of the diverse pneumococcal population, 79 additional isolates from SAVE were randomly selected as "background". These background isolates were selected using a random number generator and included three isolates of each PPV-23 vaccine serotype not already included in the eleven most common types and up to three of any other non-vaccine serotype to total 79 isolates. In an effort to control one of the many variables, isolates were selected from the ≥65-year age category where possible, as this age group had the largest and most diverse collection of isolates from which to sample.

To include isolates from a broader time span, specifically those that were collected prior to PCV-13 introduction, the remaining 30 isolates were randomly selected from the BESST study (Baseline Epidemiology of *Streptococcus pneumoniae* Serotypes in Canada prior to the Introduction of the 13-valent Pneumococcal Vaccine) (30). BESST included a random sample of 400 respiratory and 400 invasive *S. pneumoniae* isolates collected as part of the CANWARD 2007-2009 study (Canadian Ward Surveillance Study, an annual national study collecting clinical isolates from patients at 15 tertiary care centers across Canada) (87). For the current study, only isolates collected from the same provinces, specimen source and age group as the other 162 isolates were included. Overall, 44 different serotypes were represented for WGS analysis; these isolates are described in further detail in Appendix A.

4.5.2 DNA Extraction

High-quality DNA was extracted from *S. pneumoniae* isolates using the Epicentre

MasterPure™ Complete DNA and RNA Purification Kit (Mandel Scientific, Guelph, Ontario).

All reagents used were part of this kit unless otherwise specified. Bacterial isolates were cultured

from frozen stock on 5%-SBA. A full loop of bacterial culture was suspended in a solution

containing 600 µL of 2X T&C Lysis Solution, 2 µL of Proteinase K (50 µg/µL), 2 µL of

ReadyLyse™ Lysozyme (20,000U/µL, Mandel Scientific) and 2 µL of Mutanolysin (10U/µL,

Sigma Aldrich). The solution was vortexed vigorously to fully suspend the bacteria. Suspensions

were incubated for 30 minutes at 50°C in a dry bath with constant high-speed mixing, followed

by boiling at 99°C for 20 minutes with constant mixing at medium-speed. Samples were cooled

to room temperature, mixed with 2 µL of 5 mg/mL RNase A and incubated at 37°C for 30

minutes with slow mixing. To precipitate protein, samples were chilled on ice for 5 minutes prior

to adding 300 µL MPC Protein Precipitation Reagent. Protein debris was pelleted by

centrifugation at 4°C for 10 minutes at ≥10,000Xg. The supernatant was transferred to a clean

tube and the pellet discarded. The RNase A and protein precipitation steps were repeated with an

additional 2 µL of RNase A and 150 µL of MPC Protein Precipitation Reagent. The resulting

supernatant was transferred to a second clean tube for DNA precipitation.

Cold isopropanol was added to fill each of the supernatant tubes. Tubes were gently

mixed by inverting 30-40 times. DNA was pelleted by centrifugation at 4°C for 10 minutes. The

isopropanol was carefully pipetted off and discarded and the tubes were centrifuged again for 3

minutes to collect any residual isopropanol for removal. Each DNA pellet was rinsed with 75%

ethanol and left to air dry at room temperature. DNA was resuspended by adding 60 µL of

nuclease-free water and incubating for 30 minutes at 37°C with gentle mixing. Following

overnight storage at 4°C, DNA was considered suitable for quantification.

4.5.3 Quantification of DNA

To ensure there was enough DNA present for sequencing, DNA was quantified using the Qubit® 2.0 Fluorometer and Qubit® dsDNA BR Assay (Invitrogen). A master mix was prepared containing 199 μL of Qubit® dsDNA BR Buffer and 1 μL of Qubit® dsDNA BR Reagent per isolate being tested. In special thin-walled Qubit® Assay tubes, 199 μL of master mix was combined with 1 μL of DNA sample. Tubes were vortexed for 3 seconds and incubated at room temperature for 2 minutes to achieve optimal fluorescence. The fluorometer was calibrated prior to each use with Qubit® dsDNA BR Standard #1 (0 ng/μL) and Standard #2 (100 ng/μL), prepared similarly to the above samples except with 190 μL of master mix and 10 μL of standard. Concentrations given by the fluorometer were converted to ng/μL. DNA samples were considered acceptable if over 50 ng/μL.

Agarose gel electrophoresis was run in tandem with fluorometric readings. Similar to above, a 1% agarose gel was prepared by boiling 1g of agarose (Invitrogen) with 100 mL of 1X Tris-Acetate-EDTA buffer (TAE, 40 mM Tris-acetate, 1 mM EDTA [pH 8.3 ± 0.1]). SYBR® Safe DNA Gel Stain (Invitrogen) was added once the solution cooled. A 26-well gel was poured and allowed to cool. Gels were placed into an electrophoresis chamber filled with fresh 1X TAE buffer. For the middle 24 wells, 4 μL of each sample was mixed with a small amount of BlueJuice™ Gel Loading Buffer 10X (Invitrogen). Similarly, the wells on either end were loaded with 1 Kb Plus DNA Ladder (Invitrogen) mixed with loading buffer. Gels were run at 120V for 50 minutes and visualized using the Gel Doc™ XR+ System (BioRad Laboratories). Genomic DNA appeared as a dark smudge down the entire lane.

4.5.4 Sequencing and *De Novo* Assembly

Prior to sequencing, the genomic DNA was diluted to a final concentration of 50 ng/μL and a final volume of 50 μL using nuclease-free water. Multiplexed libraries were prepared using TruSeq Nano DNA HT Library Preparation Kits (Illumina, San Diego, California). Up to 300bp paired-end sequence reads were generated using the Illumina MiSeq Platform (Illumina) and 600-cycle MiSeq Reagent Kits v3 (Illumina). Library preparation and sequencing were performed by the Genomics Core Facility at PHAC-NML, who then provided the raw data for further analysis. The planned calculated sequence redundancy for the paired-end reads was ~60x coverage. Quality control was assessed using the FastQC tool (v.0.11.2, http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and poor quality reads were removed. On average, the number of reads generated per genome was 539,336 with an average genome coverage of 77x.

High-quality reads were merged using the FLASH program (88), a tool designed to extend the length of shorter reads by taking paired-end reads and overlapping them, thus improving assembly quality. Merged reads were then assembled *de novo* into contiguous sequences (contigs) using SPAdes v.3.9 (89). The average contig length generated was 59,983 bp and the average N50 contig length (a measure of assembly quality) was 121,439 bp. Prokka v.1.11 (90) was used to annotate assembled genomes.

4.5.5 Phylogenomic Analysis based on Core Single Nucleotide Variants (SNVs)

Phylogenomic analysis was conducted using SNVPhyl, a PHAC-NML custom built pipeline (91). The input for this pipeline is a set of sequence reads and a reference genome. The standard reference genome recommended by the National Center for Biotechnology Information (NCBI) for *S. pneumoniae* is R6 (see Table 4.4). To begin, repeat regions of the reference

genome were identified using MUMMer v.3.23 (92) and collected into a masking file of locations to be excluded from further analysis. MUMMer was run using a minimum length of 150 and a minimum percent identity of 90. Reads were then mapped to the reference genome using SMALT v.0.7.4 (http://www.sanger.ac.uk/science/tools/smalt), with a k-mer size of 13 and a step size of 6. Variant calling was performed using both FreeBayes v.0.9.20 (93) and SAMtools (94, 95). FreeBayes detected variants using a minimum coverage of 10, a minimum mean mapping quality of 30 and an alternate allele proportion of 0.75. SAMtools was used to confirm the variant calls made by FreeBayes. These SNVs were filtered and merged, as previously described, to construct a multiple sequence alignment (91). Filtered SNVs were designated as "coverage" (one or more isolates failed the minimum coverage threshold), "mpileup" (one or more isolates had conflicting base calls between FreeBayes and SAMtools) or "invalid" (the SNV overlapped a masked region). Isolates that mapped poorly to the reference strain (<80%) were removed and the analysis repeated (SC11-1882-P, SC11-2703-P, SC11-4503-P, SC13-2375-P). PhyML v.3.0 (96) was used to generate a maximum likelihood phylogenetic tree from this alignment, which was then visualized using FigTree software (v.1.4.3, http://tree.bio.ed.ac.uk/software/figtree). Phylogenetic clades were determined using ClusterPicker v.1.2.3 software (97) using an initial and main support threshold of 0.9, a genetic distance threshold of 4.5 and a large cluster threshold of 10.

Based on this maximum likelihood tree, isolates were broken down into 14 smaller groups for more detailed analysis. Where possible, a completed NCBI genome was selected as the reference input for analysis. Where internal references were necessary, they were chosen to include the fewest contigs and the highest N50 value. The full list of reference genomes utilized in this study can be found in Table 4.4.

TABLE 4.4: *S. pneumoniae* genomes utilized as references for WGS phylogenomic analysis.

Genome Name	NCBI Accession Number	Level	Size (Mb)	Serotype	Sequence Type
R6	NC_003098	Completed	2.03862	2	595
TIGR4 (ATCC BAA-334)	NC_003028	Completed	2.16084	4	205
70585	NC_012468	Completed	2.18468	5	289
Taiwan19F-14	NC_012469	Completed	2.11215	19F	236
G54	NC_011072	Completed	2.07895	19F	63
ATCC 700669	NC_011900	Completed	2.22132	23F	81
INV200	NC_017593	Completed	2.09332	14	9
OXC141	NC_017592	Completed	2.03687	3	180
SPN032672	NC_021003	Completed	2.13119	1	306
Genome Name	NCBI SRA Accession Number	Level	Bases in all Contigs (Mb)	Serotype	Sequence Type
PMEN1	ERR1022095	Contig	2.17882	23F	81
PMEN3	ERR1022114	Contig	2.11382	9V	156
PMEN4	ERR1022115	Contig	2.06711	23F	37
PMEN9	ERR1022120	Contig	2.05746	14	9
PMEN18	ERR1022104	Contig	2.11841	14	67
PMEN19	ERR1022105	Contig	2.17334	5	289
PMEN21	ERR1022108	Contig	2.11721	19F	177
PMEN24	ERR1022111	Contig	2.00156	35B	377
PMEN25	ERR1022112	Contig	2.06395	15A	63
PMEN26	ERR1022113	Contig	2.09393	23F	338
SC12-3017-P	This study	Contig	2.04363	33F	100
SC13-0467-P	This study	Contig	1.98860	8	1480

4.5.6 Identification of Acquired Resistance Genes

To correlate the presence of acquired resistance genes to susceptibility testing results, the ResFinder 2.1 program (98) was used to identify genes. This program is freely available for use through the Center for Genomic Epidemiology (https://cge.cbs.dtu.dk/services/ResFinder/) and only identifies acquired genes, not chromosomal mutations. The whole genomes were uploaded to the site as fasta files. Resistance genes for macrolides (*mefA, ermB*), tetracyclines (*tetM*) and chloramphenicol (*cat*) were identified. Isolates with discrepancies between genotype and phenotype had MIC testing repeated in triplicate to confirm values, with values updated as necessary.

4.5.7 Gene Extraction

Genes were extracted using a simple workflow, using a reference gene (from *S. pneumoniae* R6, see Table 4.4) and a selection of contigs as input sequence. Prodigal v.2.6.3 (99), a microbial gene prediction software, was utilized to identify genes in each contig. Next, the NCBI BLAST+ makeblastdb tool (100, 101) was used to create a database of genes for each contig. Lastly, the NCBI BLAST+ blastn tool (100, 101) was used to search the previously created nucleotide databases for the gene of interest, based on the input reference sequence. Matches were extracted from the contig and cut into separate files for further analysis.

4.5.7.1 Chromosomal Antimicrobial Resistance Genes

For antimicrobial resistance genes with common chromosomal mutations, extracted genes were aligned to the reference sequence using the ClustalW2 multiple sequence alignment program (102). For the penicillin-binding proteins, DNA sequences were translated into amino acid sequences and examined for mutations in the active site motifs of *pbp1A* (STMK, SRNVP, KTG), *pbp2B* (SVVK, SSNT, KTGTA) and *pbp2X* (STMK, AHSSNV, LKSGT), as previously

described (103, 104). *parC*, *gyrA* and *folA/P* were examined for previously described mutations that convey fluoroquinolone and trimethoprim-sulfamethoxazole resistance, respectively (36, 38). Isolates with discrepancies between genotype and phenotype had MIC testing repeated in triplicate to confirm values, with values updated as necessary.

4.5.7.2 Virulence Genes

The presence of pneumococcal surface proteins *pspA* and *pspC* and neuraminidases *nanA*, *nanB* and *nanC* were identified using the "primersearch" function provided by EMBOSS (European Molecular Biology Open Software Suite) (105). Using previously described primers (55, 106) and up to 10% mismatch (to allow for the variability of surface proteins), contigs were searched for matching DNA sequences. Primer sequences can be found in Table 4.5.

TABLE 4.5: Primers used to identify pneumococcal surface proteins with the EMBOSS "primersearch" function. Primers previously described in (55, 106).

Gene	Sub-class Primer	Primer Sequence
pspA	All types	
	LSM12* (F)	5' CAGCGTCGCTATCTTAGGGGCTGGTT 3'
	SKH02 (R)	5' CCACATACCGTTTTCTTGTTTCCAGCC 3'
	Family 1	
	LSM12* (F)	5' CAGCGTCGCTATCTTAGGGGCTGGTT 3'
	SKH63 (R)	5' TTTCTGGCTCATYAACTGCTTTC 3'
	Family 2	
	LSM12* (F)	5' CAGCGTCGCTATCTTAGGGGCTGGTT 3'
	SKH52 (R)	5' TGGGGGTGGAGTTTCTTCTTCATCT 3'
	Family 3	
	SKH41 (F)	5' CGCACAGACTTAACAGATGAAC 3'
	SKH42 (R)	5' CTTGTCCATCAACTTCATCC 3'
pspC	IF30 (F)	5' AAGATGAAGATCGCCTACGAACAC 3'
	IF43 (R)	5' AATGAGAAACGAATCCTTAGCAAT 3'
	ABW13 (F)	5' CGACGAATAGCTGAAGAGG 3'
	SKH2 (R)	5' CCACATACCGTTTTCTTGTTTCCAGCC 3'
nanA	Forward	5' ATAGACGTGCGCAAAATACAGAATCA 3'
	Reverse	5' GTCGAACTCCAAGCCAATAACTCCT 3'
nanB	Forward	5' ACTACGAGGTGTTAATCGTGAAGG 3'
	Reverse	5' CCAATACCCGCAGGCATAACATC 3'
nanC	Forward	5' TGGGGTAAGTACAAACAAGAGG 3'
	Reverse	5' CTAATGGTACTGGCGCAAAATCA 3'

* Adapted slightly from (106) by removing CCGGATC from the 5' end.

4.5.8 *In Silico* MLST

MLST genes were extracted from whole genomes using SRST2 (Short Read Sequence Typing) v.0.2.0 (107). This tool is specifically designed to extract short sequences from whole genomes of bacterial pathogens. To determine MLST STs, contigs were uploaded with a file of all allele sequences for the seven housekeeping genes outlined in Section 4.5.2, as well as a file of ST definitions. SRST2 found and extracted the seven genes for each isolate and assigned a ST. Isolates that had previously been assigned a ST by MLST were compared to the ST assignment extracted from the whole genome to determine the accuracy of *in silico* MLST in comparison to the traditional method.

4.5.9 Confirmation of Putative Capsular Switch Variants

Putative capsular switch variants were originally identified using MLST STs; isolates that demonstrated a ST normally attributed to a different serotype were included in the selection for WGS analysis. Analysis of the maximum likelihood tree generated in Section 4.5.5 indicated that there were more putative capsular switch variants than previously thought, due to inclusion of background isolates of serotypes that were not previously studied.

Variants were confirmed using a penicillin-binding protein transpeptidase domain typing scheme outlined by Metcalf *et al.* (38). Briefly, *pbp1A*, *pbp2B* and *pbp2X* were translated into amino acid sequence and trimmed to include only the transpeptidase domain. The transpeptidase domain of *pbp1A* was a 277-residue fragment beginning at Thr338, *pbp2B* was 278 residues beginning at Thr384, and *pbp2X* began at Gly229 and was 359 residues in length. Similarly to MLST, these sequences were compared to a previously generated database (38) and assigned a numeric identifier. Penicillin-susceptible reference strain TIGR4 was considered the original sequence of comparison, and designated 1A-0, 2B-0, 2X-0. Unique sequences not previously

found in the database at the time of this study were assigned the next sequential number available. Database hits identified in this study can be found in Appendix B, with new sequences marked next to the identifier.

The MLST alleles and transpeptidase identifiers of putative donor, recipient and progeny strains were visualized diagrammatically as outlined by Metcalf *et al.* (38). Putative progeny strains were confirmed as a capsular switch variant when the MLST alleles were identical to the recipient (or highly similar: at least six of seven matching alleles), when the *pbp2B* transpeptidase domain was identical to the recipient, and when the capsule was identical to the donor. Depending on the size of the fragment transferred, the progeny may also have demonstrated one, both or neither of the *pbp1A* and *pbp2X* transpeptidase domains in common with the donor.

4.5.9.1 Determination of Recombination Fragment Size

To determine the approximate length of the fragments transferred during capsular switch events, the fastq files for the donor, recipient and progeny strains were aligned to the *S. pneumoniae* R6 reference genome using the BWA-MEM alignment algorithm (108). SNVs in the alignments were visualized using IGV (Integrative Genomics Viewer) v.2.3.93 (109, 110). Based on whether the donor *pbp1A* and *pbp2X* were transferred into the progeny strain, an area of the genome inside or outside these genes was searched for an approximate region where the SNVs stopped being identical between donor and progeny and started being identical between recipient and progeny. The distance between these two regions was therefore considered the approximate length of the fragment transferred from donor to progeny, relative to the reference strain.

4.5.10 Identification of Highly Recombinant Isolates

Gubbins (<u>G</u>enealogies <u>U</u>nbiased <u>B</u>y recom<u>B</u>inations <u>I</u>n <u>N</u>ucleotide <u>S</u>equences) was utilized to determine areas of recombination in the highly variable *S. pneumoniae* genomes (111). This software identifies regions with high densities of base substitutions (areas of likely recombination) and subsequently generates a phylogeny based on putative point mutations outside of these putative areas of recombination. To achieve input files suitable for Gubbins, the SNVPhyl pipeline described in Section 4.5.5 was re-run with the SNV density filter disabled. This allowed SNVs in high density regions to be included in the SNVPhyl output. Next, to create whole-genome alignments, the SNVPhyl output was aligned to the *S. pneumoniae* R6 reference genome using a reference-mapping approach; the PHAC-NML tool "Positions to SNV invariant alignment" applied the identified SNVs to the reference genome to create alignments. Resulting Gubbins recombination predictions were visualized using Phandango (112), along with the final phylogenetic tree and *S. pneumoniae* R6 reference.

4.6 Statistical Analysis

Changes in MDR rates between 2011 and 2015 were assessed for statistical significance ($P<0.05$) using the two-tailed Fisher's exact test ($\alpha=0.05$) available from GraphPad Software (San Diego, CA).

For a detailed breakdown of which work described in this thesis was performed by the candidate, please see Appendix A.

5. RESULTS

Part 1: Antimicrobial Resistance and Multidrug Resistance of *S. pneumoniae* Collected by the SAVE 2011-2015 Study

5.1 Antimicrobial Susceptibility

The activity of select antimicrobials tested against all *S. pneumoniae* collected from SAVE 2011-15 is summarized in Table 5.1. Despite demonstrating lowered susceptibility to penicillin (89.1% by oral penicillin V and IV meningitis breakpoints), other β-lactam agents demonstrated excellent activity against *S. pneumoniae* isolates. This includes third, fourth and fifth generation cephalosporins, carbapenems and β-lactam/β-lactamase inhibitor combinations. Other agents demonstrating excellent overall activity included the fluoroquinolones and tigecycline (>99% susceptibility), as well as antibiotics of last resort such as linezolid and vancomycin (100% susceptibility). Both doxycycline and trimethoprim-sulfamethoxazole demonstrated susceptibilities lower than 90%, while clarithromycin had the worst overall activity against *S. pneumoniae* isolates at 75.1% susceptibility.

TABLE 5.1: Activity of select antimicrobials against all *S. pneumoniae* (n=6,060[a]) collected from SAVE 2011-15.

Antibiotic	MIC (µg/mL)				MIC Interpretation[b]		
	MIC_{50}	MIC_{90}	Min	Max	%S	%I	%R
Penicillin (iv, nonmeningitis)					98.5	1.5	0.05
Penicillin (iv, meningitis)	≤ 0.03	0.12	≤ 0.03	8	89.1	-	10.9
Penicillin (oral, penicillin V)					89.1	7.8	3.2
AMC[c]	≤ 0.06	≤ 0.06	≤ 0.06	16	98.2	0.8	1.0
Ceftaroline	≤ 0.008	0.015	≤ 0.008	0.25	100	-	-
Ceftriaxone (nonmeningitis)	≤ 0.12	≤ 0.12	≤ 0.12	8	99.3	0.5	0.2
Ceftriaxone (meningitis)					96.7	2.6	0.7
Cefuroxime (parenteral)	≤ 0.25	≤ 0.25	≤ 0.25	> 16	95.0	0.4	4.7
Cefuroxime (oral)					95.3	0.9	3.7
TZP[c]	≤ 1	≤ 1	≤ 1	8	ND[d]		
Ertapenem	≤ 0.06	≤ 0.06	≤ 0.06	4	98.5	1.4	0.1
Imipenem	≤ 0.03	≤ 0.03	≤ 0.03	2	96.1	2.5	1.4
Meropenem	≤ 0.06	≤ 0.06	≤ 0.06	2	95.9	2.1	2.0
Clarithromycin	≤ 0.03	4	≤ 0.03	> 32	75.1	1.6	23.3
Clindamycin	≤ 0.12	≤ 0.12	≤ 0.12	> 64	93.1	0.4	6.5
Levofloxacin	1	1	≤ 0.06	32	99.3	0.1	0.6
Moxifloxacin	0.12	0.25	≤ 0.06	8	99.4	0.3	0.3
SXT[c]	0.25	1	≤ 0.12	> 8	87.7	6.1	6.2
Doxycycline	≤ 0.25	0.5	≤ 0.25	> 16	89.5	0.8	9.6
Tigecycline	≤ 0.015	0.03	≤ 0.015	0.06	100	-	-
Chloramphenicol	2	4	≤ 0.12	32	98.3	-	1.7
Linezolid	1	2	≤ 0.12	4	100	-	-
Daptomycin	0.12	0.12	≤ 0.03	1	ND[d]		
Vancomycin	0.25	0.5	≤ 0.12	1	100	-	-

[a] n for which complete susceptibility data is available for all selected antimicrobials; [b] %S: % susceptible, %I: % intermediate, %R: % resistant; [c] AMC: amoxicillin/clavulanic acid, TZP: piperacillin/tazobactam, SXT: trimethoprim-sulfamethoxazole; [d] ND, breakpoints not defined.

In many cases where the overall susceptibilities to agents were low, the decreased susceptibilities were due to a select group of serotypes. Figure 5.1 depicts the differing susceptibilities of key antimicrobial agents against the twelve serotypes of interest to this study. Serotype 15A demonstrated the lowest penicillin susceptibility at 39%, with serotypes 35B, 19A and 6C also demonstrating reduced susceptibility. Many common serotypes were sources of macrolide resistance, particularly serotypes 33F and 15A (<25% susceptibility to clarithromycin), 12F and 19A (<40%), as well as 6C, 11A, 22F and 35B (<80%). Reduced susceptibility to clindamycin and doxycycline was seen particularly in serotype 15A (39% and 23%, respectively), as well as 19A and 33F. Trimethoprim-sulfamethoxazole susceptibility was lowest for serotype 33F (33%), with serotypes 6C, 11A and 19A also demonstrating varying levels of resistance. Serotypes 22F and 35B deviated very slightly from full susceptibility to levofloxacin, whilst serotype 3 demonstrated the lowest susceptibility (91%) to chloramphenicol.

FIGURE 5.1: Differing susceptibilities of key antimicrobial agents against the eleven most common *S. pneumoniae* serotypes, plus 35B.
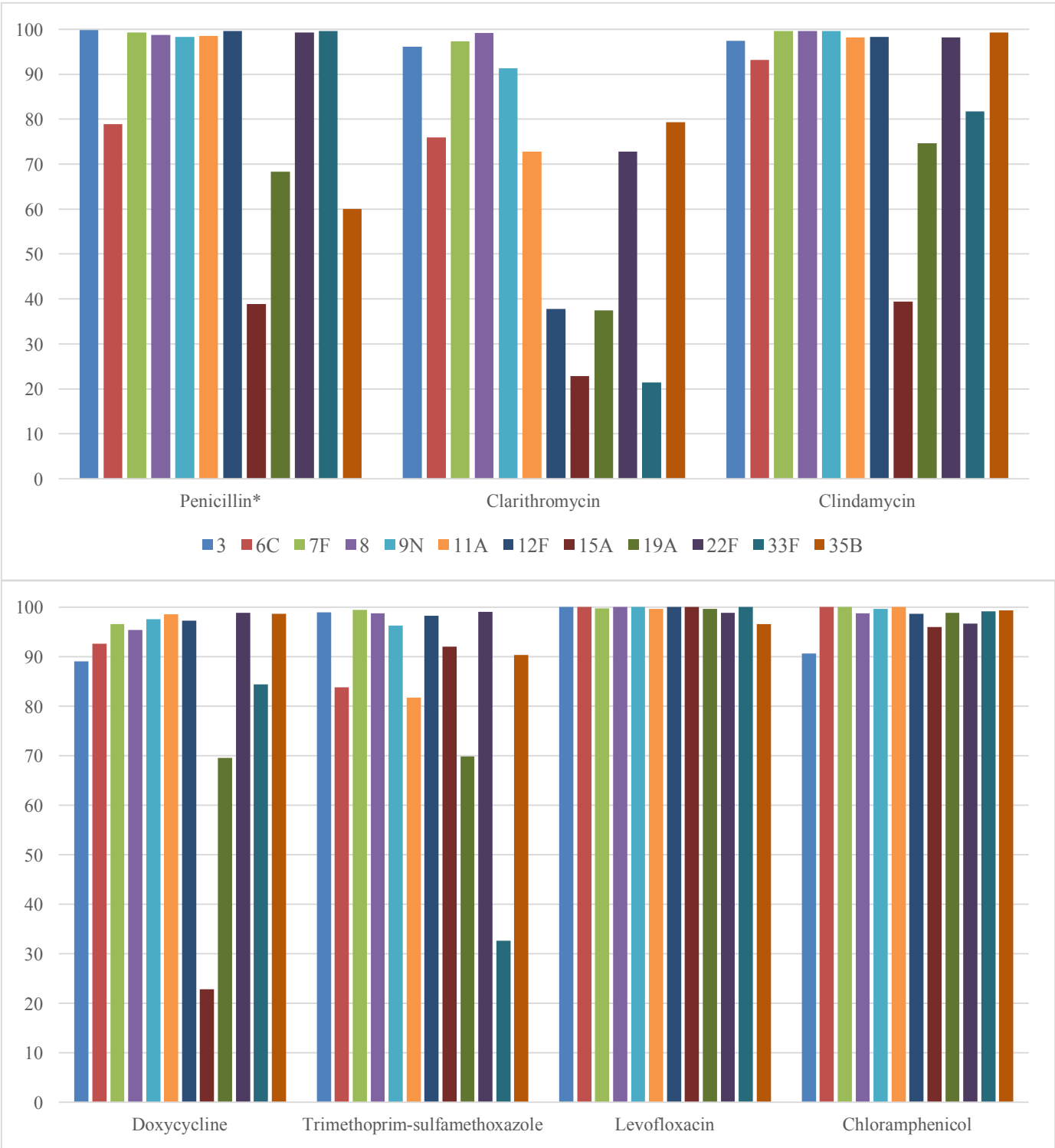
*, oral penicillin V breakpoints.

5.2 Multidrug Resistance

Overall, from 2011 to 2015, 6.2% (377/6,060) of *S. pneumoniae* isolates were MDR. However, the rate of MDR significantly decreased from 8.5% (116/1,362) in 2011 to 5.6% (64/1,150) in 2015 (P=0.0051). The only serotype of interest to demonstrate a significant change in MDR rate was serotype 15A, which decreased from 67.2% MDR in 2011 to 40.0% in 2015 (P=0.0125). The highest proportion of MDR isolates was seen in Eastern Canada (7.4%, 54/733), followed by Central Canada (6.3%, 252/4,001) and Western Canada (5.4%, 71/1,326). Infants between the ages of one and two had a significantly higher rate of MDR (12%, 24/184, P=0.0054) when compared to the overall rate of 6.2%.

MDR varied dramatically by serotype, as demonstrated in Figure 5.2. The MDR rates of the twelve serotypes of interest are listed above their respective bars, with 15A (57%, 100/175), 19A (26%, 155/600), 33F (6.3%, 14/224) and 6C (5.9%, 12/204) demonstrating the highest MDR rates among the types of interest. As depicted in Figure 5.3, each of these four serotypes had one antimicrobial resistance pattern that predominated. For serotypes 6C, 15A and 33F, the most common pattern included resistance to clarithromycin, clindamycin and doxycycline (92%, 82% and 86% of isolates, respectively). The most common resistance pattern for serotype 19A was an XDR pattern, with 61% of isolates demonstrating resistance to clarithromycin, clindamycin, doxycycline, penicillin and trimethoprim-sulfamethoxazole. Serotypes 6C and 33F only demonstrated one other MDR pattern, while serotypes 15A and 19A demonstrated many. Serotype 15A demonstrated five other patterns, including two XDR patterns. Serotype 19A isolates were highly diverse, demonstrating 13 additional resistance patterns, including two other XDR phenotypes. Several other, less frequently isolated serotypes also demonstrated MDR. Serotypes demonstrating > 5% MDR included 6B (30%, 6/20), 9V (22%, 4/18), 14 (25%, 6/24),

15F (100%, 1/1), 19F (27%, 22/82), 23F (11%, 2/18), 24F (8%, 2/26), 28A (6%, 1/17) and 35A (50%, 1/2).

FIGURE 5.2: MDR and XDR of *S. pneumoniae* isolates by serotype, 2011-15. MDR rates for the twelve serotypes of interest are listed above their respective bar. Classes included in the MDR/XDR definition include: β-lactams, macrolides, lincosamides, tetracyclines, fluoroquinolones, trimethoprim-sulfamethoxazole and chloramphenicol.
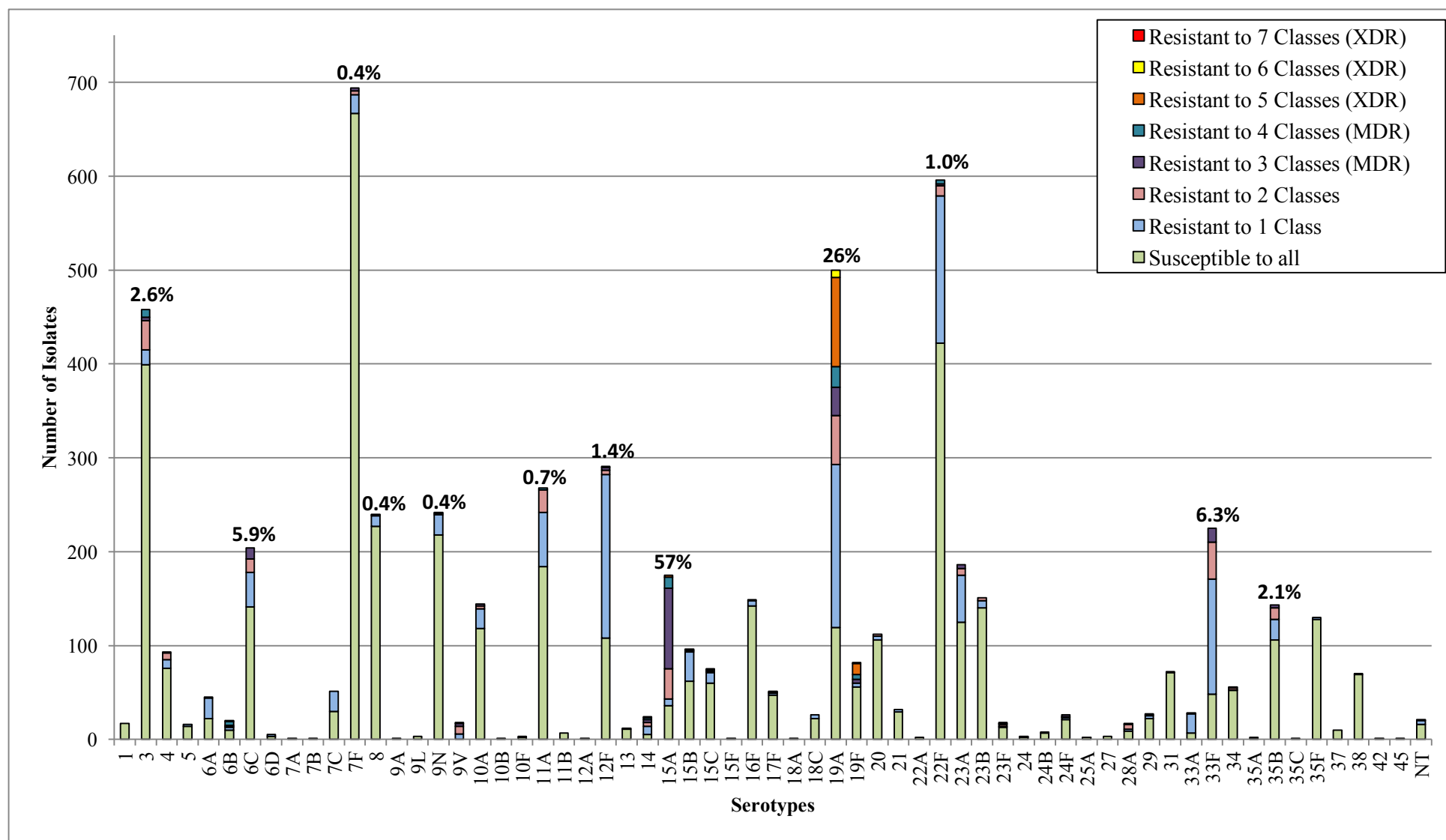
FIGURE 5.3: Antimicrobial resistance patterns demonstrated by *S. pneumoniae* serotypes 6C, 15A, 19A and 33F.
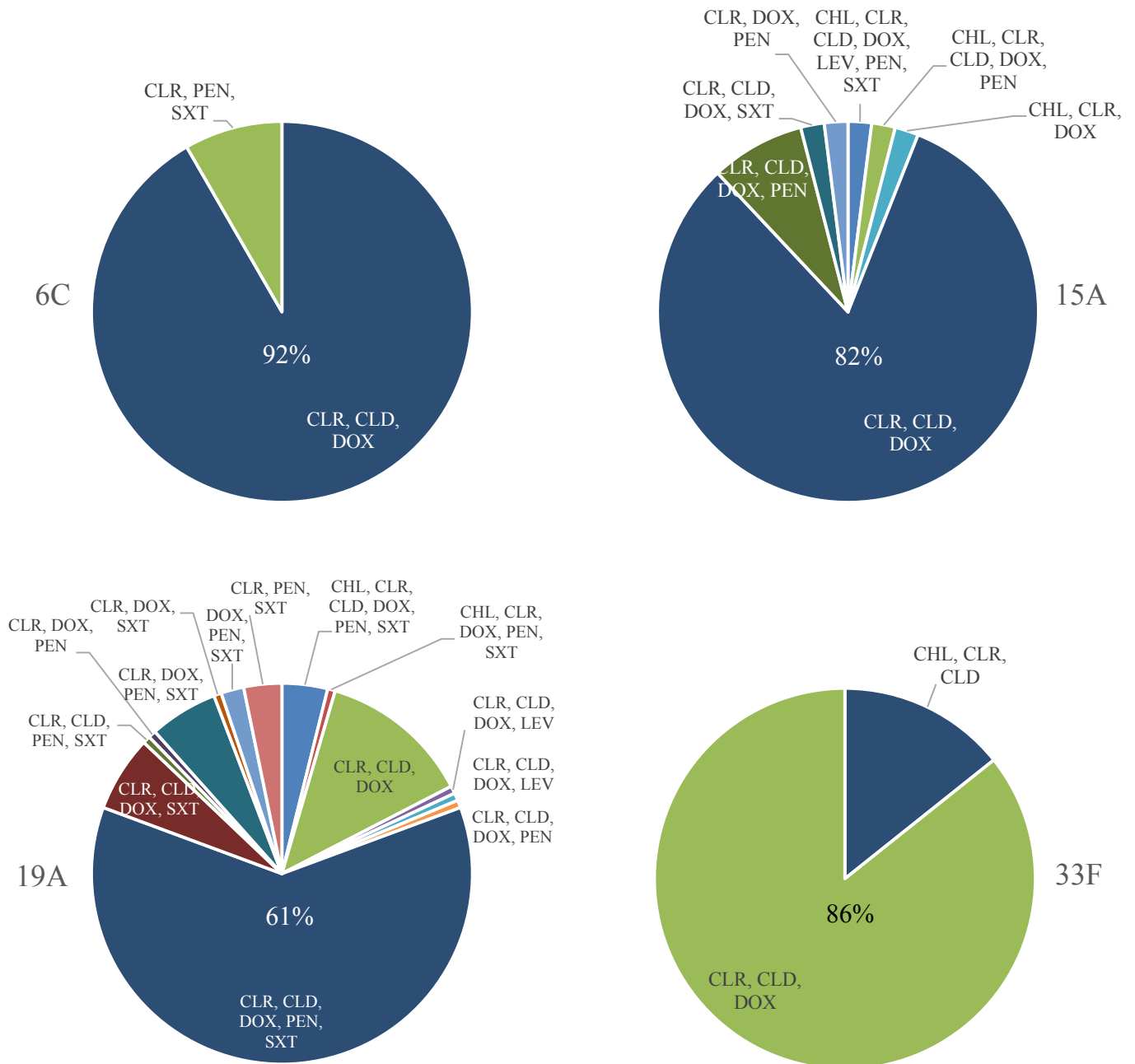
CHL, chloramphenicol; CLD, clindamycin; CLR, clarithromycin; DOX, doxycycline; LEV, levofloxacin; PEN, penicillin; SXT, trimethoprim-sulfamethoxazole.

Although the rate of MDR changed significantly over the course of the study, the rate of XDR isolates remained relatively constant at 2.0% of all *S. pneumoniae* isolates (2.5% in 2011, 1.7% in 2015, *P*=0.14). Only serotypes 6B (17% of MDR 6B isolates were XDR, 1/6), 14 (17%, 1/6), 15A (2%, 2/100), 15B (50%, 1/2), 15C (67%, 2/3), 19A (66%, 103/155), 19F (59%, 13/22) and 23F (50%, 1/2) demonstrated XDR phenotypes. The only instance of concomitant resistance to seven antimicrobial classes was the serotype 23F isolate, which demonstrated resistance to chloramphenicol, clarithromycin, clindamycin, doxycycline, levofloxacin, penicillin and trimethoprim-sulfamethoxazole.

5.3 Analysis of Antimicrobial Resistance Genes

Of the 192 isolates characterized by whole genome sequencing, 33 (see Appendix A) demonstrated discrepancies between genotype and phenotype for one or more antimicrobials. After repeating the susceptibility testing in triplicate, two isolates (SC13-2616-P and SC14-0090-P) would not grow for repeat testing and were therefore removed from the analysis of resistance genes, leaving 190 genomes available for this analysis.

5.3.1 Acquired Resistance Genes

ResFinder 2.1 identified acquired resistance genes for macrolides/lincosamides, tetracycline and chloramphenicol, as shown in Table 5.2. Both *mefA* and *ermB* macrolide resistance genes were common amongst tested isolates. Isolates carrying *mefA* were found to be 95.0% nonsusceptible to clarithromycin. *ermB* was more commonly identified in this cohort of isolates than *mefA*, and was found in a larger number of serotypes. The majority (98.0%) of isolates with *ermB* were nonsusceptible to clarithromycin, but only 82.0% demonstrated nonsusceptibility to clindamycin. Of the two macrolide resistance determinants on their own, *ermB* was more commonly associated with MDR isolates; 82.0% (41/50) of *ermB* carrying

isolates were also MDR, as opposed to 25.0% (5/20) for *mefA*. The dual *mefA*/*ermB* genotype was only present in serotype 19A and 19F isolates, specifically those demonstrating MDR and XDR phenotypes. However, this dual genotype did not necessarily convey full resistance to macrolides and lincosamides, with isolates demonstrating 94.4% and 61.1% nonsusceptibility, respectively. Interestingly, upon further inspection, six of 18 isolates that exhibited the dual *mefA*/*ermB* genotype but were clarithromycin susceptible, intermediate, or demonstrated only low-level resistance (1 µg/mL) contained a truncated version of the *ermB* gene caused by a premature stop codon at base 642.

The *tetM* gene associated with tetracycline resistance was present in 35.3% (67/190) of isolates and conferred 94.0% nonsusceptibility to doxycycline. Four isolates carried an intact *tetM* gene but were not resistant to doxycycline, perhaps indicating a nonfunctional gene. No premature stop codons or mutations in ribosomal binding sites were identified. Nineteen different serotypes were found to carry *tetM*, and 88.1% (59/67) of these isolates were also MDR.

Only a small number of isolates, predominantly serotype 3, carried the *cat* gene (8, 5.0%). However, possession of this gene invariably provided resistance to chloramphenicol (100% resistance) and all isolates carrying *cat* were also MDR.

TABLE 5.2: Acquired resistance genes identified using ResFinder 2.1 in 190 *S. pneumoniae* whole genomes.

Antibiotic Class	Resistance Gene	Count (%)	S/I/R (n)	%S	%NS	Serotypes	%MDR
Macrolide/	*mefA* only	20 (10.5)	1/2/17[a]	5.0	95.0	6ABC(6), 9V(2), 14(3), 12F(1), 15B(1), 19A(1), 22F(1), 29(1), 35B(4)	25.0 (5)
Lincosamide/	*ermB* only	50 (26.3)	1/1/48[a]	2.0	98.0	3(6), 6BC(3), 7F (1), 8(1), 9N(1), 11A(2), 12F(2), 15AB(15), 17F(2), 19A(8), 22F(3), 23AF(2), 24F(1), 33F(3)	82.0 (41)
Streptogramin			9/0/41[b]	18.0	82.0		
	Dual	18 (9.5)	1/2/15[a]	5.6	94.4	19A(17), 19F(1)	100 (18)
			7/0/11[b]	38.9	61.1		
	None	102 (53.7)	99/0/3[a]	97.1	2.9	-	2.0 (2)
			101/0/1[b]	99.0	1.0		
Tetracycline	*tetM*	67 (35.3)	4/0/63	6.0	94.0	3(6), 6BC(3), 7F(1), 8(1), 9N(1), 10A(1), 11A(1), 12F(2), 15AB (15), 17F(2), 19AF(26), 22F(2), 23F(1), 24F(1), 25F(1), 33F(3)	88.1 (59)
	None	123 (64.7)	119/2/2	96.7	3.3	-	5.7 (7)
Chloramphenicol[c]	*cat*	8 (5.0)	0/0/8	0	100	3(5), 15B(1), 19A(1), 23F(1)	100 (8)
	None	152 (95.0)	152/0/0	100	0	-	36.8 (56)

S, susceptible; I, intermediate; R, resistant; NS, nonsusceptible. [a], susceptibility to clarithromycin. [b], susceptibility to clindamycin.

[c], isolates from the BESST study were not tested for chloramphenicol susceptibility, therefore n=160 for chloramphenicol.

5.3.2 Chromosomal Mutations

5.3.2.1 Penicillin-Binding Proteins

Alterations in key motifs of PBPs were discovered in 82/190 (43%) isolates with whole genomes available (Tables 5.3 and 5.4), particularly in serotypes 19A and 15A. The most common alteration was a lone Thr451Ala mutation in the SSNT motif of *pbp2B*; this mutation was found in a variety of serotypes and resulted in 73.7% penicillin-nonsusceptibility. Conversely, isolates with solely a *pbp2X* alteration were fully susceptible to penicillin. The most frequent alteration in *pbp2X* was Thr338Ala in the STMK motif, found commonly in serotype 3 isolates; however, Ala393Thr and His394Leu in the AHSSNV motif and Leu546Ile/Val in the LKSGT motif were also seen (Table 5.4). No isolate was found to contain mutations in *pbp1A* alone.

The highest penicillin MICs were most commonly associated with mutations in all three PBPs (100% nonsusceptibility overall). Ten unique sets of alterations were observed; patterns that were identified in multiple isolates tended to be specific to one or two serotypes. The most common set of alterations (16/82 isolates) was Thr371Ser in STMK and Pro432Thr in SRNVP of *pbp1A*, Thr451Ala in SSNT and Ala624Gly in KTGTA of *pbp2B,* and Thr338Ala in STMK and Leu546Val in LKSGT of *pbp2X* (Table 5.4). This pattern of alterations was exclusively associated with serotype 19A isolates with MDR/XDR phenotypes. Interestingly, the second most common set of alterations (14/82 isolates) differed in only two respects, having Thr371Ala instead of serine in STMK of *pbp1A,* and not having Ala624Gly in KTGTA of *pbp2B*. This pattern was commonly associated with MDR serotypes 15A and 9V. Serotype 35B, which demonstrated decreased penicillin susceptibility, also exhibited a unique set of alterations, specifically Thr371Ser in STMK of *pbp1A*, Thr451Ala in SSNT of *pbp2B*, and Thr338Ala in

STMK and Leu546Val in LKSGT of *pbp2X*. This serotype also demonstrated two other sets of alterations that varied from the above pattern only in *pbp2B*; these included a pattern with no alterations in *pbp2B* and one with a unique Lys620Gln alteration in the KTGTA motif.

Although mutations in the STMK motif of *pbp1A* were common, Thr371Ser was found to be unique to serotypes 19A/F and 35B (Table 5.4). Serotypes 19A/F also had unique alterations Ala624Gly in the KTGTA motif of *pbp2B,* as well as the only noted double alteration in one motif, Thr338Ala and Met339Phe in STMK of *pbp2X*. Two serotypes that were uncommonly penicillin resistant, serotypes 12F and 33F, each demonstrated one alteration in *pbp2X* that did not confer any change in susceptibility, but were not seen in any other serotype. Serotype 12F exhibited a Ala393Thr alteration in the AHSSNV motif, while serotype 33F demonstrated Leu546Ile in the LKSGT motif.

TABLE 5.3: Chromosomal mutations conferring resistance identified in 190 *S. pneumoniae* whole genomes.

Antibiotic Class	Resistance Gene	Count (%)	S/I/R (n)	%S	%NS	Serotypes	%MDR
β-Lactam	*pbp2B* only	19 (10.0)	5/14/0	26.3	73.7	6ABC(4), 7F(1), 8(1), 10A(1), 15A(6), 19A(2), 22F(2), 23B(2)	57.9 (11)
	pbp2X only	12 (6.3)	12/0/0	100	0	3(4), 5(1), 11A(1), 12F(1), 15B(1), 16F(1), 19A(2), 33F(1)	41.7 (5)
	1A+2B	1 (0.5)	0/1/0	0	100	24F(1)	0
	1A+2X	1 (0.5)	0/1/0	0	100	35B(1)	0
	2B+2X	7 (3.7)	2/4/1	28.6	71.4	6C(2), 15A(2), 19A(3)	71.4 (5)
	1A+2B+2X	42 (22.1)	0/10/32	0	100	6B(1), 9V(4), 15AB(7), 19AF(23), 23F(1), 29(1), 35B(5)	83.3 (35)
	None	108 (56.8)	104/4/0	96.3	3.7	-	9.3 (10)
Fluoroquinolone	*parC* S79 only	4 (2.1)	2/0/2[a]	50.0	50.0	11A(1), 19A(2), 22F(1)	50.0 (2)
	gyrA S81 only	3 (1.6)	2/1/0[b]	66.7	33.3	9N(1), 19A(1), 35B(1)	33.3 (1)
	Both	8 (4.2)	0/0/8[a]	0	100	6A(2), 11A(1), 19A(1), 22F(2), 23F(2)	50.0 (4)
			0/4/4[b]	0	100		
	None	175 (92.1)	175/0/0[a]	100	0	-	33.7 (59)
			175/0/0[b]	100	0		

TMP-SMX							
TMP-SMX	*folA* I100L only	1 (0.5)	0/1/0	0	100	19A(1)	0
	folP mutation only	13 (6.8)	4/8/1	30.8	69.2	5(1), 10A(1), 15BC(4), 18C(1), 19A(1), 23B(1), 24F(1), 25F(1), 33F(2)	7.7 (1)
	Both	38 (20.0)	0/1/37	0	100	5(1), 6ABC(4), 9V(4), 10A(1), 11A(2), 15AB(2), 19AF(21), 23F(1), 35B(2)	81.6 (31)
	None	138 (72.6)	137/0/1	99.3	0.7	-	24.6 (34)

S, susceptible; I, intermediate; R, resistant; NS, nonsusceptible. [a], susceptibility to levofloxacin. [b], susceptibility to moxifloxacin.

TABLE 5.4: Penicillin-binding protein (PBP) 1A, 2B and 2X conserved amino acid motif alterations in 190 _S. pneumoniae_ whole genomes.

Number of Isolates	PBP1A			PBP2B			PBP2X			Penicillin MIC/ range (µg/mL)	% MDR	Common Serotype?
	STMK	SRNVP	KTG	SVVK	SSNT	KTGTA	STMK	AHSSNV	LKSGT			
108[a]	----	-----	---	----	----	-----	----	------	-----	≤0.03 – 0.12	9.3 (10)	NA[b]
19	----	-----	---	----	---A	-----	----	------	-----	≤0.03 – 0.25	52.6 (10)	NA[b]
16	-S--	----T	---	----	---A	----G	-A--	------	V----	0.5 – 4	100 (16)	19A
14	-A--	----T	---	----	---A	-----	-A--	------	V----	1 – 2	85.7 (12)	9V, 15A
6	----	-----	---	----	----	-----	-A--	------	-----	≤0.03 – 0.06	66.7 (4)	3, 19A
5	-S--	-----	---	----	---A	-----	-A--	------	V----	0.5 – 2	20.0 (1)	35B
3	----	-----	---	----	---A	-----	-A--	------	-----	0.06 – 2	66.7 (2)	NA[b]
3	----	-----	---	----	----	-----	----	------	V---	≤0.03	0	NA[b]
2	-S--	----T	---	----	---A	----G	-AF-	------	V----	4	100 (2)	19AF
2	----	-----	---	----	---A	-----	----	-L----	-----	0.12	100 (2)	19A
2	----	-----	---	----	---A	-----	-A--	------	V----	≤0.03 – 0.25	50.0 (1)	15A
1	-A--	----T	---	----	---A	----G	-A--	------	V----	2	100 (1)	19A
1	-A--	-----	---	----	---A	-----	-A--	------	V----	2	100 (1)	19A
1	-S--	----T	---	----	---A	-----	-A--	------	V----	2	100 (1)	19A
1	----	-----	---	----	----	-----	----	------	I---	≤0.03	100 (1)	12F
1	----	----T	---	----	---A	-----	----	-L----	V----	0.25	100 (1)	19A
1	----	-----	---	----	----	-----	----	T-----	-----	≤0.03	0	33F
1	----	-----	---	----	----	-----	----	-L----	-----	≤0.03	0	5
1	-A--	----T	---	----	---A	-----	----	------	-----	0.12	0	24F
1	-S--	-----	---	----	----	-----	-A--	------	V----	0.25	0	35B
1	-S--	-----	---	----	---A	N----	-A--	------	V----	2	0	35B

[a], the pattern demonstrated by these isolates was considered wild-type sequence. [b], NA, several different serotypes demonstrated this pattern of PBP alterations.

5.3.2.2 DNA Topoisomerase IV (*parC*) and DNA Gyrase (*gyrA*)

Mutations in the QRDR regions of *parC* and *gyrA* were uncommon, with only 15/190

isolates (7.9%) demonstrating alterations (Table 5.3). Four isolates exhibited alterations in the

QRDR of *parC*, three with Ser79Phe and one with Ser79Tyr. These mutations conferred 50%

nonsusceptibility to levofloxacin, the fluoroquinolone that preferentially targets *parC*. Similarly,

only three isolates were identified with mutations in *gyrA*, all of which were Ser81Phe. These

isolates demonstrated 33.3% susceptibility to moxifloxacin, which preferentially targets *gyrA*.

Eight isolates were determined to have mutations in the QRDR region of both *parC* and

gyrA. Four isolates contained Ser79Phe and Ser81Phe mutations in *parC* and *gyrA*, respectively,

two contained Ser79Tyr and Ser81Phe, and two contained Ser79Phe and Ser81Leu. These last

two isolates were both serotype 22F with a resistance pattern of clarithromycin, clindamycin,

doxycycline and levofloxacin. Overall, about half of the isolates demonstrating mutations in

these genes were also MDR or XDR (Table 5.3). Other than serotype 22F as mentioned above,

mutations in *parC* and *gyrA* were not particular to serotype. However, mutations in one or both

genes were essential for fluoroquinolone resistance, as isolates with neither mutation were fully

susceptible to both levofloxacin and moxifloxacin.

5.3.2.3 Dihydrofolate Reductase (*folA*) and Dihydropteroate Synthase (*folP*)

Mutations in *folA* and localized insertions in *folP* were identified in 52/190 isolates

(27.3%) (Table 5.3). Only one of the above 52 isolates possessed the Ile100Leu mutation in *folA*

with no accompanying alteration in *folP*. As expected, this isolate demonstrated intermediate

resistance to trimethoprim-sulfamethoxazole. Isolates with *folP* insertions with no corresponding

folA mutation were more common (17/52), but similarly provided intermediate resistance or

susceptibility to trimethoprim-sulfamethoxazole. Only one of the 18 isolates with a change in a

single gene was also MDR. Dual alteration of both *folA* and *folP* were more commonly identified

than either single mutation (38/52). The combination of both mutations conferred 100%

nonsusceptibility to trimethoprim-sulfamethoxazole, and were more commonly associated with

MDR (81.6%).

Interestingly, there were seven different *folP* insertions of one or two amino acids

between codons 59 and 69 that were associated with varying levels of resistance, MDR and

serotype specificity (Table 5.5). Insertion of Arg-Gly (RG) after codon 60 was specifically

associated with serogroup 15; however, this insertion was not associated with the Ile100Leu

mutation in *folA* and was therefore also associated with intermediate trimethoprim-

sulfamethoxazole MICs and no MDR. Conversely, insertion of an extra Serine around codons

61-62 was not tied to a specific serotype, however this insertion always had the corresponding

Ile100L mutation in *folA*, and was consistently associated with trimethoprim-sulfamethoxazole

resistance and MDR. The most common insertion consisted of an extra Glu-Ile (EI) after codon

66 of *folP*. This alteration was present in combination with *folA*-Ile100Leu in 17 serotype 19A

isolates, was associated with the highest trimethoprim-sulfamethoxazole MICs, and isolates were

either MDR or XDR. Upon further investigation, these serotype 19A isolates were the same as

those demonstrating the most common set of alterations in all three PBPs (Table 5.4), perhaps

indicating that these specific chromosomal mutations are a clonal property.

TABLE 5.5: One and two codon insertions identified between codons 59 and 69 of *folP* in 190 *S. pneumoniae* whole genomes. Inserted codons are displayed in red.

Number of Isolates	Amino acid sequence	SXT[a] MIC/range (µg/mL)	% MDR	Common Serotype?	*folA* I100L? (Y/N)[b]
139	58 RPGSSYVEIE 67[c]	≤0.12 – 8	24.5 (34)	NA[d]	N
8	58 RP**RP**GSSYVEI 69	0.5 – >8	50.0 (4)	NA[d]	Y (6)
4	58 RPG**RG**SSYVEIE 69	0.5 – 1	0	15BC	N
9	58 RPG**R**SSYVEIE 68	0.5 – >8	55.6 (5)	NA[d]	Y (8)
3	58 RPGSS**S**YVEIE 68	4 – 8	100 (3)	NA[d]	Y (3)
7	58 RPGSSY**SY**VEIE 69	0.25 – >8	28.6 (2)	NA[d]	Y (3)
3	58 RPGSSYVE**VE**IE 69	1 – >8	33.3 (1)	33F, 19A	Y (1)
17	58 RPGSSYVEI**EI**E 69	8 – >8	100 (17)	19A	Y (17)

[a], SXT = trimethoprim-sulfamethoxazole. [b], number in brackets indicates the number of isolates in the corresponding category that also contain the I100L mutation in *folA*. [c], wild-type sequence.

[d], NA, several different serotypes demonstrated this *folP* insertion.

Part 2: Genetic Relatedness of Select *S. pneumoniae* Collected by the SAVE 2011-2015 Study

5.4 Molecular Characterization of 12 Serotypes by PFGE, MLST and Phylogenomic Analysis

5.4.1 The Big Picture

The first method utilized to assess the genetic relatedness of the twelve *S. pneumoniae* serotypes of interest was PFGE. Due to the diversity of the organism, all *S. pneumoniae* isolates typed were generally considered to be unrelated to each other (<80% similarity). However, when clusters were identified, they were largely comprised of an individual serotype. Due to the size of the dendrogram generated by the 585 typed isolates, a PFGE figure for all isolates was not included. Dendrograms for each individual serotype will be discussed beginning in Section 5.4.2.

MLST was performed following PFGE to better discriminate relationships between isolates (Figure 5.4). Over 100 different STs were identified in the twelve serotypes of interest, generating 29 clusters of two or more isolates and 24 singletons. Of the clusters identified, 12 were related to PMEN international clones originally isolated on four different continents. These included PMEN clones 3 (Spain9V), 14 (Taiwan19F), 18 (Tennessee14), 21 (Portugal19F), 24 (Utah35B), 25 (Sweden15A), 26 (Colombia23F), 31 (Netherlands3), 33 (Netherlands8), 34 (Denmark12F), 37 (Netherlands15B) and 39 (Netherlands7F). A variety of potential capsular switch variants were also noted, including serotype 15A-ST63 to 8-ST63, 19A-ST63 and ST63-related-22F; 9V-ST156 to 11A-ST156 and ST156-related-15A; 19F-ST177 to 3-ST177 and ST177-related-33F; 14-ST67 to numerous ST67-related-9N and 12F; 15B-ST199 to 19A-ST199; and 12F-ST218 to ST218-related-9N.

FIGURE 5.4: Minimum spanning tree (generated by PHYLOViZ 2.0) of MLST sequence types demonstrated by the twelve *S. pneumoniae* serotypes of interest collected by the SAVE 2011-2015 study. Green outlines indicate a group founder; numbers indicate the number of differences between the MLST profiles of the two connected nodes (≤ 2 indicates the two nodes are part of a cluster). Clusters with relation to PMEN international clones are highlighted in grey and listed along with the representative serotype for that clone.

81
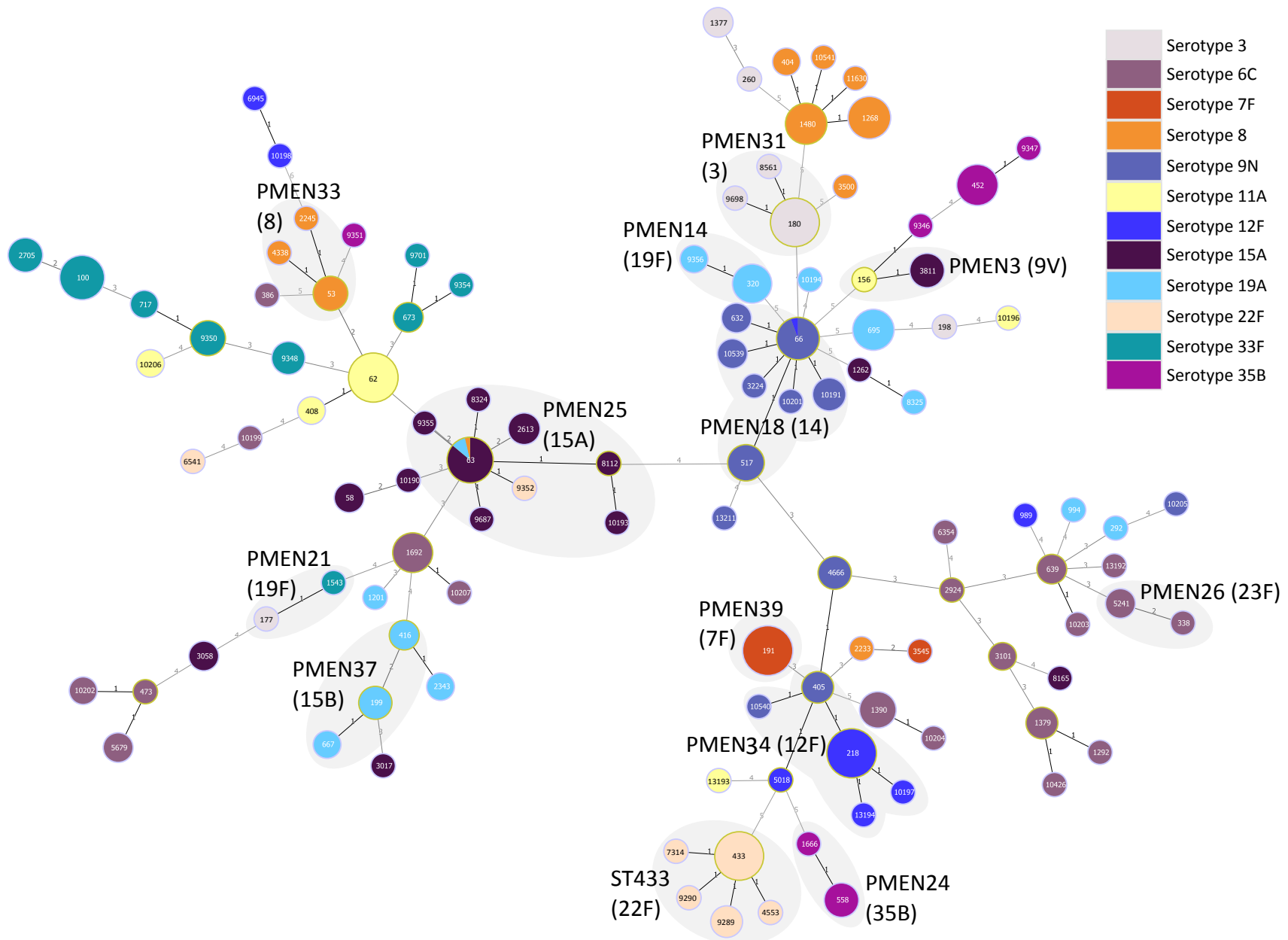
WGS was performed to obtain the most in depth information on the genetic content of a subset of the twelve serotypes of interest, as well as background and temporally diverse strains. A maximum likelihood tree of the identified SNVs of 188 isolates produced 35 different clusters of isolates, which included as few as two isolates and as many as 21 (Figure 5.5). As with MLST, many of these clusters demonstrated relatedness to PMEN clones; additional PMEN clones to those listed above include 1 (Spain23F), 4 (Tennessee23F), 19 (Colombia5), 28 (Sweden1), 30 (Greece21) and 32 (Denmark14). Most of the serotypes in these additional clone groups were considered background strains, however some potential capsular switching was noted, which will be discussed further beginning in Section 5.4.2. Using *S. pneumoniae* R6 as a reference genome resulted in only 49.37% of the positions that were valid, included and part of the whole genome being included in the creation of the phylogenetic tree; most filtered SNVs were considered "invalid", falling within regions masked from the final alignment. As this coverage was relatively low, the tree was divided into 14 smaller groups for further analysis with a more appropriate reference genome (Table 5.6). As this was the only characterization method that included background isolates, they were included in the subsequent analysis of smaller groups/serotypes of interest to provide context/additional information.

FIGURE 5.5: Whole genome core SNV maximum likelihood phylogenetic tree of 188 *S. pneumoniae* isolates rooted on NCBI reference genome SPN R6 (NC_003098). Tree was generated using PhyML and visualized using FigTree. All coloured clusters were identified by ClusterPicker. Smaller groups for detailed analysis are labelled 1-14; detailed information about these groups can be found in Table 5.6.
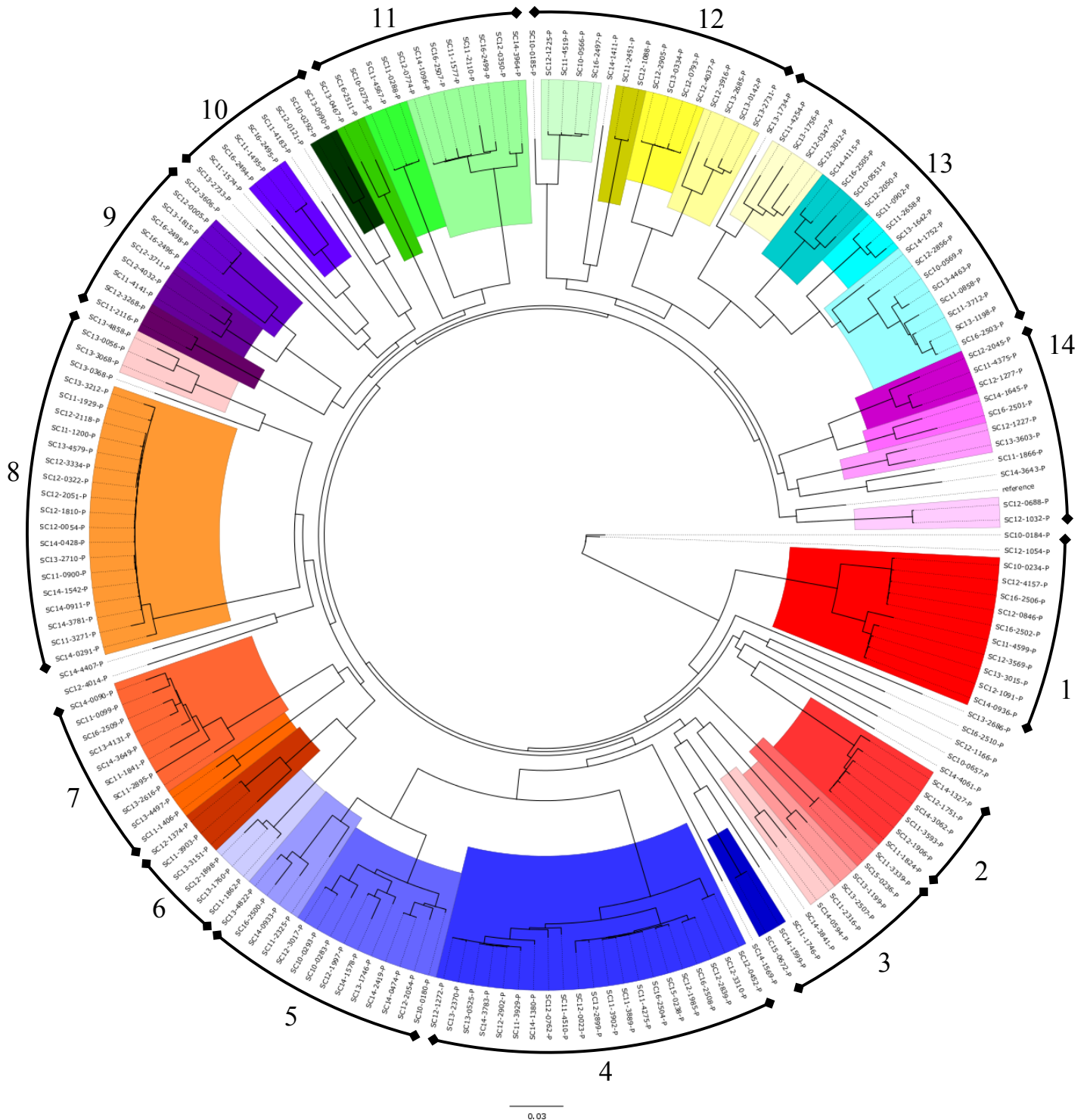
TABLE 5.6: Detailed information on the 14 clusters of *S. pneumoniae* whole genomes highlighted in Figure 5.5.

Cluster #	# of Isolates	Serotypes (n)	STs	Reference Strain[a]	Percent Coverage[b]	Corresponding Tree Figure #
1	12	3 (10)	180	OXC141	81.02	5.26
		7F (2)	191			
2	6	35B (5)	558 and variants	PMEN24	90.35	5.43
		29 (1)				
3	10	1 (2)	306	SPN032672	63.57	5.34
		21 (2)	432			
		20 (2)	1257, 6805			
		11A (1)	99			
		34 (3)	5854, 547			
4	21	15A (14)	63 and variants	G54	67.28	5.8
		19A (3)				
		22F (2)				
		7F (1)				
		8 (1)				
5	13	33F (5)	100 and variants	SC12-3017-P	61.50	5.23
		18C (4)	496			
		11A (3)	62			
		8 (1)	NF			
6	6	6A (2)	1876	PMEN21	84.25	5.14
		6B (1)				
		3 (1)	177 and variants			
		19A (1)				
		24F (1)				

7	10	9V (4)	156 and variants	70585	74.79	5.9
		14 (1)				
		15A (1)				
		19A (1)				
		35B (1)				
		5 (2)	289 and variants			
8	22	19A (17)	320 and variants	Taiwan19F-14	74.92	5.13
		19F (1)				
		6B (1)	138			
		6C (1)	338 and variants			
		23B (2)				
9	9	12F (2)	218	PMEN4	81.07	5.39
		23A (4)	42 and variants			
		23F (3)	36			
10	10	14 (2)	9 and variants	INV200	72.49	5.15
		19A (1)				
		7C (1)	1797			
		11A (1)	10196			
		19F (1)	2631			
		6C (3)	1379, 3101			
		6D (1)	1692			
11	13	22F (7)	433	SC13-0467-P	83.28	5.31
		31 (1)				
		38 (3)	393			
		8 (1)	404			

12	16	4 (4)	205 and variants	TIGR4	73.17	5.10
		16F (3)	570, 1840			
		15B (1)	1292 and variants			
		15C (3)				
		15A (2)				
		19A (1)	193 and variants			
		21 (1)				
13	20	10A (1)	3135	ATCC 700669	72.37	5.16
		17F (1)	2355			
		24F (2)	72, 230			
		19A (5)	81, 199, 319, 667, 10194			
		15B (3)	199			
		23F (1)	13190			
		9N (7)	66 and variants			
14	11	35F (3)	446 and variants	R6	78.11	5.40
		10A (2)	97, 585			
		12F (2)	6945, 10198			
		31 (2)	1766, 13187			
		17F (2)	123			
Singles	8	6A (1)	490	NA[c]	NA[c]	NA[c]
		6B (1)	90			
		11B (1)	10552			
		13 (1)	574			
		14 (1)	124			
		25F (1)	105			
		35B (1)	198			
		37 (1)	447			

[a] additional information on reference strains can be found in Table 4.4. [b] positions that were valid, included and part of the whole genome. [c] singles were not included in any additional phylogenetic analysis.

5.4.2 Diverse Serotypes

5.4.2.1 Serotype 15A

Serotype 15A isolates grouped into three different clusters by PFGE, demonstrating greater than 80% relatedness (Figure 5.6). The largest group was Cluster 2, comprising almost two-thirds of the isolates typed by PFGE. These isolates were predominantly MDR, demonstrating resistance to clarithromycin, clindamycin and doxycycline, however some only displayed resistance to two of the above antimicrobials. Three isolates near the bottom of this cluster also exhibited penicillin resistance. Isolates in cluster 2 were collected from all regions, age groups and years. Clusters 1 and 3 were two separate clusters of susceptible isolates. Cluster 1 contained isolates from all years and provinces, however cluster 3 isolates were largely isolated from Ontario in the later years of the SAVE study (2014-15), perhaps indicating an outbreak.

MLST of serotype 15A isolates generated 14 different STs (Figure 5.7). Half of these STs were identical or related to ST63, a frequently MDR international clone (PMEN25). The other seven STs were predominantly susceptible to all antimicrobials. The colouration of the minimum spanning tree demonstrates that most ST63 isolates were obtained from Central Canada, while more susceptible STs had a higher proportion of isolates obtained from Western and Eastern Canada.

Relating the MLST data back to the PFGE findings, the two methods did relatively well at differentiating between ST63 isolates and those that were unrelated. PFGE Cluster 2 was comprised of most of the ST63-related isolates. Interestingly, the small group of penicillin-resistant isolates previously noted in this cluster all typed as ST2613, a DLV of ST63. The other two clusters of susceptible isolates separated by ST, with ST3058 isolates in cluster 1 and ST3811 isolates in cluster 3.

FIGURE 5.6: PFGE dendrogram of serotype 15A *S. pneumoniae* isolates collected by the SAVE 2011-15 study. MLST sequence types are listed for comparison.

ISOLATE	YR	PROV	GDR	AGE GROUP	ST
SC13-1449-P	2013	NB	M	18 to less than 50	ST58
SC15-0018-P	2014	ON	F	65 years or greater	ST58
SC12-1745-P	2012	QC	F	65 years or greater	ST58
SC13-1438-P	2013	NB	F	65 years or greater	ST3058
SC14-0157-P	2014	SK	F	18 to less than 50	ST3058
SC14-0151-P	2013	SK	F	65 years or greater	ST3058
SC12-0969-P	2012	ON	M	65 years or greater	ST10190
SC11-4159-P	2011	QC	F	65 years or greater	ST3017
SC14-0562-P	2014	ON	M	50 to less than 65	ST8165
SC15-0364-P	2015	ON	M	65 years or greater	ST1262
SC15-0372-P	2015	ON	M	50 to less than 65	ST63
SC15-1594-P	2015	ON	M	65 years or greater	ST63
SC11-2123-P	2011	QC	F	18 to less than 50	ST63
SC12-0339-P	2012	QC	F	65 years or greater	ST63
SC12-0763-P	2012	QC	F	65 years or greater	ST63
SC12-1271-P	2012	QC	F	50 to less than 65	ST63
SC12-1992-P	2012	QC	F	65 years or greater	ST9355
SC12-4025-P	2012	NLA	M	65 years or greater	ST63
SC13-3201-P	2013	ON	F	65 years or greater	ST10193
SC13-4631-P	2013	ON	F	65 years or greater	ST63
SC13-4693-P	2013	ON	F	65 years or greater	ST63
SC13-4928-P	2013	ON	F	50 to less than 65	ST63
SC12-1978-P	2012	QC	M	50 to less than 65	ST63
SC13-4466-P	2013	ON	F	50 to less than 65	ST63
SC14-3287-P	2014	ON	F	65 years or greater	ST63
SC11-2694-P	2011	ON	M	65 years or greater	ST63
SC12-0769-P	2012	QC	F	65 years or greater	ST63
SC12-1985-P	2012	QC	M	65 years or greater	ST63
SC12-3069-P	2012	NB	M	18 to less than 50	ST63
SC13-0525-P	2013	SK	M	18 to less than 50	ST2613
SC14-3646-P	2014	ON	M	18 to less than 50	ST8112
SC15-0746-P	2015	ON	F	65 years or greater	ST63
SC12-0358-P	2012	QC	F	65 years or greater	ST63
SC12-2897-P	2012	QC	F	50 to less than 65	ST63
SC11-2062-P	2011	ON	M	50 to less than 65	ST63
SC11-2651-P	2011	QC	M	50 to less than 65	ST63
SC11-4266-P	2011	QC	M	1 to less than 2	ST2613
SC12-2902-P	2012	QC	M	65 years or greater	ST2613
SC14-0503-P	2014	SK	F	18 to less than 50	ST2613
SC13-1761-P	2013	ON	M	18 to less than 50	ST8324
SC14-0217-P	2014	ON	F	65 years or greater	ST63
SC15-2589-P	2015	QC	M	2 to less than 6	ST63
SC15-0817-P	2015	PEI	F	65 years or greater	ST9687
SC15-1312-P	2015	ON	F	65 years or greater	ST3811
SC15-3627-P	2015	ON		65 years or greater	ST3811
SC14-0197-P	2014	ON	F	65 years or greater	ST3811
SC15-0023-P	2014	ON	F	65 years or greater	ST3811
SC14-1500-P	2014	MB	F	65 years or greater	ST3811
SC15-2235-P	2015	SK	F	65 years or greater	ST3811

GDR, YR, PROV: gender, year and province of isolation. Gaps in the demographics indicate that the value was not given at the time of collection. The dashed line indicates 80% similarity. Clusters of isolates are outlined with red boxes and labelled 1-3.

FIGURE 5.7: Minimum spanning tree (generated by PHYLOViZ 2.0) of MLST sequence types demonstrated by serotype 15A isolates, coloured according to province of isolation[a]. Green outlines indicate a group founder; numbers indicate the number of differences between the MLST profiles of the two connected nodes (≤ 2 indicates the two nodes are part of a cluster). PMEN clones are marked with red circles.
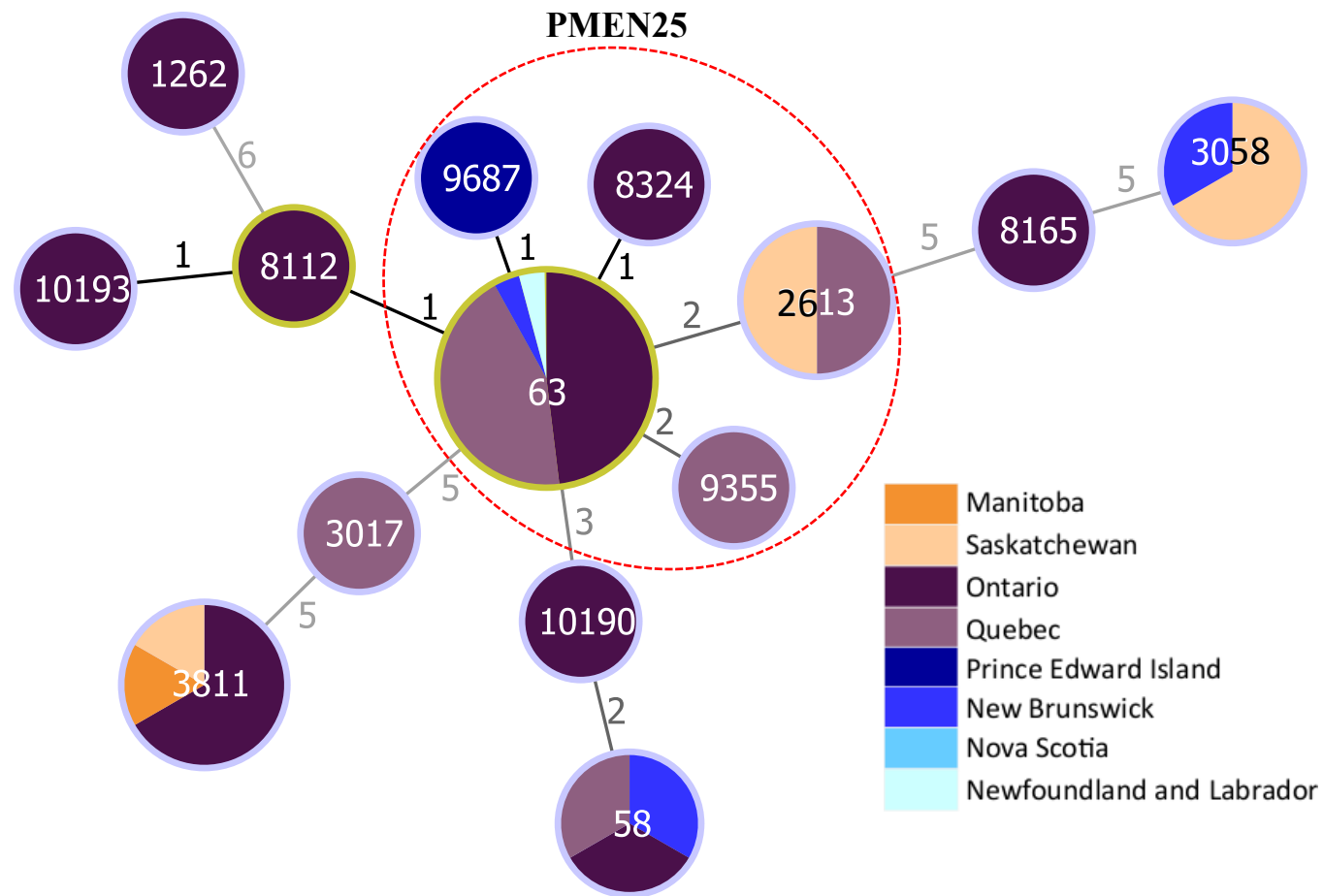
[a] Orange, Western Canada; Purple, Central Canada; Blue, Eastern Canada.

Most serotype 15A isolates grouped together in cluster 4 of the original phylogenetic analysis (Figure 5.5). These isolates were all found to be MDR and ST63 or a related type. However, looking at these isolates in more detail and with a more appropriate reference genome, there were smaller groups of serotype 15A within WGS cluster 4 (Figure 5.8). Cluster A contained those isolates previously described above that were resistant to penicillin. These strains were all ST2613 (as noted by PFGE/MLST) and possessed the second most common set of PBP mutations (as described in Section 5.3.2.1). Cluster B (though not formally a cluster picked by ClusterPicker) contained the isolates that typed as ST63; these all demonstrated the ST63 resistance pattern of clarithromycin, clindamycin and doxycycline and generally only possessed mutations in *pbp2B*. However, this group was broken down into three smaller clusters, two of which contained serotype 15A isolates. Cluster B2 contained solely ST63-15A isolates, while B3 encompassed both ST63-15A isolates and other serotypes bearing the same ST. These other serotypes, as well as cluster B1, will be discussed in future sections. Every serotype 15A isolate in this cluster possessed *ermB* and *tetM* genes.

Two other serotype 15A isolates were included in the phylogenetic analysis and each fell into a different, ST63-unrelated cluster. As shown in Figure 5.9, one isolate typed as ST3811 and fell into cluster 7, which was comprised of isolates identical and related to ST156 (PMEN3). As ST3811 is only a DLV of ST156, it fell just outside of a group of more closely related isolates. However, the similarity between this isolate and the 9V-ST156 isolates in the cluster may indicate a capsular switch event. Another potential capsular switch event was noted in cluster 12 (Figure 5.10), where the serotype 15A isolate was most closely related to isolates typing as, or related to, PMEN30 (serotype 21). Interestingly, this serotype 15A isolate was more related to the uncommon serotype 21 than the other serogroup 15 isolates (B, C) in the adjacent cluster A.

FIGURE 5.8: Whole genome core SNV maximum likelihood phylogenetic tree of *S. pneumoniae* "cluster 4" rooted on reference genome *S. pneumoniae* G54 (NC_011072). Tree was generated using PhyML and visualized using FigTree. Coloured clusters were identified by ClusterPicker. Clusters of interest are labeled A-B3.

Isolate	Year	Serotype	ST	Res. Pattern
SC12-0452-P	2012	22F	ST9352	CLR, CLD, DOX, LEV
SC12-3310-P	2012	22F	ST9352	CLR, CLD, DOX, LEV
reference	G54	19F	ST63	
SC14-1380-P	2014	15A	ST374	CLR, CLD, DOX
SC11-3929-P	2011	15A	ST2613	CLR, CLD, DOX, PEN
SC12-2902-P	2012	15A	ST2613	CLR, DOX, PEN
SC13-0525-P	2013	15A	ST2613	CLR, CLD, DOX
SC14-3783-P	2014	15A	ST2613	CLR, CLD, DOX, PEN
SC12-1272-P	2012	15A	ST2613	CLR, CLD, DOX, PEN
SC13-2370-P	2012	15A	ST2613	CLR, DOX, PEN
SC12-2899-P	2012	15A	ST63	CLR, CLD, DOX
SC12-0023-P	2011	15A	ST63	CLR, CLD, DOX, SXT
SC12-2839-P	2012	19A	ST63	CLR, CLD, DOX
SC16-2508-P	2008	19A	ST63	CLR, CLD, DOX
SC12-1985-P	2012	15A	ST63	CLR, CLD, DOX
SC15-0238-P	2014	15A	ST9355	CLR, CLD, DOX
SC11-3889-P	2011	7F	ST63	CLR, CLD, DOX
SC11-3902-P	2011	8	ST63	CLR, CLD, DOX
SC11-4275-P	2011	15A	ST63	CLR, CLD, DOX
SC16-2504-P	2008	15A	ST63	CLR, CLD, DOX
SC12-0762-P	2011	15A	ST63	CLR, DOX
PMEN25		15A	ST63	
SC11-4510-P	2011	19A	ST63	CLR, CLD, DOX

0.07

92

FIGURE 5.9: Whole genome core SNV maximum likelihood phylogenetic tree of *S. pneumoniae* "cluster 7" rooted on reference genome *S. pneumoniae* 70585 (NC_012468). Tree was generated using PhyML and visualized using FigTree. Coloured clusters were identified by ClusterPicker.

Isolate	Year	Serotype	ST	Res. Pattern
SC13-4497-P	2013	5	ST4219	
reference	SPN70585	5	ST289	
SC13-2616-P	2013	15A	ST3811	
SC11-2895-P	2011	19A	ST9345	CLR, PEN, SXT
SC14-0090-P	2013	14	ST156	DOX, LEV, PEN
SC11-0099-P	2011	9V	ST156	CLR, PEN, SXT
SC11-1841-P	2011	35B	ST9346	CLR, PEN, SXT
PMEN3		9V	ST156	
SC14-3649-P	2014	9V	ST156	CLR, PEN, SXT
SC13-4131-P	2013	9V	ST156	PEN, SXT
SC16-2509-P	2008	9V	ST1556	SXT
SC11-1406-P	2011	5	ST289	
PMEN19		5	ST289	

Spain9V-3

0.2
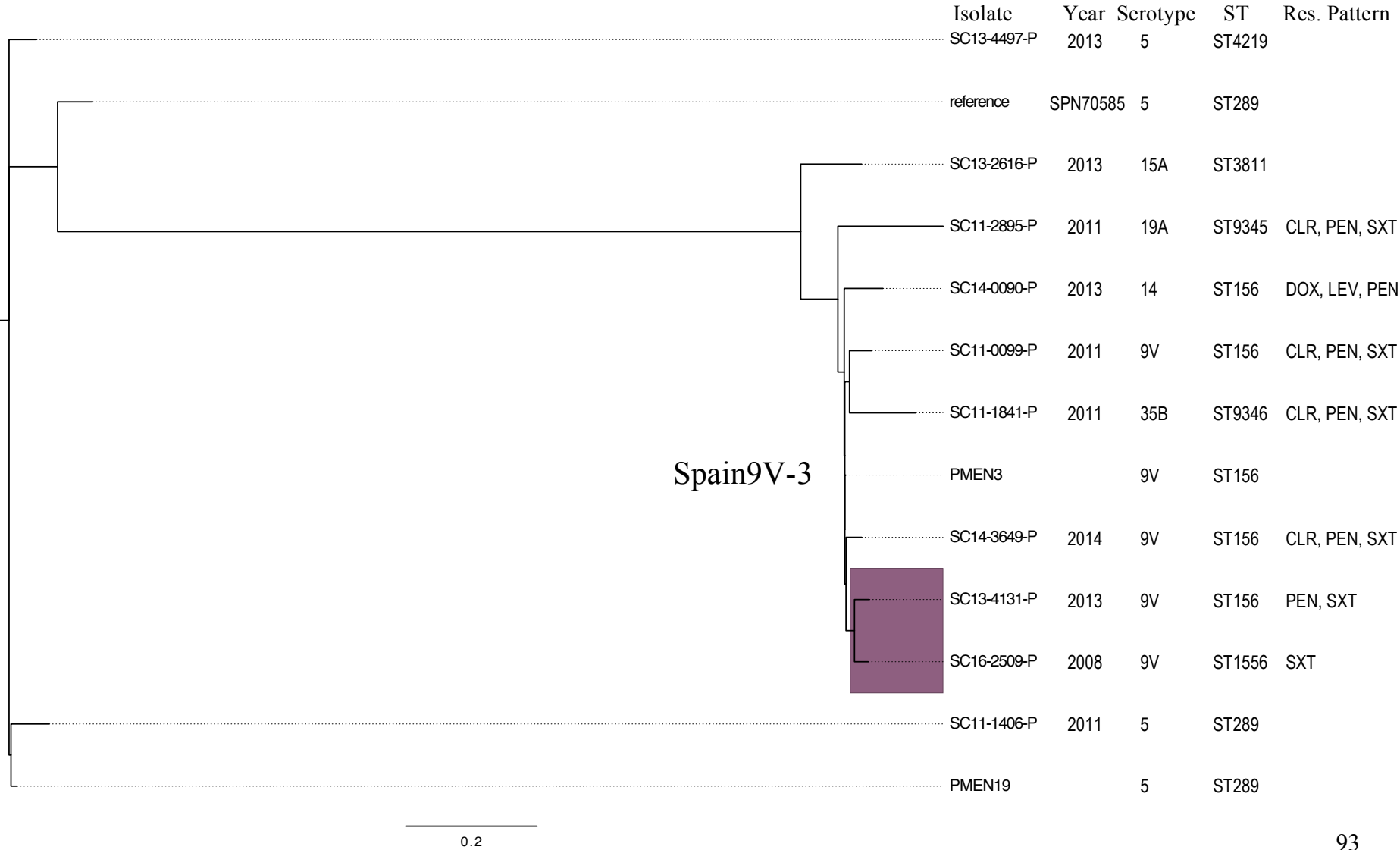
FIGURE 5.10: Whole genome core SNV maximum likelihood phylogenetic tree of *S. pneumoniae* "cluster 12" rooted on reference genome *S. pneumoniae* TIGR4 (NC_003028). Tree was generated using PhyML and visualized using FigTree. Coloured clusters were identified by ClusterPicker. Clusters of interest are labelled A-C.

Isolate	Year	Serotype	ST	Res. Pattern
SC11-2451-P	2011	16F	ST1840	
SC12-1088-P	2012	16F	ST1840	
SC14-1411-P	2014	16F	ST570	
SC12-0793-P	2012	15C	ST8325	
SC12-4037-P	2012	15C	ST1262	
SC12-2905-P	2012	15C	ST1262	
SC13-0334-P	2012	15B	ST1262	
SC12-3916-P	2012	15A	ST410	
SC13-2685-P	2013	19A	ST193	
SC13-2731-P	2013	21	ST193	
SC13-0142-P	2013	19A	ST193	
SC10-0185-P	2009	19A	ST695	
SC12-1225-P	2012	4	ST205	
SC10-0566-P	2009	4	ST3004	CLR
SC16-2497-P	2007	4	ST3004	
SC11-4519-P	2011	4	ST205	
reference	TIGR4	4	ST205	

A

B – Greece21-30

C – Sweden4-38

0.08

5.4.2.2 Serotype 19A

Serotype 19A isolates grouped into three different clusters by PFGE, demonstrating greater than 80% relatedness (Figure 5.11). Cluster 1 was comprised of almost 30% of the typed isolates, which were isolated from Ontario and Nova Scotia over the course of the study period. These isolates were also characterized by clarithromycin resistance. The ten isolates in cluster 2 were collected from Ontario and Eastern Canada and were largely susceptible to antimicrobials. Cluster 3 contained 13 isolates from across the country that were similar in their possession of an XDR phenotype. Most of these isolates demonstrated resistance to clarithromycin, clindamycin, doxycycline, penicillin and trimethoprim-sulfamethoxazole, with a few also displaying resistance to a sixth antimicrobial (levofloxacin or chloramphenicol).

Thirteen STs were identified by MLST, indicating high diversity in this serotype (Figure 5.12). The most common type identified was ST695, which was associated with susceptibility to all antimicrobials except for clarithromycin. Two other clusters were related to PMEN clones; a larger group of isolates related to PMEN14, predominantly XDR ST320, and a smaller cluster of isolates related to PMEN37, including STs 199, 416 and 667. This clone was originally identified in serotype 15B, indicating potential capsular switching. Interestingly, only the ST320 isolates were collected from across the country; all other STs were collected from Central and Eastern Canada.

PFGE again was reasonably successful at clustering isolates together by ST. Isolates that were typed as ST695, were clarithromycin resistant and were only isolated from Ontario and Nova Scotia represent cluster 1 of the dendrogram. The XDR ST320 isolates with widespread collection comprised cluster 3, while the susceptible isolates in cluster 2 were all ST199 or related.

FIGURE 5.11: PFGE dendrogram of serotype 19A *S. pneumoniae* isolates collected by the SAVE 2011-15 study. MLST sequence types are listed for comparison.

ISOLATE	YR	PROV	GDR	AGE GROUP	ST
SC12-1625-P	2012	ON	M	50 to less than 65	ST8325
SC14-0792-P	2014	ON	M	65 years or greater	ST416
SC12-1713-P	2012	QC	M	2 to less than 6	ST416
SC12-1426-P	2012	ON	M	18 to less than 50	ST695
SC14-1330-P	2014	ON	F	65 years or greater	ST695
SC14-3960-P	2014	ON	F	65 years or greater	ST695
SC12-1305-P	2012	ON	M	6 to less than 18	ST695
SC13-4635-P	2013	ON		1 to less than 2	ST695
SC15-0012-P	2014	ON	F	65 years or greater	ST695
SC11-0090-P	2011	ON	M	65 years or greater	ST695
SC13-2697-P	2013	NS	F	50 to less than 65	ST695
SC15-0940-P	2015	ON	F	65 years or greater	ST695
SC15-2505-P	2015	NS	M	50 to less than 65	ST695
SC15-4061-P	2015	ON	M	18 to less than 50	ST695
SC15-1308-P	2015	ON	F	50 to less than 65	ST695
SC12-1048-P	2012	ON	M	0 to less than 1	ST695
SC13-4697-P	2013	ON	F	50 to less than 65	ST695
SC11-4568-P	2011	NS	F	65 years or greater	ST695
SC15-0256-P	2014	NS	M	18 to less than 50	ST695
SC12-0953-P	2012	NB	M	65 years or greater	ST63
SC15-2813-P	2015	ON	M	65 years or greater	ST63
SC12-3012-P	2012	QC	M		ST10194
SC15-1582-P	2015	NB	M	65 years or greater	ST667
SC11-1777-P	2011	ON	M	18 to less than 50	ST416
SC13-0346-P	2013	NS	F	65 years or greater	ST1201
SC14-0192-P	2014	ON	F	65 years or greater	ST2343
SC14-0470-P	2014	ON	F	65 years or greater	ST2343
SC11-1806-P	2011	ON	F	50 to less than 65	ST199
SC11-4398-P	2011	NB	M	50 to less than 65	ST199
SC12-0438-P	2012	ON	M	65 years or greater	ST199
SC14-0085-P	2013	ON	M	65 years or greater	ST199
SC14-0091-P	2013	ON	M	65 years or greater	ST667
SC14-2721-P	2014	NB	F	65 years or greater	ST199
SC13-4611-P	2013	NB	F	50 to less than 65	ST994
SC15-2381-P	2015	NLA	F	65 years or greater	ST320
SC15-3550-P	2015	MB	F	6 to less than 18	ST320
SC11-0740-P	2011	SK	F	18 to less than 50	ST9356
SC11-1534-P	2011	ON	M	1 to less than 2	ST320
SC11-2214-P	2011	ON	F	65 years or greater	ST320
SC12-2060-P	2012	MB	M	50 to less than 65	ST320
SC12-3982-P	2012	QC		2 to less than 6	ST320
SC12-4251-P	2012	NLA	M	65 years or greater	ST320
SC13-2411-P	2013	NB	M	65 years or greater	ST9356
SC14-0874-P	2014	ON	M	65 years or greater	ST320
SC14-3781-P	2014	SK	F	65 years or greater	ST320
SC13-0659-P	2013	ON	F	18 to less than 50	ST320
SC13-1133-P	2013	ON	M	6 to less than 18	ST320
SC13-3257-P	2013	ON	F	50 to less than 65	ST320
SC15-4117-P	2015	ON	M	50 to less than 65	ST292

GDR, YR, PROV: gender, year and province of isolation. Gaps in the demographics indicate that the value was not given at the time of collection. The dashed line indicates 80% similarity. Clusters of isolates are outlined with red boxes and labelled 1-3.

FIGURE 5.12: Minimum spanning tree (generated by PHYLOViZ 2.0) of MLST sequence types demonstrated by serotype 19A isolates, coloured according to province of isolation[a]. Green outlines indicate a group founder; numbers indicate the number of differences between the MLST profiles of the two connected nodes (≤ 2 indicates the two nodes are part of a cluster). PMEN clones are marked with red circles.
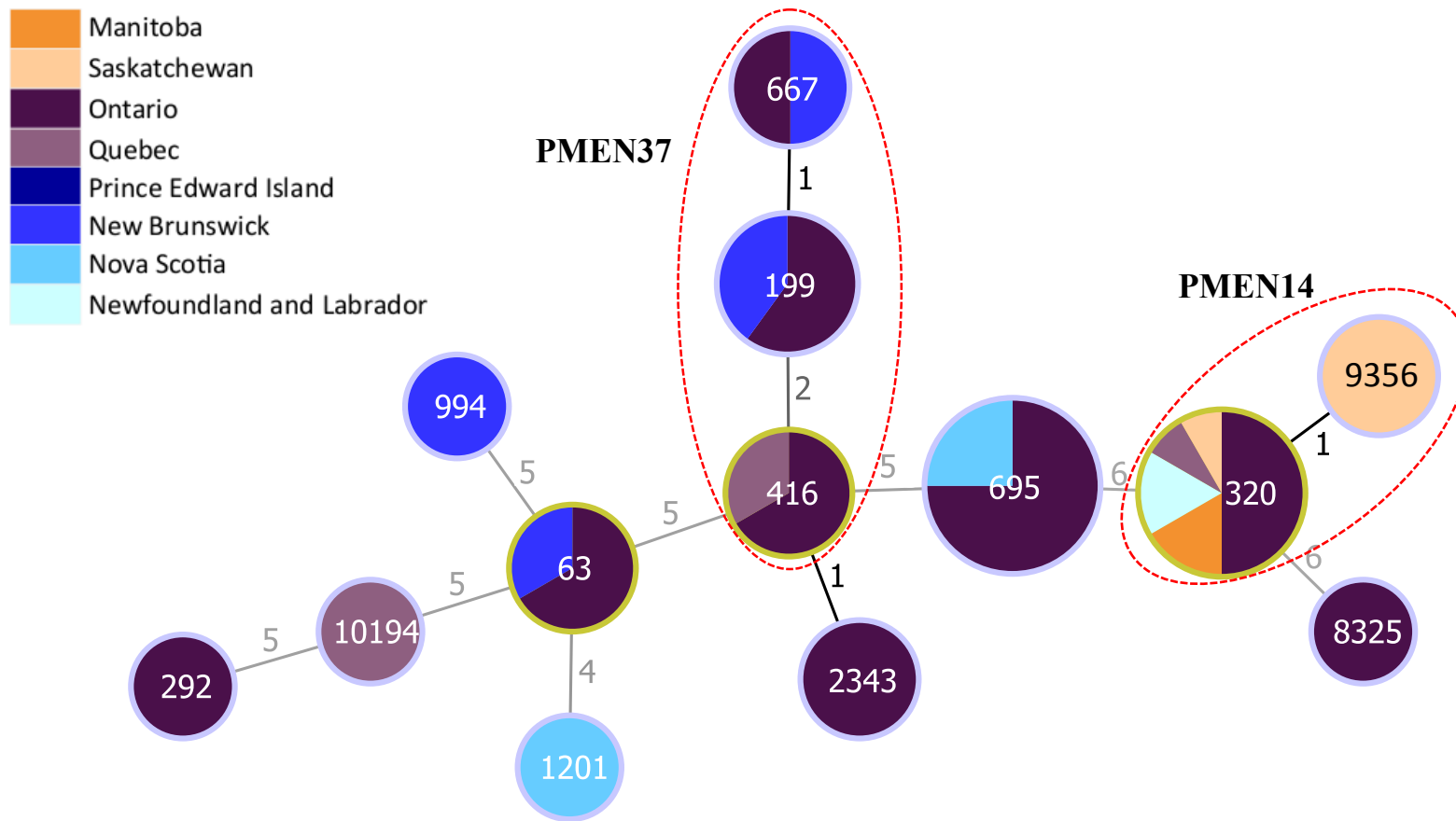
[a] Orange, Western Canada; Purple, Central Canada; Blue, Eastern Canada.

As depicted in Figure 5.13, over half of the serotype 19A isolates clustered together in the phylogenetic analysis. Cluster 8 contained 17 serotype 19A isolates that typed as ST320 (related to PMEN14) and 16 of these were almost identical by phylogenetic analysis. Each isolate was resistant to at least three antimicrobials, with two demonstrating resistance to six different antimicrobials. All isolates possessed the dual *mefA/ermB* genotype, though those isolates with no or low clarithromycin resistance lacked the fully functional *ermB*, instead possessing the truncated version described in Section 5.3.1. Isolates also possessed *tetM*, numerous mutations in all three PBPs, an Ile100Leu mutation in *folA*, and a duplication of Glu-Ile (EI) after codon 66 of *folP*.

Aside from this one large cluster, serotype 19A isolates demonstrated relatedness to several different serotypes, resulting in the remaining 14 isolates being distributed throughout many smaller clusters of background isolates. Three isolates in cluster 4 typed as ST63 and clustered with the other PMEN25 isolates (Figure 5.8), also demonstrating the same *ermB/tetM* genotype and resistance pattern of clarithromycin, clindamycin and doxycycline. An isolate in cluster 6, while demonstrating a MDR pattern identified in many ST320 strains (conferred by *ermB*, *tetM*, and three altered PBPs) instead clustered with isolates related to ST177 (PMEN21) (Figure 5.14). An isolate displayed in Figure 5.15 clustered with serotype 14 isolates related to PMEN9, while one isolate in cluster 7 (Figure 5.9) was related to non-serotype 9V isolates related to ST156 (PMEN3). This resistant serotype 19A isolate differed from the XDR isolates above in that possessed only *mefA* (as opposed to *ermB* and *tetM*), in addition to PBP mutations, *folA*-Ile100Leu and an additional Val-Glu (VE) in *folP*. As depicted in Figure 5.10, serotype 19A isolates fell into two separate lineages in cluster 12. Cluster B of this phylogeny demonstrates two isolates related to uncommon serotype 21 associated with PMEN30, while cluster C depicts

one isolate distantly related to PMEN38. Lastly, Figure 5.16 demonstrates relatedness of five serotype 19A isolates to three different groups within cluster 13. Two MDR isolates in cluster A possessing *ermB*, *tetM*, numerous PBP mutations, *folA*-Ile100Leu and a duplication of Ser-Tyr (SY) in *folP* demonstrated relatedness to PMEN32, a less common clone of serotype 14. In cluster B, one XDR isolate grouped together with XDR serotype 23F and 15B isolates related to ST81 (PMEN1). This isolate possessed *cat*, *ermB*, *tetM*, three altered PBPs, *folA*-Ile100Leu and a serine duplication in *folP*, while the related serotype 23F isolate had additional Ser79Phe and Ser81Phe mutations in *parC* and *gyrA*, respectively, to make it the only isolate in the SAVE study to be resistant to seven classes of antimicrobials. An additional two isolates, despite demonstrating similar MLST types to serotype 15B-PMEN37, were separated from their 15B counterparts in the phylogenetic analysis. The relatedness that these serotype 19A isolates show to numerous other serotypes, STs and clones, and the variety of resistance determinants and alterations they possess, indicates that many isolates may have been the products of recombination generating capsular switch variants.

FIGURE 5.13: Whole genome core SNV maximum likelihood phylogenetic tree of *S. pneumoniae* "cluster 8" rooted on reference genome *S. pneumoniae* Taiwan19F-14 (NC_012469). Tree was generated using PhyML and visualized using FigTree. Coloured clusters were identified by ClusterPicker.

Isolate	Year	Serotype	ST	Res. Pattern
SC13-0056-P	2012	23B	ST338	
SC13-4858-P	2013	6C	ST338	
SC13-3068-P	2013	23B	ST1373	
SC13-0368-P	2013	6B	ST138	CLR, CLD, DOX, SXT
reference Taiwan19F-14		19F	ST236	
SC14-0291-P	2014	19F	ST271	CLR, CLD, DOX, PEN, SXT
SC11-3271-P	2011	19A	ST320	CLR, DOX, PEN, SXT
SC14-0911-P	2014	19A	ST320	CLR, CLD, DOX, PEN, SXT
SC13-3212-P	2013	19A	ST320	CLR, CLD, DOX, SXT
SC11-1929-P	2011	19A	ST320	CLR, CLD, DOX, PEN, SXT
SC12-2051-P	2012	19A	ST320	CLR, CLD, DOX, PEN, SXT
SC11-1200-P	2011	19A	ST320	CLR, CLD, DOX, LEV, PEN, SXT
SC12-2118-P	2012	19A	ST320	DOX, PEN, SXT
SC13-4579-P	2013	19A	ST320	CLR, CLD, DOX, PEN, SXT
SC12-0322-P	2012	19A	ST320	CLR, DOX, SXT
SC12-3334-P	2012	19A	ST320	CLR, DOX, PEN, SXT
SC14-0428-P	2014	19A	ST320	CHL, CLR, CLD, DOX, PEN, SXT
SC14-1542-P	2014	19A	ST320	CLR, CLD, DOX, SXT
SC11-0900-P	2011	19A	ST320	DOX, PEN, SXT
SC13-2710-P	2013	19A	ST320	DOX, PEN, SXT
SC14-3781-P	2014	19A	ST320	CLR, DOX, PEN, SXT
SC12-0054-P	2011	19A	ST320	CLR, CLD, PEN, SXT
SC12-1810-P	2012	19A	ST320	CLR, CLD, DOX, SXT

Colombia23F-26

Taiwan19F-14

0.2

FIGURE 5.14: Whole genome core SNV maximum likelihood phylogenetic tree of *S. pneumoniae* "cluster 6" rooted on reference genome PMEN21 (ERR1022108). Tree was generated using PhyML and visualized using FigTree. Coloured clusters were identified by ClusterPicker.

Isolate	Year	Serotype	ST	Res. Pattern
SC12-1898-P	2012	6A	ST1876	LEV
SC13-1760-P	2013	6A	ST1876	
SC11-1862-P	2011	6B	ST1876	CLR
SC11-3903-P	2011	3	ST177	
SC12-1374-P	2012	24B	ST177	
reference	PMEN21	19F	ST177	
SC13-3151-P	2013	19A	ST179	CLR, CLD, DOX, PEN

Portugal19F-21

0.3
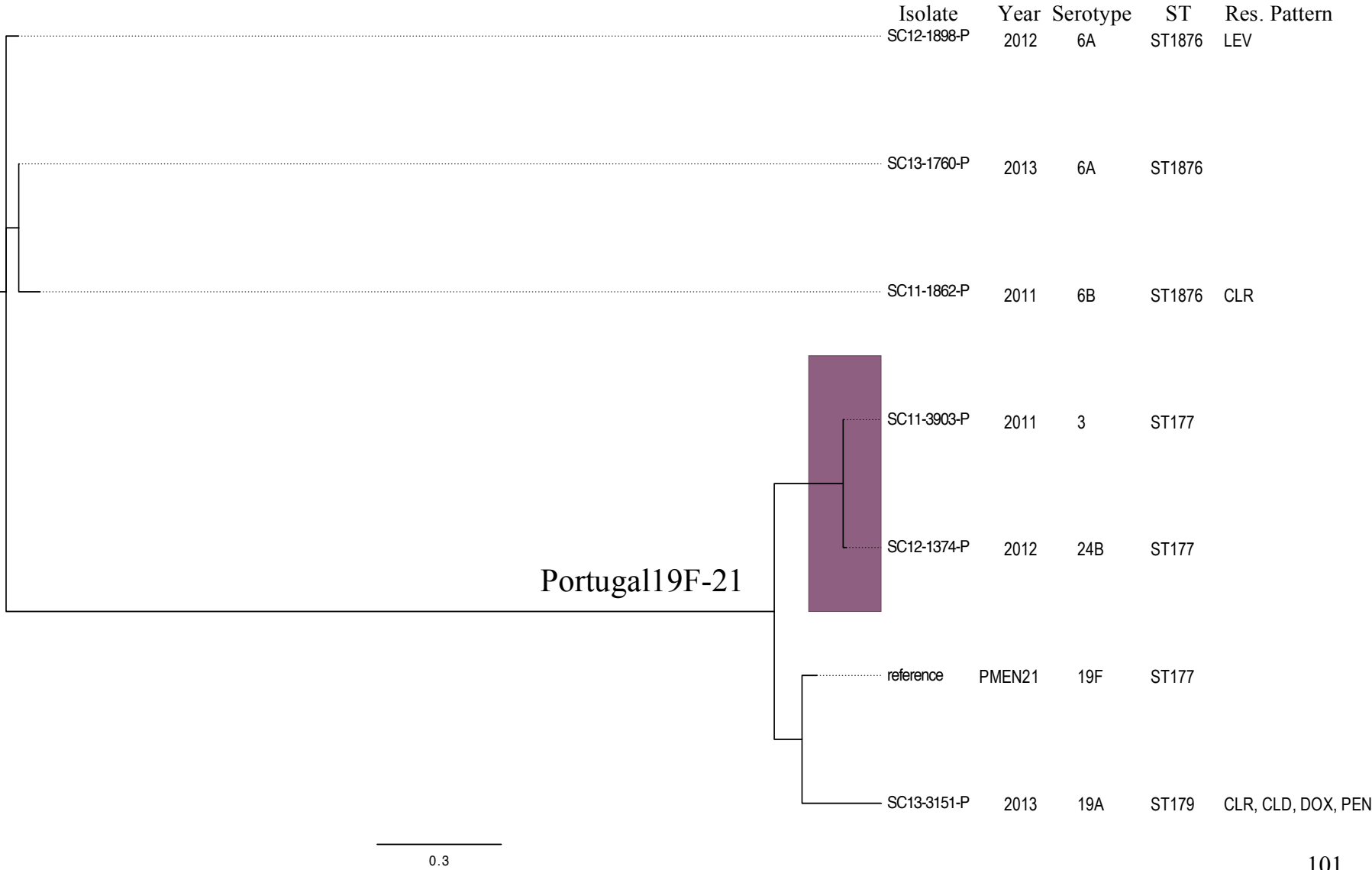
FIGURE 5.15: Whole genome core SNV maximum likelihood phylogenetic tree of *S. pneumoniae* "cluster 10" rooted on reference genome *S. pneumoniae* INV200 (NC_017593). Tree was generated using PhyML and visualized using FigTree. Coloured clusters were identified by ClusterPicker.

Isolate	Year	Serotype	ST	Res. Pattern
SC10-0292-P	2008	6C	ST3101	CLR
SC13-0990-P	2013	6C	ST1379	CLR, PEN, SXT
SC12-0121-P	2011	6C	ST1390	
SC11-4183-P	2011	6D	ST1692	
SC12-3606-P	2012	19F	ST2631	
SC13-2733-P	2013	11A	ST10196	LEV, SXT
SC11-1574-P	2011	7C	ST1797	
SC16-2494-P	2007	19A	ST13191	CLR, CLD
PMEN9		14	ST9	
SC11-1495-P	2011	14	ST9	CLR
reference	INV200	14	ST9	
SC16-2495-P	2007	14	ST9	CLR

England14-9

0.06
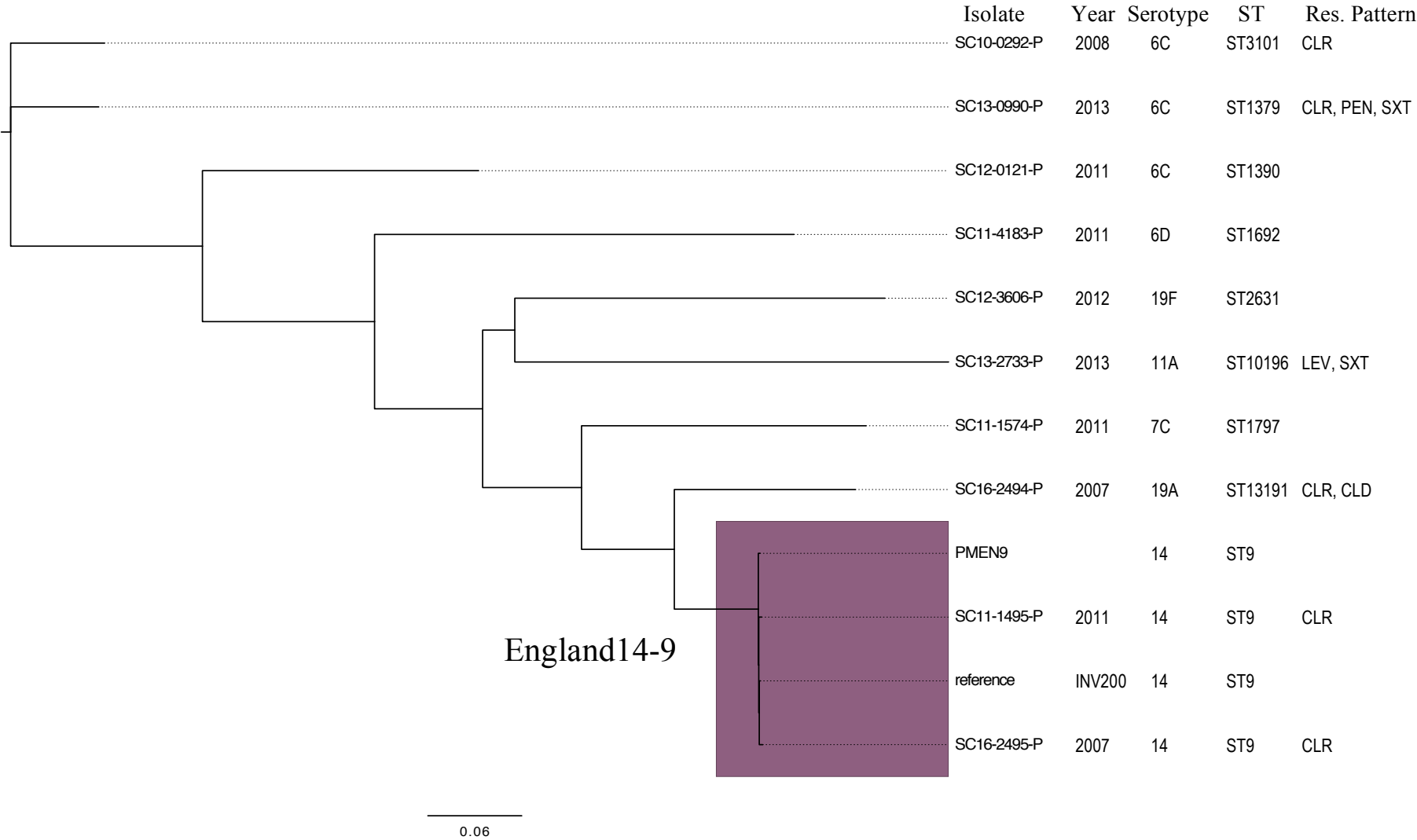
FIGURE 5.16: Whole genome core SNV maximum likelihood phylogenetic tree of *S. pneumoniae* "cluster 13" rooted on reference genome *S. pneumoniae* ATCC700669 (NC_011900). Tree was generated using PhyML and visualized using FigTree. Coloured clusters were identified by ClusterPicker. Clusters of interest are labelled A-C.

Isolate	Year	Serotype	ST	Res. Pattern
SC14-4115-P	2014	15B	ST199	CLR
SC16-2505-P	2008	15B	ST199	
SC10-0551-P	2009	19A	ST199	
SC12-2050-P	2012	19A	ST667	
SC13-1756-P	2013	17F	ST2355	CLR, CLD, DOX
SC12-0347-P	2011	24F	ST230	CLR
SC12-3012-P	2012	19A	ST10194	CLR, DOX, PEN
SC11-4254-P	2011	19A	ST319	CLR, CLD, DOX, SXT
SC13-1734-P	2013	10A	ST3135	DOX, SXT
SC14-1752-P	2014	24F	ST72	
SC11-0902-P	2011	19A	ST81	CHL, CLR, DOX, PEN, SXT
PMEN1		23F	ST81	
reference	ATCC70066	23F	ST81	
SC13-1642-P	2013	23F	ST13190	CHL, CLR, CLD, DOX, LEV, PEN, SXT
SC11-2658-P	2011	15B	ST83	CHL, CLR, CLD, DOX, PEN, SXT
SC12-2856-P	2012	9N	ST3224	CLR, DOX
PMEN18		14	ST67	
SC10-0569-P	2009	9N	ST10191	
SC13-4463-P	2013	9N	ST10191	
SC11-0858-P	2011	9N	ST66	
SC11-3712-P	2011	9N	ST66	
SC13-1198-P	2013	9N	ST517	
SC16-2503-P	2008	9N	ST517	

Netherlands15B-37

A – Denmark14-32

B – Spain23F-1

C – Tennessee14-18

0.08

5.4.2.3 Serotype 6C

A PFGE dendrogram of serotype 6C isolates is presented in Figure 5.17. Serotype 6C was very diverse by PFGE. Four clusters of four or more isolates were identified, along with numerous smaller clusters and unique isolates. Isolates within these clusters were collected from various provinces and study years. There were no defining characteristics for clusters 1, 2 and 4; however, cluster 3 contained most of the tested serotype 6C isolates with a MDR phenotype. Isolates in this cluster were resistant to clarithromycin, clindamycin and doxycycline; additional single MDR isolates were present in cluster 2, and in a doublet just above cluster 3.

Serotype 6C demonstrated 20 different STs by MLST, the most of any serotype in this study. As opposed to a large cluster of closely related types, serotype 6C had a few smaller clusters containing two or three STs, with many others differing by three or more alleles (Figure 5.18). The most common STs, including STs 1379, 1390 and 1692, were predominantly isolated from Central Canada. Conversely, other types such as ST3101 and ST639 and variants were only obtained from Western and Eastern Canada. Despite the large number of STs, only two were related to an international clone: ST338, an international clone (PMEN26) originally associated with serotype 23F, and ST5241, a DLV of the same clone. Much of the MDR in the typed serotype 6C isolates was attributed to this cluster, specifically ST5241, which demonstrated resistance to clarithromycin, clindamycin and doxycycline.

Relating the MLST data back to the PFGE dendrogram, it is evident that cluster 3 containing MDR isolates was made up of ST5241 strains. In general, PFGE clustered most other isolates (that originally had no defining characteristics) by ST; cluster 1 contained most of the ST1692 isolates, cluster 2 was made up of ST1390 isolates, while cluster 4 contained the less common ST639.

FIGURE 5.17: PFGE dendrogram of serotype 6C *S. pneumoniae* isolates collected by the SAVE 2011-15 study. MLST sequence types are listed for comparison.

ISOLATE	YR	PROV	GDR	AGE GROUP	ST
SC12-3155-P	2012	ON	M	6 to less than 18	ST5679
SC13-3444-P	2013	ON	M	65 years or greater	ST5679
SC12-0734-P	2012	ON	M	50 to less than 65	ST5679
SC12-1880-P	2012	ON	M	50 to less than 65	ST10202
SC15-3260-P	2015	ON	F	65 years or greater	ST10202
SC11-2957-P	2011	ON	M	65 years or greater	ST473
SC11-4514-P	2011	MB	F	65 years or greater	ST10199
SC13-1811-P	2013	ON	F	65 years or greater	ST1692
SC15-1159-P	2015	ON	M	65 years or greater	ST1692
SC12-0044-P	2011	ON	F	65 years or greater	ST1692
SC12-0544-P	2012	MB	M	50 to less than 65	ST6354
SC12-3078-P	2012	ON	F	65 years or greater	ST1692
SC13-2057-P	2013	ON	M		ST1692
SC13-2359-P	2013	ON		65 years or greater	ST10207
SC11-0239-P	2011	ON	M	65 years or greater	ST1692
SC11-0906-P	2011	ON	M	50 to less than 65	ST1692
SC11-1876-P	2011	ON	M	65 years or greater	ST1692
SC14-4408-P	2014	QC	M	65 years or greater	ST1692
SC14-4409-P	2014	QC	M	65 years or greater	ST1692
SC12-4321-P	2012	ON	M	65 years or greater	ST1692
SC14-3904-P	2014	ON	M	65 years or greater	ST1692
SC15-0877-P	2015	ON	F	50 to less than 65	ST1692
SC15-2609-P	2015	ON	M	65 years or greater	ST2924
SC12-1573-P	2012	ON	M	65 years or greater	ST1390
SC13-0053-P	2012	SK	M	18 to less than 50	ST10204
SC14-0211-P	2014	ON	M	50 to less than 65	ST1390
SC14-0665-P	2014	ON	F	65 years or greater	ST1390
SC14-1103-P	2014	ON	M	50 to less than 65	ST1390
SC12-0121-P	2011	MB	M	65 years or greater	ST1390
SC14-1754-P	2014	ON	F	65 years or greater	ST1292
SC15-3460-P	2015	PEI	M	65 years or greater	ST1390
SC11-2696-P	2011	ON	M	65 years or greater	ST1390
SC13-0990-P	2013	ON	M	50 to less than 65	ST1379
SC15-0376-P	2015	ON	M	50 to less than 65	ST1379
SC11-4414-P	2011	ON	F	65 years or greater	ST1390
SC14-2223-P	2014	ON	M	50 to less than 65	ST10426
SC13-3626-P	2013	NB	F	50 to less than 65	ST3101
SC13-3743-P	2013	NB	F	65 years or greater	ST3101
SC13-2696-P	2013	NS	F	65 years or greater	ST1379
SC15-0235-P	2014	NS	M	50 to less than 65	ST1379
SC15-0619-P	2015	QC	F	50 to less than 65	ST13192
SC15-2173-P	2015	QC	M	50 to less than 65	ST386
SC11-0111-P	2011	ON	M	50 to less than 65	ST5241
SC12-3340-P	2012	ON	M	65 years or greater	ST5241
SC13-1980-P	2013	ON	M	50 to less than 65	ST5241
SC13-4858-P	2013	NLA	M	50 to less than 65	ST338
SC12-3797-P	2012	SK	M	18 to less than 50	ST10203
SC14-1268-P	2014	NLA	M	65 years or greater	ST639
SC15-1487-P	2015	MB	F	65 years or greater	ST639
SC15-2832-P	2015	SK	F	1 to less than 2	ST639

GDR, YR, PROV: gender, year and province of isolation. Gaps in the demographics indicate that the value was not given at the time of collection. The dashed line indicates 80% similarity. Clusters of isolates are outlined with red boxes and labelled 1-4.

FIGURE 5.18: Minimum spanning tree (generated by PHYLOViZ 2.0) of MLST sequence types demonstrated by serotype 6C isolates, coloured according to province of isolation[a]. Green outlines indicate a group founder; numbers indicate the number of differences between the MLST profiles of the two connected nodes (≤ 2 indicates the two nodes are part of a cluster). PMEN clones are marked with red circles.
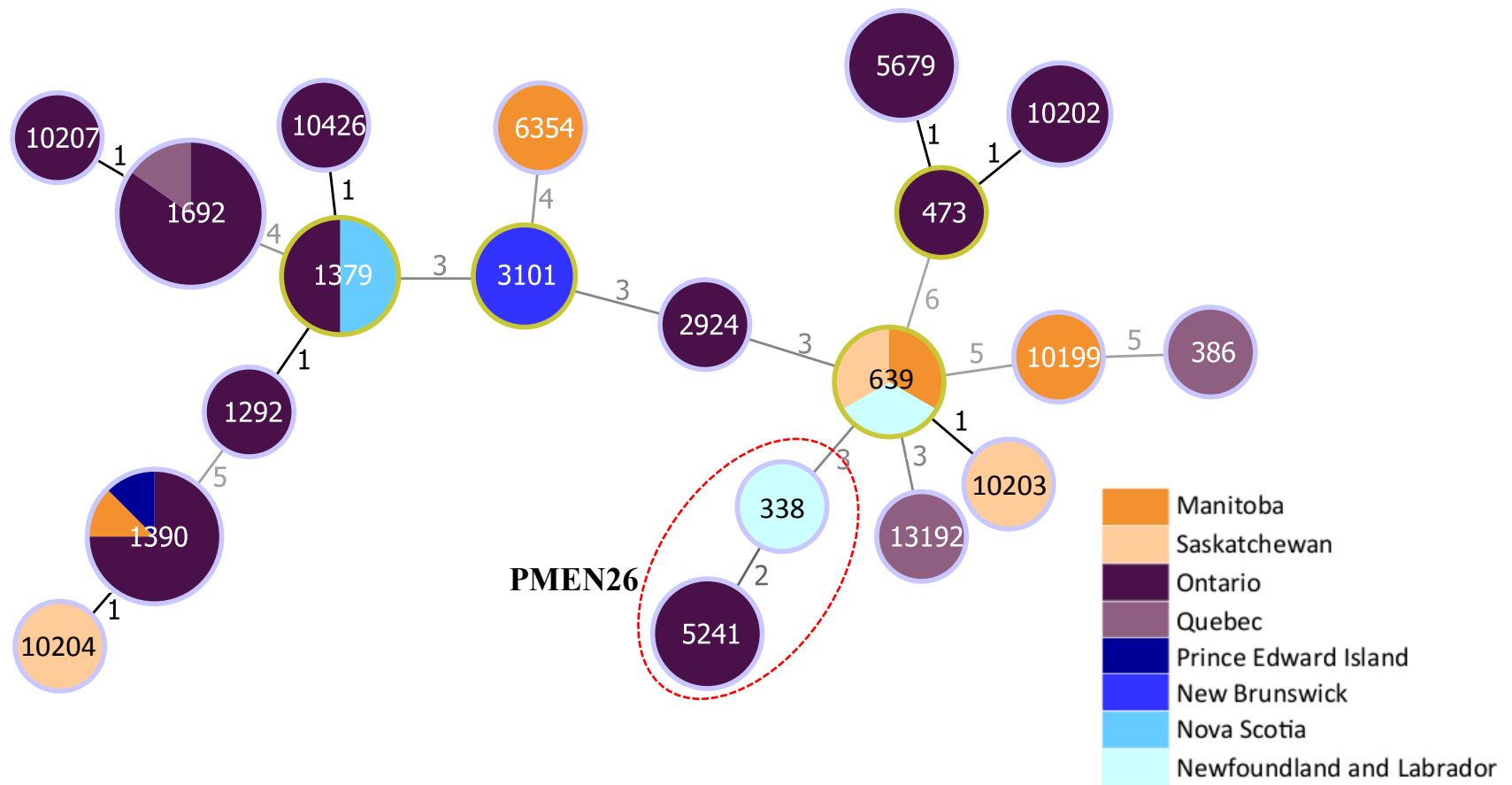
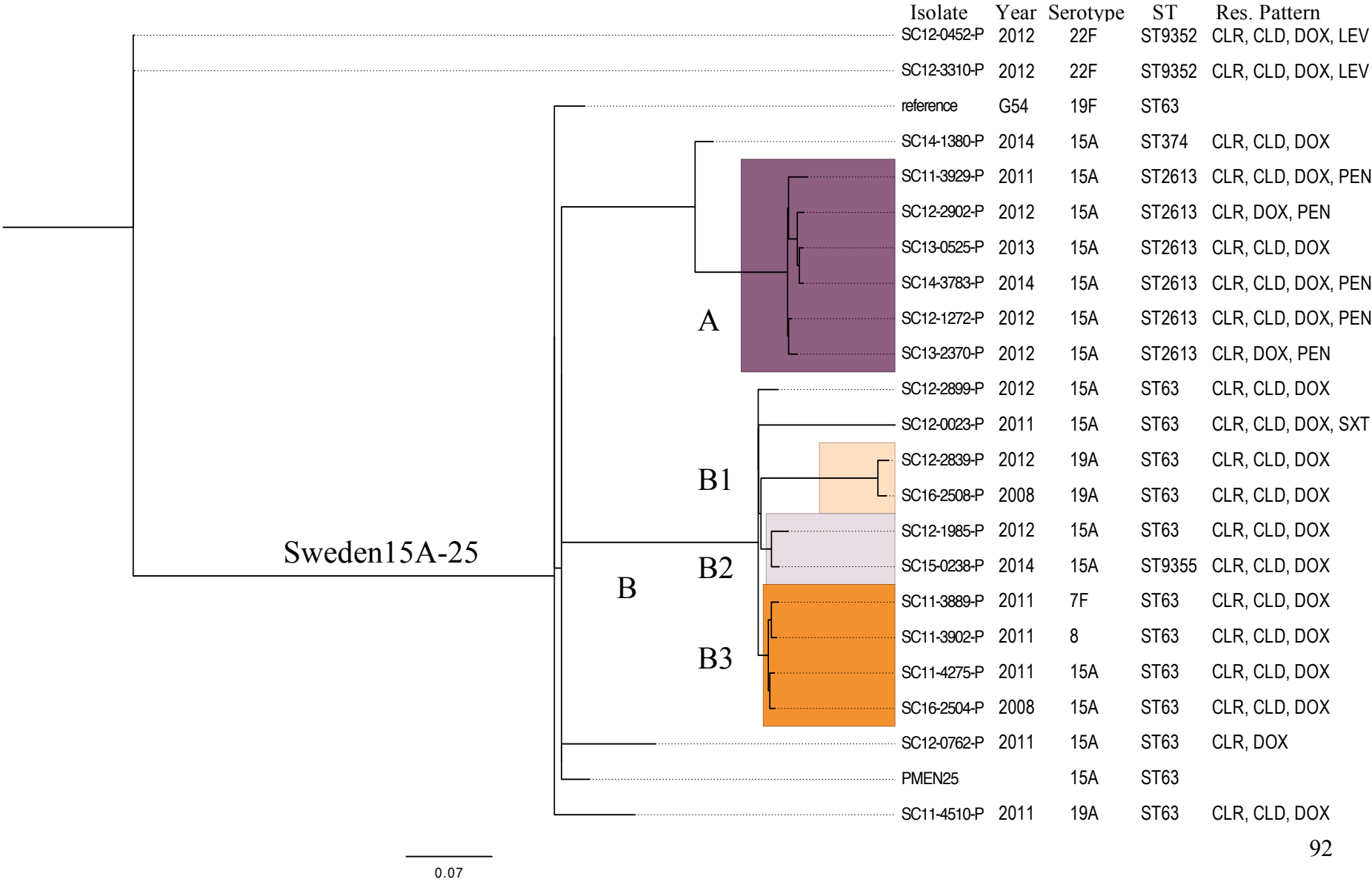[a] Orange, Western Canada; Purple, Central Canada; Blue, Eastern Canada.

Five serotype 6C isolates were included in the selection for WGS; as one demonstrated poor mapping to the reference strain, only four were included in the subsequent phylogenetic analysis. Three of the four remaining serotype 6C isolates located to cluster 10, depicted in Figure 5.15. These isolates typed as previously described STs 1390, 1397 and 3101. Although not clustering as closely as some previous clones (PMEN14, for example), the isolates clearly demonstrate greater relatedness to each other than to other serotypes; the next most closely related isolate was a serotype 6D isolate with a ST also previously described above (ST1692). This could be the result of a capsular switch, or perhaps an accumulation of mutations in the 6C capsule.

The last serotype 6C isolate was identical to PMEN26, and therefore clustered separately in Cluster 8 (Figure 5.13). Typing as ST338, this isolate demonstrated relatedness to two serotype 23B background strains, one with the same ST and the other a variant. As PMEN26 was originally isolated as a PCV-13-serotype 23F strain, this serotype 6C isolate was likely the product of a vaccine escape recombination event.

5.4.2.4 Serotype 9N

PFGE of serotype 9N isolates yielded four clusters with >80% relatedness (Figure 5.19).

Clusters 1-3 contained 22, 6, and 4 isolates, respectively; however, if the definition of related is

relaxed, these three clusters, plus an additional four isolates become one large cluster

demonstrating 74% relatedness. Cluster 4, consisting of 10 isolates, was <70% related to the

other large cluster. As the only MDR isolate fell below cluster 4 (demonstrating resistance to

clarithromycin, clindamycin, doxycycline and penicillin) and few isolates were antimicrobial

resistant, clusters could not be distinguished using that criteria. Isolates within each cluster were

obtained from across the country from all years of the study.

By MLST, serotype 9N demonstrated 12 STs. Although this large number of STs gave

the appearance of diversity, seven were contained in one large cluster of related isolates founded

on ST66, a SLV of PMEN18 (Figure 5.20). Interestingly, while the PMEN18 clone is known for

its resistance to antimicrobials, these related serotype 9N isolates were fully susceptible. Isolates

related to PMEN18 were predominantly contained within clusters 1-3 of the PFGE dendrogram.

A small number of isolates were variants of PMEN34, commonly associated with serotype 12F;

included were STs 405, 4666 and 10540. These isolates corresponded to cluster 4 of the PFGE

dendrogram.

Phylogenetic analysis of serotype 9N isolates presented very similar results to MLST.

Depicted as part of cluster 13C (Figure 5.16), isolates were further separated into two groups by

ClusterPicker: ST10191 and STs 66 and 517. Only one isolate possessed resistance determinants

(*ermB* and *tetM*) and was resistant to antimicrobials (clarithromycin, doxycycline); it appeared

more distantly related to both the tested isolates and the PMEN18 reference strain.

FIGURE 5.19: PFGE dendrogram of serotype 9N *S. pneumoniae* isolates collected by the SAVE 2011-15 study. MLST sequence types are listed for comparison.

ISOLATE	YR	PROV	GDR	AGE GROUP	ST	Cluster
SC11-0908-P	2011	ON	F	18 to less than 50	ST66	1
SC11-4422-P	2011	ON	M	18 to less than 50	ST517	1
SC11-4500-P	2011	ON	M	65 years or greater	ST517	1
SC12-1318-P	2012	ON		18 to less than 50	ST10201	1
SC13-1198-P	2013	ON	F	65 years or greater	ST517	1
SC13-4929-P	2013	ON	M	65 years or greater	ST517	1
SC11-2515-P	2011	ON	M	18 to less than 50	ST10191	1
SC11-4504-P	2011	ON	M	65 years or greater	ST66	1
SC12-0311-P	2012	ON	F	65 years or greater	ST517	1
SC13-3636-P	2013	QC	M	65 years or greater	ST66	1
SC14-3762-P	2014	MB	F	50 to less than 65	ST66	1
SC12-0313-P	2012	ON	F	65 years or greater	ST517	1
SC14-2775-P	2014	ON	F	65 years or greater	ST66	1
SC12-3887-P	2012	PEI	M	0 to less than 1	ST66	1
SC14-1442-P	2014	ON	F	65 years or greater	ST10191	1
SC11-0894-P	2011	ON	F	65 years or greater	ST66	1
SC14-3693-P	2014	ON	M	50 to less than 65	ST66	1
SC15-0251-P	2014	NS	F	50 to less than 65	ST517	1
SC13-1193-P	2013	ON	F	50 to less than 65	ST10191	1
SC13-4463-P	2013	ON	M	50 to less than 65	ST10191	1
SC15-0607-P	2015	ON	F	18 to less than 50	ST10191	1
SC15-0871-P	2015	ON	F	65 years or greater	ST66	1
SC12-3180-P	2012	MB	M	18 to less than 50	ST66	2
SC12-3888-P	2012	MB	M	18 to less than 50	ST66	2
SC14-3243-P	2014	NB	F	6 to less than 18	ST10539	2
SC13-1775-P	2013	ON	M		ST632	2
SC15-0279-P	2015	ON	M	18 to less than 50	ST632	2
SC14-2720-P	2014	NB	M	18 to less than 50	ST10539	2
SC11-4378-P	2011	ON	M	18 to less than 50	ST66	3
SC12-2856-P	2012	ON	M	50 to less than 65	ST3224	3
SC11-3712-P	2011	SK	M	65 years or greater	ST66	3
SC13-1758-P	2013	ON	M	18 to less than 50	ST66	3
SC15-2538-P	2015	ON		65 years or greater	ST66	
SC15-2653-P	2015	ON	M	65 years or greater	ST632	
SC15-0431-P	2015	NB	M	50 to less than 65	ST10539	
SC15-1158-P	2015	ON	F	50 to less than 65	ST66	
SC12-1224-P	2012	MB	M		ST66	
SC15-2499-P	2015	NS	M	50 to less than 65	ST4666	4
SC15-3143-P	2015	MB	F	0 to less than 1	ST405	4
SC15-0205-P	2014	SK	F	65 years or greater	ST405	4
SC12-0047-P	2011	ON	F	50 to less than 65	ST405	4
SC13-0175-P	2013	NB	M	50 to less than 65	ST4666	4
SC11-2182-P	2011	NS	M	50 to less than 65	ST4666	4
SC14-3503-P	2014	ON	M	50 to less than 65	ST10540	4
SC14-4363-P	2014	ON	F	65 years or greater	ST405	4
SC13-2025-P	2013	MB	M		ST4666	4
SC12-0216-P	2012	MB	F	50 to less than 65	ST4666	4
SC13-0111-P	2013	ON	F	65 years or greater	ST517	
SC15-1591-P	2015	ON	M	65 years or greater	ST13211	
SC13-0054-P	2012	SK	F	65 years or greater	ST10205	

GDR, YR, PROV: gender, year and province of isolation. Gaps in the demographics indicate that the value was not given at the time of collection. The dashed lines indicate 74% and 80% similarity, respectively. Clusters of isolates are outlined with red boxes labelled 1-4.

FIGURE 5.20: Minimum spanning tree (generated by PHYLOViZ 2.0) of MLST sequence types demonstrated by serotype 9N isolates, coloured according to province of isolation[a]. Green outlines indicate a group founder; numbers indicate the number of differences between the MLST profiles of the two connected nodes (≤ 2 indicates the two nodes are part of a cluster). PMEN clones are marked with red circles.
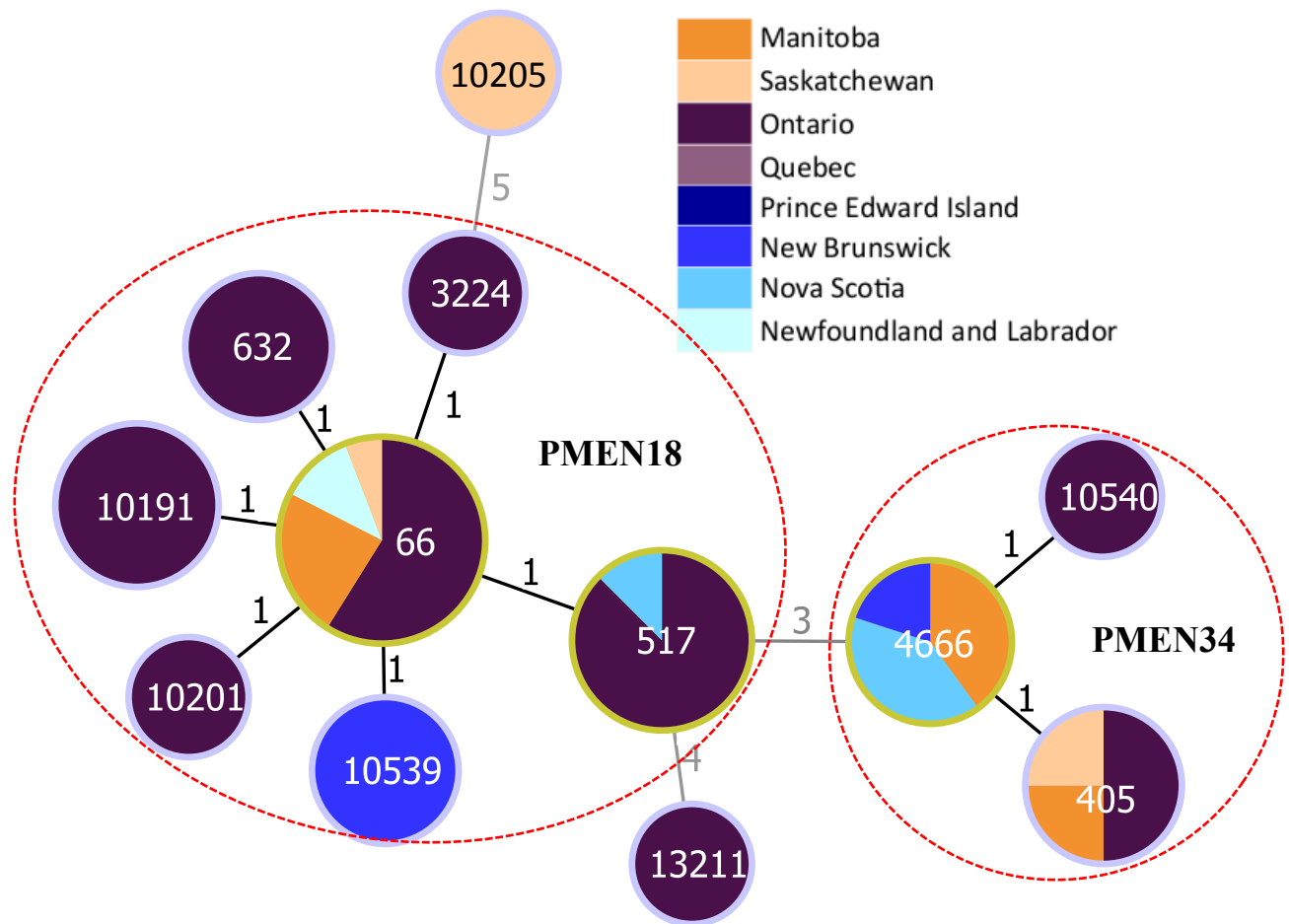
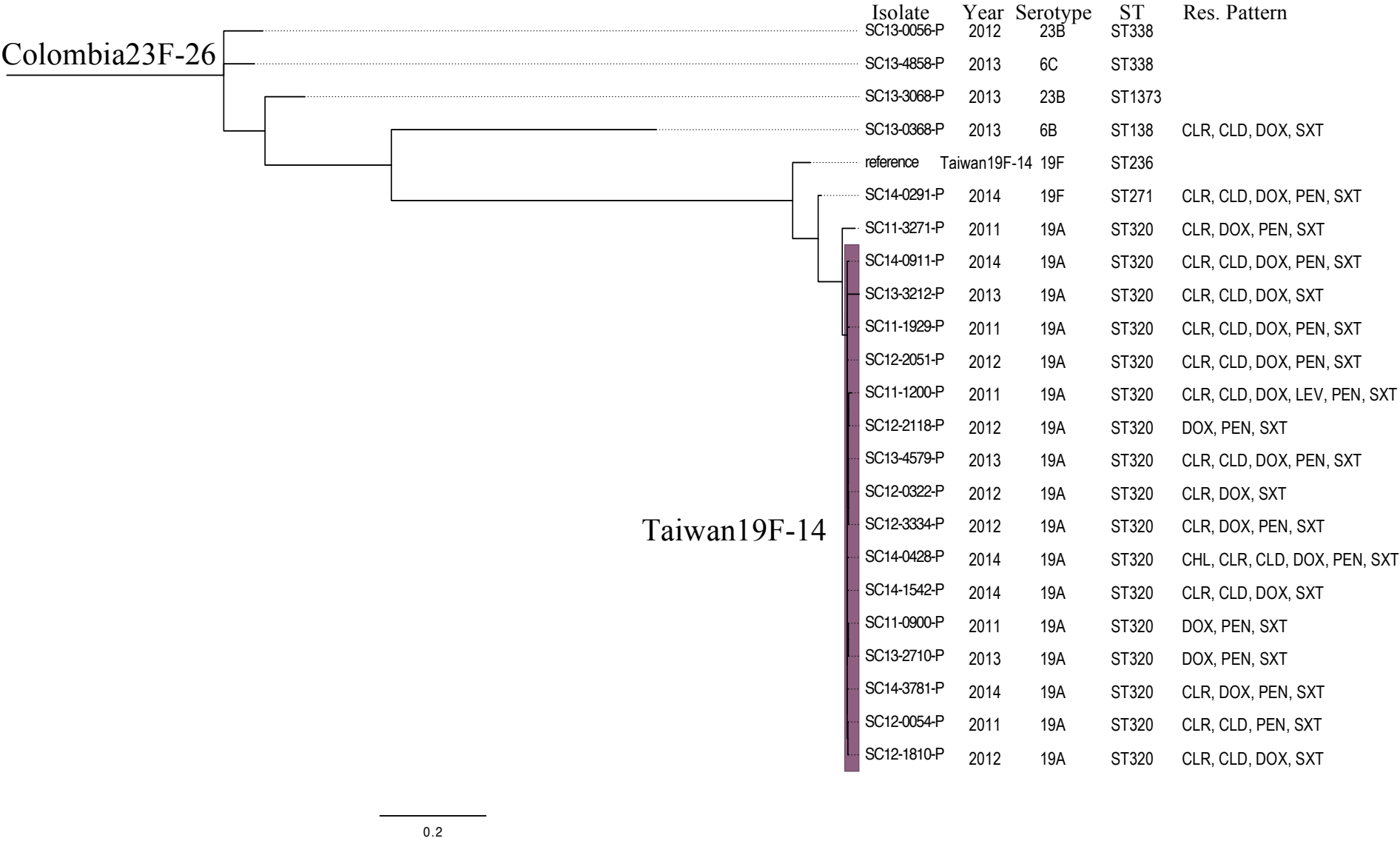[a] Orange, Western Canada; Purple, Central Canada; Blue, Eastern Canada.

5.4.2.5 Serotype 33F

Serotype 33F initially appeared to be a clonal serotype by PFGE (Figure 5.21). Isolates fell into one of two clusters, principally depending on resistance profile. Forty-four isolates that spanned the study period were part of cluster 1, which were all resistant to at least one of clarithromycin or trimethoprim-sulfamethoxazole. Isolates resistant to two or more classes were predominantly found in sub-clusters 1A and 1B; noted resistance was to at least clarithromycin and clindamycin, with four isolates demonstrating an MDR pattern including either chloramphenicol or doxycycline. Isolates within these sub-clusters were all obtained from Quebec. A second, smaller cluster of isolates was located at the bottom of the dendrogram and included those isolates that were fully susceptible to all tested antimicrobials.

Conversely to PFGE, serotype 33F showed an increased level of diversity by MLST (Figure 5.22). Eight STs were identified, four of which were newly assigned during the study period. MLST was much more discriminatory than PFGE, which broke the isolates down into five distinct clusters separated by at least three allelic differences. The group of isolates with founder ST673 was fully susceptible to all antimicrobials, corresponding to cluster 2 by PFGE. The cluster containing STs 100 and 2705 was comprised of isolates resistant to one or both of clarithromycin or trimethoprim-sulfamethoxazole. All MDR isolates were either ST9348 or ST9350; however, these STs differed enough by MLST that they were not considered one cluster. PFGE grouped these isolates together in cluster 1 as >80% related, despite there being at least four allelic differences between some STs. Even in sub-clusters 1A and 1B, some of the MDR isolates fell outside of these groups, making it hard to discriminate differences between strains. The last ST was related to PMEN21, represented by serotype 19F. This was the only isolate that did not fall into a cluster by PFGE, and is likely a vaccine escape recombinant.

FIGURE 5.21: PFGE dendrogram of serotype 33F *S. pneumoniae* isolates collected by the SAVE 2011-15 study. MLST sequence types are listed for comparison.

ISOLATE	YR	PROV	GDR	AGE GROUP	ST
SC14-2209-P	2014	MB	F	50 to less than 65	ST2705
SC14-3363-P	2014	ON	M	50 to less than 65	ST2705
SC15-0623-P	2015	QC	F	1 to less than 2	ST100
SC12-0354-P	2011	QC	M	2 to less than 6	ST9348
SC15-0414-P	2015	QC	F	65 years or greater	ST9350
SC15-0435-P	2015	QC	M	65 years or greater	ST717
SC15-1688-P	2015	QC	F	65 years or greater	ST100
SC14-2911-P	2014	QC	F	65 years or greater	ST9350
SC14-4158-P	2014	QC	M	1 to less than 2	ST9350
SC12-0338-P	2011	QC	F	65 years or greater	ST100
SC12-0442-P	2012	ON	F	2 to less than 6	ST100
SC12-3162-P	2012	QC	F	65 years or greater	ST100
SC12-3760-P	2012	ON	M	18 to less than 50	ST100
SC12-4339-P	2012	PEI	M	18 to less than 50	ST100
SC13-2499-P	2013	MB	F	6 to less than 18	ST100
SC13-2845-P	2013	MB	F	65 years or greater	ST100
SC13-3009-P	2013	ON	M	65 years or greater	ST100
SC13-3106-P	2013	MB	F	65 years or greater	ST100
SC13-3361-P	2013	ON		18 to less than 50	ST717
SC14-1040-P	2014	SK	M	65 years or greater	ST100
SC14-1835-P	2014	PEI	M	65 years or greater	ST100
SC15-1428-P	2015	QC	M	50 to less than 65	ST100
SC11-0897-P	2011	ON	M	50 to less than 65	ST100
SC11-4259-P	2011	ON	M	18 to less than 50	ST100
SC12-0315-P	2012	ON	F	65 years or greater	ST100
SC14-1127-P	2014	QC	F	6 to less than 18	ST100
SC14-1575-P	2014	QC	F	65 years or greater	ST100
SC13-4125-P	2013	QC	M	1 to less than 2	ST9350
SC14-3387-P	2014	MB	M	0 to less than 1	ST100
SC12-1007-P	2012	NB	F	65 years or greater	ST2705
SC12-1010-P	2012	NB	F	65 years or greater	ST2705
SC12-0156-P	2012	ON	F	65 years or greater	ST2705
SC15-3783-P	2015	NB	M	0 to less than 1	ST2705
SC13-3147-P	2013	QC	M	2 to less than 6	ST9348
SC13-3421-P	2013	QC	M	65 years or greater	ST9350
SC11-2126-P	2011	QC	M	18 to less than 50	ST9348
SC11-4269-P	2011	QC	F	18 to less than 50	ST9350
SC14-1578-P	2014	QC	F	65 years or greater	ST9348
SC12-1981-P	2012	QC	M	1 to less than 2	ST9348
SC15-0966-P	2015	QC	F	65 years or greater	ST9350
SC11-2621-P	2011	ON		50 to less than 65	ST100
SC13-2509-P	2013	QC	F	6 to less than 18	ST100
SC15-4238-P	2015	NB	M	18 to less than 50	ST100
SC12-3021-P	2012	SK	M	65 years or greater	ST100
SC15-0418-P	2015	QC	M	18 to less than 50	ST1543
SC13-3155-P	2013	QC	F	1 to less than 2	ST9701
SC15-3617-P	2015	ON	F	18 to less than 50	ST673
SC12-2896-P	2012	QC	F	50 to less than 65	ST673
SC12-3080-P	2012	ON	F	65 years or greater	ST673
SC11-1936-P	2011	MB	F	2 to less than 6	ST9354

Clusters labelled: 1A, 1B, 1, 2

GDR, YR, PROV: gender, year and province of isolation. Gaps in the demographics indicate that the value was not given at the time of collection. The dashed line indicates 80% similarity. Clusters of isolates are outlined with red boxes and labelled 1-2.

112

FIGURE 5.22: Minimum spanning tree (generated by PHYLOViZ 2.0) of MLST sequence types demonstrated by serotype 33F isolates, coloured according to province of isolation[a]. Green outlines indicate a group founder; numbers indicate the number of differences between the MLST profiles of the two connected nodes (≤ 2 indicates the two nodes are part of a cluster). PMEN clones are marked with red circles.
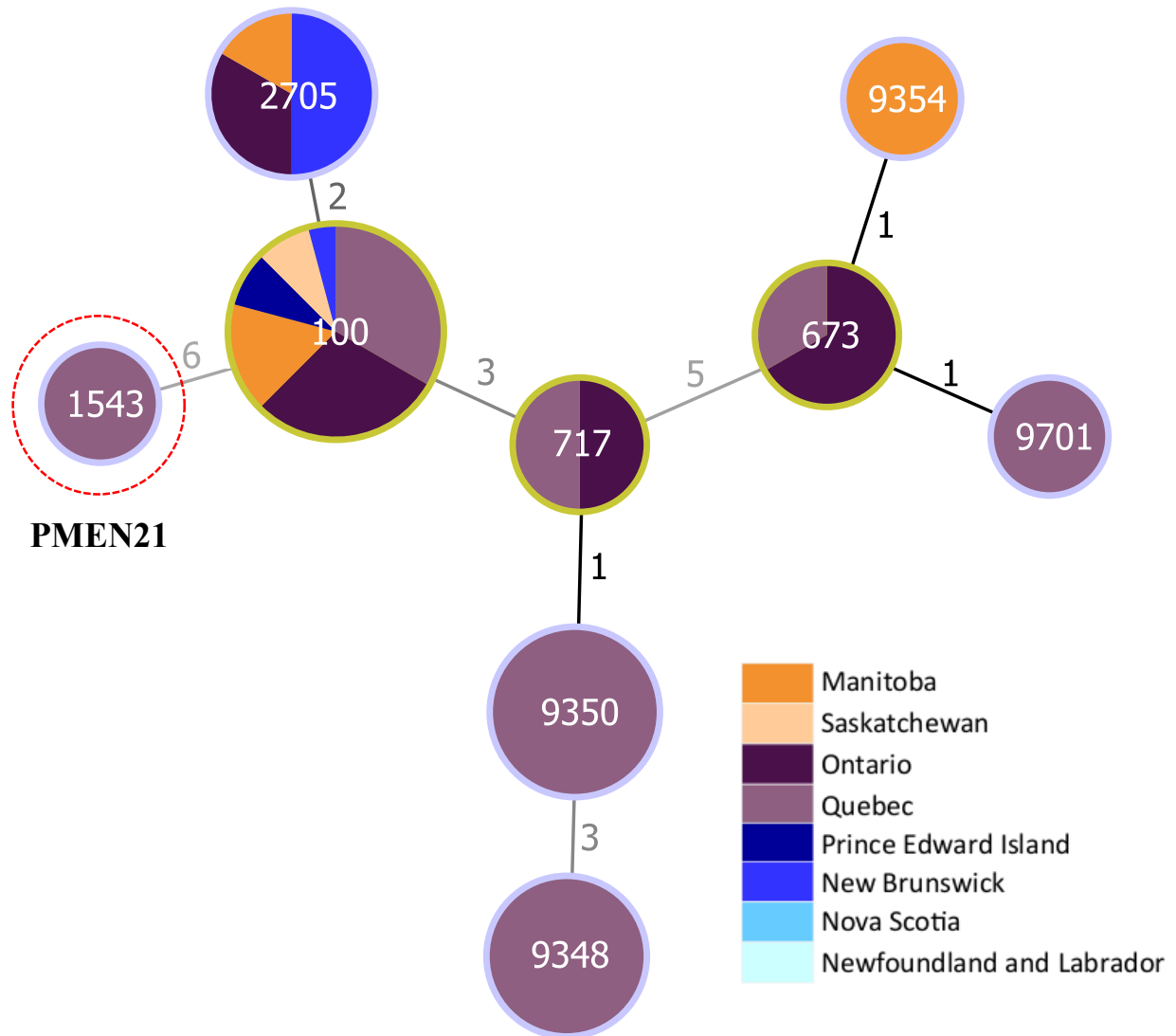
[a] Orange, Western Canada; Purple, Central Canada; Blue, Eastern Canada.

Despite the variability in ST, all serotype 33F isolates clustered closely by phylogenetic analysis. In the original iteration of the analysis using *S. pneumoniae* R6 as the reference genome, all six serotype 33F isolates clustered together, regardless of ST, and were also closely related to serotype 18C, a background type not studied in detail (Figure 5.5). When broken down into smaller groups for analysis, one of the serotype 33F isolates was used as an internal reference; this version of the analysis yielded slightly different results (Figure 5.23). Cluster 5A was comprised of two serotype 33F isolates, plus the reference, that were collected during 2007-2012, were generally susceptible to antimicrobials and related to ST100. Two isolates demonstrated intermediate resistance to trimethoprim-sulfamethoxazole, as conferred by a Val-Glu (VE) duplication in *folP*. Cluster B2 also contained serotype 33F isolates, except these were resistant to at least two antimicrobials, were novel STs and collected from 2012 onwards. These isolates all expressed *ermB*, however despite all possessing *tetM*, only one of the three expressed it in the form of doxycycline resistance. Contrary to the initial analysis, these resistant isolates were more closely related to the serotype 18C background strains in cluster B2 than they were to the other serotype 33F isolates.

FIGURE 5.23: Whole genome core SNV maximum likelihood phylogenetic tree of *S. pneumoniae* "cluster 5" rooted on reference genome SC12-3017-P (this study). Tree was generated using PhyML and visualized using FigTree. Coloured clusters were identified by ClusterPicker. Clusters of interest are labelled A-B2.
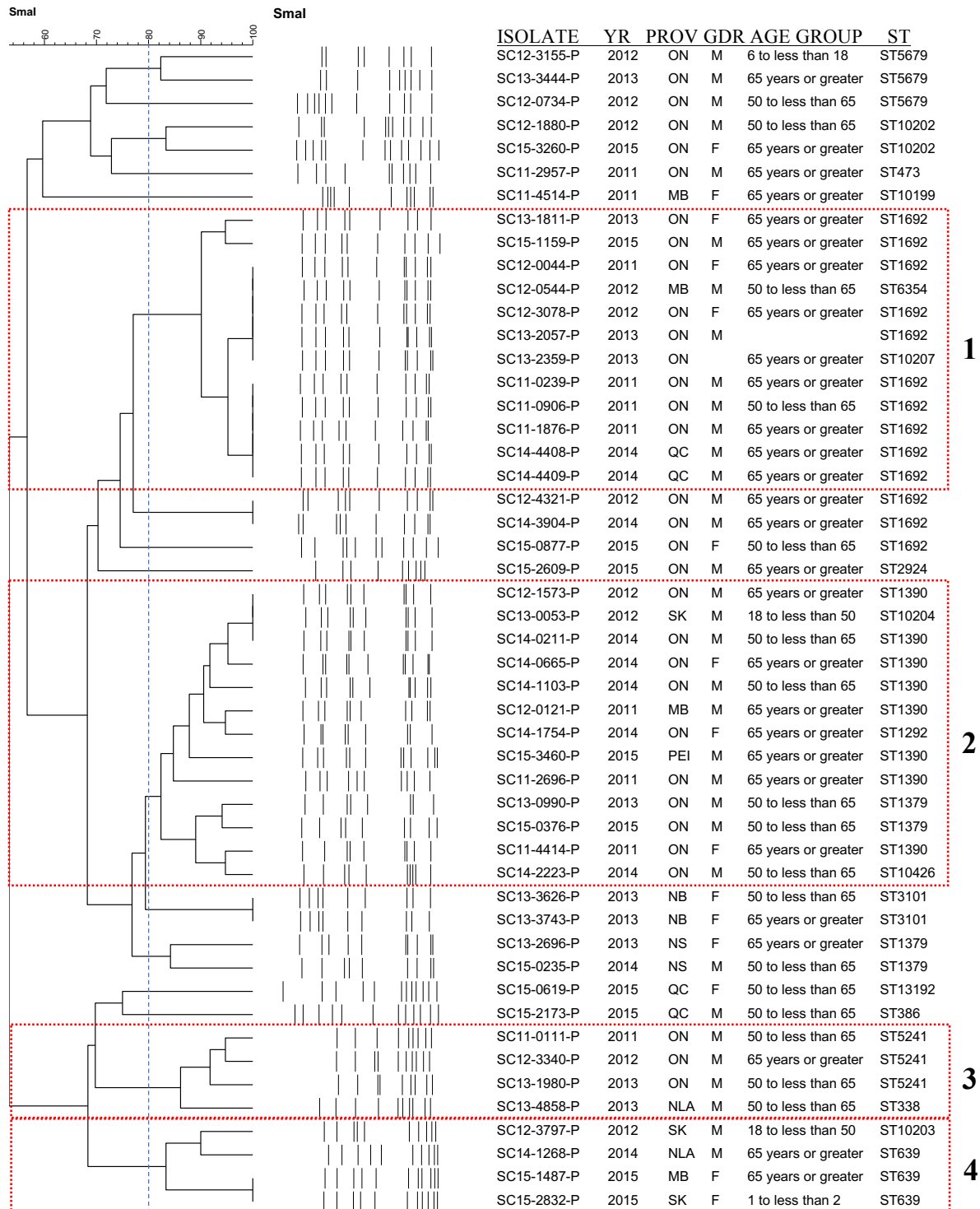
Isolate	Year	Serotype	ST	Res. Pattern
SC13-4822-P	2013	11A	ST62	
SC16-2500-P	2007	11A	ST62	
SC14-0933-P	2014	11A	ST62	CLR, CLD
SC11-2325-P	2011	8		
SC10-0283-P	2007	33F	ST100	
SC10-0293-P	2008	33F	ST13186	
reference SC12-3017-P		33F	ST100	
SC10-0180-P	2009	18C	ST496	
SC12-2054-P	2012	18C	ST496	SXT
SC14-0474-P	2014	18C	ST496	CHL
SC14-2419-P	2014	18C	ST496	
SC13-1746-P	2013	33F	ST9689	CLR, CLD
SC12-1997-P	2012	33F	ST9350	CLR, CLD, DOX
SC14-1578-P	2014	33F	ST9348	CLR, CLD

0.1

115

5.4.3 Clonal Serotypes

5.4.3.1 Serotype 3

In general, serotype 3 isolates were highly related by PFGE (Figure 5.24). Roughly 75% of the isolates typed clustered with 80% similarity and 70% clustered with 90% similarity. All typed isolates that were resistant to two or more classes of antimicrobials were located at the top of the dendrogram above the large cluster, which contained susceptible isolates. Isolates were obtained from across the country from all years of the study, with most of the tested strains being collected from adults.

MLST of serotype 3 isolates generated seven different STs (Figure 5.25). Isolates were highly clonal, with almost 90% typing as ST180 (PMEN31) or a SLV. Four other unrelated STs were identified, including two that are less commonly seen but still attributed to serotype 3 in the MLST database (ST260 and ST1377). The other two STs are indicative of potential capsular switch variants, as ST198 is largely associated with serotype 35B and ST177 is an international clone (PMEN21) with 19F as its representative serotype.

Relating the MLST data back to the PFGE findings, the two methods were comparable for differentiating ST180 isolates from those outside of the clonal cluster (CC). All isolates bearing a type outside of CC180 fell at the bottom of the PFGE dendrogram, outside of the cluster demonstrating 90% relatedness. PFGE, however, clustered the resistant isolates separately from the susceptible; as all isolates demonstrating resistance to two or more classes were also ST180, the minimum spanning tree of MLST data was unable to visualize these differences in phenotype.

FIGURE 5.24: PFGE dendrogram of serotype 3 *S. pneumoniae* isolates collected by the SAVE 2011-15 study. MLST sequence types are listed for comparison.

ISOLATE	YR	PROV	GDR	AGE GROUP	ST
SC13-2434-P	2013	NLA	M		ST180
SC13-3015-P	2013	ON	M	65 years or greater	ST180
SC14-0936-P	2014	ON	M	65 years or greater	ST180
SC13-1157-P	2013	SK	M	50 to less than 65	ST180
SC11-4599-P	2011	ON	F	50 to less than 65	ST180
SC15-4001-P	2015	ON	F	50 to less than 65	ST180
SC14-2134-P	2014	PEI	F	65 years or greater	ST180
SC11-0361-P	2011	SK	F	65 years or greater	ST180
SC11-2360-P	2011	ON	M	18 to less than 50	ST180
SC12-0846-P	2012	SK	F	65 years or greater	ST180
SC12-1362-P	2012	ON	F	65 years or greater	ST180
SC12-3821-P	2012	MB	F	18 to less than 50	ST180
SC12-4058-P	2012	SK	F	18 to less than 50	ST180
SC12-4433-P	2012	ON	F	65 years or greater	ST180
SC13-0018-P	2012	ON	F	6 to less than 18	ST180
SC13-4837-P	2013	NS	M	18 to less than 50	ST180
SC14-0409-P	2014	ON	M	18 to less than 50	ST180
SC14-1160-P	2014	ON	M	50 to less than 65	ST180
SC12-1544-P	2012	MB	F		ST180
SC13-2097-P	2013	SK	F	18 to less than 50	ST180
SC14-2829-P	2014	NB	M	50 to less than 65	ST180
SC11-1487-P	2011	NLA	M	65 years or greater	ST180
SC11-1815-P	2011	ON	M	18 to less than 50	ST8561
SC12-1098-P	2012	ON	M	18 to less than 50	ST180
SC13-1777-P	2013	ON	M	18 to less than 50	ST180
SC13-3300-P	2013	MB	M	1 to less than 2	ST180
SC14-0080-P	2013	ON	M	50 to less than 65	ST180
SC14-1632-P	2014	ON	F		ST180
SC14-2222-P	2014	ON	M	65 years or greater	ST180
SC14-3508-P	2014	ON	M	50 to less than 65	ST180
SC12-1911-P	2012	PEI	M	65 years or greater	ST180
SC14-1869-P	2014	ON	F	65 years or greater	ST180
SC15-0456-P	2015	ON	F	65 years or greater	ST180
SC15-1153-P	2015	ON	F	50 to less than 65	ST9698
SC15-1453-P	2015	SK	M	50 to less than 65	ST180
SC15-3048-P	2015	ON	M	65 years or greater	ST180
SC15-3797-P	2015	MB	M	50 to less than 65	ST180
SC15-3625-P	2015	ON		65 years or greater	ST180
SC11-0565-P	2011	ON	F	65 years or greater	ST180
SC15-2171-P	2015	NB	M	50 to less than 65	
SC15-3262-P	2015	ON	M	0 to less than 1	ST180
SC15-3876-P	2015	ON	F	65 years or greater	ST180
SC11-0096-P	2011	ON	M	50 to less than 65	ST1377
SC13-2056-P	2013	ON	M		ST1377
SC14-4056-P	2014	NLA	F	65 years or greater	ST260
SC11-3907-P	2011	ON	M	18 to less than 50	ST198
SC11-3903-P	2011	SK	M	0 to less than 1	ST177
SC12-0549-P	2012	SK	M	50 to less than 65	ST180
SC11-0833-P	2011	SK	F	65 years or greater	ST180
SC13-0117-P	2013	ON	F	65 years or greater	ST1377

GDR, YR, PROV: gender, year and province of isolation. Gaps in the demographics indicate that the value was not given at the time of collection. The dashed lines indicate 80% and 90% similarity, respectively. Clusters of isolates are outlined with red boxes.

FIGURE 5.25: Minimum spanning tree (generated by PHYLOViZ 2.0) of MLST sequence types demonstrated by serotype 3 isolates, coloured according to province of isolation[a]. Green outlines indicate a group founder; numbers indicate the number of differences between the MLST profiles of the two connected nodes (≤ 2 indicates the two nodes are part of a cluster). PMEN clones are marked with red circles.

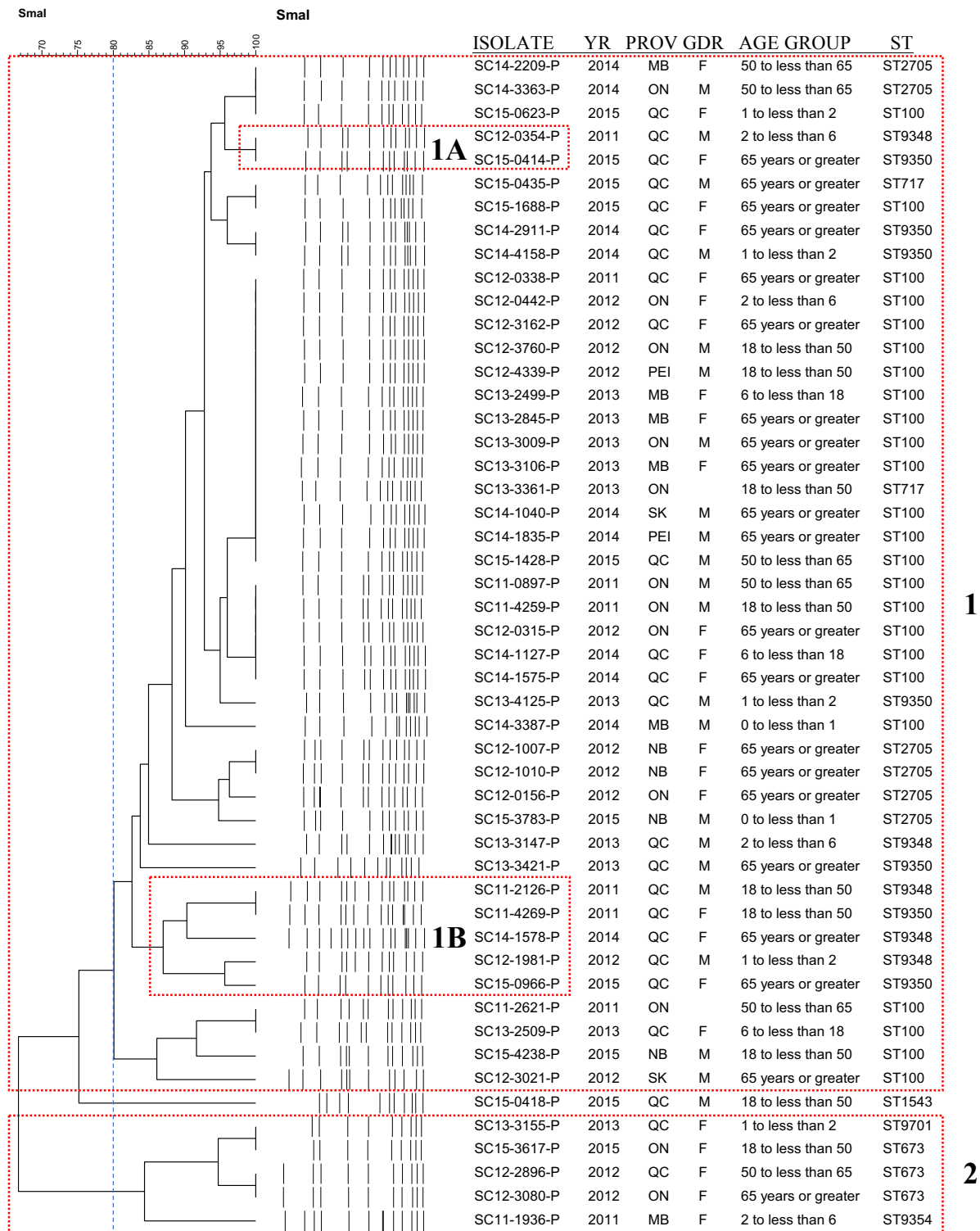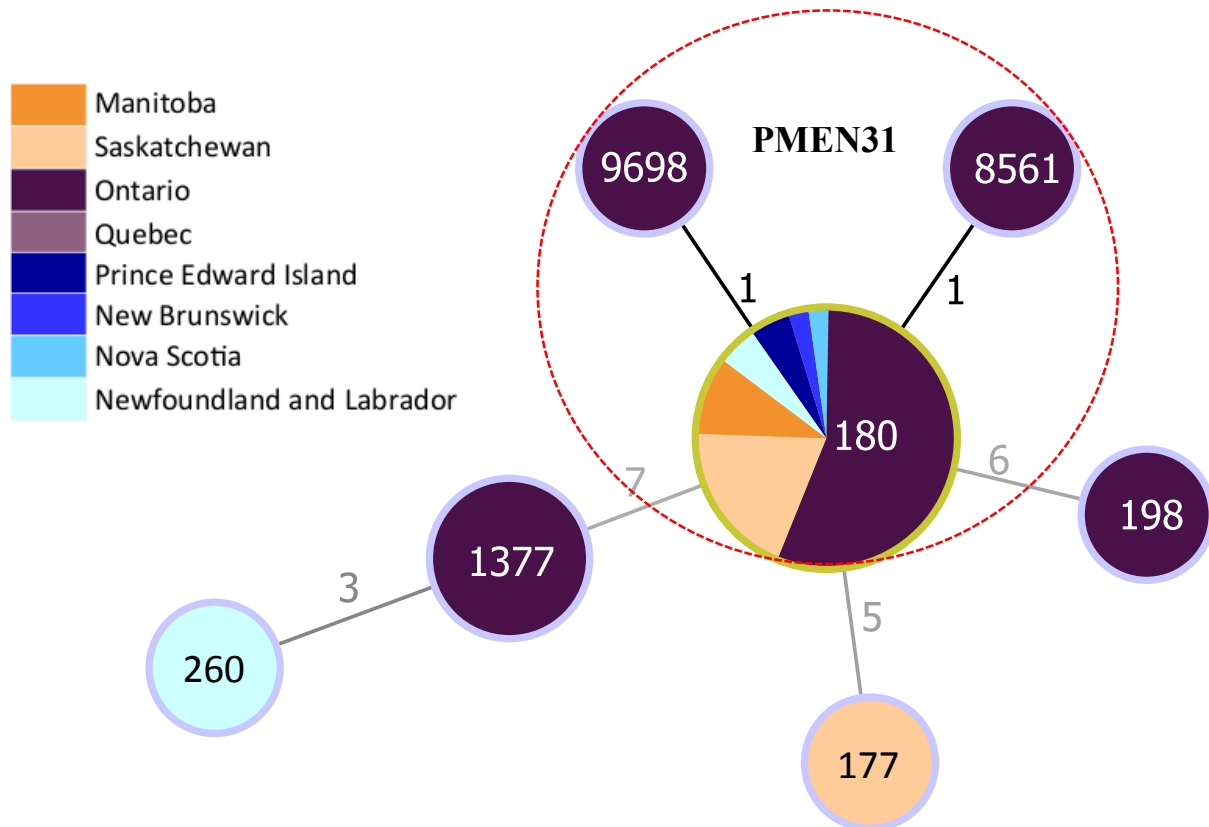[a] Orange, Western Canada; Purple, Central Canada; Blue, Eastern Canada.

Phylogenetic analysis of serotype 3 isolates included in the WGS sampling is presented in Figures 5.26 and 5.14, representing Clusters 1 and 6, respectively (as described in Table 5.5). All serotype 3 isolates but one grouped together in Cluster 1. *In silico* MLST identified each of these isolates as ST180, again demonstrating the clonal nature of this serotype. However, when examining the phylogeny of these isolates, ST180 was comprised of two distinct clades. Clade I was generally susceptible to all antimicrobials, except for one isolate that possessed *ermB* and *tetM* and demonstrated resistance to clarithromycin, clindamycin and doxycycline. Conversely, all isolates in clade II were MDR, demonstrating resistance to chloramphenicol, clarithromycin, doxycycline and often clindamycin. Along with *ermB* and *tetM*, these serotype 3-clade II isolates were the only serotype/group of isolates in the WGS subset to consistently possess the chloramphenicol resistance determining gene, *cat*. Interestingly, isolates within clade I were collected from 2008-2012, while clade II isolates were collected from 2011-2014.

The last serotype 3 isolate typed as ST177, and thus grouped separately from the other isolates of the same serotype, falling in Cluster 6. Cluster 6 included all isolates that typed as identical or similar to PMEN clone Portugal19F-21. The phylogenetic analysis revealed that this serotype 3 isolate was virtually identical to a serotype 24B background isolate sharing the same ST; it was also closely related to the PMEN21 reference strain (serotype 19F) and a SLV with serotype 19A. This similarity indicates that isolates in this group may have undergone recombination to become capsular switch variants.

FIGURE 5.26: Whole genome core SNV maximum likelihood phylogenetic tree of *S. pneumoniae* "cluster 1" rooted on reference genome *S. pneumoniae* OXC141 (NC_017592). Tree was generated using PhyML and visualized using FigTree. Coloured clusters were identified by ClusterPicker.

Isolate	Year	Serotype	ST	Res. Pattern
SC10-0184-P	2009	7F	ST191	
SC12-1054-P	2012	7F	ST191	
SC11-4599-P	2011	3	ST180	CHL, CLR, CLD, DOX
SC12-3569-P	2012	3	ST180	CHL, CLR, CLD, DOX
SC13-3015-P	2013	3	ST180	CHL, CLR, CLD, DOX
SC12-1091-P	2012	3	ST180	CHL, CLR, DOX
SC14-0936-P	2014	3	ST180	CHL, CLR, CLD, DOX
reference	OXC141	3	ST180	
SC10-0234-P	2009	3	ST180	
SC12-4157-P	2012	3	ST180	CLR, CLD, DOX
SC16-2506-P	2008	3	ST180	
SC12-0846-P	2012	3	ST180	
SC16-2502-P	2008	3	ST180	

Clade II

Clade I

0.2

5.4.3.2 Serotype 7F

Serotype 7F, the most commonly isolated serotype in the SAVE study, was also the most clonal of the serotypes studied. Almost all isolates (46/50) clustered together with >80% relatedness (Figure 5.27). Isolates in this cluster were from all provinces, ages and study years. Few serotype 7F isolates demonstrated resistance to antimicrobials; those that did were a part of cluster 1A, a subgroup of cluster 1, and demonstrated >90% relatedness. Isolates were either resistant to clarithromycin and doxycycline or doxycycline alone.

All but one isolate typed as ST191 by MLST, identical to the PMEN39 international clone originally isolated in the Netherlands (Figure 5.28). This included all isolates from cluster 1A that displayed resistance to antimicrobials. The lone isolate not related to PMEN39 typed as ST3545 and was obtained from Western Canada; in the PFGE dendrogram, this isolate fell just below cluster 1, with which it demonstrated approximately 70% relatedness.

Interestingly, the ST191 isolates typed by WGS were the least related to the rest of the population, clustering distantly from all other isolates (Figures 5.5 and 5.26). One serotype 7F isolate demonstrating resistance to clarithromycin, clindamycin and doxycycline typed as ST63 and clustered accordingly with the MDR 15A isolates (cluster 4-B3, Figure 5.8). Based on the resistance pattern and clustering, this isolate is likely another product of capsular switching.

FIGURE 5.27: PFGE dendrogram of serotype 7F *S. pneumoniae* isolates collected by the SAVE 2011-15 study. MLST sequence types are listed for comparison.

ISOLATE	YR	PROV	GDR	AGE GROUP	ST
SC12-2890-P	2012	QC	M	50 to less than 65	ST191
SC13-2363-P	2013	ON	M	50 to less than 65	ST191
SC13-2674-P	2013	NS	F	50 to less than 65	ST191
SC11-0330-P	2011	MB	M	65 years or greater	ST191
SC11-0391-P	2011	MB	M	18 to less than 50	ST191
SC11-0519-P	2011	MB	M	65 years or greater	ST191
SC13-3167-P	2013	ON	F	65 years or greater	ST191
SC11-0301-P	2011	ON	F	6 to less than 18	ST191
SC11-4282-P	2011	QC	M	2 to less than 6	ST191
SC14-4238-P	2014	ON	M	18 to less than 50	ST191
SC14-4281-P	2014	ON	M	65 years or greater	ST191
SC14-0074-P	2013	ON	M	18 to less than 50	ST191
SC11-2112-P	2011	QC	F	65 years or greater	ST191
SC12-1228-P	2012	SK	M	50 to less than 65	ST191
SC12-1359-P	2012	ON	F	65 years or greater	ST191
SC13-2100-P	2013	SK	F	18 to less than 50	ST191
SC14-0662-P	2014	ON	M	50 to less than 65	ST191
SC14-0671-P	2014	ON	M	50 to less than 65	ST191
SC14-1155-P	2014	ON	M	18 to less than 50	ST191
SC14-1439-P	2014	ON	F	18 to less than 50	ST191
SC14-2335-P	2014	ON	F	18 to less than 50	ST191
SC13-1808-P	2013	ON	F	50 to less than 65	ST191
SC12-1184-P	2012	ON	M	2 to less than 6	ST191
SC11-1205-P	2011	NLA	M	65 years or greater	ST191
SC11-1220-P	2011	ON	F	50 to less than 65	ST191
SC11-1857-P	2011	ON	M	0 to less than 1	ST191
SC13-0483-P	2013	ON	F	18 to less than 50	ST191
SC14-3055-P	2014	ON	F	50 to less than 65	ST191
SC12-0975-P	2012	ON	M	65 years or greater	ST191
SC12-3166-P	2012	QC	F	0 to less than 1	ST191
SC13-2461-P	2013	ON	M	18 to less than 50	ST191
SC11-3995-P	2011	ON	F	50 to less than 65	ST191
SC15-0732-P	2015	NB	F	18 to less than 50	ST191
SC15-0939-P	2015	ON	F	65 years or greater	ST191
SC15-1702-P	2015	MB	M	65 years or greater	ST191
SC15-2706-P	2015	NLA	F	65 years or greater	ST191
SC12-3083-P	2012	NLA	F	65 years or greater	ST191
SC15-0467-P	2015	ON	M	50 to less than 65	ST191
SC15-3620-P	2015	ON	F	65 years or greater	ST191
SC15-2259-P	2015	ON		50 to less than 65	ST191
SC15-3019-P	2015	SK	F	0 to less than 1	ST191
SC15-3258-P	2015	ON	F	2 to less than 6	ST191
SC14-0567-P	2014	ON	M	18 to less than 50	ST191
SC14-2735-P	2014	ON	M	6 to less than 18	ST191
SC12-3657-P	2012	PEI	M	6 to less than 18	ST191
SC15-4127-P	2015	SK	F	65 years or greater	ST3545
SC12-3508-P	2012	QC	M	6 to less than 18	
SC13-1045-P	2013	ON	M		ST191
SC12-0550-P	2012	SK	M	18 to less than 50	ST191

GDR, YR, PROV: gender, year and province of isolation. Gaps in the demographics indicate that the value was not given at the time of collection. The dashed lines indicate 80% and 90% similarity, respectively. Clusters of isolates are outlined with red boxes labelled 1-1A.

FIGURE 5.28: Minimum spanning tree (generated by PHYLOViZ 2.0) of MLST sequence types demonstrated by serotype 7F isolates, coloured according to province of isolation[a]. Green outlines indicate a group founder; numbers indicate the number of differences between the MLST profiles of the two connected nodes (≤ 2 indicates the two nodes are part of a cluster). PMEN clones are marked with red circles.

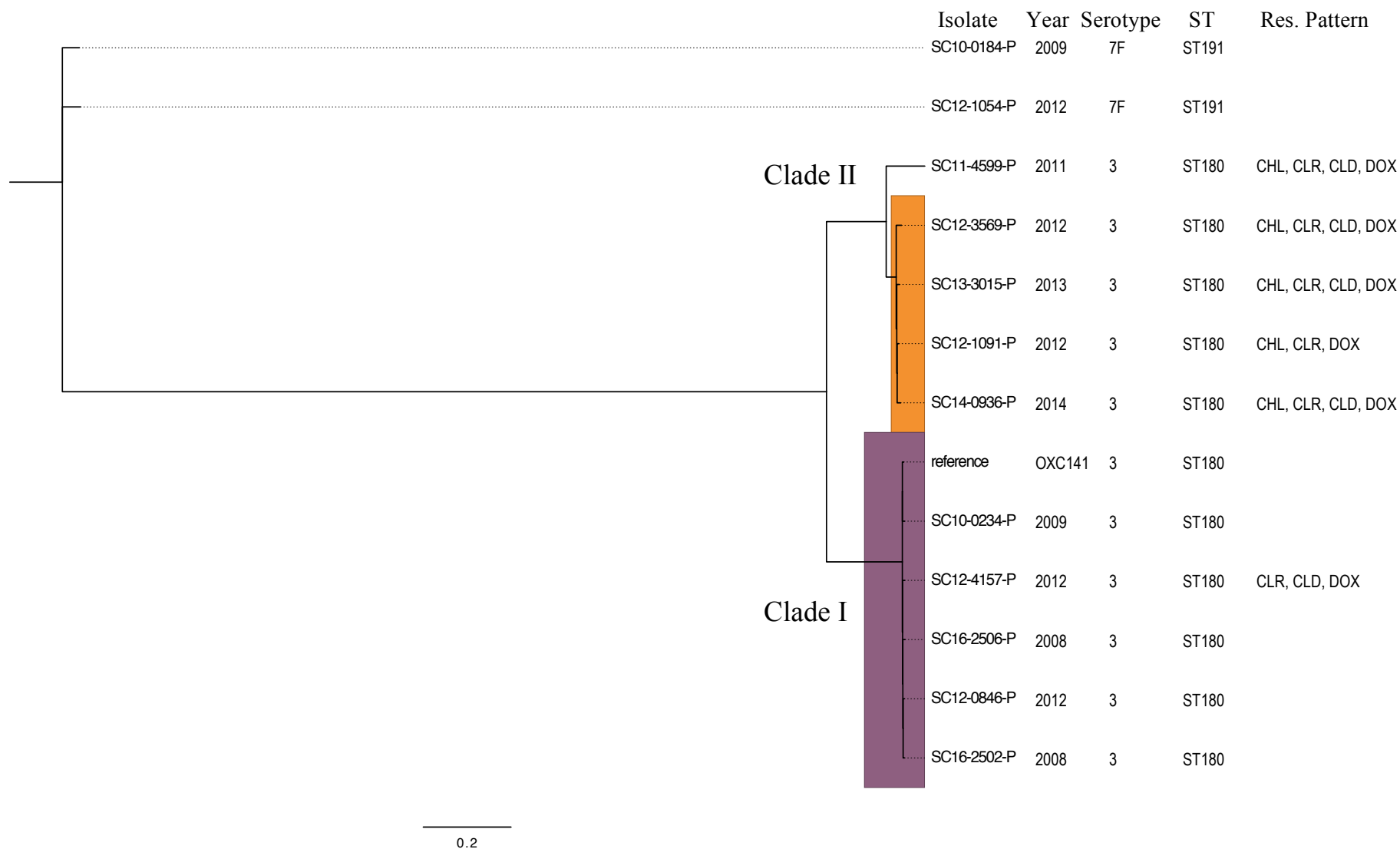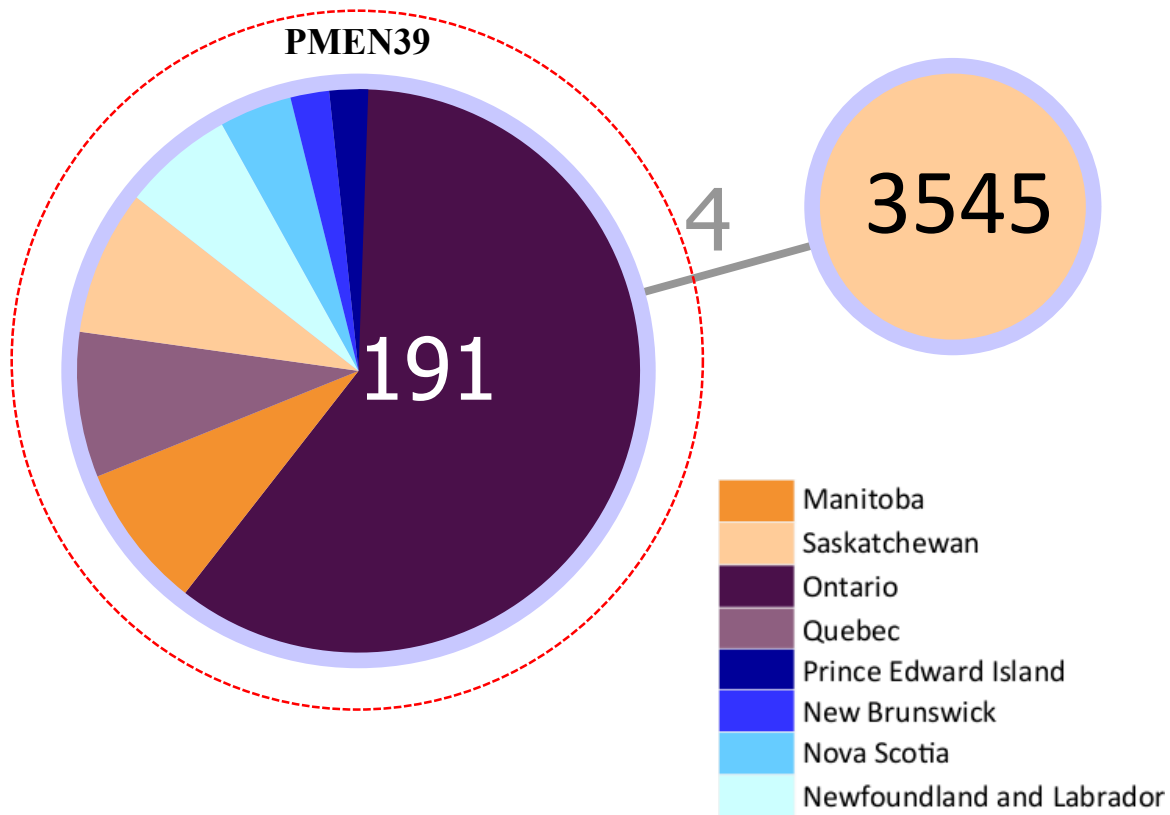[a] Orange, Western Canada; Purple, Central Canada; Blue, Eastern Canada.

5.4.3.3 Serotype 22F

Serotype 22F isolates were closely related by PFGE (Figure 5.29). As with serotype 7F, most of the 22F isolates (42/50) clustered in one large group demonstrating >80% relatedness. Isolates in cluster 1 demonstrated variable resistance to antimicrobials; many demonstrated no resistance, while some were resistant to clarithromycin. The clarithromycin-resistant isolates were largely contained in sub-cluster 1A, however others were scattered throughout cluster 1. Four isolates fell into cluster 2, which had the defining characteristic of predominantly being collected in 2014 from New Brunswick females over the age of 65. Three MDR serotype 22F isolates were identified in different parts of the dendrogram. One with resistance to clarithromycin, clindamycin and doxycycline was part of cluster 1; a second had additional resistance to chloramphenicol and clustered just outside of cluster 1. The last MDR isolate had the pattern above with additional resistance to levofloxacin; this isolate was the least like the other serotype 22F isolates, demonstrating approximately 70% relatedness.

Over 95% of serotype 22F isolates typed by MLST fell into a cluster founded by ST433 (Figure 5.30); this cluster included five of the seven identified STs. ST433 was identified across the country, however STs 9289 and 9290 were collected from New Brunswick and Saskatchewan, respectively, while STs 4533 and 7314 were only collected from Ontario. All of these STs were present in cluster 1 of the above dendrogram except ST9289, which comprised cluster 2. The two STs that were unrelated to ST433 were both MDR; ST9352 was a SLV of ST63 and demonstrated resistance to clarithromycin, clindamycin, doxycycline and levofloxacin (the least related isolate from the dendrogram), while ST6541 was the isolate from just below dendrogram cluster 1 that was resistant to chloramphenicol, clarithromycin, clindamycin and doxycycline.

FIGURE 5.29: PFGE dendrogram of serotype 22F *S. pneumoniae* isolates collected by the SAVE 2011-15 study. MLST sequence types are listed for comparison.

ISOLATE	YR	PROV	GDR	AGE GROUP	ST
SC13-0031-P	2012	ON	M	65 years or greater	ST433
SC13-3839-P	2013	ON	M	15-64 years	ST4553
SC14-3336-P	2014	ON		18 to less than 50	ST433
SC13-0361-P	2013	ON	M	50 to less than 65	ST433
SC14-1567-P	2014	ON	M	50 to less than 65	ST433
SC14-2767-P	2014	ON	F	65 years or greater	ST433
SC11-1531-P	2011	ON	M	65 years or greater	ST433
SC11-1538-P	2011	ON	F	65 years or greater	ST433
SC11-2098-P	2011	ON	M	18 to less than 50	ST433
SC11-4247-P	2011	ON	M	50 to less than 65	ST433
SC13-2391-P	2013	NLA	F	65 years or greater	ST433
SC15-1447-P	2015	SK	M	65 years or greater	ST433
SC15-2565-P	2015	SK	M	65 years or greater	ST433
SC12-3886-P	2012	NLA	M	0 to less than 1	ST433
SC14-1331-P	2014	ON	F	50 to less than 65	ST433
SC15-3212-P	2015	ON	M	18 to less than 50	ST433
SC15-3214-P	2015	ON		50 to less than 65	ST433
SC15-2519-P	2015	NS	M	65 years or greater	ST433
SC11-1530-P	2011	ON	M	65 years or greater	ST433
SC12-2908-P	2012	QC	M	50 to less than 65	ST433
SC15-1317-P	2015	ON	F	65 years or greater	ST433
SC15-3863-P	2015	NB	M	50 to less than 65	ST433
SC12-4394-P	2012	ON	M	50 to less than 65	ST433
SC13-1054-P	2013	ON	M	65 years or greater	ST433
SC12-3999-P	2012	ON	F	50 to less than 65	ST433
SC13-0196-P	2013	MB	M	50 to less than 65	ST433
SC14-1154-P	2014	ON	M	65 years or greater	ST433
SC13-2050-P	2013	ON	M	6 to less than 18	ST433
SC11-4384-P	2011	NB	F	65 years or greater	ST433
SC13-3849-P	2013	ON	F	65 years or greater	ST7314
SC12-1574-P	2012	ON	M	65 years or greater	ST433
SC11-4291-P	2011	NLA	M	65 years or greater	ST433
SC12-0774-P	2012	QC	M	65 years or greater	ST433
SC15-3050-P	2015	MB	F	50 to less than 65	
SC11-0576-P	2011	ON	M	18 to less than 50	ST433
SC13-0446-P	2013	MB	M	6 to less than 18	ST433
SC11-1854-P	2011	ON	M	50 to less than 65	ST433
SC12-3022-P	2012	SK	F	18 to less than 50	ST9290
SC12-4001-P	2012	ON	M	50 to less than 65	ST433
SC13-1778-P	2013	ON	F	1 to less than 2	ST433
SC13-3016-P	2013	ON	F	65 years or greater	ST433
SC12-1572-P	2012	ON	M	65 years or greater	ST433
SC14-3633-P	2014	ON	M	18 to less than 50	ST433
SC15-1593-P	2015	ON	M	50 to less than 65	ST6541
SC11-4423-P	2011	ON		65 years or greater	ST433
SC14-4263-P	2014	NB	F	65 years or greater	ST9289
SC14-3861-P	2014	NB	F	65 years or greater	ST9289
SC14-0042-P	2014	NB	F	65 years or greater	ST9289
SC15-0754-P	2015	NB	F	65 years or greater	ST9289
SC14-0467-P	2014	ON	M	65 years or greater	ST9352

GDR, YR, PROV: gender, year and province of isolation. Gaps in the demographics indicate that the value was not given at the time of collection. The dashed lines indicate 80% and 90% similarity, respectively. Clusters of isolates are outlined with red boxes and labelled 1-2.

FIGURE 5.30: Minimum spanning tree (generated by PHYLOViZ 2.0) of MLST sequence types demonstrated by serotype 22F isolates, coloured according to province of isolation[a]. Green outlines indicate a group founder; numbers indicate the number of differences between the MLST profiles of the two connected nodes (≤ 2 indicates the two nodes are part of a cluster). PMEN clones are marked with red circles.
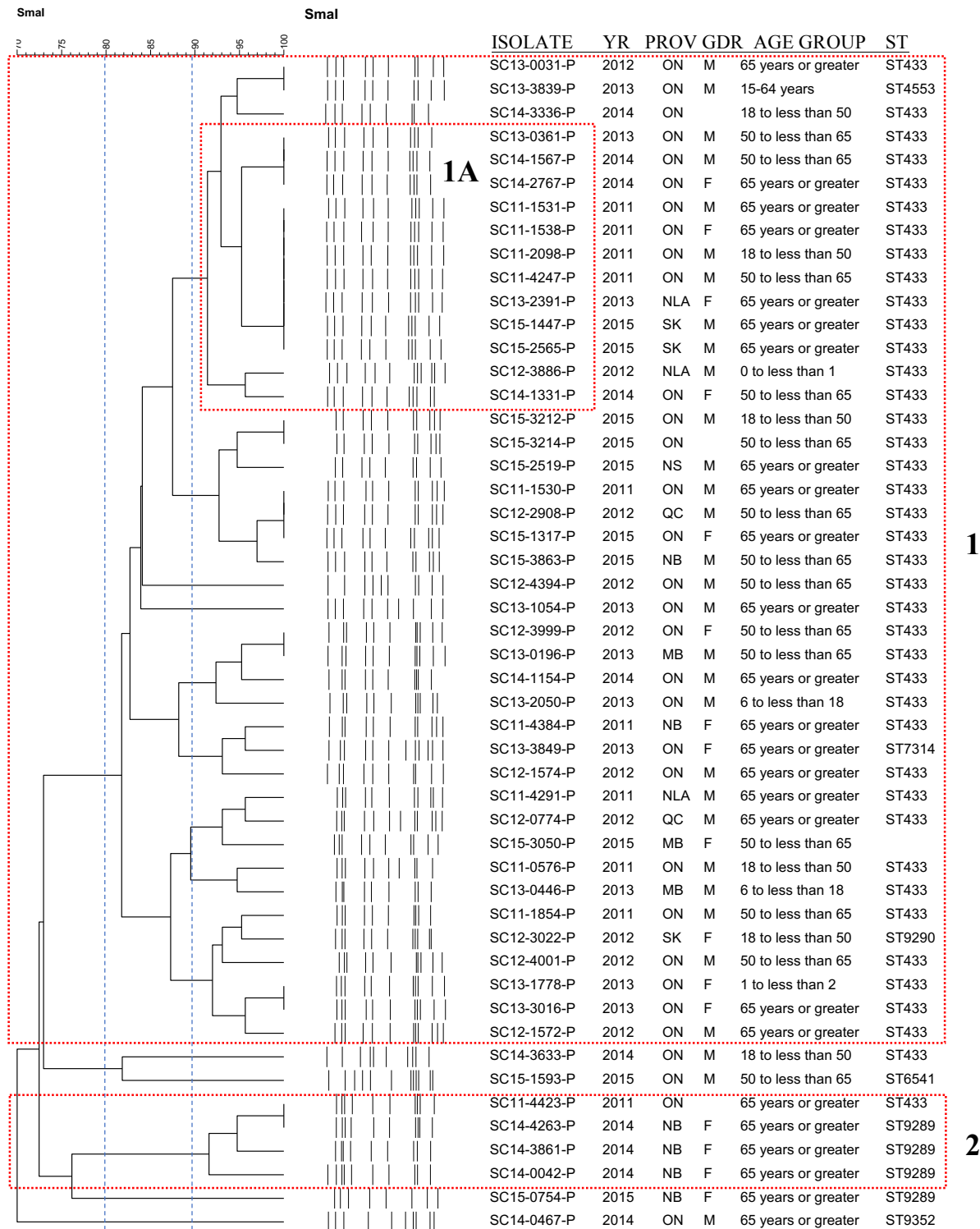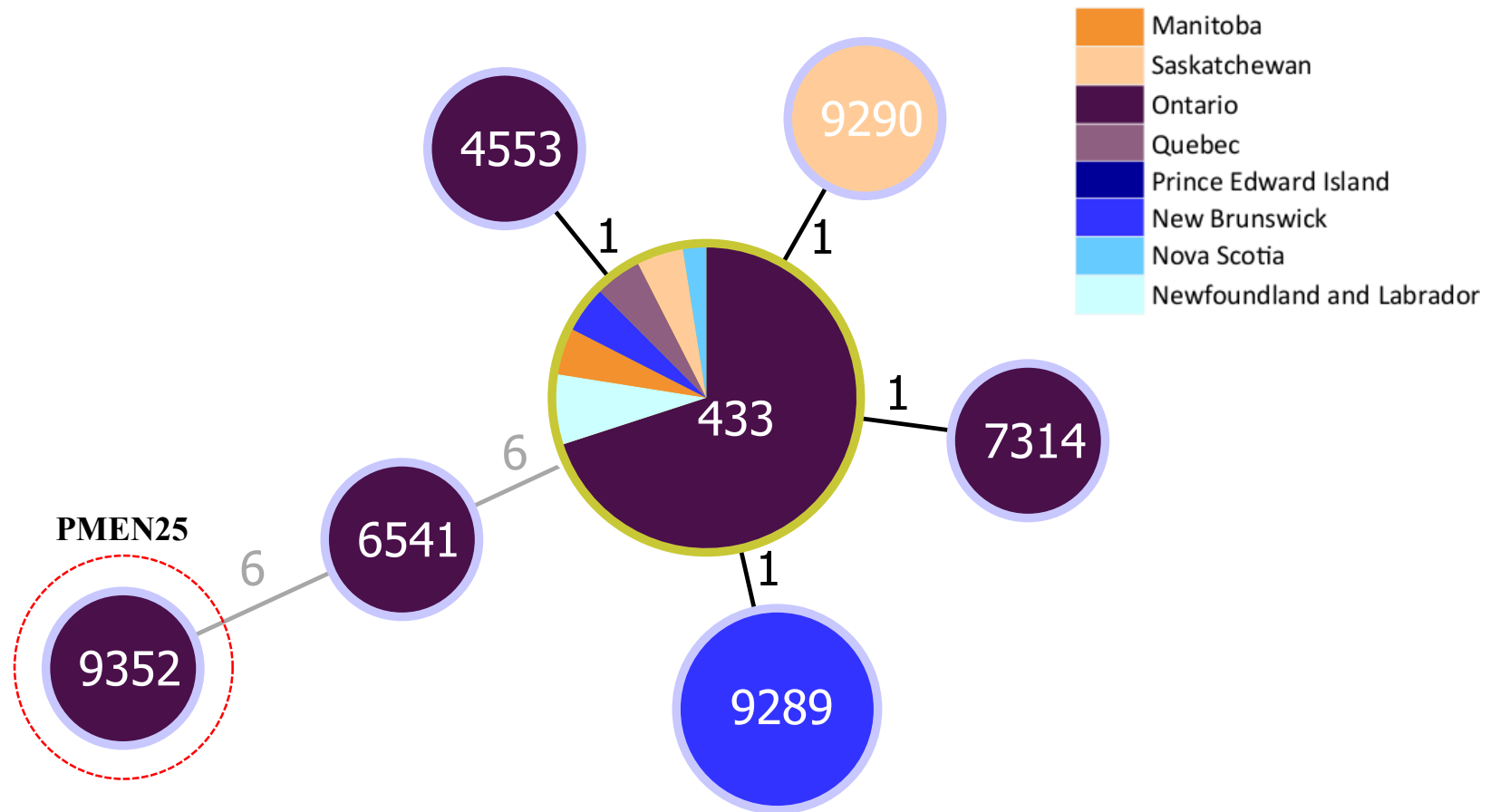

[a] Orange, Western Canada; Purple, Central Canada; Blue, Eastern Canada.

Similarly to MLST, most serotype 22F isolates (7/9) were highly related in the phylogenetic analysis. Depicted in Figure 5.31 as part of cluster 11, these isolates were most closely related to background serotype 38. The cluster of highly related serotype 22F isolates was further divided by ClusterPicker into A1 and A2; A2 contained most isolates, with a range of collection dates from 2007-2014. Many of these were considered almost identical in terms of SNVs, regardless of antimicrobial resistance pattern. Cluster A1 contained two isolates, one serotype 22F-ST433 and the second serotype 31-ST433. The relatedness of these two isolates in comparison to others of the same ST indicates that they likely were the recipient and progeny, respectively, of a capsular switch event.

Isolates that typed as ST9352 by MLST were also related by phylogenetic analysis. As shown in Figure 5.8, the two ST9352 isolates included were almost identical to each other, though not as related to their ST63 founders as would be predicted from their "SLV" status. Along with possessing *ermB* and *tetM*, these isolates demonstrated mutations in both genes responsible for fluoroquinolone resistance, specifically Ser79Phe in *parC* and Ser81Leu in *gyrA*. Despite the difference in resistance pattern between these ST9352 isolates and their founders, they are still possible candidates as capsular switch progeny due to their phylogenetic similarity.

FIGURE 5.31: Whole genome core SNV maximum likelihood phylogenetic tree of *S. pneumoniae* "cluster 11" rooted on reference genome SC13-0467-P (this study). Tree was generated using PhyML and visualized using FigTree. Coloured clusters were identified by ClusterPicker. Clusters of interest are labelled A1 and A2.

Isolate	Year	Serotype	ST	Res. Pattern
SC16-2511-P	2009	8	404	
reference SC13-0467-P		8	1480	
SC11-0288-P	2011	38	393	
SC11-4567-P	2011	38	393	
SC10-0275-P	2007	38	393	
SC12-0350-P	2011	22F	433	
SC14-3964-P	2014	31	433	
SC11-2110-P	2011	22F	433	LEV
SC16-2499-P	2007	22F	433	CLR
SC16-2507-P	2008	22F	433	
SC11-1577-P	2011	22F	433	
SC14-1096-P	2014	22F	433	CHL, CLR, CLD
SC12-0774-P	2012	22F	433	CLR, CLD, DOX

0.2

128

5.4.3.4 Serotype 11A

Serotype 11A was highly related by PFGE (Figure 5.32). Forty isolates clustered together

with >80% relatedness, with only three of 50 not clustering with >75% relatedness. Within this

large cluster were four sub-clusters that contained isolates resistant to one or two antimicrobials.

Cluster 1A contained isolates resistant to trimethoprim-sulfamethoxazole, while cluster 1B

largely contained clarithromycin-resistant isolates. Clusters 1C and 1D contained those isolates

that were resistant to both of the antimicrobials mentioned above. Of the three isolates that did

not cluster with >75% relatedness, all were resistant to trimethoprim-sulfamethoxazole; in

addition, one was also penicillin-resistant while another was resistant to doxycycline.

Serotype 11A was also clonal by MLST (Figure 5.33). Most isolates typed as ST62 or a

variant type. Some of these isolates were resistant to clarithromycin, trimethoprim-

sulfamethoxazole, or both. One isolate typed as ST156, identical to commonly antimicrobial

resistant international clone PMEN3; the serotype 11A isolate in question demonstrated

resistance to penicillin and trimethoprim-sulfamethoxazole and was therefore one of the isolates

that did not cluster by PFGE. Two isolates with newly identified STs also demonstrated

resistance; ST10196 was resistant to levofloxacin and trimethoprim-sulfamethoxazole, while

ST13193 was resistant to the latter plus doxycycline.

FIGURE 5.32: PFGE dendrogram of serotype 11A *S. pneumoniae* isolates collected by the SAVE 2011-15 study. MLST sequence types are listed for comparison.

ISOLATE	YR	PROV	GDR	AGE GROUP	ST
SC11-0264-P	2011	SK	M	65 years or greater	ST62
SC14-0986-P	2014	ON	F	18 to less than 50	ST156
SC13-2362-P	2013	ON	F	65 years or greater	ST62
SC15-0625-P	2015	QC	M	50 to less than 65	ST62
SC15-0971-P	2015	QC	M	18 to less than 50	ST408
SC15-0616-P	2015	QC	F	65 years or greater	ST62
SC13-2090-P	2013	SK	F	6 to less than 18	ST10206
SC15-1979-P	2015	ON	F	50 to less than 65	ST62
SC15-1987-P	2015	ON	F	65 years or greater	ST62
SC12-3993-P	2012	ON	M	65 years or greater	ST62
SC12-4039-P	2012	ON	F	65 years or greater	ST62
SC12-3250-P	2012	ON	M	65 years or greater	ST62
SC14-1257-P	2014	SK	M	50 to less than 65	ST62
SC13-2008-P	2012	ON			ST62
SC15-0252-P	2014	NS	M	18 to less than 50	ST62
SC11-0382-P	2011	ON	F	65 years or greater	ST62
SC11-1292-P	2011	MB	M	0 to less than 1	ST62
SC11-2225-P	2011	QC	M	50 to less than 65	ST62
SC11-2676-P	2011	QC	F	65 years or greater	ST62
SC11-4184-P	2011	ON	M	18 to less than 50	ST62
SC12-1509-P	2012	SK	M	18 to less than 50	ST62
SC12-1980-P	2012	QC	M	6 to less than 18	ST62
SC12-4336-P	2012	ON	F	65 years or greater	ST62
SC13-3644-P	2013	ON	F	50 to less than 65	ST62
SC13-4132-P	2013	MB	M	50 to less than 65	ST62
SC14-1977-P	2014	SK	F	65 years or greater	ST62
SC14-2160-P	2014	ON	M	1 to less than 2	ST62
SC14-4198-P	2014	ON	M	50 to less than 65	ST62
SC15-2567-P	2015	SK	M	18 to less than 50	ST62
SC14-1129-P	2014	QC	F	65 years or greater	ST62
SC11-0132-P	2011	MB	M	0 to less than 1	ST62
SC11-0370-P	2011	ON	M	65 years or greater	ST62
SC11-4139-P	2011	ON	F	65 years or greater	ST62
SC11-1542-P	2011	ON	M	65 years or greater	ST62
SC12-0148-P	2012	ON	M	65 years or greater	ST62
SC12-3478-P	2012	QC	M	2 to less than 6	ST62
SC13-3445-P	2013	ON	M	65 years or greater	ST62
SC12-3946-P	2012	ON	F	65 years or greater	ST62
SC13-3215-P	2013	SK	M	18 to less than 50	ST62
SC15-0373-P	2015	ON	M	65 years or greater	ST62
SC13-3239-P	2013	MB	M	18 to less than 50	ST62
SC15-2134-P	2015	ON	F	18 to less than 50	ST62
SC13-1821-P	2013	ON	M	65 years or greater	ST62
SC13-2512-P	2013	QC	M	65 years or greater	ST62
SC15-0618-P	2015	QC	M	50 to less than 65	ST62
SC13-2733-P	2013	MB	M	65 years or greater	ST10196
SC14-3271-P	2014	QC	M	65 years or greater	ST62
SC14-2969-P	2014	SK	M	65 years or greater	ST408
SC14-2426-P	2014	SK	M	50 to less than 65	ST10206
SC15-1202-P	2015	ON	F	50 to less than 65	ST13193

GDR, YR, PROV: gender, year and province of isolation. Gaps in the demographics indicate that the value was not given at the time of collection. The dashed lines indicate 70% and 80% similarity, respectively. Clusters of isolates are outlined with red boxes and labelled 1-1D.

FIGURE 5.33: Minimum spanning tree (generated by PHYLOViZ 2.0) of MLST sequence types demonstrated by serotype 11A isolates, coloured according to province of isolation[a]. Green outlines indicate a group founder; numbers indicate the number of differences between the MLST profiles of the two connected nodes (≤ 2 indicates the two nodes are part of a cluster). PMEN clones are marked with red circles.
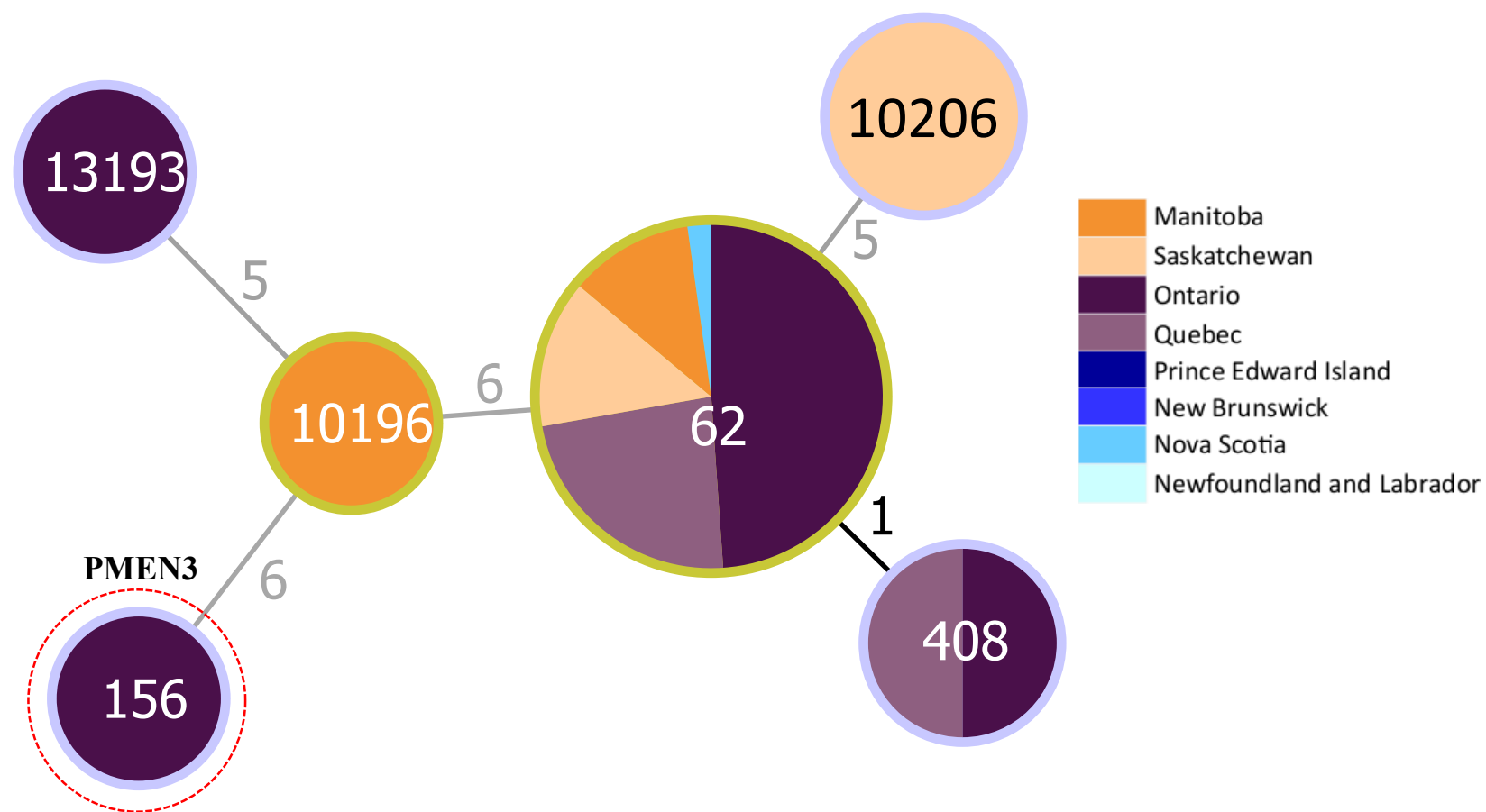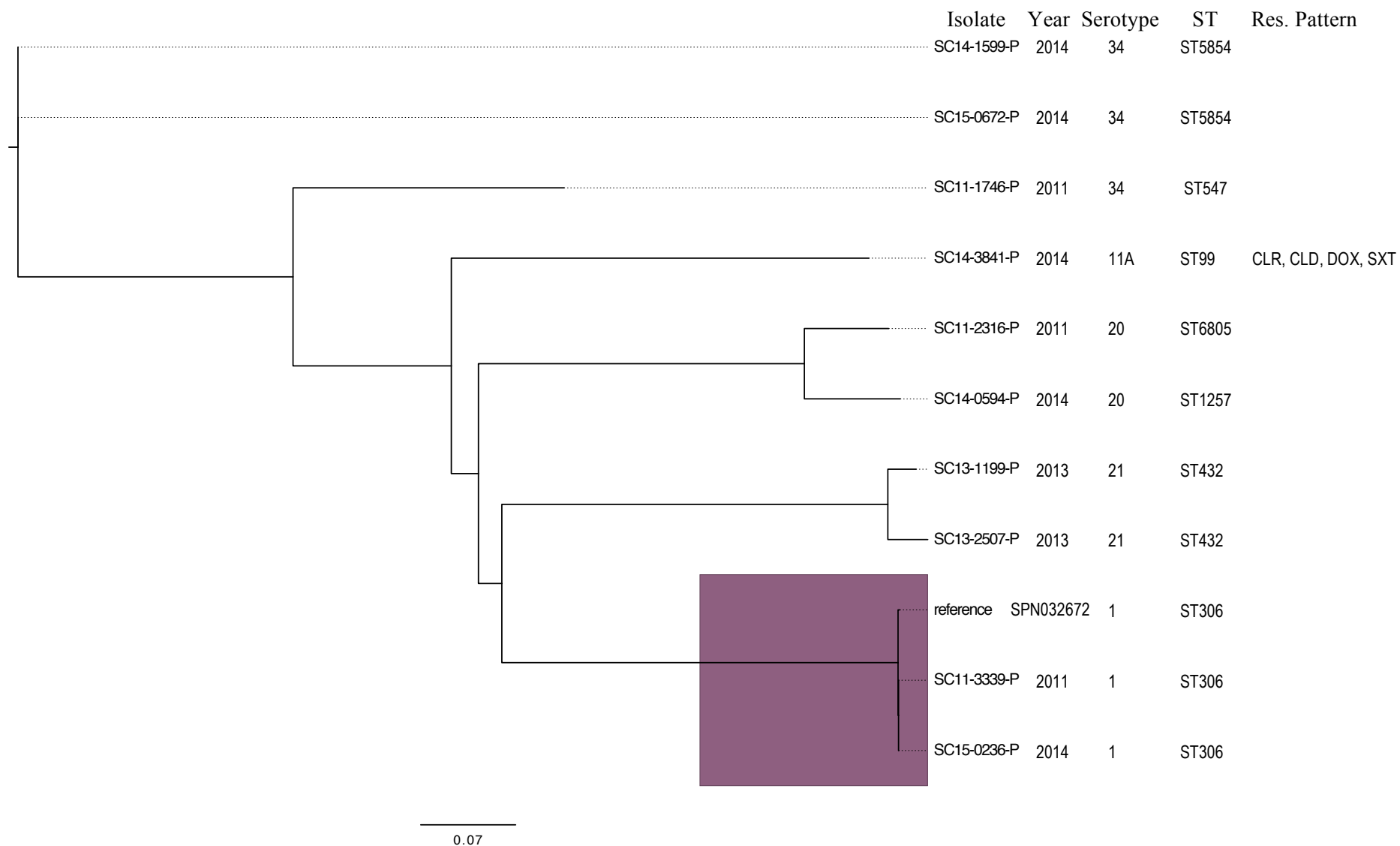
[a] Orange, Western Canada; Purple, Central Canada; Blue, Eastern Canada.

Interestingly, serotype 11A isolates fell in three different clusters by phylogenetic analysis. Three isolates were a part of cluster 5, demonstrating relatedness to serotype 8, 18C and 33F isolates (Figure 5.23). Despite all being the same serotype and typing as ST62, these isolates were not as closely related as other isolates in the same situation (for example, the two highlighted isolates in Figure 5.23, cluster 5-B1; they are of the same serotype and ST, but are virtually identical). In this case, the presence of resistance genes may have played a role in the analysis, as the isolate possessing *ermB* was the least related to the other isolates.

Two other serotype 11A isolates were highly unique by phylogenetic analysis. One, which had previously typed as ST10196 by MLST and had demonstrated levofloxacin and trimethoprim-sulfamethoxazole resistance fell in cluster 10 (Figure 5.15). This isolate possessed many chromosomal mutations in key resistance genes, including *parC*-Ser79Tyr, *gyrA*-Ser81Phe, *folA*-Ile100Leu and a Arg-Pro (RP) duplication in *folP*. Despite this, the isolate was most closely related to a serotype 19F isolate possessing no resistance to antimicrobials or resistance determinants and only three MLST alleles in common. The last serotype 11A isolate clustered into a group of background isolates (Figure 5.34). Along with this isolate, cluster 3 was composed of serotypes 1, 20, 21 and 34. The serotype 11A isolate possessed *ermB*, *tetM*, *folA*-Ile100Leu and an additional arginine in *folP*; as the isolates of other serotypes possessed no resistance determinants and were of very different STs, this isolate was not considered to be highly related to anything it grouped with in this phylogenetic analysis.

FIGURE 5.34: Whole genome core SNV maximum likelihood phylogenetic tree of *S. pneumoniae* "cluster 3" rooted on reference genome *S. pneumoniae* SPN032672 (NC_021003). Tree was generated using PhyML and visualized using FigTree. Coloured clusters were identified by ClusterPicker.

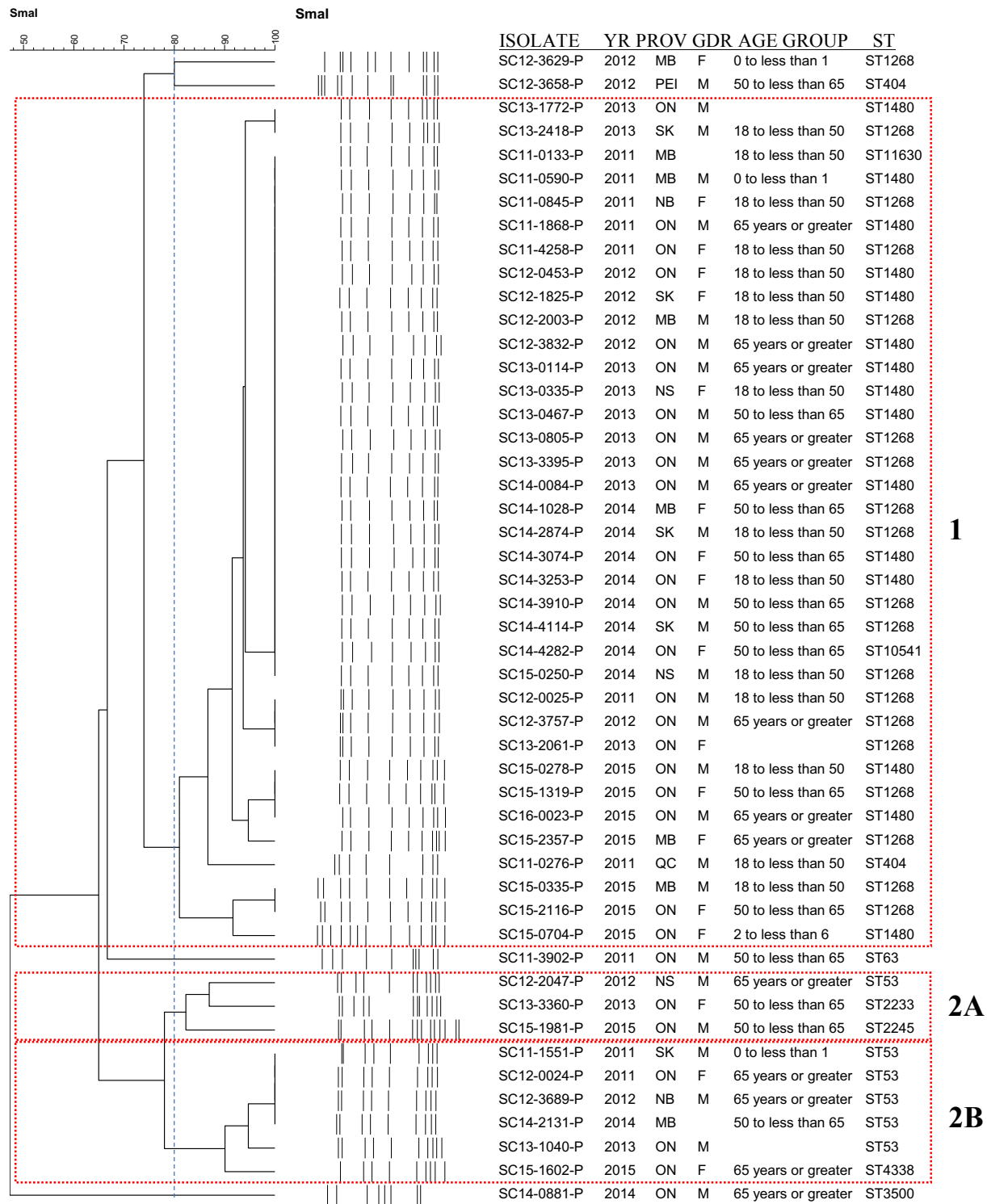| Isolate	Year	Serotype	ST	Res. Pattern
SC14-1599-P	2014	34	ST5854	
SC15-0672-P	2014	34	ST5854	
SC11-1746-P	2011	34	ST547	
SC14-3841-P	2014	11A	ST99	CLR, CLD, DOX, SXT
SC11-2316-P	2011	20	ST6805	
SC14-0594-P	2014	20	ST1257	
SC13-1199-P	2013	21	ST432	
SC13-2507-P	2013	21	ST432	
reference SPN032672		1	ST306	
SC11-3339-P	2011	1	ST306	
SC15-0236-P	2014	1	ST306	

0.07

5.4.4 Serotypes with Two Major Clones

These serotypes were distinguished from the other nine of interest because they demonstrated two distinct clusters of isolates by MLST: one related to an international clone and the other not.

5.4.4.1 Serotype 8

Serotype 8 isolates fell into two distinct clusters by PFGE (Figure 5.35). The first contained 36 isolates from 2011-2015 that demonstrated >80% relatedness. These isolates included the breadth of the demographic parameters and did not demonstrate any antimicrobial resistance. Though clusters 2A and 2B each demonstrated >80% relatedness separately, together they were approximately 77% related. One isolate that was resistant to doxycycline was found in cluster 2A, with the two other resistant isolates falling outside of the clusters; one MDR isolate, resistant to clarithromycin, clindamycin and doxycycline, fell just below cluster 1, while the second was the least related to the other isolates (<50%) and was resistant to chloramphenicol and doxycycline.

By MLST, the largest cluster of serotype 8 isolates was not related to an international clone (Figure 5.36). Less than 20% of isolates typed by MLST were part of the PMEN33 cluster (ST53), corresponding predominantly with clusters 2A and 2B from PFGE. Instead, most serotype 8 isolates (38/50) clustered around founder ST1480, a type that is six alleles different from ST53. This ST and its four variants comprised PFGE cluster 1. The lone MDR isolate typed as ST63, demonstrating the characteristic resistance pattern previously mentioned. The other two isolates from PFGE with noted resistance typed as obscure and largely unrelated STs 2233 and 3500.

FIGURE 5.35: PFGE dendrogram of serotype 8 *S. pneumoniae* isolates collected by the SAVE 2011-15 study. MLST sequence types are listed for comparison.

ISOLATE	YR	PROV	GDR	AGE GROUP	ST
SC12-3629-P	2012	MB	F	0 to less than 1	ST1268
SC12-3658-P	2012	PEI	M	50 to less than 65	ST404
SC13-1772-P	2013	ON	M		ST1480
SC13-2418-P	2013	SK	M	18 to less than 50	ST1268
SC11-0133-P	2011	MB		18 to less than 50	ST11630
SC11-0590-P	2011	MB	M	0 to less than 1	ST1480
SC11-0845-P	2011	NB	F	18 to less than 50	ST1268
SC11-1868-P	2011	ON	M	65 years or greater	ST1480
SC11-4258-P	2011	ON	F	18 to less than 50	ST1268
SC12-0453-P	2012	ON	F	18 to less than 50	ST1480
SC12-1825-P	2012	SK	F	18 to less than 50	ST1480
SC12-2003-P	2012	MB	M	18 to less than 50	ST1268
SC12-3832-P	2012	ON	M	65 years or greater	ST1480
SC13-0114-P	2013	ON	M	65 years or greater	ST1480
SC13-0335-P	2013	NS	F	18 to less than 50	ST1480
SC13-0467-P	2013	ON	M	50 to less than 65	ST1480
SC13-0805-P	2013	ON	M	65 years or greater	ST1268
SC13-3395-P	2013	ON	M	65 years or greater	ST1268
SC14-0084-P	2013	ON	M	65 years or greater	ST1480
SC14-1028-P	2014	MB	F	50 to less than 65	ST1268
SC14-2874-P	2014	SK	M	18 to less than 50	ST1268
SC14-3074-P	2014	ON	F	50 to less than 65	ST1480
SC14-3253-P	2014	ON	F	18 to less than 50	ST1480
SC14-3910-P	2014	ON	M	50 to less than 65	ST1268
SC14-4114-P	2014	SK	M	50 to less than 65	ST1268
SC14-4282-P	2014	ON	F	50 to less than 65	ST10541
SC15-0250-P	2014	NS	M	18 to less than 50	ST1268
SC12-0025-P	2011	ON	M	18 to less than 50	ST1268
SC12-3757-P	2012	ON	M	65 years or greater	ST1268
SC13-2061-P	2013	ON	F		ST1268
SC15-0278-P	2015	ON	M	18 to less than 50	ST1480
SC15-1319-P	2015	ON	F	50 to less than 65	ST1268
SC16-0023-P	2015	ON	M	65 years or greater	ST1480
SC15-2357-P	2015	MB	F	65 years or greater	ST1268
SC11-0276-P	2011	QC	M	18 to less than 50	ST404
SC15-0335-P	2015	MB	M	18 to less than 50	ST1268
SC15-2116-P	2015	ON	F	50 to less than 65	ST1268
SC15-0704-P	2015	ON	F	2 to less than 6	ST1480
SC11-3902-P	2011	ON	M	50 to less than 65	ST63
SC12-2047-P	2012	NS	M	65 years or greater	ST53
SC13-3360-P	2013	ON	F	50 to less than 65	ST2233
SC15-1981-P	2015	ON	M	50 to less than 65	ST2245
SC11-1551-P	2011	SK	M	0 to less than 1	ST53
SC12-0024-P	2011	ON	F	65 years or greater	ST53
SC12-3689-P	2012	NB	M	65 years or greater	ST53
SC14-2131-P	2014	MB		50 to less than 65	ST53
SC13-1040-P	2013	ON	M		ST53
SC15-1602-P	2015	ON	F	65 years or greater	ST4338
SC14-0881-P	2014	ON	M	65 years or greater	ST3500

GDR, YR, PROV: gender, year and province of isolation. Gaps in the demographics indicate that the value was not given at the time of collection. The dashed line indicates 80% similarity. Clusters of isolates are outlined with red boxes and labelled 1-2B.

FIGURE 5.36: Minimum spanning tree (generated by PHYLOViZ 2.0) of MLST sequence types demonstrated by serotype 8 isolates, coloured according to province of isolation[a]. Green outlines indicate a group founder; numbers indicate the number of differences between the MLST profiles of the two connected nodes (≤ 2 indicates the two nodes are part of a cluster). PMEN clones are marked with red circles.
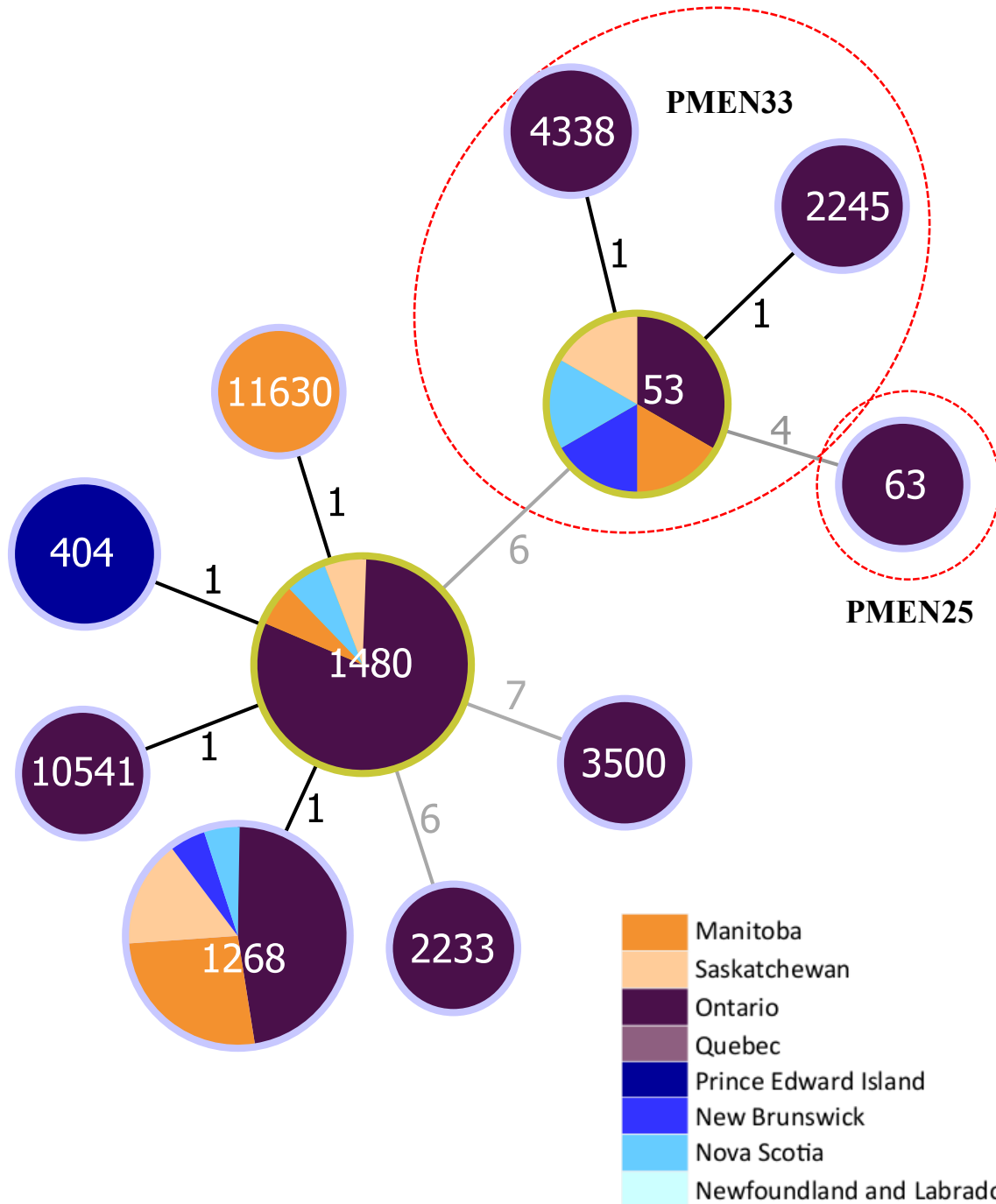
[a] Orange, Western Canada; Purple, Central Canada; Blue, Eastern Canada.

Only four serotype 8 isolates were included in the phylogenetic analysis; none had a type closely related to the PMEN33 clone, however two were related to the more common ST1480. One was used as the internal reference for this cluster, as it had the best contig assembly stats of the isolates contained in cluster 11 (Figure 5.31); however, these isolates did cluster with close relatedness in the maximum likelihood tree, with the small differences potentially due to the minor variation in ST. Interestingly, the most closely related serotypes to this serotype 8 clone were 38 and 22F. A third serotype 8 isolate clustered closely with serotype 11A in cluster 5 (Figure 5.23); this isolate, though demonstrating a ST not yet assigned at the time of writing, shared four alleles in common with PMEN33. As depicted in the original MLST spanning tree (Figure 5.5), ST53 and serotype 11A clone ST62 are DLVs. This indicates that if ST53 isolates had been included in the analysis, they would likely have clustered with serotype 11A.

The ST63 isolate described in the above PFGE/MLST analysis was also included; it clustered accordingly in the phylogenetic analysis (Figure 5.8). As with the other isolates that appear to be the result of a capsular switch event with serotype 15A-ST63, this isolate also possessed *ermB* and *tetM* resistance determinants.

5.4.4.2 Serotype 12F

By PFGE, Serotype 12F isolates clustered according to their resistance pattern (Figure 5.37). Cluster 2, comprising 33 isolates, contained strains from across the country and study period that were resistant to clarithromycin. Clusters 1 and 3 contained susceptible isolates that clustered by region of collection; cluster 1 isolates were obtained from Western Canada, while cluster 3 isolates were from Central Canada. Three MDR isolates fell outside of the clusters; two with resistance to clarithromycin, clindamycin and doxycycline were located above cluster 1, while the last was approximately 55% related to the other serotype 12Fs and was resistant to chloramphenicol, clarithromycin, doxycycline and trimethoprim-sulfamethoxazole.

The largest serotype 12F cluster by MLST was founded on ST218, a PMEN clone from Denmark (PMEN34) (Figure 5.38). Thirty-three of 46 ST218 and related isolates were resistant to clarithromycin; however, regardless of resistance, all ST218 isolates were found within PFGE clusters 1-3. In sharp contrast, the other cluster of two STs shared no MLST alleles in common with other serotype 12F isolates and contained MDR isolates with resistance to clarithromycin, clindamycin and doxycycline. The last two isolates were also highly unrelated; one typed as ST989 and matched to the least related isolate by PFGE that demonstrated resistance to four antimicrobials. The last typed as ST66 and was therefore a SLV of PMEN18. This ST is often attributed to serotype 9N, making this serotype 12F isolate the likely product of a capsular switch.

MLST was unable to segregate ST218 isolates into those that were clarithromycin-resistant and those that were not. Also, those that were not resistant were divided by region of isolation; this indicates that PFGE was more discriminatory for this serotype, both in identifying resistant isolates as well as regional differences within the same ST.

FIGURE 5.37: PFGE dendrogram of serotype 12F *S. pneumoniae* isolates collected by the SAVE 2011-15 study. MLST sequence types are listed for comparison**.**

ISOLATE	YR	PROV	GDR	AGE GROUP	ST
SC12-1227-P	2012	SK	F	50 to less than 65	ST10198
SC13-3603-P	2013	NB	M	65 years or greater	ST6945
SC12-0403-P	2012	MB	F	18 to less than 50	ST218
SC13-2620-P	2013	SK	M	18 to less than 50	ST218
SC13-2525-P	2013	MB		18 to less than 50	ST218
SC15-3802-P	2015	SK	M	18 to less than 50	ST218
SC14-3785-P	2014	SK	M	18 to less than 50	ST5018
SC14-4117-P	2014	SK	M	1 to less than 2	ST218
SC15-0200-P	2014	SK	F	18 to less than 50	ST218
SC15-1742-P	2015	SK	F	50 to less than 65	ST218
SC13-4359-P	2013	MB	F	50 to less than 65	ST218
SC13-4591-P	2013	MB	M	6 to less than 18	ST218
SC14-3818-P	2014	MB	M	18 to less than 50	ST218
SC11-0831-P	2011	SK	M	18 to less than 50	ST218
SC11-1898-P	2011	MB	F	18 to less than 50	ST218
SC11-2390-P	2011	MB	M	50 to less than 65	ST218
SC11-3577-P	2011	MB	M	6 to less than 18	ST218
SC11-4340-P	2011	MB	F	18 to less than 50	ST218
SC12-3218-P	2012	SK	M	65 years or greater	ST218
SC12-3367-P	2012	MB	F	50 to less than 65	ST218
SC12-3500-P	2012	MB	F	50 to less than 65	ST218
SC12-4009-P	2012	MB	F	50 to less than 65	ST218
SC13-0073-P	2012	MB	F	50 to less than 65	ST218
SC13-2098-P	2013	SK	M	18 to less than 50	ST218
SC13-2283-P	2013	ON	M	50 to less than 65	ST218
SC13-2617-P	2013	SK	F	18 to less than 50	ST218
SC13-3406-P	2013	SK	F	65 years or greater	ST218
SC14-0218-P	2014	ON	F	50 to less than 65	ST218
SC14-0634-P	2014	MB	F	65 years or greater	ST218
SC14-0915-P	2014	MB	F	18 to less than 50	ST218
SC14-1981-P	2014	SK	F	50 to less than 65	ST218
SC14-2942-P	2014	MB	M	50 to less than 65	ST218
SC12-3268-P	2012	MB	F	65 years or greater	ST218
SC13-2468-P	2013	MB	M	50 to less than 65	ST218
SC15-1451-P	2015	SK	M	18 to less than 50	ST13194
SC15-1652-P	2015	MB	F	18 to less than 50	ST218
SC15-1238-P	2015	MB	M	18 to less than 50	ST218
SC11-3938-P	2011	ON		18 to less than 50	ST218
SC15-2233-P	2015	SK	M	18 to less than 50	ST218
SC12-3710-P	2012	ON	M	18 to less than 50	ST10197
SC12-3822-P	2012	MB	F	65 years or greater	ST218
SC15-2703-P	2015	MB	F	65 years or greater	ST218
SC15-3668-P	2015	MB	F	18 to less than 50	ST218
SC14-3637-P	2014	ON	F	18 to less than 50	ST218
SC15-4226-P	2015	QC	F	65 years or greater	ST218
SC11-4273-P	2011	QC	F	18 to less than 50	ST218
SC11-4281-P	2011	QC	M	18 to less than 50	ST218
SC11-2780-P	2011	ON	M	50 to less than 65	ST66
SC11-1483-P	2011	MB	M	18 to less than 50	ST218
SC15-0461-P	2015	ON		50 to less than 65	ST989

GDR, YR, PROV: gender, year and province of isolation. Gaps in the demographics indicate that the value was not given at the time of collection. The dashed line indicates 80% similarity. Clusters of isolates are outlined with red boxes and labelled 1-3.

139

FIGURE 5.38: Minimum spanning tree (generated by PHYLOViZ 2.0) of MLST sequence types demonstrated by serotype 12F isolates, coloured according to province of isolation[a]. Green outlines indicate a group founder; numbers indicate the number of differences between the MLST profiles of the two connected nodes (≤ 2 indicates the two nodes are part of a cluster). PMEN clones are marked with red circles.
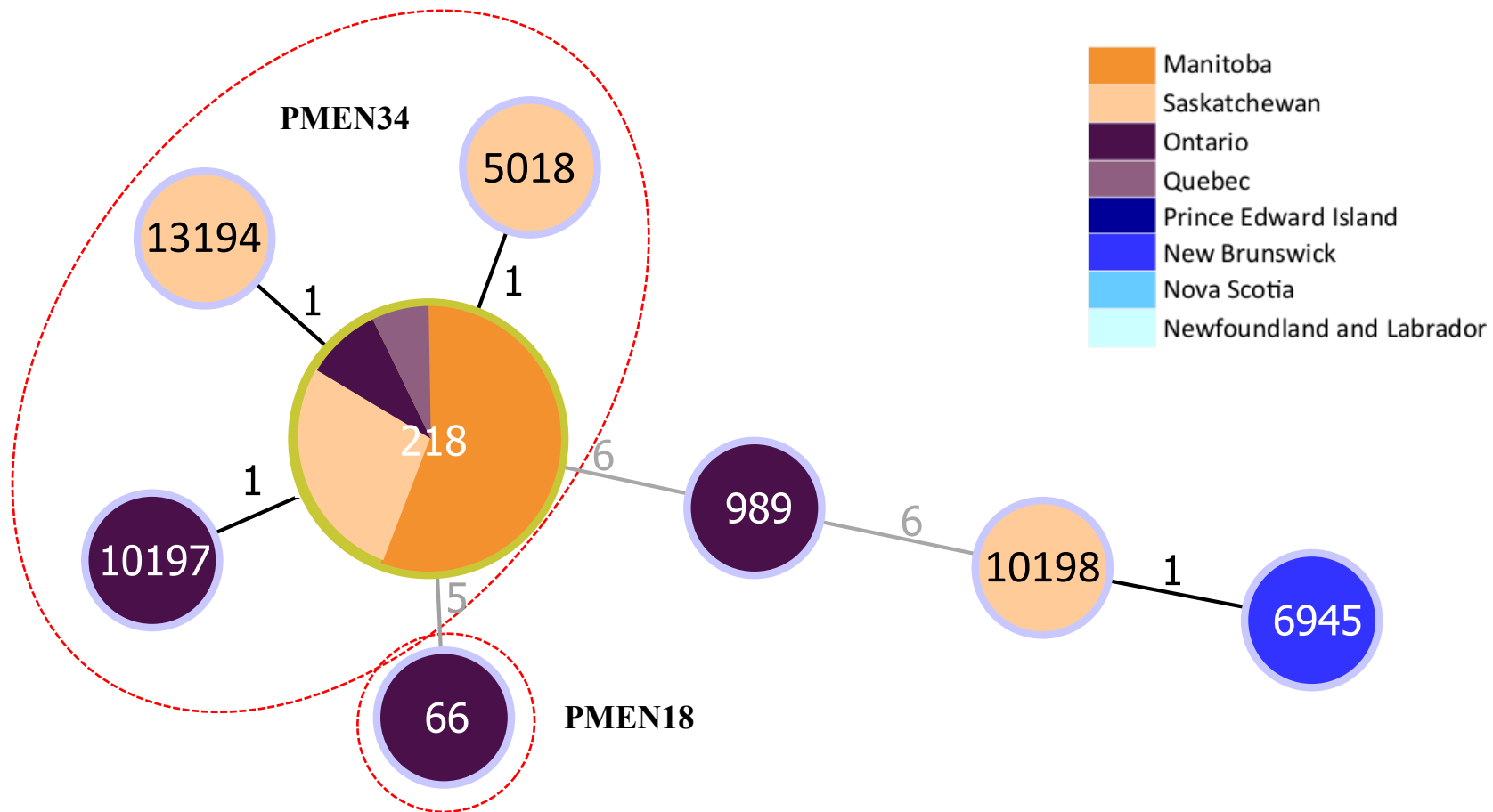
[a] Orange, Western Canada; Purple, Central Canada; Blue, Eastern Canada.

Phylogenetic analysis was able to highlight the genetic differences between the two

clusters of serotype 12F isolates. ST218 isolates fell into cluster 9 of the detailed phylogenetic

analyses (Figure 5.39) and were found to be most closely related to a series of serotype 23A and

23F isolates related to PMEN4. The two ST218 isolates were virtually identical in this analysis,

despite one possessing *mefA* and thus clarithromycin resistance. The two MDR isolates,

previously described above, were only distantly related to the others, clustering near several

background serotypes, including 31, 10A and 35F (Figure 5.40). Interestingly, unlike the ST218

isolate that possessed *mefA*, these MDR serotype 12F isolates instead possessed *ermB* (in

combination with *tetM*), thus providing higher level macrolide resistance.

FIGURE 5.39: Whole genome core SNV maximum likelihood phylogenetic tree of *S. pneumoniae* "cluster 9" rooted on reference genome PMEN4 (ERR1022115). Tree was generated using PhyML and visualized using FigTree. Coloured clusters were identified by ClusterPicker.

Isolate	Year	Serotype	ST	Res. Pattern
SC11-2116-P	2011	12F	218	
SC12-3268-P	2012	12F	218	CLR
SC11-4141-P	2011	23A	190	
SC12-4032-P	2011	23F	36	
SC12-3711-P	2012	23A	42	CLR, CLD
SC16-2498-P	2012	23A	42	
SC12-0005-P	2013	23F	36	CLR, LEV
SC13-1815-P	2007	23A	1839	
SC16-2496-P	2007	23F	36	
reference	PMEN4	23F	37	

Denmark12F-34

Tennessee23F-4

0.2
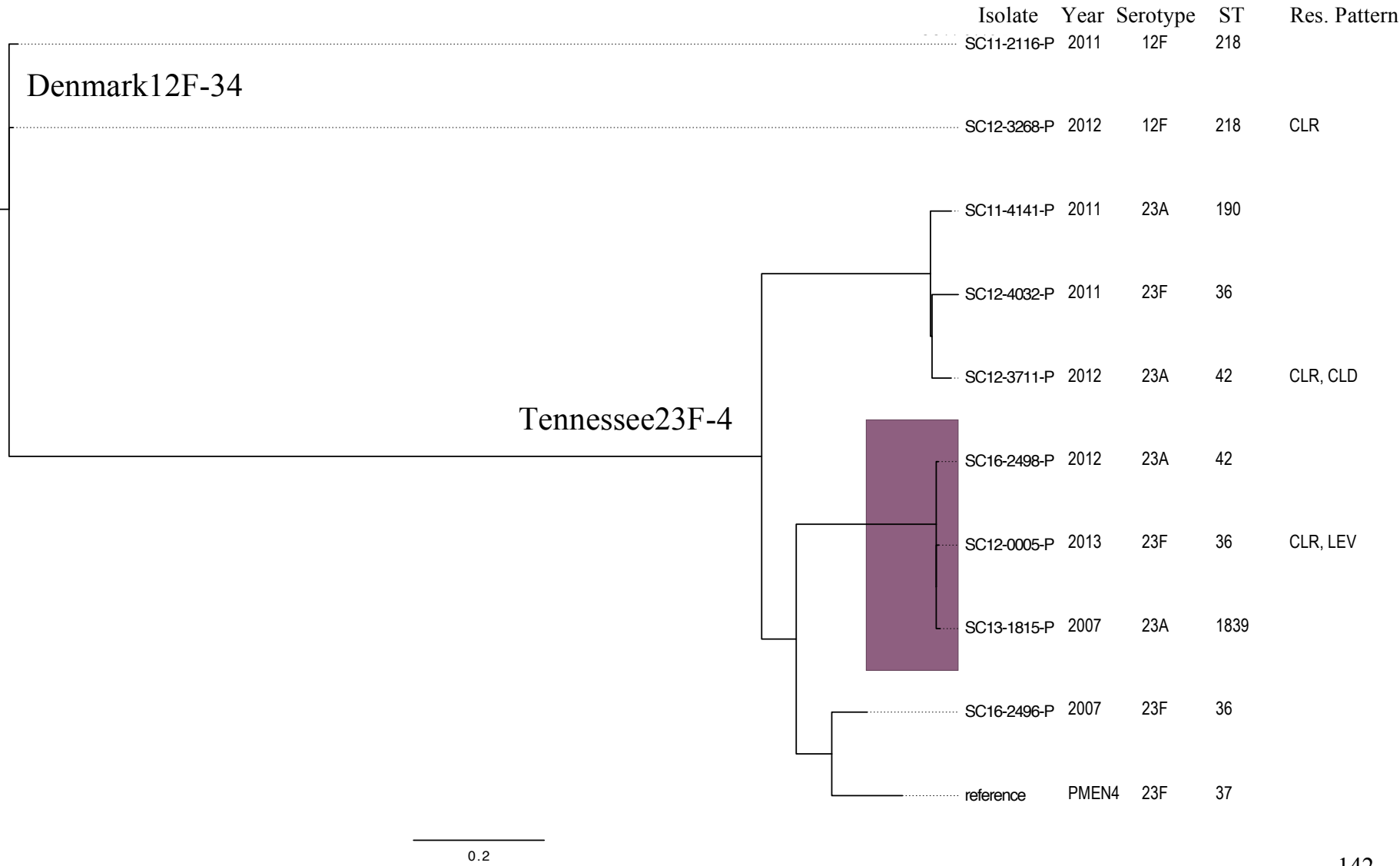
FIGURE 5.40: Whole genome core SNV maximum likelihood phylogenetic tree of *S. pneumoniae* "cluster 14" rooted on reference genome *S. pneumoniae* R6 (NC_003098). Tree was generated using PhyML and visualized using FigTree. Coloured clusters were identified by ClusterPicker.

Isolate	Year	Serotype	ST	Res. Pattern
SC12-0688-P	2011	17F	ST123	
SC12-1032-P	2012	17F	ST123	CLD
reference	R6	2	ST595	
SC14-1645-P	2014	10A	ST97	
SC16-2501-P	2007	10A	ST585	
SC12-2045-P	2012	35F	ST498	
SC11-4375-P	2011	35F	ST446	
SC12-1277-P	2012	35F	ST13189	
SC12-1227-P	2012	12F	ST10198	CLR, CLD, DOX
SC13-3603-P	2013	12F	ST6945	CLR, CLD, DOX
SC11-1866-P	2011	31	ST13187	
SC14-3643-P	2014	31	ST1766	

0.06
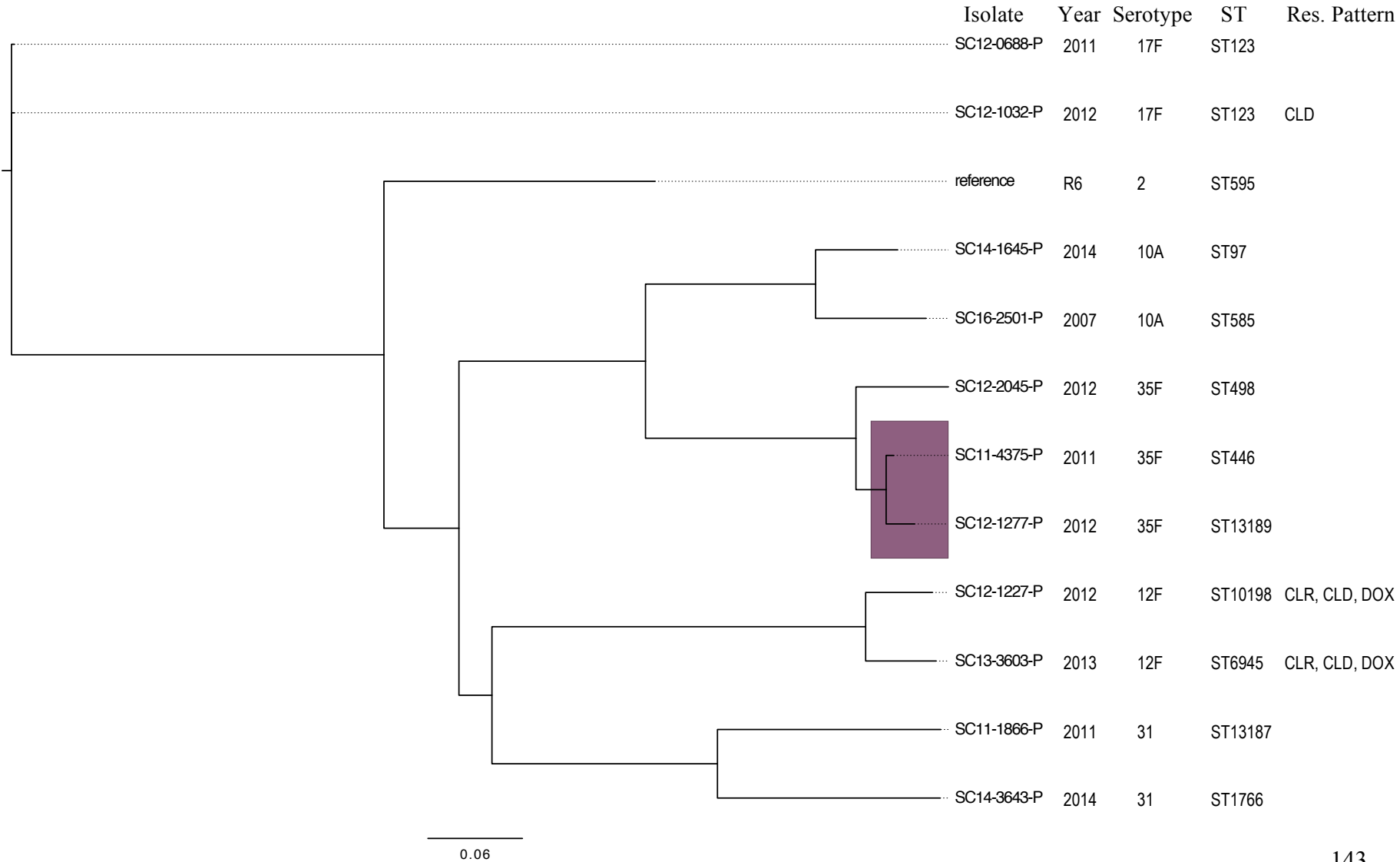
5.4.4.3 Serotype 35B

Serotype 35B isolates fell into two major clusters by PFGE (Figure 5.41). Cluster 1 contained 16 isolates with >80% relatedness that were predominantly collected during the early years of the SAVE study. Apart from two isolates resistant to levofloxacin, cluster 1 isolates were susceptible to antimicrobials. Clusters 2A and 2B each demonstrated >80% relatedness separately, however together they were approximately 77% related. As a whole, cluster 2 contained those isolates that were penicillin-resistant or demonstrated intermediate resistance to penicillin; some exhibited additional resistance to clarithromycin or trimethoprim-sulfamethoxazole. Two MDR isolates with a resistance pattern of clarithromycin, penicillin and trimethoprim-sulfamethoxazole fell between the clusters, demonstrating slightly more relatedness to cluster 1 than cluster 2.

Isolates clustered similarly by MLST (Figure 5.42), with most typing as ST452 or a variant. ST452 isolates largely made up PFGE cluster 1, as well as related ST9347 and unrelated ST9351. Isolates from PFGE cluster 2 were variants of PMEN24, a clone known for penicillin resistance. Only one of the MDR isolates was typed by MLST and was determined to be ST9346. This is a SLV of commonly antimicrobial-resistant clone PMEN3 (associated with serotype 9V).

FIGURE 5.41: PFGE dendrogram of serotype 35B *S. pneumoniae* isolates collected by the SAVE 2011-15 study. MLST sequence types are listed for comparison.

ISOLATE	YR	PROV	GDR	AGE GROUP	ST
SC12-0149-P	2012	ON	F	65 years or greater	ST452
SC15-1322-P	2015	ON		65 years or greater	
SC12-2016-P	2012	NS	F	50 to less than 65	ST452
SC11-1537-P	2011	ON	F	50 to less than 65	ST452
SC11-2132-P	2011	QC	M	50 to less than 65	ST452
SC13-1647-P	2013	NB	M	65 years or greater	ST452
SC13-2506-P	2013	QC	F	65 years or greater	ST452
SC11-0592-P	2011	MB		65 years or greater	ST452
SC15-2246-P	2015	NB	F	65 years or greater	
SC13-1742-P	2013	QC	M	18 to less than 50	ST452
SC11-4650-P	2011	MB	F	65 years or greater	ST9351
SC13-2032-P	2013	MB	M		ST452
SC13-2524-P	2013	MB		65 years or greater	ST452
SC15-0266-P	2015	ON	M	1 to less than 2	
SC11-1398-P	2011	ON	M	65 years or greater	ST452
SC11-2107-P	2011	QC	F	65 years or greater	ST9347
SC13-4121-P	2013	QC	M	50 to less than 65	ST452
SC13-3145-P	2013	QC	M	65 years or greater	ST452
SC11-1841-P	2011	ON	F	65 years or greater	ST9346
SC15-3778-P	2015	ON	F	18 to less than 50	
SC15-3181-P	2015	SK	M	65 years or greater	
SC15-3185-P	2015	SK	M	65 years or greater	
SC15-2187-P	2015	ON	F	18 to less than 50	
SC15-0983-P	2015	QC	M	50 to less than 65	
SC13-3293-P	2013	ON	F	65 years or greater	ST558
SC13-3362-P	2013	ON	F	50 to less than 65	ST558
SC11-1824-P	2011	NB	M	65 years or greater	ST1666
SC12-0841-P	2012	SK	M	6 to less than 18	ST558
SC13-2353-P	2013	ON	F	65 years or greater	ST558
SC12-1751-P	2012	QC	F	65 years or greater	ST558
SC11-3593-P	2011	QC		6 to less than 18	ST558
SC15-0471-P	2015	ON	M	18 to less than 50	
SC13-0564-P	2013	ON	M	65 years or greater	ST452
SC13-0820-P	2013	SK	F	50 to less than 65	ST452

GDR, YR, PROV: gender, year and province of isolation. Gaps in the demographics indicate that the value was not given at the time of collection. The dashed line indicates 80% similarity. Clusters of isolates are outlined with red boxes and labelled 1-2B.

145

FIGURE 5.42: Minimum spanning tree (generated by PHYLOViZ 2.0) of MLST sequence types demonstrated by serotype 35B isolates, coloured according to province of isolation[a]. Green outlines indicate a group founder; numbers indicate the number of differences between the MLST profiles of the two connected nodes (≤ 2 indicates the two nodes are part of a cluster). PMEN clones are marked with red circles.
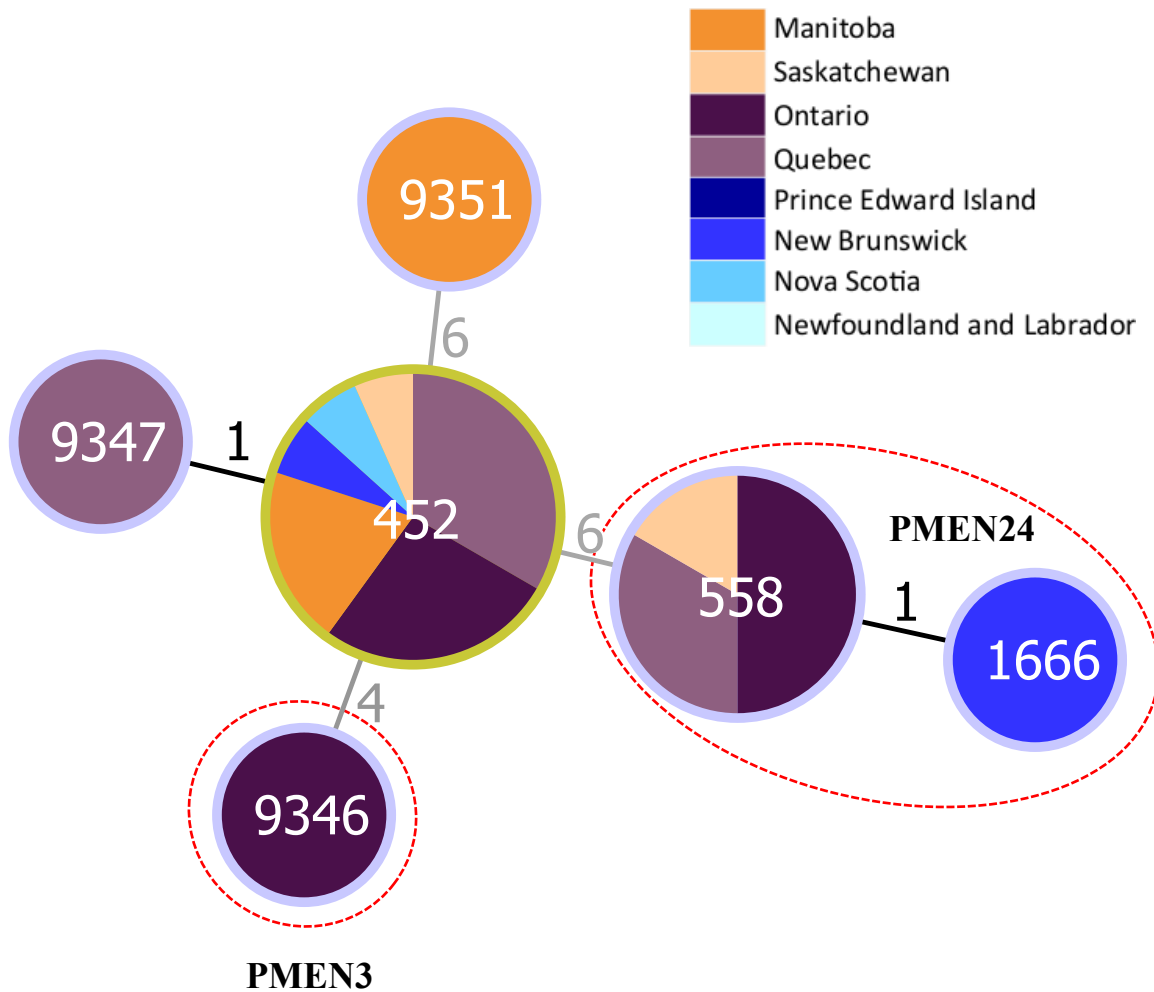
[a] Orange, Western Canada; Purple, Central Canada; Blue, Eastern Canada.

Seven serotype 35B isolates were included in the phylogenetic analysis. One was a singleton and therefore not included in further analysis. Five of these isolates, contained in cluster 2 (Figure 5.43), were highly related to each other and PMEN24. Three were resistant to clarithromycin due to the presence of *mefA*; one also demonstrated levofloxacin resistance conferred by a lone Ser81Phe mutation in *gyrA*, while another was trimethoprim-sulfamethoxazole resistant due to *folA*-Ile100Leu and a Ser-Tyr (SY) duplication in *folP*. Two of five isolates were resistant to penicillin, with the rest demonstrating intermediate resistance. Despite this, all but one of the isolates demonstrated mutations in all three PBP genes (Table 5.3). Interestingly, one serotype 29 background isolate also fell into this group; it was a DLV of PMEN24 and similarly possessed the same PBP alterations and *mefA*-mediated clarithromycin resistance. As mentioned previously, serotype 35B is one of few serotypes in this study to exhibit the Thr371Ser mutation in the STMK motif of *pbp1A*. The serotype 29 isolate in this group also possessed this uncommon mutation, which strengthens the case that this isolate is a capsular switch variant.

An additional serotype 35B isolate was likely a capsular switch variant generated by recombination with a serotype 9V related to PMEN3 (ST156). As depicted in Figure 5.9, this isolate clustered closely with serotype 9Vs (and other putative switch variants) of related ST and possessing a similar resistance pattern. Isolates in this cluster demonstrated similar resistance mechanisms, including clarithromycin resistance due to *mefA* and trimethoprim-sulfamethoxazole resistance due to *folA*-Ile100Leu and a specific arginine insertion in *folP*. Though all were penicillin-resistant or –intermediate and demonstrated alterations in all PBPs, the serotype 35B progeny retained the unique Thr371Ser mutation in the STMK motif of *pbp1A*, while the serotype 9V putative recipients demonstrated Thr371Ala. Interestingly, while the

original MLST typed this isolate as ST9346, a SLV of ST156, *in silico* MLST identified it as identical to ST156.

FIGURE 5.43: Whole genome core SNV maximum likelihood phylogenetic tree of *S. pneumoniae* "cluster 2" rooted on reference genome PMEN24 (ERR1022111). Tree was generated using PhyML and visualized using FigTree.

Isolate	Year	Serotype	ST	Res. Pattern
SC11-3593-P	2011	35B	ST558	PEN
SC12-1751-P	2012	35B	ST558	CLR, SXT
SC11-1824-P	2011	35B	ST1666	
SC12-1906-P	2012	35B	ST558	CLR, PEN
SC14-1327-P	2014	29	ST7486	CLR
SC14-3062-P	2014	35B	ST558	CLR, LEV
reference	PMEN24	35B	ST377	

0.05

5.5 Analysis of Capsular Switch Variants

Putative donor, recipient and progeny strains were deduced using MLST and WGS data, as well as PBP transpeptidase domain typing. It is important to note that the donor and recipient strains in each scenario are not the *exact* strains that participated in recombination to produce the progeny strain; true donor and recipient strains are unknown, but they are likely genetically identical or similar to those used in this study.

5.5.1 Vaccine Escape Recombinants

5.5.1.1 9V → 35B

Most serotype 35B isolates in this study were related to PMEN24 (ST558); however, one was a variant of PMEN3 (9V-ST156) and therefore suspected of being a capsular switch variant. The relationship between these isolates was strengthened by their many identical antimicrobial resistance determinants and mutations, as noted in Section 5.4.4.3. The isolate in question, SC11-1841-P, was originally typed as a SLV (ST9346) by MLST, and subsequently typed as ST156 by *in silico* MLST. An examination of the *spi* genes of the isolates in question revealed no SNV difference between SC11-1841-P and ST156 reference strains; accordingly, the isolate was assumed to be ST156 for this part of the analysis. As visualized in Figure 5.44, the PBP-typing profiles were 1A-4, 2B-7/16/79, 2X-7 for putative serotype 35B donors, 1A-15, 2B-12, 2X-18 for serotype 9V-ST156 recipients and 1A-4, 2B-12, 2X-7 for the serotype 35B-ST156 progeny. This indicates that *pbp1A* and *pbp2X* were likely transferred from donor to progeny along with the capsule during the recombination event. Analysis of the SNVs between putative donor, recipient and progeny strains indicated that the recombination fragment was approximately 35.7kb relative to the reference (*S. pneumoniae* R6), spanning from hypothetical protein spr0300 upstream of *pbp2X* to hypothetical protein spr0332 downstream of *pbp1A*.

FIGURE 5.44: Visualization of a vaccine escape recombination event involving a non-vaccine donor (35B) and PCV-7/13 recipient (9V)[a]. Identification was performed using a penicillin-binding protein transpeptidase domain typing scheme outlined by Metcalf *et al* (38). Similarly to MLST, PBPs were assigned arbitrary allele numbers based on individual nucleotide differences from the penicillin-susceptible NCBI reference strain TIGR4 (NC_003028). The donor strain has a red capsule and black genes; the recipient strain has a purple capsule and blue genes; progeny strain is a combination of both.
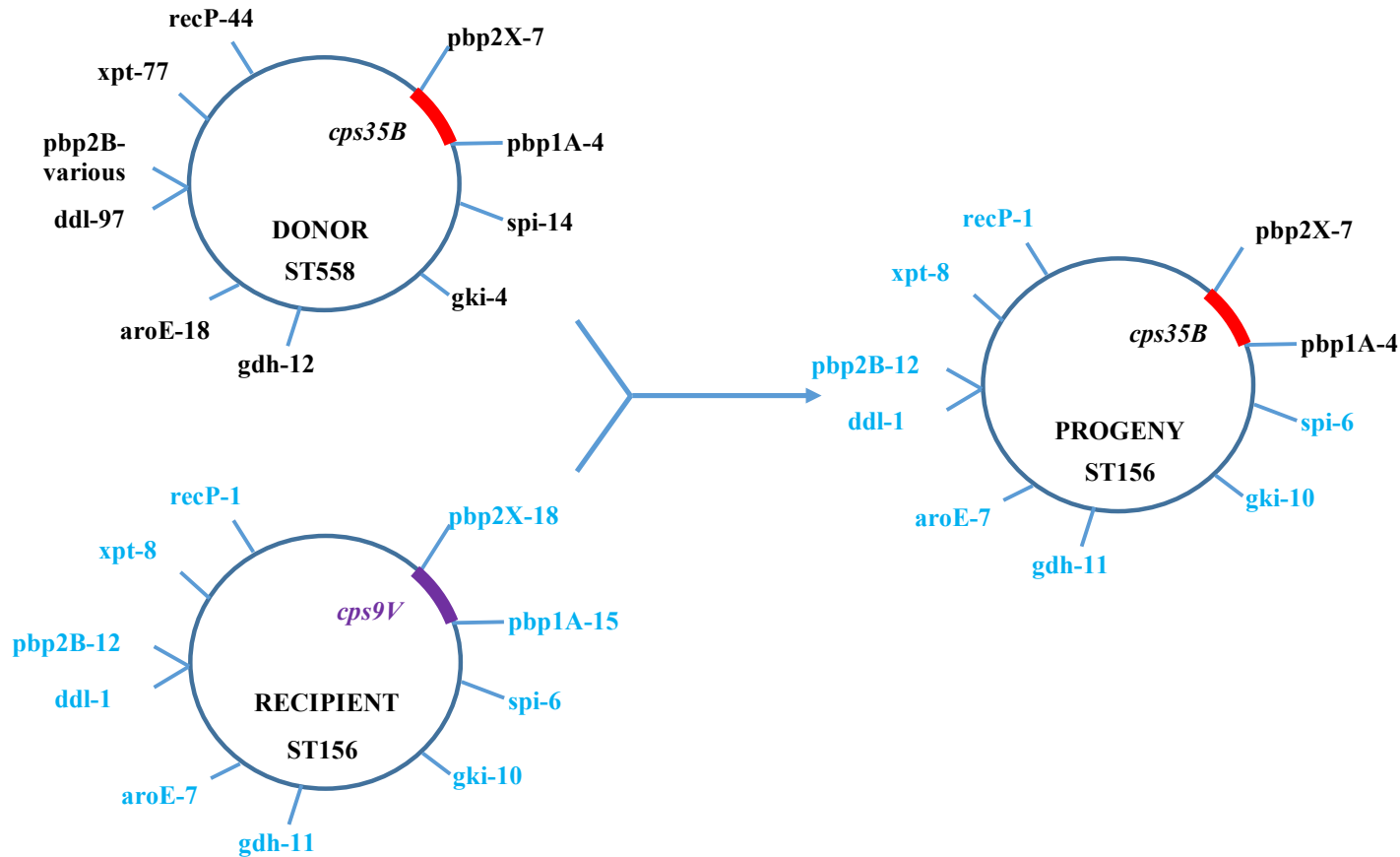
[a] Donor, SC11-3593-P; Recipient, SC11-0099-P; Progeny, SC11-1841-P; difference of 510 SNVs between recipient and progeny.

151

5.5.1.2 23F → 6C, 19A, 15B

Three isolates of very different serotype were related to serotype 23F in the phylogenetic

analysis; the first was a serotype 6C identical to PMEN26 (serotype 23F-ST338). Isolated in late

2013, SC13-4858-P demonstrated a PBP profile of 1A-19, 2B-1, 2X-1. As there was no serotype

23F isolate typed in this study that demonstrated the correct ST, the PMEN26 reference strain

was used (NCBI ERR1022113) as a recipient strain. The recipient had a PBP-typing profile of

1A-12, 2B-1, 2X-1, possessing two alleles in common with the progeny. The only serotype 6C

isolate with the appropriate PBP profile to be a donor was ST1379, with a typing pattern of 1A-

19, 2B-31, 2X-8. This indicates that *pbp1A* was transferred with the capsule during

recombination, while *pbp2X* was retained from the recipient (Figure 5.45). Analysis of the SNVs

between putative donor, recipient and progeny strains indicated that the recombination fragment

was approximately 32.5kb relative to the reference (*S. pneumoniae* R6), spanning from

hypothetical protein spr0309, just upstream of the capsule, to *cbpF*, located downstream of

pbp1A.

The second and third isolates related to serotype 23F were one each of serotypes 19A

(SC11-0902-P) and 15B (SC11-2628-P), which were identical and related to PMEN1 (serotype

23F-ST81), respectively. Like PMEN1, SC11-0902-P and SC11-2628-P were XDR,

demonstrating the presence of *ermB*, *tetM*, *cat*, *folA*-Ile100Leu and serine insertion in *folP*; the

main difference between the two was their ST and PBP-typing profiles. As depicted in Figure

5.46, SC11-0902-P was identical to PMEN1 and had the profile 1A-13, 2B-12, 2X-16. The

NCBI reference *S. pneumoniae* ATCC 700669 (NCBI NC_011900) was used as the putative

recipient strain for this event, as it matched in ST and had a PBP profile of 1A-15, 2B-12, 2X-18.

A serotype 19A-ST320 was chosen as the donor strain, as its *pbp1A* and *pbp2X* matched the

progeny, indicating both were transferred during recombination. Analysis of the SNVs between putative donor, recipient and progeny strains indicated that the recombination fragment was extremely large, approximately 70kb relative to the reference (*S. pneumoniae* R6), spanning from *sulD* upstream of the capsule, to *gnd* downstream.

The second isolate related to PMEN1, SC11-2628-P, demonstrated a slightly different PBP typing pattern of 1A-70, 2B-12, 2X-18 (Figure 5.47). Comparison of the progeny pattern to that of the ATCC 700669 recipient above (1A-15, 2B-12, 2X-18) indicated that *pbp2X* was not a part of the recombination fragment; detailed analysis of *pbp1A* revealed that 1A-15 and 1A-70 differed by only one nucleotide change. A G→A transition in 1A-15 resulted in a Gly→Ser change that created 1A-70. For this reason, the difference between the *pbp1A* genes was considered negligible; therefore, because both binding proteins were retained from the recipient, a common serotype 15B-ST199 was chosen as the donor strain. Analysis of the SNVs between putative donor, recipient and progeny strains indicated that the recombination fragment was approximately 19.2kb relative to the reference (*S. pneumoniae* R6), spanning from just outside the capsule locus in *dexB*, to hypothetical protein spr0328, located just upstream of *pbp1A*. Interestingly, the *spi* locus did not demonstrate any relatedness between donor, recipient or progeny. Comparison of the progeny *spi*-6 allele with the donor and recipient yielded over 15 different SNVs, suggesting that the progeny *spi* gene may have been obtained through recombination with another, unknown, isolate.

FIGURE 5.45: Visualization of a vaccine escape recombination event involving a non-vaccine donor (6C) and PCV-7/13 recipient (23F)[a]. Identification was performed using a penicillin-binding protein transpeptidase domain typing scheme outlined by Metcalf *et al* (38). Similarly to MLST, PBPs were assigned arbitrary allele numbers based on individual nucleotide differences from the penicillin-susceptible NCBI reference strain TIGR4 (NC_003028). The donor strain has a red capsule and black genes; the recipient strain has a purple capsule and blue genes; progeny strain is a combination of both.
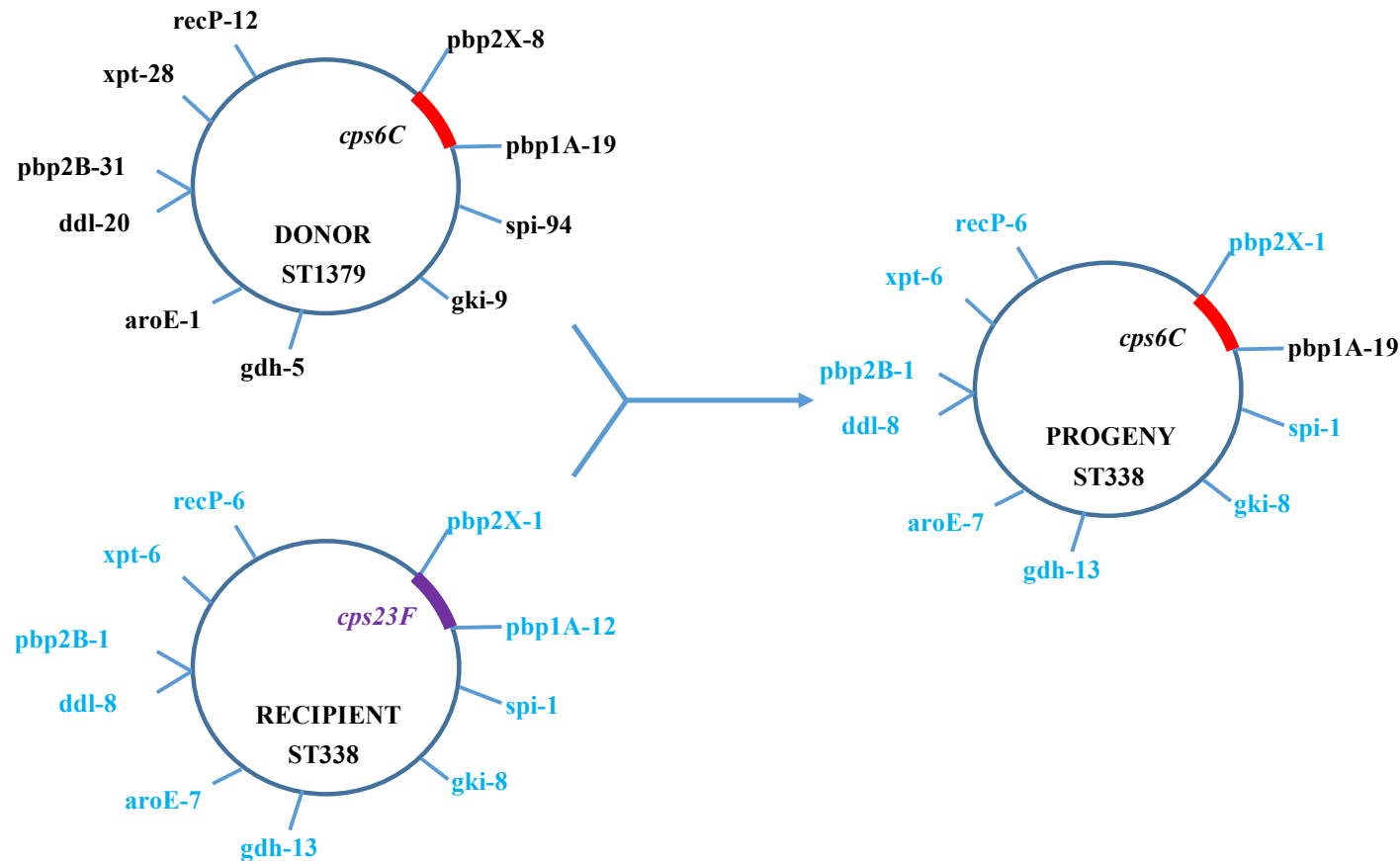
[a]Donor, SC13-0990-P; Recipient, PMEN26 (NCBI ERR1022113); Progeny, SC13-4858-P.

154

FIGURE 5.46: Visualization of a vaccine escape recombination event involving a PCV-13 donor (19A) and PCV-7 recipient (23F)[a]. Identification was performed using a penicillin-binding protein transpeptidase domain typing scheme outlined by Metcalf *et al* (38). Similarly to MLST, PBPs were assigned arbitrary allele numbers based on individual nucleotide differences from the penicillin-susceptible NCBI reference strain TIGR4 (NC_003028). The donor strain has a red capsule and black genes; the recipient strain has a purple capsule and blue genes; progeny strain is a combination of both.
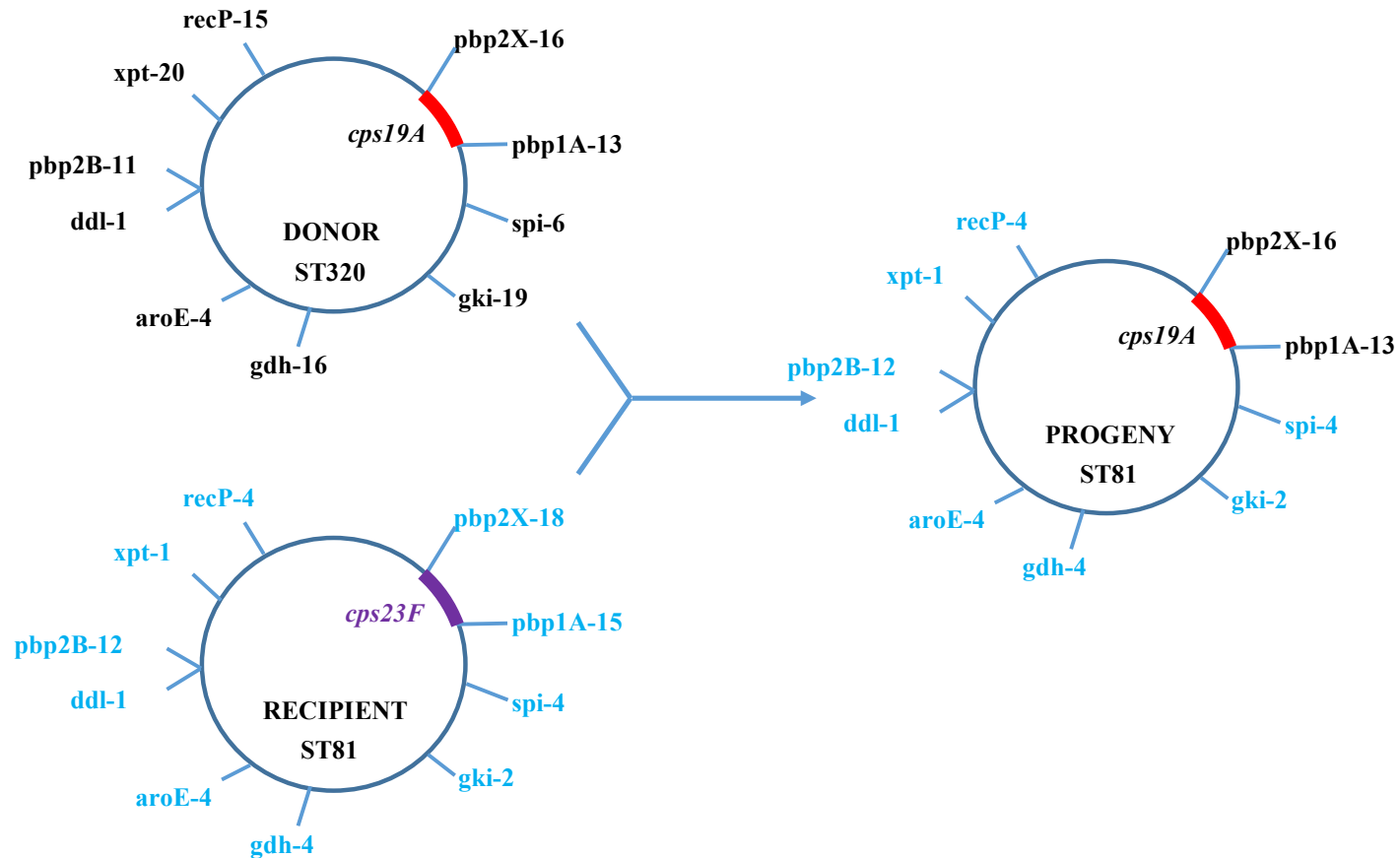
[a]Donor, SC11-0900-P; Recipient, ATCC 700669 (NCBI NC_011900); Progeny, SC11-0902-P.

155

FIGURE 5.47: Visualization of a vaccine escape recombination event involving a non-vaccine donor (15B) and PCV-7/13 recipient (23F)[a]. Identification was performed using a penicillin-binding protein transpeptidase domain typing scheme outlined by Metcalf *et al* (38). Similarly to MLST, PBPs were assigned arbitrary allele numbers based on individual nucleotide differences from the penicillin-susceptible NCBI reference strain TIGR4 (NC_003028). The donor strain has a red capsule and black genes; the recipient strain has a purple capsule and blue genes; progeny strain is a combination of both. Orange italicized genes indicate variants from either parent strain.
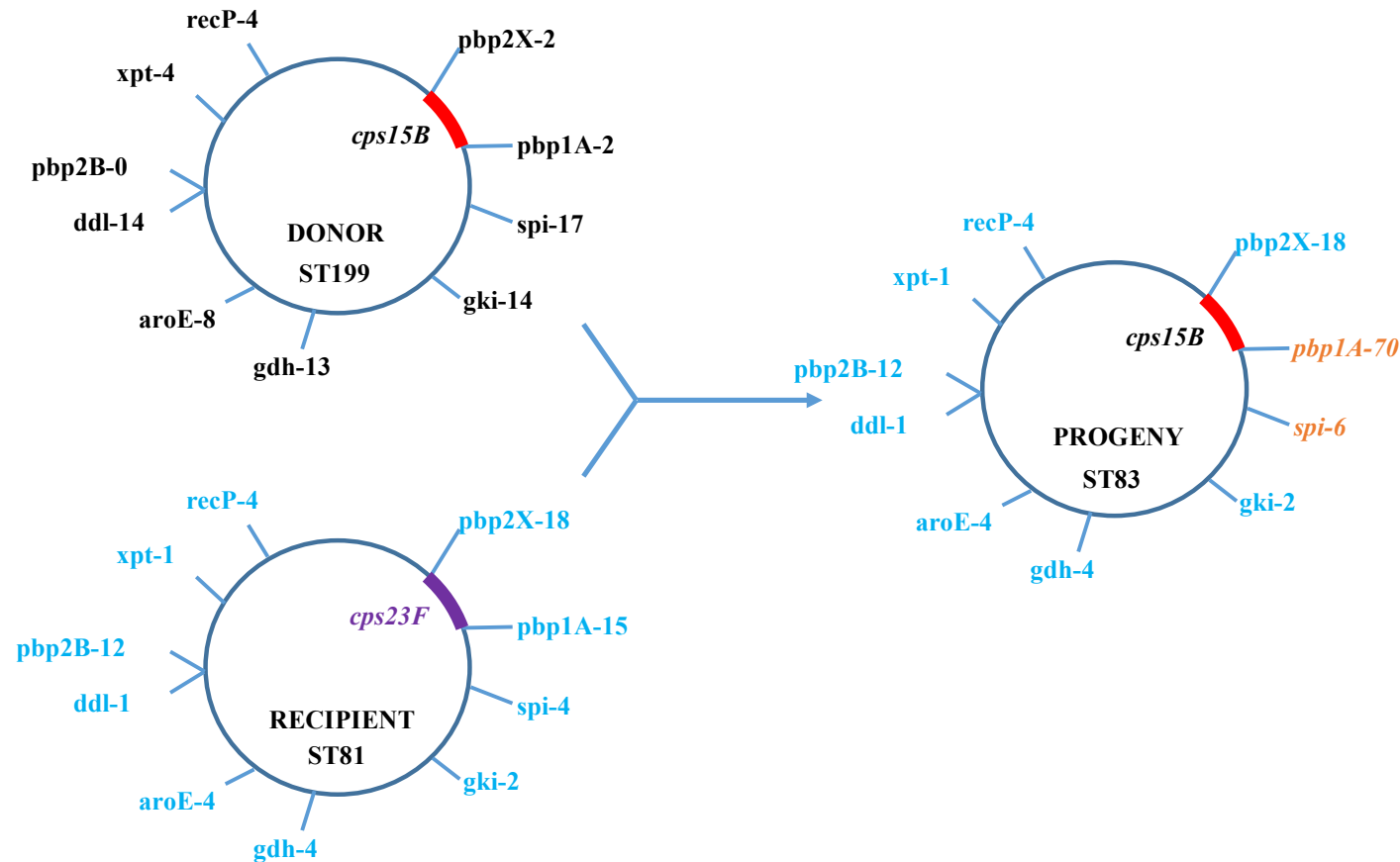

[a]Donor, SC16-2505-P; Recipient, ATCC 700669 (NCBI NC_011900); Progeny, SC11-2658-P.

5.5.2 Events Expanding Antimicrobial Resistance

5.5.2.1 15A → 7F, 8, 22F

PMEN25 (ST63) and related isolates made up a large cluster in the phylogenetic analysis. Predominantly comprised of serotype 15A isolates, this cluster was often MDR, expressing resistance to clarithromycin, clindamycin and doxycycline. It is apparent that isolates in this cluster are also often recipients for capsular switching with other serotypes that are not frequently MDR, including 7F, 8 and 22F.

In the phylogenetic analysis, one isolate each of serotypes 7F and 8 (which each had an overall %MDR rate of less than 0.5) were ST63 and MDR. Both isolates possessed *ermB* and *tetM*, as well as the same PBP-typing pattern as serotype 15-ST63 isolates of 1A-24, 2B-27, 2X-28. This indicates that the *pbp1A* and *pbp2x* of the recipient were retained by the progeny strains (7F, SC11-3889-P and 8, SC11-3902-P). A ST191 isolate was chosen as the donor for serotype 7F (Figure 5.48), and a ST404 for serotype 8 (Figure 5.49). Analysis of the SNVs between putative donor, recipient and progeny strains indicated that the recombination fragment was small relative to the reference (*S. pneumoniae* R6), only spanning the length of the capsule locus (≤12kb, relative to the reference).

FIGURE 5.48: Visualization of a capsular switch recombination event involving a susceptible donor (7F) and MDR, non-vaccine recipient (15A)[a]. Identification was performed using a penicillin-binding protein transpeptidase domain typing scheme outlined by Metcalf *et al* (38). Similarly to MLST, PBPs were assigned arbitrary allele numbers based on individual nucleotide differences from the penicillin-susceptible NCBI reference strain TIGR4 (NC_003028). The donor strain has a red capsule and black genes; the recipient strain has a purple capsule and blue genes; progeny strain is a combination of both**.**
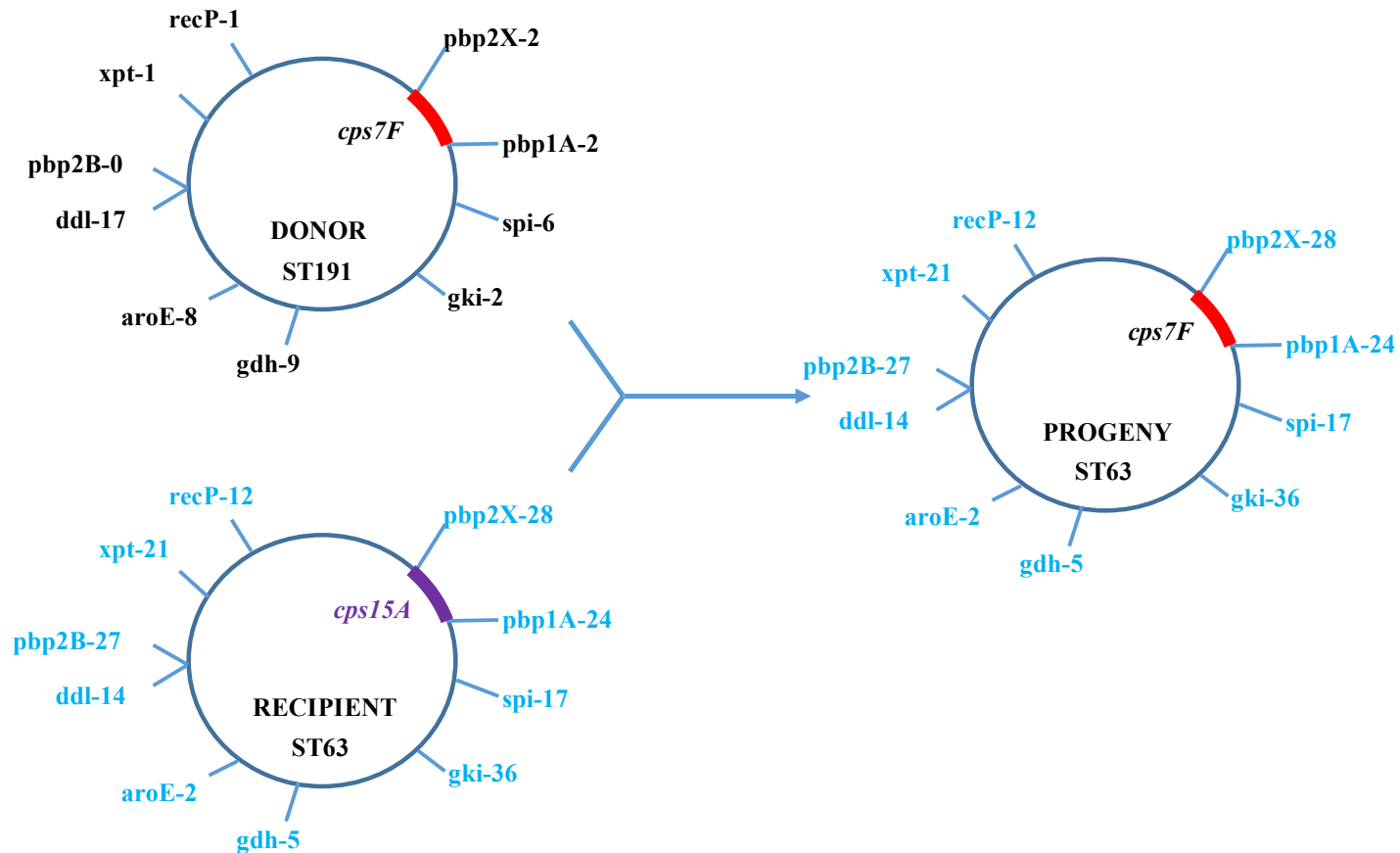
[a]Donor, SC10-0184-P; Recipient, SC11-4275-P; Progeny, SC11-3889-P; difference of 16 SNVs between recipient and progeny.

FIGURE 5.49: Visualization of a capsular switch recombination event involving a susceptible donor (8) and MDR, non-vaccine recipient (15A)[a]. Identification was performed using a penicillin-binding protein transpeptidase domain typing scheme outlined by Metcalf *et al* (38). Similarly to MLST, PBPs were assigned arbitrary allele numbers based on individual nucleotide differences from the penicillin-susceptible NCBI reference strain TIGR4 (NC_003028). The donor strain has a red capsule and black genes; the recipient strain has a purple capsule and blue genes; progeny strain is a combination of both.
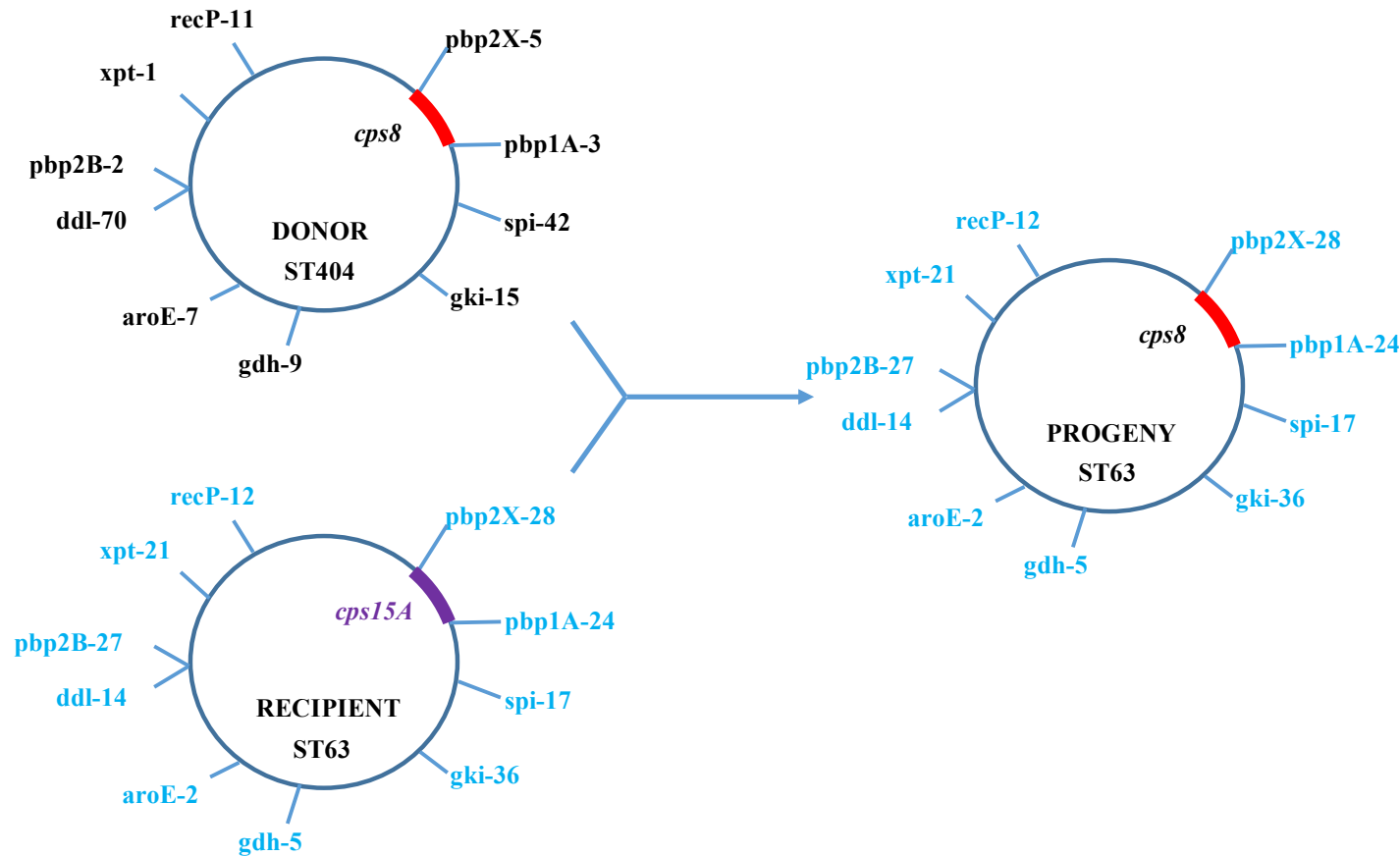
[a]Donor, SC16-2511-P; Recipient, SC11-4275-P; Progeny, SC11-3902-P; difference of 14 SNVs between recipient and progeny.

Two serotype 22F isolates demonstrated relatedness to the PMEN25 cluster (SC12-0452-P and SC12-3310-P). These MDR isolates typed as ST9352, a SLV of ST63 newly identified over the course of this study and only identified in Canadian isolates since. These STs differed at the *gdh* allele and only by one individual nucleotide change. Similarly to ST63 isolates, SC12-0452-P and SC12-3310-P possessed *ermB* and *tetM*, in addition to *parC*-Ser79Phe and *gyrA*-Ser81Leu mutations. As no potential serotype 15A recipient strains from this study demonstrated fluoroquinolone resistance, either the true recipient was fluoroquinolone resistant, or these stepwise mutations developed spontaneously in late 2011-early 2012, when the first ST9352 isolate was collected (SC12-0452-P). The PBP-typing pattern of the putative progeny strains was 1A-1, 2B-27, 2X-2. In comparison to the serotype 15A-ST63 recipient (1A-24, 2B-27, 2X-28), only *pbp2B* was the same, suggesting that the progeny *pbp1A* and *pbp2X* were obtained from the donor. As all other serotype 22F isolates were ST433 with appropriate PBPs (1A-1, 2B-2, 2X-2), this clone was chosen as donor (Figure 5.50). Analysis of the SNVs between putative donor, recipient and progeny strains indicated that the recombination fragment was approximately 40.7kb relative to the reference (*S. pneumoniae* R6), spanning from the beginning of *pbp2X* upstream of the capsule to the end of *cbpF* downstream.

FIGURE 5.50: Visualization of a capsular switch recombination event involving a susceptible donor (22F) and MDR, non-vaccine recipient (15A)[a]. Identification was performed using a penicillin-binding protein transpeptidase domain typing scheme outlined by Metcalf *et al* (38). Similarly to MLST, PBPs were assigned arbitrary allele numbers based on individual nucleotide differences from the penicillin-susceptible NCBI reference strain TIGR4 (NC_003028). The donor strain has a red capsule and black genes; the recipient strain has a purple capsule and blue genes; progeny strain is a combination of both. Orange italicized genes indicate variants from either parent strain.
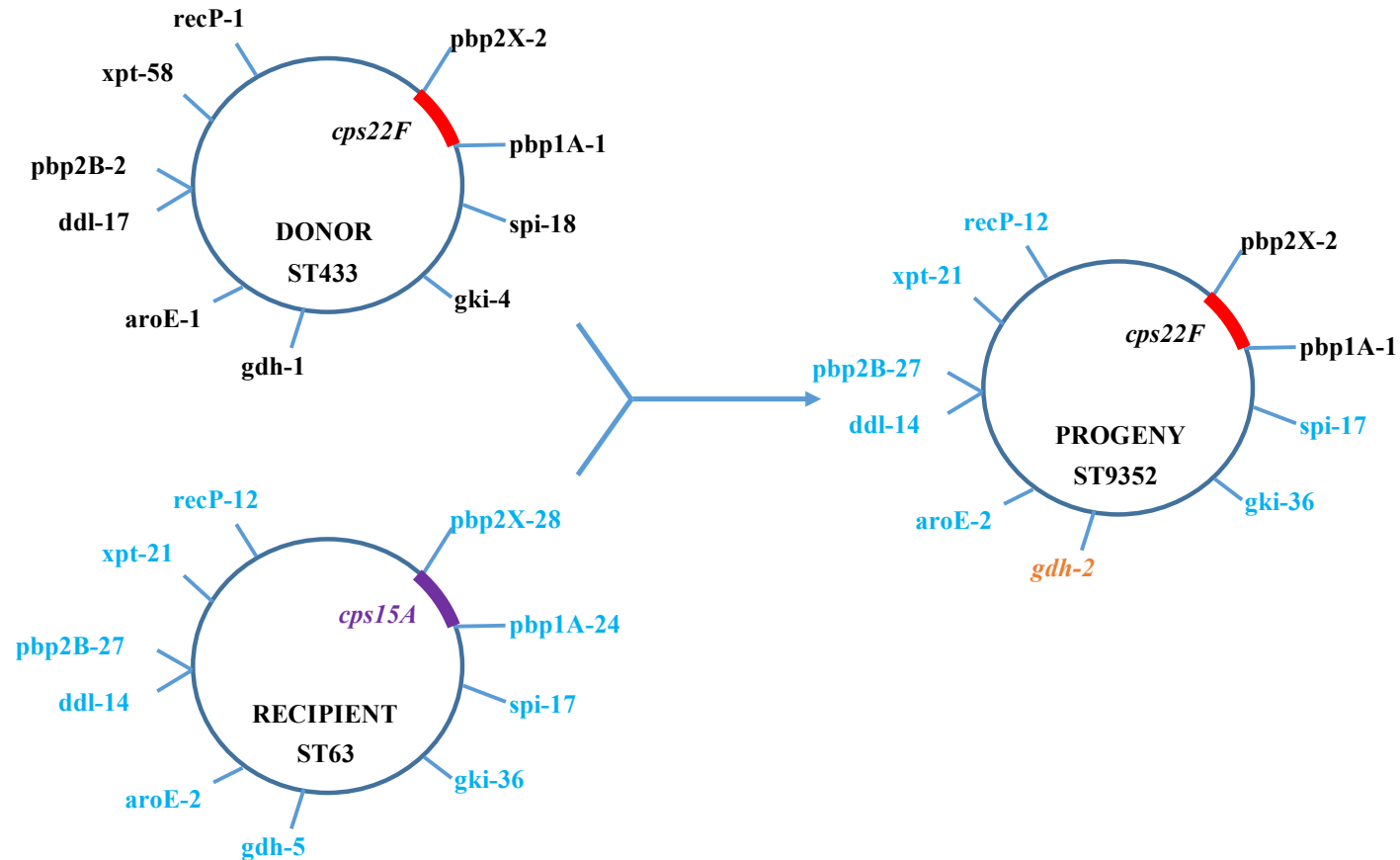
[a]Donor, SC16-2507-P; Recipient, SC12-2899-P; Progeny, SC12-0452-P; difference of 514 SNVs between recipient and progeny.

5.5.2.2 35B → 29

Serotype 29 was not a particularly common type isolated in the SAVE study and it is not well known for antimicrobial resistance; however, the one isolate randomly selected as a background strain was related to serotype 35B. SC14-1327-P demonstrated a MLST type of ST7486, a SLV of ST558 commonly seen in serotype 35B isolates; it also possessed *mefA*, a gene commonly possessed by ST558 isolates. Because of this, a recipient strain was selected from those serotype 35B-ST558 isolates also possessing *mefA*. The putative recipient and progeny strains demonstrated the same PBP-typing pattern of 1A-4, 2B-7, 2X-7, indicating that the PBPs were retained from the recipient (Figure 5.51). As no other serotype 29 isolates from this study were characterized enough to be considered as a donor strain, an isolate from NCBI was used (ERR057775). This isolate was ST10444, with a PBP typing pattern of 1A-0, 2B-4, 2X-130. Analysis of the SNVs between putative donor, recipient and progeny strains indicated that the recombination fragment was small relative to the reference (*S. pneumoniae* R6), only spanning the length of the capsule locus (≤12kb). The progeny strain contained a different *recP* allele than the recipient strain; *recP*-2 and *recP*-44 differ by four nucleotides, likely indicating that SC14-1347-P obtained this gene from another recombination event.

FIGURE 5.51: Visualization of a capsular switch recombination event involving a susceptible donor (29) and MDR, non-vaccine recipient (35B)[a]. Identification was performed using a penicillin-binding protein transpeptidase domain typing scheme outlined by Metcalf *et al* (38). Similarly to MLST, PBPs were assigned arbitrary allele numbers based on individual nucleotide differences from the penicillin-susceptible NCBI reference strain TIGR4 (NC_003028). The donor strain has a red capsule and black genes; the recipient strain has a purple capsule and blue genes; progeny strain is a combination of both. Orange italicized genes indicate variants from either parent strain.
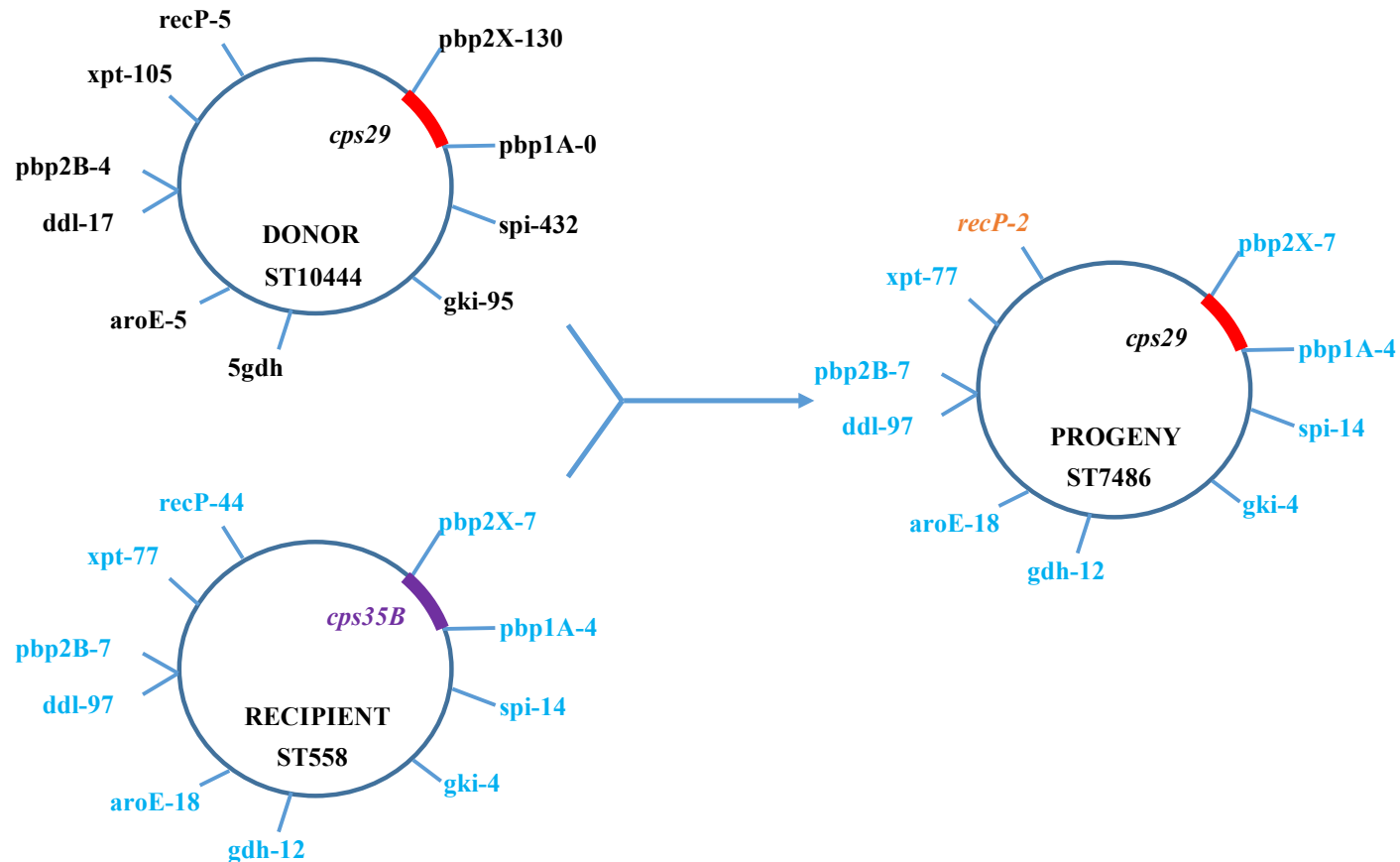
[a]Donor, NCBI ERR057775; Recipient, SC14-3062-P; Progeny, SC14-1347-P; difference of 200 SNVs between recipient and progeny.

5.5.2.3 9V → 14

Serotype 14 isolates related to PMEN9 (ST9) were commonly resistant to clarithromycin in this study. However, one serotype 14 isolate (SC14-0090-P) was MDR and related to PMEN3 isolates also demonstrating MDR patterns. This isolate was unfortunately excluded from the bulk of the antimicrobial resistance gene analysis, as it did not grow for triplicate susceptibility testing. However, as SC14-0090-P appeared to be a switch variant, it was typed using the PBP scheme; the resulting pattern of 1A-15, 2B-12, 2X-18 was identical to that of other serotype 9V-ST156 isolates, indicating that *pbp1A* and *pbp2X* were retained from the recipient (Figure 5.52). A ST9 isolate with a PBP typing pattern of 1A-2, 2B-0, 2X-64 was chosen as the capsule donor. Analysis of the SNVs between putative donor, recipient and progeny strains indicated that the recombination fragment was approximately 31.2kb relative to the reference (*S. pneumoniae* R6), spanning from within *dexB* upstream of the capsule to within *pbp1A* downstream.

FIGURE 5.52: Visualization of a capsular switch recombination event involving a susceptible PCV-7/13 donor (14) and MDR, PCV-7/13 recipient (9V)[a]. Identification was performed using a penicillin-binding protein transpeptidase domain typing scheme outlined by Metcalf *et al* (38). Similarly to MLST, PBPs were assigned arbitrary allele numbers based on individual nucleotide differences from the penicillin-susceptible NCBI reference strain TIGR4 (NC_003028). The donor strain has a red capsule and black genes; the recipient strain has a purple capsule and blue genes; progeny strain is a combination of both.
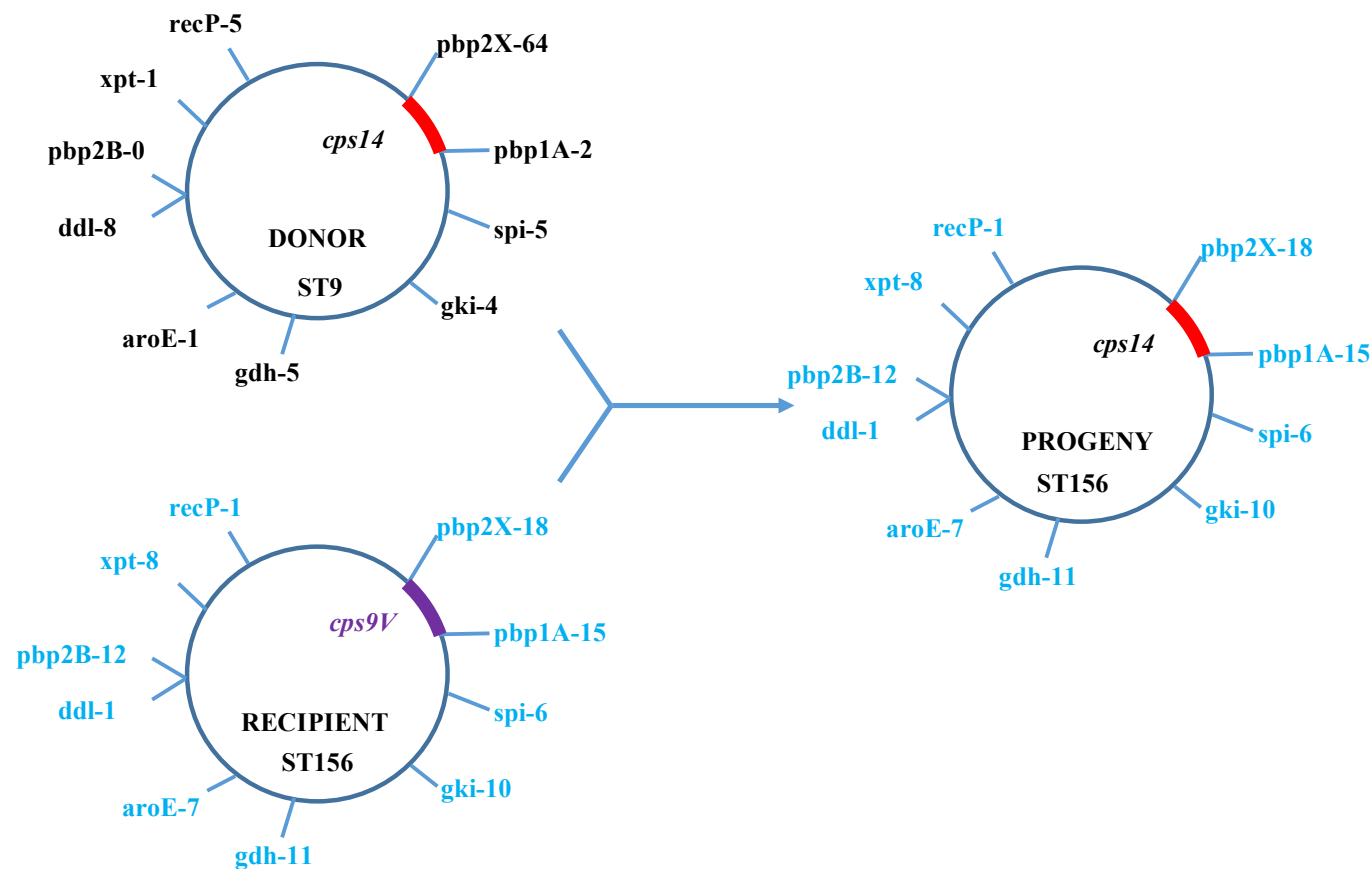
[a]Donor, SC13-2375-P; Recipient, SC11-0099-P; Progeny, SC14-0090-P.

165

5.5.3 Events Increasing Serotype Diversity

5.5.3.1 22F → 31

Serotype 31 is another type that was not particularly common in the SAVE study. One background isolate randomly selected for this study typed as ST433, identical to the most common clone of serotype 22F. Progeny strain SC14-3964-P demonstrated a PBP-typing pattern of 1A-1, 2B-2, 2X-0; this was similar to serotype 22F-ST433 recipient candidates, which all had the pattern 1A-1, 2B-2, 2X-2. Available serotype 31 genomes were ideal donors, bearing the pattern 1A-12, 2B-0, 2X-0. This indicated that *pbp1A* was retained from the recipient strain, while *pbp2X* was transferred with the capsule locus from the donor (Figure 5.53). Analysis of the SNVs between putative donor, recipient and progeny strains indicated that the recombination fragment was approximately 26.2kb relative to the reference (*S. pneumoniae* R6), spanning from the beginning of *pbp2X* upstream of the capsule to hypothetical protein spr0328 downstream. As serotype 31 is not included in a vaccine formulation, this is not a vaccine escape event; similarly, as serotype 22F is not particularly antimicrobial resistant or MDR, this recombination does not serve to spread resistance.

FIGURE 5.53: Visualization of a capsular switch recombination event where a non-vaccine recipient (22F) increased the diversity of an uncommon donor (31)[a]. Identification was performed using a penicillin-binding protein transpeptidase domain typing scheme outlined by Metcalf *et al* (38). Similarly to MLST, PBPs were assigned arbitrary allele numbers based on individual nucleotide differences from the penicillin-susceptible NCBI reference strain TIGR4 (NC_003028). The donor strain has a red capsule and black genes; the recipient strain has a purple capsule and blue genes; progeny strain is a combination of both.
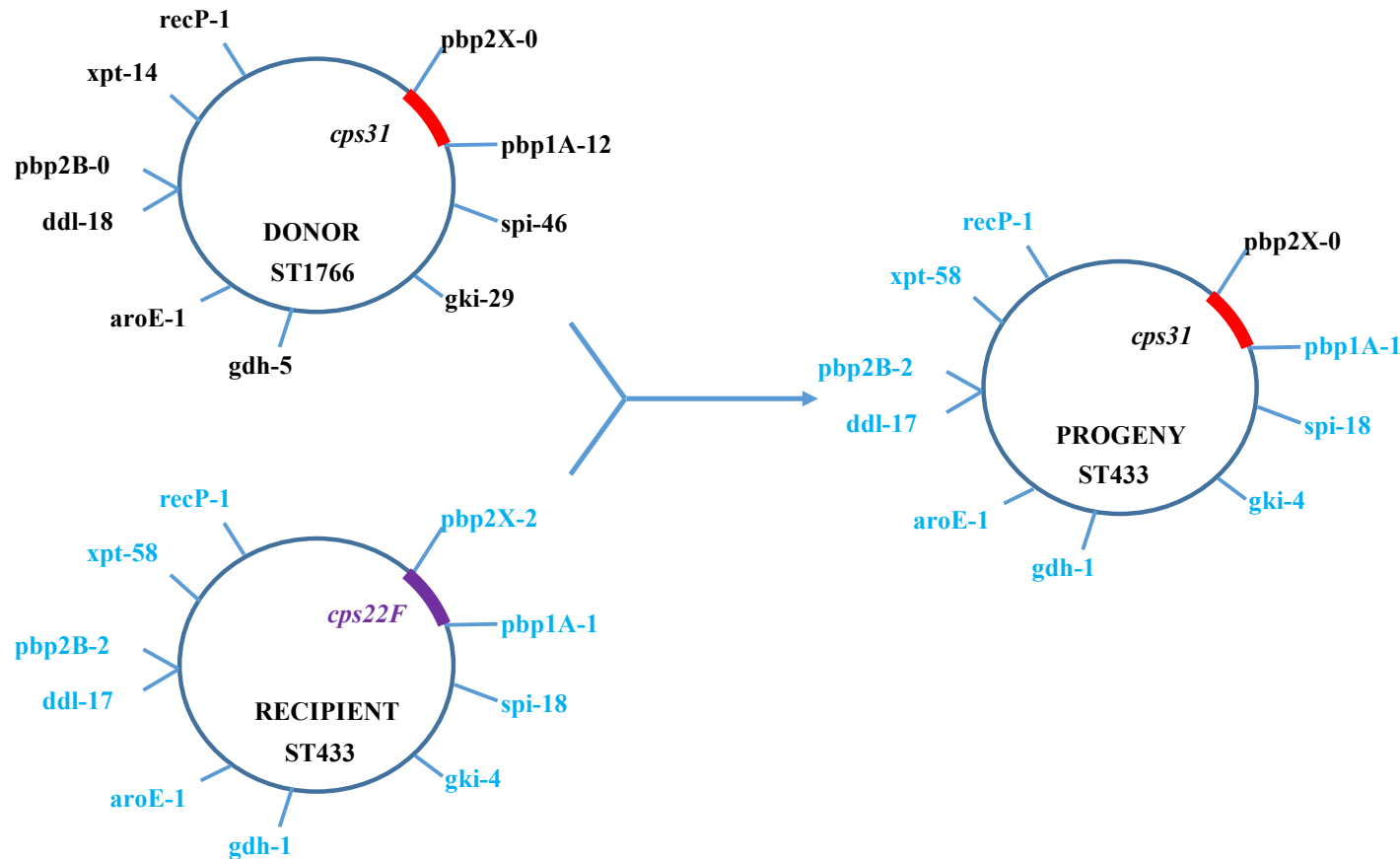
[a]Donor, SC14-3643-P; Recipient, SC11-1577-P; Progeny, SC14-3964-P; difference of 386 SNVs between recipient and progeny.

5.5.3.2 21, 15A → 19A

As previously described in Section 5.4.2.2, serotype 19A is highly diverse, demonstrating

relatedness to numerous other serotypes and clones. Phylogenetic analysis identified three

serotype 19A isolates identical to PMEN25 (ST63), as well as two identical to PMEN30

(ST193). As serotype 19A is included in the PCV-13 vaccine formulation, these progeny strains

are not vaccine escape recombinants. PMEN30 was originally isolated as serotype 21, an

uncommon serotype that did not demonstrate significant antimicrobial resistance in this study;

progeny strains such as SC13-2685-P therefore also did not demonstrate MDR, but instead

increased the breadth of genetic diversity associated with serotype 19A. Progeny strain SC13-

2685-P demonstrated a PBP-typing pattern of 1A-0, 2B-0, 2X-11; the likely donor strain, a

serotype 19A-ST199, shared 1A-0 and 2X-11 in common, indicating that *pbp1A* and *pbp2X* were

transferred along with the capsule in the recombination event (Figure 5.54). Analysis of the

SNVs between putative donor, recipient and progeny strains indicated that the recombination

fragment was approximately 48.9kb relative to the reference (*S. pneumoniae* R6), spanning from

kdgA upstream of *pbp2X,* to hypothetical protein spr0334 downstream.

Serotype 19A is commonly MDR or XDR, therefore the recombination event involving a

MDR serotype 15A-ST63 isolate did not necessarily serve to increase the resistance

demonstrated by serotype 19A, just give it a different pattern. Progeny strain SC11-4510-P

demonstrated the same presence of *ermB* and *tetM* as other ST63 strains, with a PBP-typing

pattern of 1A-0, 2B-27, 2X-28. All potential serotype 15A-ST63 recipients held the pattern 1A-

24, 2B-27, 2X-28, indicating that *pbp2X* was retained from the recipient (Figure 5.55). Using the

same donor strain as above (1A-0, 2B-29, 2X-11) would allow for *pbp1A* to be transferred from

the donor to progeny. Analysis of the SNVs between putative donor, recipient and progeny

strains indicated that the recombination fragment was approximately 29.5kb relative to the

reference (*S. pneumoniae* R6), spanning from hypothetical protein spr0309 upstream of the

capsule locus, to hypothetical protein spr0334 downstream of *pbp1A*.

FIGURE 5.54: Visualization of a capsular switch recombination event where an uncommon recipient (21) increased the diversity of an already diverse donor (19A)[a]. Identification was performed using a penicillin-binding protein transpeptidase domain typing scheme outlined by Metcalf *et al* (38). Similarly to MLST, PBPs were assigned arbitrary allele numbers based on individual nucleotide differences from the penicillin-susceptible NCBI reference strain TIGR4 (NC_003028). The donor strain has a red capsule and black genes; the recipient strain has a purple capsule and blue genes; progeny strain is a combination of both.
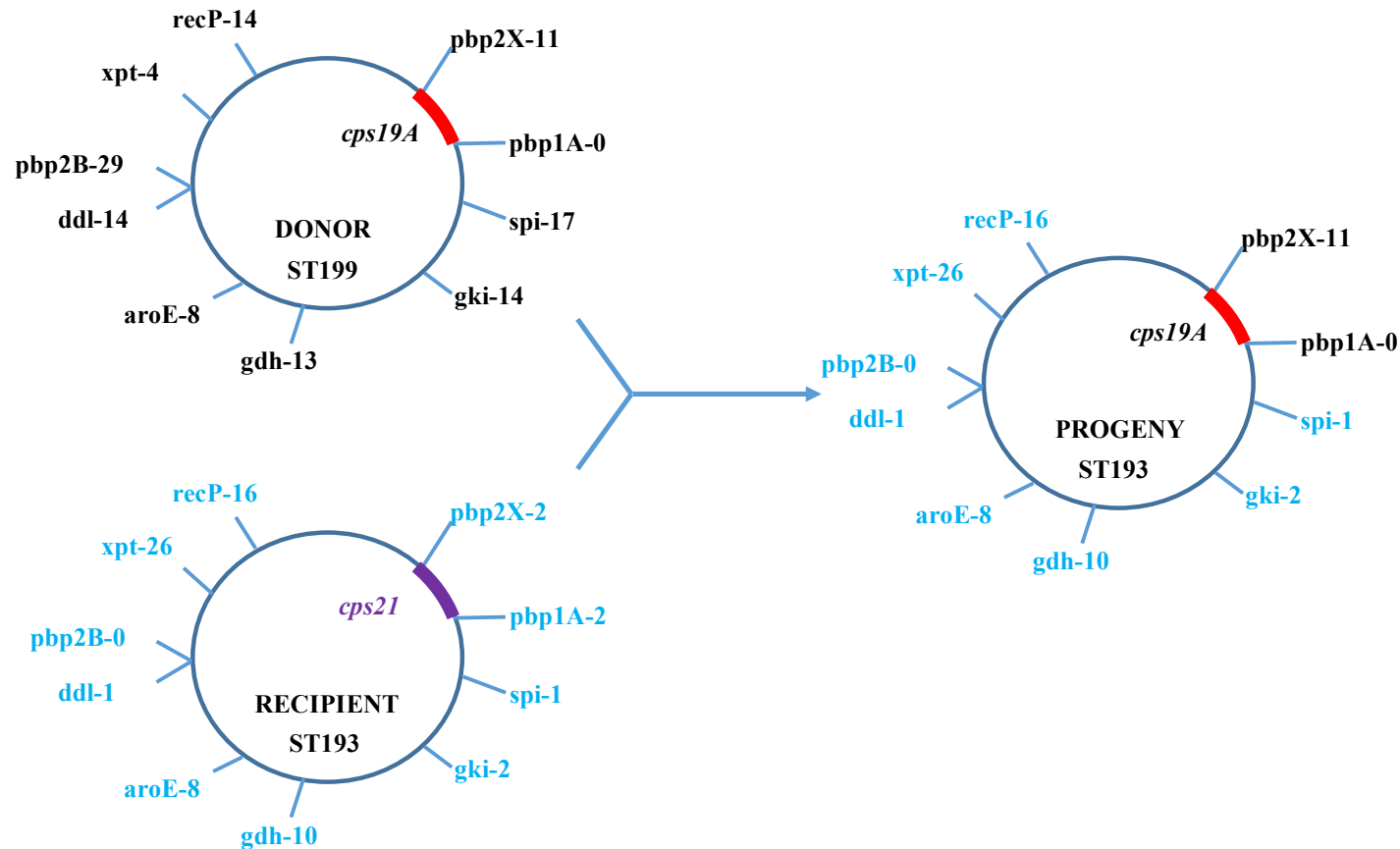
[a]Donor, SC10-0551-P; Recipient, SC13-2731-P; Progeny, SC13-2685-P; difference of 139 SNVs between recipient and progeny.

FIGURE 5.55: Visualization of a capsular switch recombination event where a common, MDR recipient (15A) increased the diversity of an already diverse donor (19A)[a]. Identification was performed using a penicillin-binding protein transpeptidase domain typing scheme outlined by Metcalf *et al* (38). Similarly to MLST, PBPs were assigned arbitrary allele numbers based on individual nucleotide differences from the penicillin-susceptible NCBI reference strain TIGR4 (NC_003028). The donor strain has a red capsule and black genes; the recipient strain has a purple capsule and blue genes; progeny strain is a combination of both.
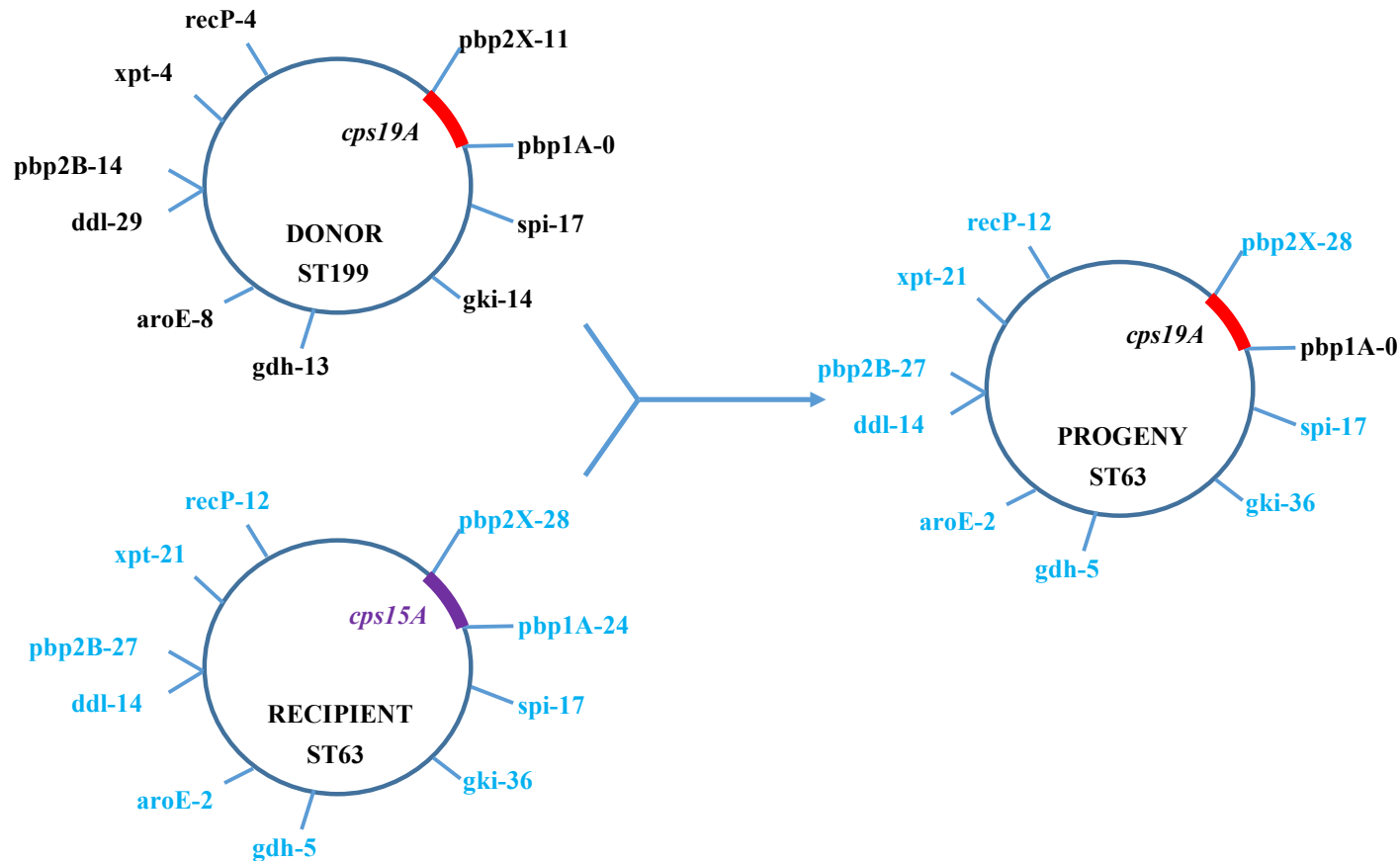
[a]Donor, SC10-0551-P; Recipient, SC11-4275-P; Progeny, SC11-4510-P; difference of 303 SNVs between recipient and progeny.

5.5.4 Putative Variants

Many additional putative variants were identified through both MLST, PBP-typing and phylogenetic analysis. These variants were unable to be confirmed for two reasons; either the isolates were not sequenced by WGS, therefore there were no PBP sequences to confirm the event; or, a donor or recipient strain of high-quality and suitable ST and PBP pattern was not available for confirmation. These switch events are detailed in Table 5.7, as well as the reason for why each event was not confirmed.

TABLE 5.7: Putative capsular switch events identified in SAVE 2011-15 isolates that could not be confirmed.

Putative Donor Serotype	Putative Recipient	Putative Progeny	Type of Recombination	Reason for not Confirming
10A	PMEN34	10A-ST230	Vaccine Escape	Unable to find a suitable recipient
24F	(14-ST230)	24F-ST230	Vaccine Escape	
19A		19A-ST319 (SLV)	Increase Diversity	
33F	PMEN21	33F-ST177		MLST only
24B	(19F-ST177)	24B-ST177	Vaccine Escape	Unable to find a suitable donor
3	35B-ST198	3-ST198	Increased Diversity	MLST only
11A	9V-ST156	11A-ST156	Vaccine Escape	MLST only
12F	PMEN18 or var. (9N-ST66)	12F-ST66	Increased Diversity	MLST only
9N	PMEN34 (12F-ST218)	9N-ST405 (SLV) 9N-ST10540 (DLV)	Increased Diversity	MLST only

5.6 Identification of Highly Recombinant Isolates

Gubbins software was used to predict areas of recombination within 192 *S. pneumoniae* whole genomes. The results for all isolates together are displayed in Figure 5.56. As predicted, there were numerous blocks of recombination detected in all isolates, both unique (blue blocks) and obtained through common descent (red blocks); this indicates that while many isolates are unique due to the recombinogenic nature of *S. pneumoniae*, many areas of the *S. pneumoniae* genome are conserved in all isolates, regardless of serotype. Areas containing a high density of recombinations were often associated with various synthases, transferases and permeases, amongst others. Of note, within clusters, recombinations pertaining to the capsule (specifically *pbp1A* and *pbp2X*) were predominantly shared through common descent; however, many isolates had unique recombinations in that region.

Many large clusters demonstrated interesting characteristics. Serotype 3-ST180 and serotype 15A-ST63 isolates were noteworthy because half of the cluster possessed certain shared recombinations while others did not (Figure 5.56). Serotype 19A-ST320 isolates were the opposite, with little difference between isolates and few unique regions. These serotypes will be discussed in more detail in Section 5.6.1.

To achieve better resolution of potential recombination events and differences in recombination frequency between common serotypes, Gubbins was re-run individually with each of the twelve serotypes of interest.
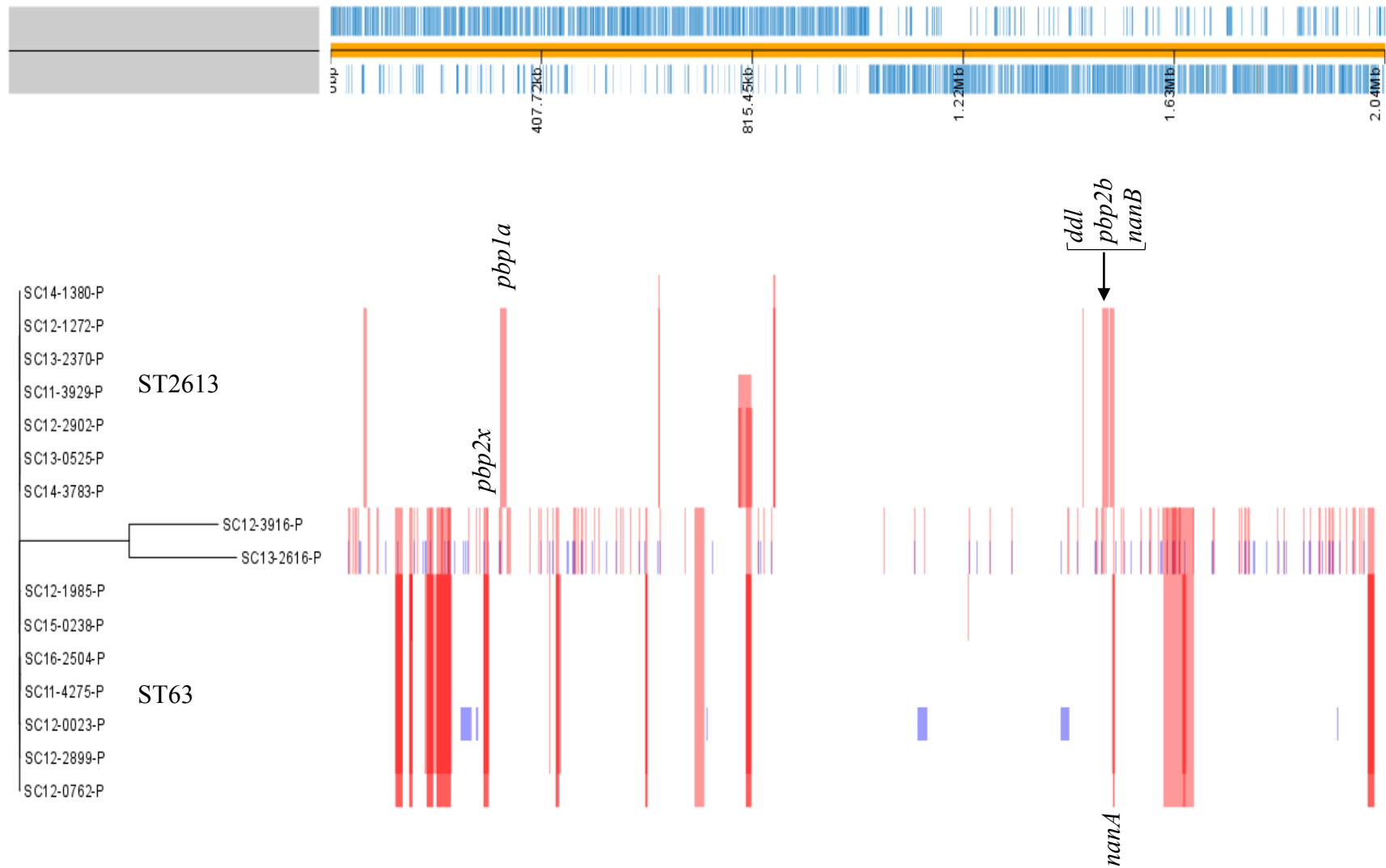
FIGURE 5.56: Phylogeny and recombination events for 192 *S. pneumoniae* isolates, predicted by Gubbins and visualized using Phandango and *S. pneumoniae* R6 reference genome. Red blocks indicate areas of recombination that are shared by multiple isolates through common descent. Blue blocks indicate areas of recombination that are unique to that isolate. Clusters of interest are outlined in purple, with specific recombination blocks of interest outlined in yellow.
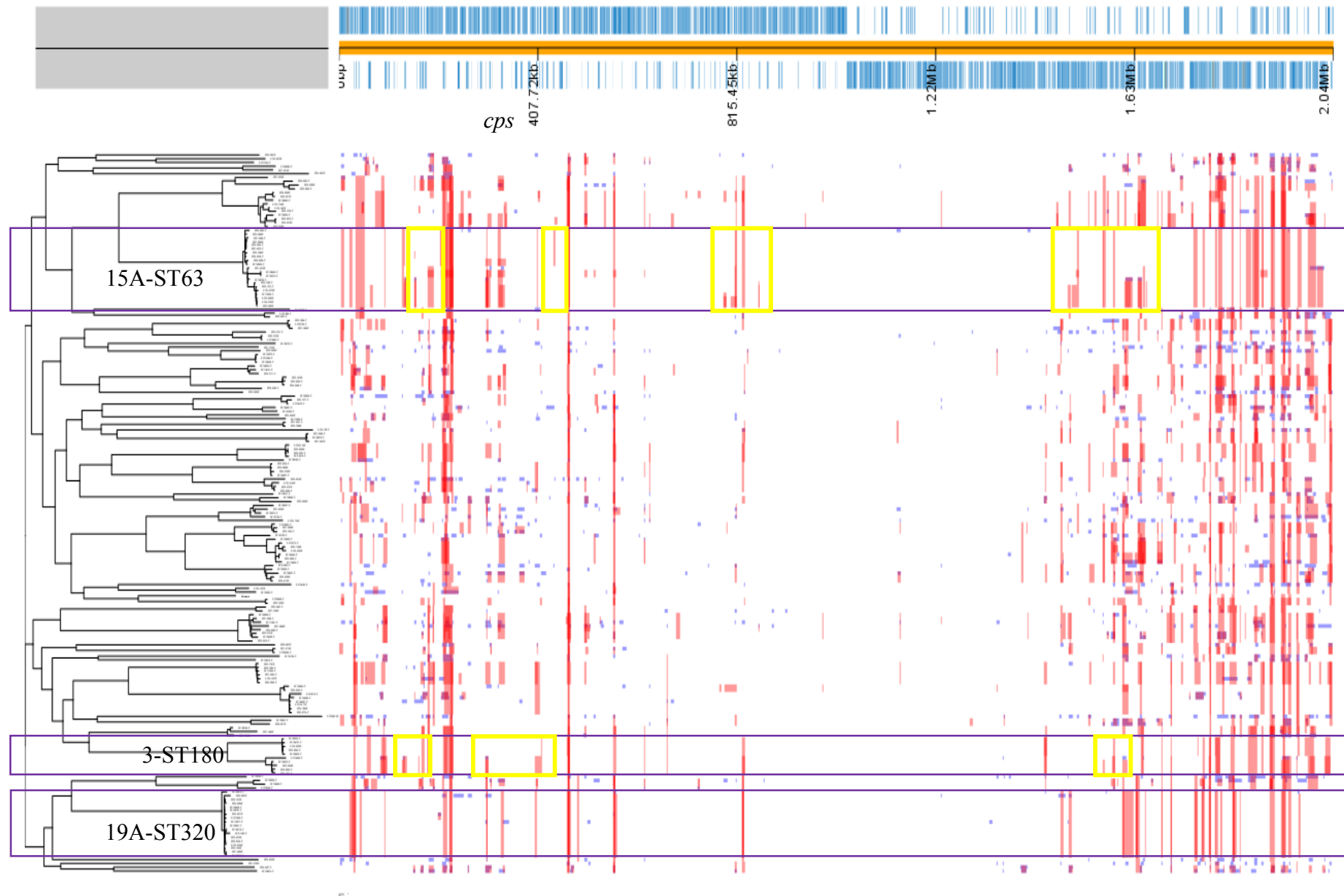
5.6.1 Serotypes 3, 15A and 19A

Examination of the putative recombination events of serotype 3 isolates helped to clarify the results presented in Section 5.4.3.1. The two clades of ST180 isolates, apparently only differing in year of isolation and antimicrobial resistance, can now be seen in Figure 5.57 to possess very different recombination histories. Clade II isolates possessed numerous blocks of recombination events obtained through common descent; clade I isolates, however, demonstrated no recombination at all. One isolate, SC11-4599-P, was the only isolate to possess unique areas of recombination in addition to those obtained through common descent; these unique areas could indicate that this isolate is a new variant of clade II. Areas of interest in the clade II isolates included variable surface proteins *pspA* and *pspC*, as well as *pbp2X*. Further investigation of *pbp2X* revealed that ST180-clade II isolates possessed an altered transpeptidase domain for this protein, specifically 2X-111 (1A-2, 2B-0), as opposed to 2X-2 (1A-2, 2B-3) which is present in ST180-clade I isolates. Surface proteins *pspA* and *pspC* will be discussed in Results Part 3.

FIGURE 5.57: Phylogeny and recombination events for serotype 3 isolates, predicted by Gubbins and visualized using Phandango and *S. pneumoniae* R6 reference genome. Red blocks indicate areas of recombination that are shared by multiple isolates through common descent. Blue blocks indicate areas of recombination that are unique to that isolate.
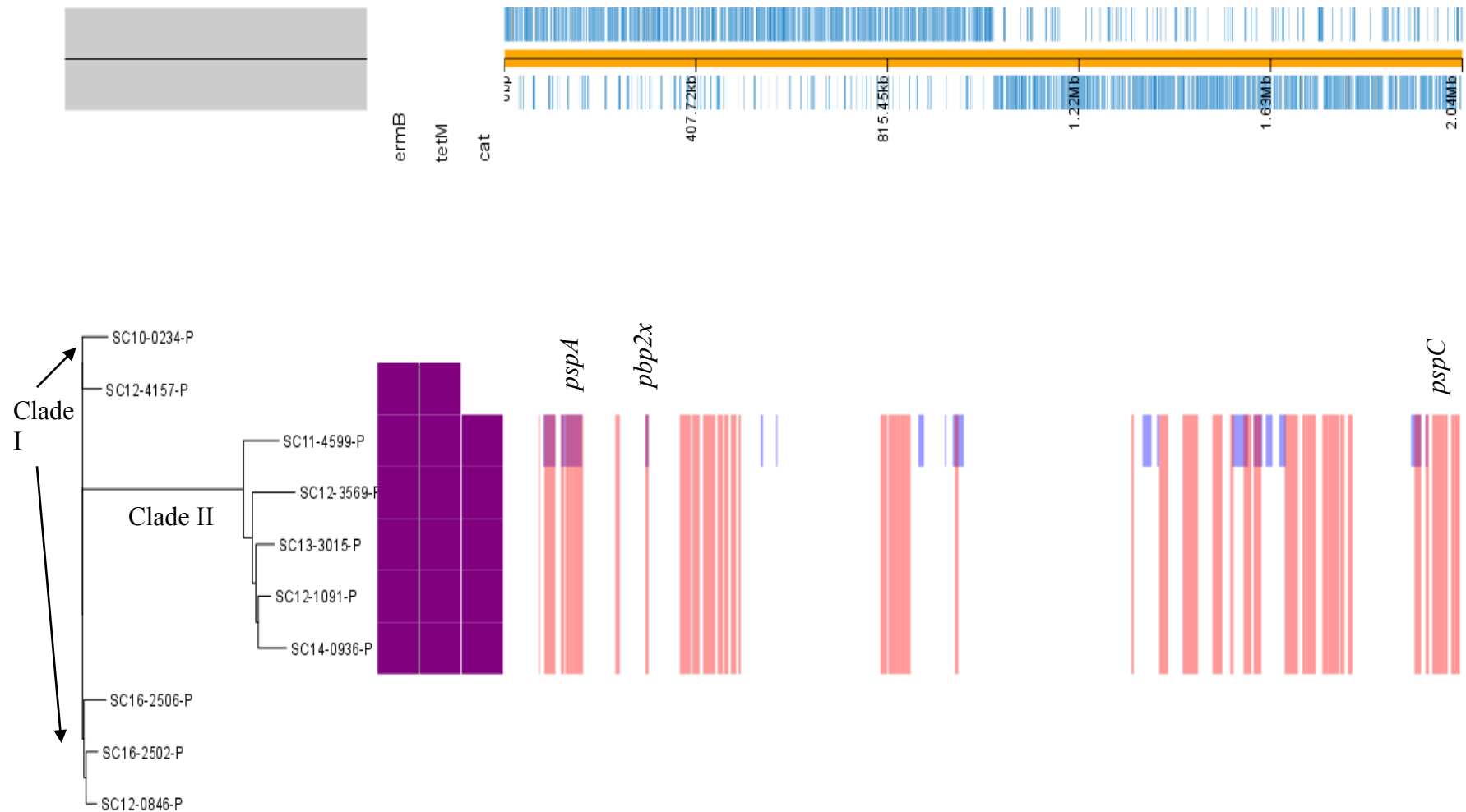
Serotype 15A was similar to serotype 3 in that the type split into two groups of varying recombination patterns; interestingly, in this case, isolates split into these two groups based on ST, with ST63 possessing different blocks of recombination than DLV ST2316 (Figure 5.58). Specific recombination blocks of interest included all three PBPs; ST63 isolates demonstrated recombination in *pbp2X* shared through common descent, while ST2613 had similar blocks for *pbp1A* and *pbp2B*. This is notable as ST2613 isolates were all penicillin-resistant, with numerous mutations in all three genes. *ddl* is located very near *pbp2B* and was also included in a recombination block that was common to ST2613; it is likely that these genes were recombined at the same time, as all ST2613 demonstrate the same mutations in *pbp2B* and the same altered *ddl* locus in comparison to ST63. Other genes of note included neuraminidase genes *nanA* and *nanB*. ST63 isolates shared a recombination block containing *nanA*, while ST2613 isolates had a *nanB* recombination block shared by common descent. These genes will be discussed further in Results Part 3.

As serotype 19A demonstrated so much variability in the previously described molecular and phylogenetic characterization, it was expected that it would also contain vast numbers of recombination blocks. Interestingly, ST320, commonly thought to have arisen from multiple recombination events, demonstrated few unique blocks (Figure 5.59). Instead, it shared a few blocks of recombination with all other 19A isolates, obtained through common descent. Only two isolates demonstrated any unique recombination blocks, with the most interesting including *pspA*. All other non-ST320 serotype 19A isolates demonstrated many blocks of unique recombinations spaced throughout the genome.

FIGURE 5.58: Phylogeny and recombination events for serotype 15A isolates, predicted by Gubbins and visualized using Phandango and *S. pneumoniae* R6 reference genome. Red blocks indicate areas of recombination that are shared by multiple isolates through common descent. Blue blocks indicate areas of recombination that are unique to that isolate.

FIGURE 5.59: Phylogeny and recombination events for serotype 19A isolates, predicted by Gubbins and visualized using Phandango and *S. pneumoniae* R6 reference genome. Red blocks indicate areas of recombination that are shared by multiple isolates through common descent. Blue blocks indicate areas of recombination that are unique to that isolate.
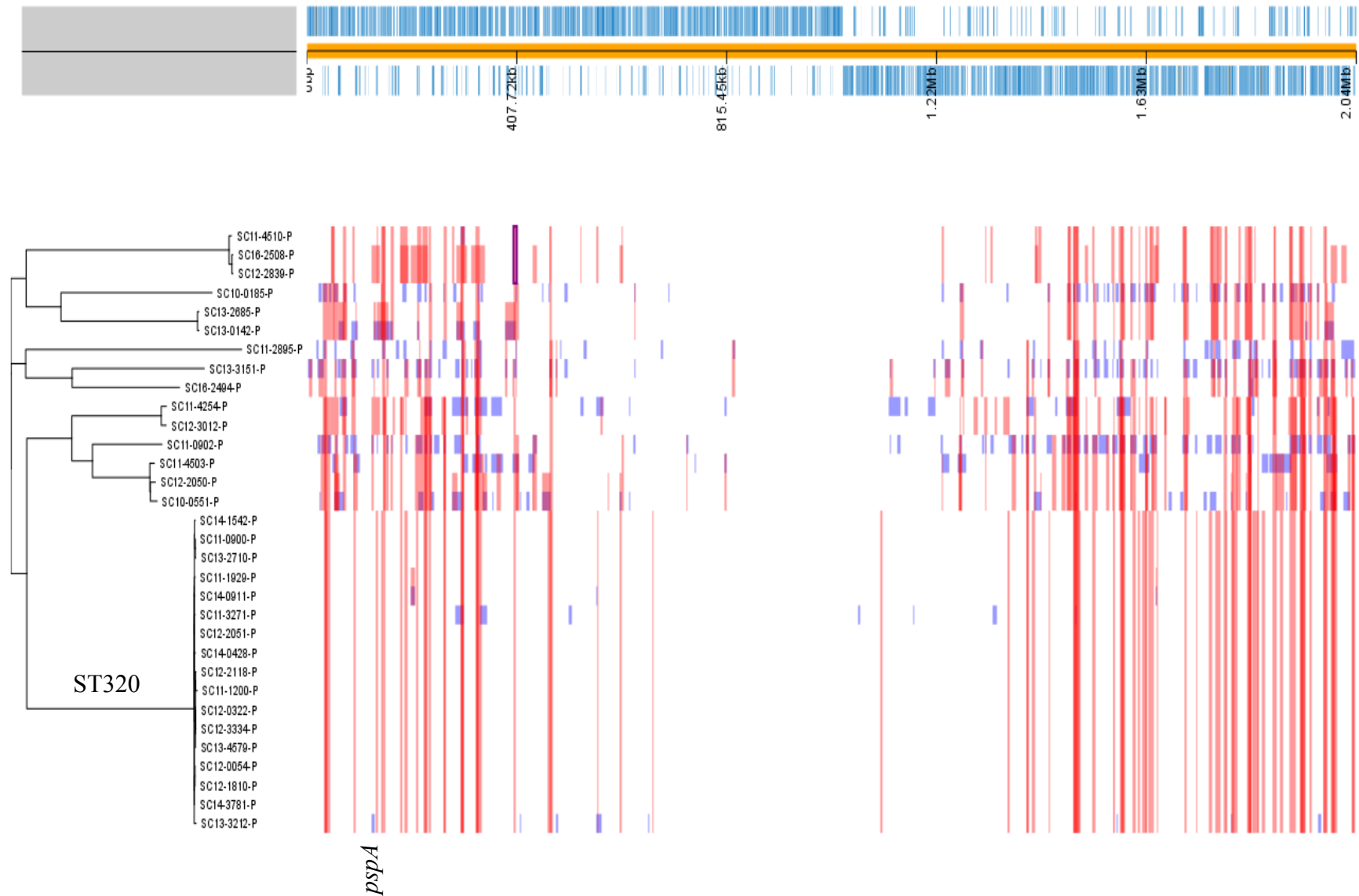
5.6.2 Serotypes 33F, 6C and 9N

Recombination analysis for serotype 33F provided additional evidence that there is a temporal trend for this type. Figure 5.60 depicts how most of the predicted recombination in this serotype was identified in isolates collected from 2012 onwards, that possessed antimicrobial resistance determinants and new STs. Recombination blocks that had shared ancestry for all serotype 33F isolates contained genes for phosphorylases, kinases and many pertaining to folate synthesis. SC14-1578-P contained the most unique recombination blocks, corresponding to the most recently collected isolate in the analysis.

Gubbins analysis of serotypes 9N and 6C produced numerous areas of predicted recombination. Serotype 6C, which demonstrated overall high diversity in the previous analyses, predictably had many areas of recombination in all isolates shared through common descent, as well as two isolates filled with unique regions (Figure 5.61, bottom panel). Serotype 9N, which had previously demonstrated high diversity within CC66, broke down into smaller clusters based on ST in the recombination analysis (Figure 5.61, top panel). ST517 isolates shared the most ancestral recombination blocks and had no unique areas. The ST66 isolates shared some of the same common areas as ST517, but only one of the two isolates demonstrated unique recombinations. ST10191 isolates shared no recombination obtained through common descent and only the isolate collected in 2013 demonstrated any unique recombinations.

FIGURE 5.60: Phylogeny and recombination events for serotype 33F isolates, predicted by Gubbins and visualized using Phandango and *S. pneumoniae* R6 reference genome. Red blocks indicate areas of recombination that are shared by multiple isolates through common descent. Blue blocks indicate areas of recombination that are unique to that isolate.
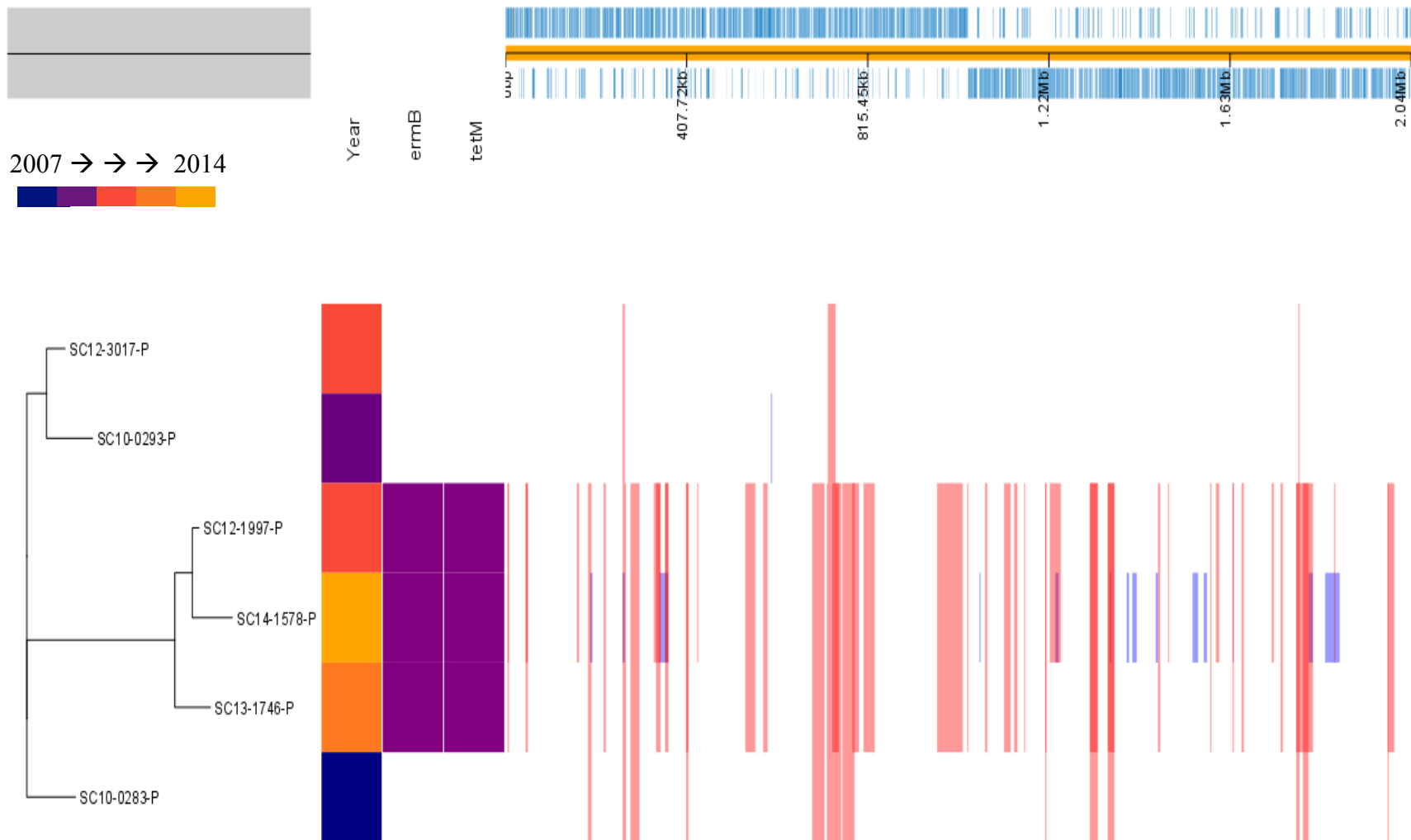
FIGURE 5.61: Phylogeny and recombination events for serotype 9N (top panel) and 6C (bottom panel) isolates, predicted by Gubbins and visualized using Phandango and *S. pneumoniae* R6 reference genome. Red blocks indicate areas of recombination that are shared by multiple isolates through common descent. Blue blocks indicate areas of recombination that are unique to that isolate.
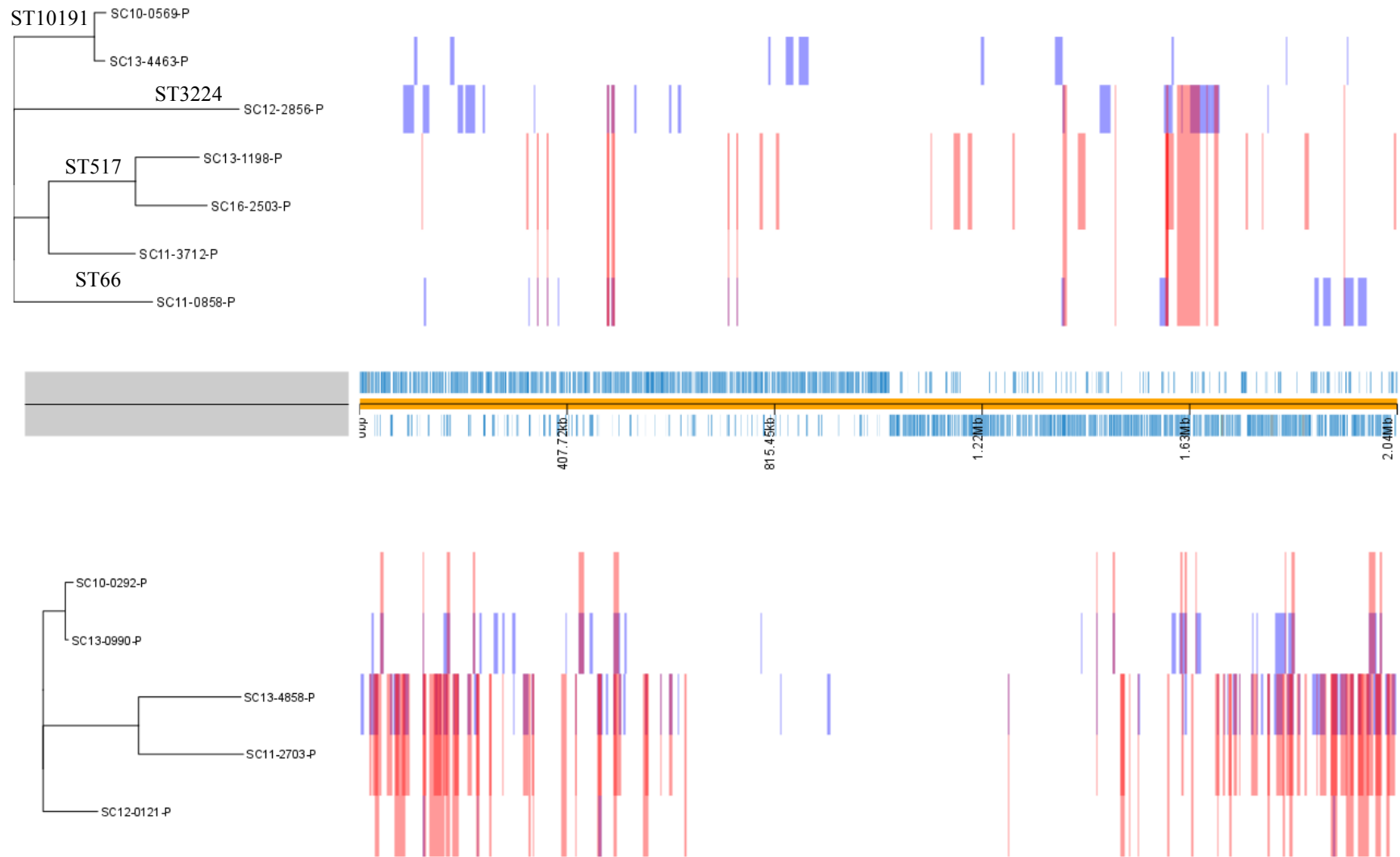
5.6.3 Other Serotypes of Interest

Serotypes 7F, 12F, 22F and 35B only demonstrated significant areas of recombination when the isolate was a ST not normally associated with the serotype (data not shown). Specifically, isolates of STs 191, 218, 433 and 558 within serotypes 7F, 12F, 22F and 35B, respectively, did not demonstrate recombination. Products of capsular switches within serotypes 7F, 22F and 35B demonstrated numerous blocks of recombination, both unique and those shared through common descent. Interestingly, the serotype 12F isolates that demonstrated recombination did not appear to be the products of capsular switches; these isolates were MDR and belonged to infrequently isolated or new STs.

Serotypes 8 and 11A were interesting in that every isolate included in the analysis demonstrated areas of predicted recombination across almost the entire genome. Given the previous data indicating that these serotypes are relatively clonal, it is possible that there was either not enough data included in the analysis to achieve accurate results, or the reference strain chosen for the reference-mapping alignment was inappropriate for these serotypes.

Part 3: Virulence Assessment of *S. pneumoniae* Collected by the SAVE 2011-2015 Study

5.7 Pneumococcal Pili

Overall, 4,016 isolates with a serotype of interest had full PCR results for both PI-1 and PI-2 (Table 5.8). Roughly two-thirds of these isolates did not demonstrate the presence of pilus genes; in particular, several full cohorts of serotypes had little to no association with either pilus, including serotypes 3, 8, 9N, 12F, 22F and 33F.

Four serotypes demonstrated PI-1 presence in >5% of the isolates tested; these were serotypes 6C, 15A, 19A and 35B. Isolates that possessed only PI-1 genes were not commonly associated with MDR; of the 16 serotype 6C isolates (8.5%) and 52 serotype 35B isolates (37.7%) that possessed PI-1, only one of each serotype was MDR. Despite being a commonly MDR serotype, 15A isolates containing PI-1 (61, 28.8%) expressed little resistance and were not associated with MDR (ST58, ST3811). Serotype 19A isolates demonstrated just over 5% MDR in isolates containing PI-1 (257, 44.5%), including ST416 and ST695.

PI-2 was the more common pilus type overall, likely due to 97.4% (683/701) of serotype 7F-ST191 isolates (the most common serotype over the study period) possessing PI-2. 7F was the only serotype where almost the entire cohort demonstrated one specific pilus-positive genotype. Over half of serotype 11A-ST62 isolates (67.2%) possessed PI-2; neither of these two serotypes demonstrated an appreciable amount of MDR when PI-2 genes were present. Few serotype 19A and 35B isolates demonstrated PI-2 alone, (five and one, respectively); however, of these isolates, two and one (40% and 100%, respectively), were MDR.

TABLE 5.8: Pneumococcal pilus presence demonstrated by the twelve serotypes of interest collected by the SAVE 2011-2015 study.

Serotype (n*)	Genotype	% with Genotype (n)	% with Genotype that are MDR (n)
7F (701)	PI-1	0	0
	PI-2	97.4 (683)	0.3 (2)
	Dual	0	0
	None	2.6 (18)	5.6 (1)
19A (578)	PI-1	44.5 (257)	5.1 (13)
	PI-2	0.9 (5)	40.0 (2)
	Dual	21.1 (122)	95.9 (117)
	None	33.6 (194)	9.3 (18)
22F (584)	PI-1	0	0
	PI-2	0	0
	Dual	0	0
	None	100 (584)	1.0 (6)
3 (480)	PI-1	0.4 (2)	0
	PI-2	0	0
	Dual	0	0
	None	99.6 (478)	2.5 (12)
12F (276)	PI-1	0	0
	PI-2	0	0
	Dual	0	0
	None	100 (276)	1.4 (4)
11A (241)	PI-1	1.2 (3)	0
	PI-2	67.2 (162)	0
	Dual	0	0
	None	31.5 (76)	2.6 (2)

9N (198)	PI-1	0.5 (1)	0
	PI-2	0	0
	Dual	0	0
	None	99.5 (197)	0.5 (1)
8 (217)	PI-1	0.5 (1)	0
	PI-2	0	0
	Dual	0	0
	None	99.5 (216)	0.5 (1)
33F (203)	PI-1	0	0
	PI-2	0	0
	Dual	0.5 (1)	0
	None	99.5 (202)	6.9 (14)
15A (212)	PI-1	28.8 (61)	0
	PI-2	0	0
	Dual	0	0
	None	71.2 (151)	63.6 (96)
6C (188)	PI-1	8.5 (16)	6.3 (1)
	PI-2	0	0
	Dual	0	0
	None	91.5 (172)	3.5 (6)
35B (138)	PI-1	37.7 (52)	1.9 (1)
	PI-2	0.7 (1)	100 (1)
	Dual	0	0
	None	61.6 (85)	0
All (4,016)	PI-1	9.8 (393)	3.8 (15)
	PI-2	21.2 (851)	0.6 (5)
	Dual	3.1 (123)	95.1 (117)
	None	66.0 (2,649)	6.1 (161)

*, n with complete results for both PCR reactions. Isolates that maintained double positive or double negative results after repeating were excluded.

The clearest association of pneumococcal pili with MDR was found with the dual PI-1/PI-2 genotype demonstrated by serotypes 19A and 19F (19F data below). Of the serotype 19A isolates tested, 21.1% demonstrated the dual genotype and 95.9% of these isolates were MDR or XDR.

5.7.1 Pneumococcal Pili in Other Serotypes

For background information, PCR to detect the presence of pneumococcal pili was also performed on isolates with a PCV-13 serotype. Not including the vaccine types mentioned above, this includes serotypes 1, 4, 5, 6A, 6B, 9V, 14, 18C, 19F and 23F. PI-1 genes were discovered in several of the PCV-13 serotypes, including 4 (91.2%), 6A (25.8%), 6B (68.4%), 9V (100%) and 14 (33.3%). PI-2 genes were identified in 100% of serotype 1 isolates. As mentioned above, serotype 19F demonstrated the presence of both pilus genes. The dual PI-1/PI-2 genotype was identified in 48.3% of serotype 19F isolates; in addition, 21.6% of 19F isolates possessed PI-1 genes alone, and 5.0% possessed PI-2 genes alone. No pilus genes were found in serotypes 5, 18C or 23F.

5.8 Pneumococcal Surface Proteins

The presence of pneumococcal surface proteins A and B (*pspA*, *pspC)* was assessed for 192 isolates using *in silico* PCR.

5.8.1 *PspA*

Of the 192 isolates tested for *pspA*, four were negative for the presence of any *pspA* family. Family 1 *pspA* variants were identified in 41.5% (78/188) of isolates. Family 1 *pspA* variants were associated with PMEN clones 2 (Spain6B), 4 (Tennessee23F), 9 (England14), 16 (Poland23F), 18 (Tennessee14), 19 (Colombia5), 26 (Colombia23F), 28 (Sweden1), 32 (Denmark14), 34 (Denmark12F) and 41 (Portugal6A). Isolates related to PMEN3 (Netherlands3)

possessed *pspA* variants from this family if they were MDR and part of ST180-clade II.

Serotypes that were not related to international clones, but possessed family 1 variants included

10A, 12F, 18C, 20, 22F, 25F, 31, 33F, 34, 35F, 38 and most serogroup 6 isolates.

Family 2 *pspA* variants were identified in 58.0% (109/188) of isolates. International

clones that possessed family 2 *pspA* variants included PMEN 1 (Spain23F), 3 (Spain9V), 14

(Taiwan19F), 21 (Portugal19F), 24 (Utah35B), 25 (Sweden15A), 30 (Greece21), 35

(Netherlands14), 36 (Netherlands18C), 37 (Netherlands15B), 38 (Sweden4) and 39

(Netherlands7F). Isolates related to PMEN3 (Netherlands3) possessed family 2 *pspA* variants if

they were susceptible to antimicrobials and part of ST180-clade I. Serotypes that were not

related to international clones, but possessed family 2 variants were 8, 13, 16F, 21, 29 and most

isolates within serogroups 11 and 15.

Interestingly, one isolate was positive for a family 3 *pspA* variant. This was the only

serotype 37 isolate included in the WGS analyses; it was not related to an international clone and

did not possess antimicrobial resistance.

5.8.2 *PspC*

Of the 192 isolates tested, 166 resulted in a product for one or both *pspC* primer sets. The

remaining 26 isolates were examined manually for a PCR product; it was found that the product

was split between two contigs, which the EMBOSS-primersearch function was unable to detect.

Of these, only one was negative for *pspC*, resulting in 99.5% (191/192) of the isolates tested

possessing *pspC*.

The manual search of contigs for *pspC* revealed 21 isolates with an abnormally long PCR

product of over 6kb. Sequences demonstrated 99% identity to a *pspC* variant containing an

IS*1167* transposase and an IS*3*-Spn1 putative transposase (NCBI accession no. AF154034), for a

total of 3.7kb of additional sequence. All isolates that possessed this gene variant were related to PMEN25 (ST63), including serotypes 7F, 8, 15A, 19A and 22F.

5.9 Neuraminidases

The presence of neuraminidases A-C (*nanA*, *nanB*, *nanC*) was assessed for 192 isolates using *in silico* PCR. *NanA* was present in 100% of isolates, while *nanB* was present in 88.5% (170/192) of isolates. Isolates of note that did not possess *nanB* were those related to PMEN30 (Greece21), serotype 12F strains that were unrelated to PMEN34 (Denmark12F), serotype 38 and half of the serogroup 6 isolates. Only 44% (85/192) of isolates possessed *nanC*. International clones that possessed this neuraminidase include PMEN 1 (Spain 23F), 9 (England14), 18 (Tennessee14), 25 (Sweden15A), 26 (Colombia23F), 28 (Sweden1), 30 (Greece21), 34 (Denmark12F) and 38 (Sweden4). Some isolates related to PMEN37 (Netherlands15B) also possessed *nanC,* but only those that were the original clone type of serotype 15B and not a variant.

Interestingly, though the expected fragment size of the *nanB* reaction was 492 bp, six isolates had a product that was over 1.7kb. Sequences were found to contain a 1,238 bp insertion with 99% identity to an IS*1239* transposase (NCBI accession no. FQ312030) and 100% identity to an unknown transposase (CP002176). All isolates that possessed this altered *nanB* gene were serotype 15A isolates of ST2613; congruently, ST2613 isolates demonstrated a large ancestral block of recombination containing *nanB* (Section 5.6.1).

6. DISCUSSION

The work presented in this thesis focused on twelve predominant serotypes isolated from invasive *S. pneumoniae* infections in Canada (3, 6C, 7F, 8, 9N, 11A, 12F, 15A, 19A, 22F, 33F and 35B). We hypothesized that these serotypes would frequently be antimicrobial resistant and/or MDR, genetically similar within the serotype and virulent due to the presence of pneumococcal pilus genes. In addition, we hypothesized that capsular switching events would contribute to the increase in non-vaccine serotypes circulating in Canada. We believed that WGS would provide greater discriminatory power over PFGE and MLST to test these hypotheses.

6.1 Connecting the Dots: Associating Antimicrobial Resistance and MDR with Serotype, Genotype and Recombination Events

6.1.1 Serotype Diversity and MDR

In general, serotypes that demonstrated high levels of diversity in this study were also those that had the highest rates of MDR, and demonstrated strong associations with a specific resistance pattern. Serotypes 6C, 15A, 19A and 33F were commonly resistant to antimicrobials (clarithromycin, clindamycin and doxycycline resistance was shared by all four serotypes), with resistance mediated by the acquisition of foreign resistance determinants *ermB* and *tetM*. A recent study of pneumococcal recombination by Croucher *et al*. identified a correlation between serotype diversity and the total number of recombination events experienced (113). Gubbins analysis of these four serotypes noted numerous recombination events, particularly in serotypes 19A, 6C and a subset of 33F. The entire cohort of serotype 19A isolates demonstrated numerous areas of recombination, even in the few strains that were not MDR. Serotype 33F isolates demonstrated many recombination events only if they also possessed resistance determinants. The number of predicted recombination events was generally high in serotype 6C isolates, both

unique and of common descent, regardless of resistance phenotype. This indicates that this serotype frequently participates in recombination, and has the potential to become increasingly MDR in the future; a higher frequency of recombination events would lead to increased chances of obtaining acquired resistance determinants. Fewer recombination events were identified in serotype 15A, however the dataset available largely constituted CC63 isolates; therefore, it could not be ascertained whether susceptible strains demonstrated less recombination than MDR strains, or if isolates belonging to other clonal complexes were more (or less) recombinant.

Interestingly, despite being highly variable and related to the antimicrobial resistant PMEN18 clone, serotype 9N was not commonly resistant or MDR. However, unlike other serotypes where variability was evident through multiple, unrelated STs, the variability in serotype 9N was limited to numerous STs within one clonal complex (CC66). Gubbins analysis revealed few predicted areas of recombination in isolates representing four members of CC66, revealing that in comparison to those types discussed above, there has likely been fewer chances for this clonal complex of serotype 9N to acquire resistance genes through recombination. A recent study determined that the degree of recombination of any *S. pneumoniae* lineage increases with capsule size and duration of carriage (49). As visualization of serotype 9N capsules determined that the type demonstrated a lower degree of encapsulation than other diverse types (19A, serogroup 6) (47), this may account for why the recombination within this lineage of serotype 9N was limited to small changes resulting in SLVs or DLVs of a common founder, as opposed to large acquisitions including resistance genes.

6.1.2 The Continuing Problem of Serotype 19A

Since the introduction and widespread use of PCV-7 in North America, serotype 19A has been associated with antimicrobial resistance, particularly to penicillin. Despite overall decreases

in IPD incidence, studies in the mid-2000s noted a steep increase in serotype 19A-IPD in the years following PCV-7 introduction in the United States (2000) and Canada (2002-2005), accompanied by an increase in penicillin resistance (63, 114). It was for this reason that serotype 19A was included in the PCV-13 formulation; the most recent SAVE study analyses have noted significant decreases in both serotype 19A prevalence and penicillin resistance since this vaccine was put into use in Canada (2010) (77, 115). However, despite decreases in prevalence, serotype 19A was still one of the most common serotypes isolated in the SAVE study (77, 116), was particularly diverse and was often MDR or XDR.

One of the most commonly studied strains of serotype 19A is ST320. This strain was originally identified as a vaccine escape recombinant in the post-PCV-7 era and has continued to be a successful clone well into the use of PCV-13, despite serotype 19A being included in the formulation (117). A Canadian study followed the development of serotype 19A from 1993-2008 and concluded that the emergence of ST320 was the combinatory result of vaccine selection pressure, antimicrobial pressure and the propensity of *S. pneumoniae* to undergo recombination (118). Approximately 25% of serotype 19A isolates typed by MLST in the current study were ST320, and possessed the dual *mefA/ermB* genotype, *tetM,* alterations in all three PBPs, *folA* mutations, *folP* insertions and genes for both pneumococcal pili. Interestingly, a subset of six ST320 isolates possessed a truncated *ermB* variant that resulted in the demonstration of much lower clarithromycin MICs than the usual >32 μg/mL, and full susceptibility to clindamycin. This specific truncated *ermB* variant was recently identified in an individual isolate in the United States, however the serotype or ST of this isolate was not given (119). All six isolates still possessed *mefA*, which functioned reasonably well in three strains, conferring low level resistance to clarithromycin, while in the other three strains it only conferred intermediate MICs.

It speaks to the success of ST320 that these strains have been isolated multiple times in both Western and Central Canada in 2011-14 despite missing a common resistance phenotype in its arsenal. Even though ST320 strains exhibited resistance to most classes of antimicrobials, isolates collected in this study were rarely fluoroquinolone resistant. It is likely that the high fitness cost associated with mutations in DNA replication enzymes prevents fluoroquinolone resistance from propagating within ST320; therefore, at least for the time being, ST320 isolates collected in Canada remain susceptible to one key antimicrobial class.

Gubbins analysis of ST320 isolates identified few unique blocks of predicted recombination, demonstrating instead several regions acquired through common descent. This indicates that ST320 isolates demonstrate little diversity within the clonal complex. The lack of unique regions of recombination suggests that this clonal complex is particularly well adapted as a virulent, resistant and successful clone, and that additional recombination is unnecessary. However, serotype 19A isolates that did not fall within this clonal complex demonstrated numerous regions of predicted recombination, both unique and ancestral; this is consistent with the fact that many of the isolates were capsular switch variants, and perhaps suggests that more recombination is necessary for these variants to become successfully circulating clones. Of the diverse serotypes noted in the previous section, 19A appeared to participate most frequently in capsular switching. In this study alone, serotype 19A isolates were the putative donor strains for multiple capsule recombinations involving serotypes 14, 15A, 15B, 19F, 21 and 23F. Several whole genome studies of *S. pneumoniae* in other countries have noted similar recombinations, including Bulgaria, Germany, Russia and the United States (117, 120–122). Importantly, capsular switching often involves the transfer of whole or partial PBPs, as two (*pbp1A*, *pbp2X*) are located on either side of the genes encoding capsular polysaccharide. A study of *S.*

pneumoniae isolates from an East Asian population noted that recombination has facilitated the consistent spread of β-lactam resistance amongst the pneumococcal population. Similarly, *folA* genes demonstrated manifestations of recombination; interestingly, there was no association between *folP* insertions and recombination (123). As serotype 19A isolates are commonly resistant to penicillin and/or trimethoprim-sulfamethoxazole, and commonly participate in recombination, both capsular and otherwise, this serotype is a key reservoir for the spread of resistance throughout the *S. pneumoniae* population.

6.1.2.1 Rise of Penicillin-Nonsusceptible and MDR Serotype 35B

As previously described, penicillin resistance has commonly been associated with serotype 19A in recent years; before the emergence of 19A, penicillin nonsusceptibility was attributable to five of the seven serotypes included in PCV-7. In the late 1990s, 78% of penicillin-nonsusceptible IPD strains in the United States were serotypes 6B, 9V, 14, 19F and 23F (124). As the use of conjugate vaccines has reduced the number of IPD cases attributed to the above serotypes, other types have become more significant sources of penicillin nonsusceptibility, particularly serotype 35B. Studies in the United States and France have identified penicillin nonsusceptibility in upwards of 80% of serotype 35B IPD isolates collected, which was attributed to the expansion of ST558 (79, 125).

ST558 was identified and described in the United States in 2002 in a cohort of isolates collected as early as 1995; at this time, 69% of serotype 35B isolates were penicillin-nonsusceptible and attributed to ST558, while the rest were predominantly antimicrobial susceptible ST452 (126). A recent publication of American data from 2015-2016 noted that ST452 had decreased during the 2000s, accompanied by a steady increase in ST558 (127). Along with this, there was a trend of increasing MDR serotype 35B isolates associated with the

expansion of ST156 (127, 128). Originally identified as serotype 9V, ST156 was one of the original international clones associated with penicillin-resistance (23). This type has shown a significant ability to persist in the pneumococcal population, despite serotype 9V virtually disappearing as a cause of IPD due to PCV-7 use; successful capsular switch recombination resulting in ST156 has been noted for serotypes 14, 19A, 19F and now 35B (23, 127, 128).

Although this study identified similar STs, resistance patterns and recombination in serotype 35B, it was not to the same extent as has been described in the United States. Approximately 40% of serotype 35B isolates in this study were penicillin-nonsusceptible, with only 2.1% demonstrating a MDR phenotype; in addition, half of the serotype 35B isolates with a known ST were antimicrobial susceptible ST452. In comparison with the most recent American study, which noted 84% ST558, 11% ST156 (assumed to be predominantly MDR) and 5% ST452 amongst serotype 35B isolates (127), our values for nonsusceptible/resistant STs were relatively low and the population skewed towards ST452. It is reasonable to suggest that ST558 will continue to increase in prevalence in Canada; as PCV-13 use continues, serotype 35B is readily available to fill the penicillin-resistant niche vacated by serotype 19A. It is a successful IPD serotype, recently reported in Europe as the serotype most commonly associated with IPD mortality (129). Similarly to 19A, serotype 35B is also a successful colonizer of the nasopharynx, likely due to the presence of pneumococcal pili and the potential to produce significant quantities of biofilm (130). As recombination most frequently occurs in the nasopharyngeal niche, long-term colonization is likely how ST156 came to acquire a 35B capsule. As only one or two ST156-related isolates were identified for serotype 35B in this study, one confirmed in 2011 and one potential from 2015, it is likely that these were isolated cases and therefore not a significant concern in Canada for the time being.

Interestingly, 35B isolates related to ST558 possessed Thr371Ser alterations in the STMK motif of *pbp1A*. This alteration was only seen in one other isolate group in this study: serotype 19A-ST320. ST320 isolates also possessed a Pro432Thr alteration in the SRNVP motif of this gene, which resulted in the highest penicillin MICs ($\geq$4 μg/mL) identified in isolates within the WGS cohort; comparatively, isolates that possessed the Thr371Ala alteration in the STMK region demonstrated MICs $\leq$2 μg/mL. As point mutations in PBPs occur in a stepwise manner, it is possible that in the future, a ST558 strain may acquire the additional *pbp1A* mutation (Pro432Thr in SRNVP); this could result in ST558 strains acquiring higher level resistance to penicillin. If successful expansion of this clone occurred, serotype 35B-ST558 could become as big of a treatment issue as serotype 19A was in the post-PCV-7 era.

6.1.3 MDR, Virulent and Successful Variants of CC63

From a treatment standpoint, one of the more problematic serotypes that became common following PCV-13 introduction was 15A. Serotype 15A was not as prevalent as some of the other types characterized in this study, however it was associated with the highest rate of MDR at 57%. The bulk of the resistance associated with this serotype was due to isolates within CC63, a finding reflected in other countries such as France, Germany and the United Kingdom (125, 131, 132). A study of ten years' worth of serotype 15A isolates collected by Public Health England (2005-2014) noted that 31.5% demonstrated "triple resistance", defined by the authors as resistance to macrolides and tetracyclines coupled with penicillin nonsusceptibility (131). While our study did not use this definition to describe isolates (instead opting to only include full resistance in the MDR definition), approximately 60% of serotype 15A isolates were nonsusceptible to penicillin. Full penicillin resistance was only identified in ST2613, a DLV of ST63. While Sheppard *et al*. did identify "triple resistance" in some SLVs and DLVs of ST63, it

was unclear if any were ST2613 and if any were fully resistant to penicillin (131). A recent study in Ontario also found no ST63 isolates demonstrating full penicillin resistance (133). It is evident from these studies, and the data obtained in our own, that CC63 is not yet associated with full penicillin resistance, with only one variant demonstrating this characteristic.

Interestingly, this study identified a large insertion downstream of *pspC* in CC63 isolates. Iannelli *et al.* originally described this gene variant as "*pspC7*", identified in NCBI reference genome G54 (NC_011072) (134). This variant included both IS*1167* and IS*3*-SpnI, both of which showed more similarity to regions associated with *Streptococcus agalactiae* than to other *pspC* genes from *S. pneumoniae* (134). *S. agalactiae* is rarely seen as a nasopharyngeal colonizer outside of the neonatal period; a study from France reported that only 4% of healthy people ≥11 years of age were colonized with this pathogen (135, 136). Despite this, it is likely that this *S. agalactiae* genomic material was obtained and recombined into ST63 during nasopharyngeal carriage, potentially in a neonate, long before serotype 15A became a common cause of IPD (as the annotated draft genome of *S. pneumoniae* G54 was published in 2001) (137).

One of the most interesting findings pertaining to CC63 was the stark difference between the founder, ST63, and DLV ST2613. Initial testing revealed that the STs differed only by MLST allele profile and penicillin susceptibility; however, WGS analyses provided much more information on these differences. ST63 possessed PBP 1A-24 while ST2613 possessed 1A-34, with the latter also demonstrating a region of recombination including *pbp1A*. While 1A-24 was not associated with motif changes, allele 34 was associated with PMEN2 (Spain6B), one of the original penicillin-nonsusceptible and MDR international clones (23); this recombination event is likely the main reason that ST2613 is commonly penicillin-resistant, while ST63 remains penicillin-intermediate or susceptible. ST2613 also had a predicted recombination area including

ddl, *pbp2B* and *nanB*, three genes located within a 1.5kb region. While the two STs possessed different *pbp2B* genes, they did not result in any motif changes; however, this gene was almost certainly included in the recombination event that resulted in ST2613 possessing *ddl*-4 (associated with PMEN2, amongst others) instead of *ddl*-14 (one of the two MLST allele variations from ST63). Likely also included in this recombination was *nanB*, which varied greatly between STs; ST2613 isolates were found to contain an IS*1239* element in the middle of *nanB*, which BLAST hits linked to both PMEN2 and PMEN9 (England14). This data suggests that a recombination fragment of at least 1.5kb originated from a ST90 strain related to PMEN2; while *ddl* and *nanB* were identical, the difference between *pbp2B* in donor (2B-81) and progeny (2B-75) was minimal, where a lone point mutation resulted in a one amino acid change in the transpeptidase domain.

The initial identification of ST2613 was in a 2005 carriage isolate from Spain (https://pubmlst.org/spneumoniae). Serotype 6B (including the PMEN2 strain) was common prior to the introduction of PCV-7, but rapidly decreased throughout the 2000s following widespread vaccine use (63); serotype 15A was, and still is in many countries, often associated with nasopharyngeal carriage and respiratory infection (132, 138). It is likely that the original acquisition of PMEN2 genes occurred in the early-mid 2000s, while serotype 15A was a prominent carriage strain and prior to PCV-7 effectively removing serotype 6B as a source of pneumococcal disease. Eight ST2613 isolates were identified and characterized in this study: two by MLST alone, two by WGS alone and four by both MLST and WGS. A search of the MLST database noted that most cases of ST2613 have been identified in Canada; in addition to our isolates and the initial Spanish isolate noted above, ST2613 has been deposited only three other times (from Germany, Italy and Spain, https://pubmlst.org/spneumoniae). Although not as

common as ST63, ST2613 was collected from across Canada during all SAVE study years; this suggests that ST2613 was not associated with an isolated event or outbreak, but is rather a successful variant of an equally successful clone that circulates at a low level in Canada.

6.1.3.1 Serotype 8-ST63: A Growing Concern?

Numerous studies have identified serotype 8 to be particularly common in adults and generally susceptible to antimicrobials, despite possessing a high invasive capacity (133, 139–141). The current study identified little antimicrobial resistance in serotype 8, with <1% of isolates demonstrating a MDR phenotype. In general, isolates were related to PMEN33 (Netherlands8), or more commonly ST1480, a finding illustrated in a second study of Canadian IPD isolates (133). However, one serotype 8 isolate was ST63, the clone most commonly associated with MDR serotype 15A. Like other ST63 isolates, the serotype 8 strain in question also demonstrated resistance to macrolides, lincosamides and tetracyclines. Interestingly, Spain has experienced clonal expansion of this MDR variant of serotype 8 related to ST63 (140). These variants were originally isolated in 2004 and restricted to HIV-positive patients in Madrid; however over the course of five years, the clone spread through adults in nine other regions of Spain (140). At the time the authors penned their manuscript, they were not aware of any serotype 8-ST63 isolated outside of Spain; though collected in 2011, the one isolate analyzed in our Canadian study was not deposited as public information in the MLST database until early 2015 (https://pubmlst.org/spneumoniae).

The serotype 8-ST63 isolate identified in this study was virtually identical to the putative serotype 15A donor; only 14 SNVs differed between donor and recipient in the original phylogenetic analysis. This isolate also highlighted the clonal nature of certain pneumococcal surface proteins; the variant serotype 8 possessed the same IS*1167* and IS*3*-SpnI in *pspC* unique

to ST63, and tested positive for the presence of *nanB* and *nanC* neuraminidases, while all other serotype 8 isolates were negative for these virulence genes. A notable difference between the isolate identified in our study and those spreading throughout Spain is fluoroquinolone resistance: the serotype 8 variants causing disease in Spain had, at a minimum, a *parC*-Ser79Phe mutation conferring low-level resistance to ciprofloxacin (140). Regrettably, an isolate with this mutation was responsible for a fatal levofloxacin failure in the Southern region of Spain; treatment failure was associated with the emergence of an additional *gyrA* mutation after four days of therapy (142). As this serotype 8-ST63 strain rapidly adapted to the pressure of fluoroquinolone treatment, it is fortunate that the Canadian strain did not possess alterations in *parC* and *gyrA*. If this clone were to spread throughout Canada as it did in Spain, MDR serotype 8-ST63 could become a strain of concern in normally healthy adult patients. However, since the initial identification of this serotype 8-ST63 isolate in 2011, there have been no other serotype 8 isolates that fit the MDR pattern associated with ST63. As the isolate was collected from Ontario, a populous province with one of the busiest Canadian airports for European travel, it is possible that this isolate was a one-time case brought back from Spain. The fact that the isolate did not possess the fluoroquinolone resistance associated with the Spanish isolates may also suggest that the variant was the result of an independent capsular switch, the result of which did not become successful in the Canadian *S. pneumoniae* population.

6.1.4 The Two Sides of Serotype 3-ST180

Serotype 3 has long been a type of interest as it is associated with the lowest vaccine effectiveness post-licensure of all serotypes included in PCV-13 (143). Additionally, it has not sustained significant reductions in prevalence post-PCV-13, as with serotypes 7F and 19A; this trend has been noted in numerous countries, including Denmark, Portugal and the United

Kingdom (131, 144, 145). In this study, serotype 3 demonstrated little diversity by MLST, with most isolates belonging to the predominant ST180 clonal complex (PMEN31). However, phylogenetic analysis and subsequent Gubbins analysis of serotype 3 revealed that ST180 isolates fell into two different clades: clade I isolates were susceptible to antimicrobials and demonstrated no regions of predicted recombination, while clade II isolates were resistant to antimicrobials and possessed numerous blocks of predicted recombination. These findings were mirrored in a 2013 study of a small international collection of CC180 isolates, which found that most were unaffected by recombination, having little diversity and appearing "frozen" from an evolutionary standpoint (clade I). However, other CC180 isolates in this collection exhibited significant accumulation of genetic variation, although little antimicrobial resistance was seen in this cohort (clade II) (146). In the current study, ST180 isolates belonging to clade II often possessed three acquired resistance determinants conferring resistance to four different antimicrobials; notably, the clade II group was the only cluster of isolates in this study to consistently possess the chloramphenicol resistance gene *cat.* As the 2013 study discussed isolates collected during 1993-2007, it is possible that the later collection date of the clade II isolates in this study (2011-14) allowed increased time to acquire resistance genes through recombination events.

A more recent study by Azarian *et al.* included isolates collected from 24 different countries from 1993-2014. It was determined that 19% of CC180 isolates belonged to clade II, and that approximately 26% of clade II isolates possessed *ermB* and *tetM*, in comparison to one-half of tested ST180 isolates belonging to clade II and 100% possessing *ermB* and *tetM* in the current study (147). It is important to note, however, that the sample of ST180 isolates characterized in this study was not fully selected at random, as many MDR isolates were chosen

specifically for that reason; therefore, the true percentage of clade II-ST180 isolates in Canada is considerably lower than 50%. Interestingly, Azarian *et al.* also determined that clade I and clade II ST180 isolates differed in their surface protein antigens, most notably *pspA*. Clade I isolates possessed family 2 *pspA* variants, while clade II isolates possessed family 1 variants (147); this finding was confirmed in our study. An additional gene that separated the two clades of ST180 was *pbp2X*. Clade II-ST180 isolates possessed the PBP-typing pattern 1A-2, 2B-0, 2X-111, while clade I isolates demonstrated the pattern 1A-2, 2B-3, 2X-2. While the change in *pbp2B* did not result in any modifications to key amino acid motifs, the alteration of *pbp2X* from allele 2 to 111 resulted in a Thr338Ala mutation in the STMK motif. This alteration was not seen in the American study by Metcalf *et al.* which described the PBP-typing scheme utilized in our project; ST180 isolates that likely belonged to clade I demonstrated the same PBP-typing pattern as above, while MDR isolates that were likely part of clade II possessed a hybrid pattern of 1A-2, 2B-0, 2X-2 (38). It is possible that the clade II isolates included in the American study have become successful without acquiring a different *pbp2X*. However, individual amino acid alterations are the first step in acquiring penicillin-nonsusceptibility; therefore, the clade II isolates with altered *pbp2X* genes obtained in this Canadian study are more worrisome, as they are closer to acquiring resistance to a fifth antimicrobial class.

Interestingly, only one serotype 3 isolate possessed unique areas of recombination in addition to those obtained through common descent. This isolate clustered as a part of clade II by phylogeny, and demonstrated the same resistance pattern as other clade II isolates; however, it also possessed the *pspA* family and *pbp2X* transpeptidase domain variants of a clade I isolate. The possession of traits from both clades could indicate that this is a hybrid isolate that does not truly belong to either clade; however, the presence of unique recombination regions could

indicate that this strain is diversifying into a new variant of ST180.

The prevalence of clade II isolates, frequent antimicrobial resistance, varying surface antigen profiles and potential to diversify indicates the need for additional screening of serotype 3 isolates in Canada. However, as MLST does not discriminate between isolates of the same ST, WGS will be crucial in separating these very different clades of ST180.

6.1.5 Serotype 12F and IPD Outbreaks

Serotype 12F is a type of *S. pneumoniae* that is not normally associated with nasopharyngeal carriage in healthy individuals, but is instead a common cause of IPD outbreaks (148). In recent decades, serotype 12F has been responsible for a number of IPD outbreaks in the United States, and most recently, in Winnipeg, Manitoba, Canada in 2008-11 (148–150). As collection of this serotype in Western Canada has been significantly higher than other Canadian regions since the outbreak (139), it was of particular interest for further study. Most outbreaks involving serotype 12F have been associated with ST218, an international clone originally isolated in Denmark in 1995; prior to 2008, the outbreaks in the United States were caused by strains that demonstrated susceptibility to antimicrobials (150). Interestingly, the Canadian outbreak introduced a clone of ST218 possessing *mefA*-mediated macrolide resistance that has since begun circulating in the United States (150). In the current study, <40% of serotype 12F isolates were susceptible to clarithromycin, and over 90% of typed isolates were related to ST218. Despite initially being isolated in Manitoba, macrolide-resistant ST218 was collected from Central Canada starting in 2011 (the beginning of the SAVE study). The continual presence of the macrolide-resistant clone indicates that it was, and continues to be, a successful cause of IPD in the adult population, and has been gradually replacing the susceptible version as the dominant clone in the populations of Canada and the Northern United States.

In contrast to serotype 12F in Canada, a recent study in Europe identified few macrolide resistant strains of the outbreak-prone serotype. A detailed study of emerging serotype 12F in France found that less than 20% of tested isolates were related to ST218, with many of the rest belonging to clonal clusters that (to our knowledge) have not been identified in Canada. In addition, antimicrobial resistance was instead to chloramphenicol and tetracycline; as no macrolide resistance was evident, the authors attributed this dual resistance to the presence of a composite Tn*5253*-related transposon (125). The clone responsible for this dual resistance was ST989, which has been growingly associated with MDR around the world. A recent report by The Global Pneumococcal Sequencing Project (GPS) indicated that ST989 was the most common serotype 12F-associated clone in their collection of isolates, and demonstrated resistance phenotypes concordant with the presence of *cat*, *tetM* and *folA* alterations (151). One isolate in our study typed as ST989, and demonstrated resistance to chloramphenicol, clarithromycin, doxycycline and trimethoprim-sulfamethoxazole; unfortunately, this isolate was not included in WGS analysis therefore the genotype is unknown. Like the Janoir *et al*. study from France, the GPS report did not describe macrolide resistance in their pool of ST989 isolates (125, 151); it is possible that resistance to this antimicrobial class is unique to Canadian or North American isolates, as the varying STs collected in the French study suggests a continental difference in serotype 12F clones causing IPD.

Interestingly, two serotype 12F isolates analyzed in the current study were MDR and unrelated to both ST218 and ST989; resistance to clarithromycin, clindamycin and doxycycline was mediated by *ermB* and *tetM*. A recent study of serotype 12F in Canada postulated that, based on the antimicrobial susceptibility testing results of a 2013 report from PHAC-NML, strains with clindamycin and doxycycline resistance were the result of macrolide-resistant ST218 acquiring

additional resistance (150); however, the results of our current study suggest that this proposal is unlikely. STs 6945 and 10198 differ from ST218 at all seven MLST loci; a search of the MLST database found that both MDR STs are SLVs of ST1527, a clone originally associated with the first wave of a serotype 12F-associated IPD outbreak in Alaska (148). Additionally, our study found macrolide resistance in ST218 to be mediated by *mefA*, while resistance to both clarithromycin and clindamycin would generally suggest the presence of *ermB*. It is therefore likely that the novel, MDR STs of serotype 12F identified in this study evolved from the ST1527 outbreak strain to possess both *ermB* and *tetM* together in a Tn*2010*-related transposon. An additional factor that differed between the two types of isolates was the presence of the neuraminidase B gene; ST218 isolates possessed *nanB* (along with A and C) while the MDR strains did not. Neuraminidase A has been proven an essential factor for adherence in nasopharyngeal colonization, but recent studies have noted that *nanB* deficient mutants demonstrate reduced *nanA* activity; this suggests *nanB* may have a role in regulating the expression of *nanA*, and thus the ability of a strain to colonize hosts (57). This suggests that MDR serotype 12F strains related to ST1527 may be even less likely to be associated with nasopharyngeal carriage and more so with invasive disease. As serotype 12F is a known outbreak type, these new MDR clones are of particular interest for future study in Canada.

The results discussed in Section 6.1 were inconsistent with the original hypothesis that prevalent serotypes would be genetically similar within the serotype. This study instead demonstrated that while many serotypes are genetically similar, just as many are incredibly diverse. WGS analyses identified several serotypes that have a predicted rich history of recombination, thus strains within these serotypes possessed different resistance mechanisms, surface proteins and STs than other isolates of the same type.

6.2 Clinical Relevance of Capsular Switches

Capsular switching is a regular occurrence in the pneumococcal population and was occurring long before the advent and introduction of conjugate vaccines. A study of a historical sample of *S. pneumoniae* isolates noted that over 90% of capsular switch variants were isolated prior to the introduction of PCV-7 in the 2000s, and variants were evenly distributed through time (67). It is clear that both known (vaccine use, antimicrobial use) and unknown selection pressures favour specific combinations of genetic variables causing a natural fluctuation in the population (4). This study categorized capsular switches based on the properties of the strains involved: either a vaccine escape recombination, an event increasing serotype diversity, or an event spreading antimicrobial resistance.

Vaccine escape recombinations are one of the most worrisome types of capsular switching events, as these recombinations have the potential to seriously undermine the effectiveness of conjugate vaccines. One of the first well characterized vaccine escape events was noted in the United States, where PCV-7 vaccine serotype 4-ST695 isolates obtained a serotype 19A capsule and were thus no longer covered by the vaccine (66). A number of vaccine escape recombinants were identified in this study, most of which can be attributed to the pressure of a specific conjugate vaccine. PCV-7 and PCV-13 serotype 23F was commonly involved in recombination events in this study. A serotype 6C variant of 23F-ST338 was collected in late 2013, a few years after PCV-13 was introduced in Canada, therefore the vaccine escape event was likely driven by PCV-13 use. A serotype 19A variant of 23F-ST81 was collected in early 2011; of these two serotypes, only serotype 23F was included in PCV-7, while 23F and 19A are both included in PCV-13. This suggests that the recombination event was driven by PCV-7 and occurred very early on in PCV-13 use, while 19A was still a viable serotype to "escape" to.

Lastly, the 9V-ST156→35B-ST156 variant identified in this study was isolated in early 2011, during the very early stages of PCV-13 use in Canada. As serotype 9V is included in both PCV-7 and PCV-13, it is likely that this vaccine escape event was driven by use of PCV-7. As previously discussed, this variant is not unique to Canada; it has become widespread and concerning in the United States due to its rapid expansion and MDR phenotype. Interestingly, the initial isolation of this variant occurred two years earlier in the United States than in Canada. As PCV-7 was introduced in the United States at minimum two years earlier than in Canada, it is possible that vaccine pressure at an earlier date caused this variant to arise proportionally earlier (127, 128). This is likely the most concerning vaccine escape event described by this study; serotype 19A was eventually included in PCV-13, a vaccine which also provides a degree of cross-protection against serotype 6C (152, 153). As this variant is already common in one large country, and as serotype 35B is not included in any current pneumococcal vaccine formulation, this escape poses a great threat to vaccination efforts.

In the years following PCV-7 introduction, studies in the United States found that although vaccine use had decreased antimicrobial-resistant infections, the relative prevalence of antibiotic resistance had not fallen as dramatically. This was attributed to the fact that certain MDR clones were growingly becoming associated with multiple capsular types (117). In the current study, this was seen particularly with MDR CC63; commonly associated with serotype 15A, this study identified four other serotypes within the cluster exhibiting the same MDR phenotype and genotype. As previously discussed, serotype 8-ST63 was noted to emerge and spread throughout Spain, possessing fluoroquinolone resistance not normally associated with CC63 in Canada (140). Serotype 19A-ST63 has been noted at low levels in a number of different countries over the past decade (38, 154, 155). To the best of our knowledge, serotype 7F-ST63

has not been described in any other countries. The CC63 variant serotype 22F-ST9352 has only been described in previous publications by our group (156, 157). Another interesting case of increased resistance was in a serotype 29 isolate that recombined with serotype 35B-ST558 to obtain its macrolide and intermediate penicillin resistance. Similar progeny strains were noted in both France and the United States, bearing either full or intermediate resistance to penicillin and variable resistance to macrolides (https://pubmlst.org/spneumoniae). These isolates ranged in collection date from 1998-2011, while the progeny strain described in this study was identified in mid-2014. This is more than three years after PCV-13 introduction in Canada and may therefore indicate that the recombination event was the result of the rise in prevalence of serotype 35B. These variants could pose serious problems for treatment of disease associated with these serotypes; as they are normally not associated with resistance, there is an increased chance of treatment failure, morbidity and mortality due to the organism being resistant to the empiric treatment.

Some of the putative variants noted in this study that we were not able to fully confirm have been noted by authors in other countries. Switches with serotype 14-ST230 have been noted to generate diversity in serotype 19A in Bulgaria (121); the same clone was seen to escape vaccine action in the early 2000s in Italy by switching to serotype 24F (158). Serotype 11A-ST156 was recently identified in Spain, adding credence to the concept that ST156 is one of the most variable and enduring STs (159). The diverse range of serotypes that participated in capsular switching events underlines the importance of sequencing a broad range of serotypes as part of routine surveillance; many of the switches identified in this study would have gone unnoticed if not for the selection of random background isolates.

It is important to note that while these strains were characterized as variants that could

pose a threat to vaccine effectiveness or antimicrobial treatment regimens, they may not necessarily become successful clones in the Canadian healthcare setting. It is difficult to say whether more than one or two isolates of a certain variant clone were isolated during the study, as only a subset were fully characterized with molecular testing. Therefore, while these results do partially confirm the hypothesis that capsular switching contributed to the expansion of non-vaccine serotypes, it is with the caveat that the extent of the role could not be quantified in this study. Further surveillance and extensive molecular characterization are required to determine to what extent capsular switching occurs in Canada, and whether the variant strains identified in this study continue to cause disease.

6.3 Pneumococcal Pili: Association with Serotype, Sequence Type and Potential Impact on Disease Burden

Initial studies reporting the identification of pneumococcal pili have described the genes as a clonal property (54, 160). In this study, there was a clear correlation between pilus gene presence and several predominant clones. This included serotype 7F-ST191 (PI-2), 19A-ST416 and ST695 (PI-1) and 19A-ST320 (PI-1 and PI-2), all of which have been previously noted in studies performed in Italy, Portugal and the United States (38, 54, 160, 161). Though other studies have also observed the lack of pili in MDR serotype 15A-ST63 isolates, as well as PI-1 presence in 15A-ST3811 (38), this is one of few studies to note that susceptible 15A-ST58 isolates often possessed PI-1. It is interesting to note that while serotype 15A often possessed pilus genes in this cohort of isolates, it appears that piliation in this serotype conforms to an inverse relationship with antimicrobial resistance; no MDR serotype 15A isolate was piliated, with the opposite also being true. This fact was true for all isolates within CC63, regardless of serotype. It is unlikely that the lack of pilus genes in these MDR isolates is related to a fitness

cost; serotype 19A-ST320 is almost invariably XDR, possesses both pilus islands and is an extremely successful clone. As previously described, studies have noted that serotype 15A is still a common colonizer and cause of noninvasive infection, therefore one could speculate that possession of a pilus may be advantageous. It remains unclear why MDR CC63 isolates lack pilus genes; perhaps it is just a matter of the clone not yet obtaining the correct DNA for recombination.

Interestingly, this study identified comparatively more serotype 11A isolates containing PI-2 genes, and fewer 6C isolates containing PI-1 genes than a recent Active Bacterial Core surveillance (ABCs) study in the United States (38). The American study found that only 38% of serotype 11A-ST62 isolates contained PI-2 in comparison to almost 70% in the current study. In general, piliation in 11A-ST62 isolates has been variable depending on the study; Zahner *et al*. noted that this variability in PI-2 presence indicates that piliation is not essential for serotype 11A to cause invasive disease (54); as serotype 11A has been demonstrated to be relatively clonal, this statement can likely be applied to the vast majority of 11A isolates. Conversely, while the ABCs study identified PI-1 genes in approximately 40% of serotype 6C isolates, the current study found <10% with this trait. However, in both studies, the piliated serotype 6C isolates also possessed at least one determinant or mutation conferring resistance to antimicrobials (38). In addition to being a common cause of IPD, serotype 6C has been reported to be a common colonizer and source of respiratory infections in both Canada and the United States (162, 163); it is possible that piliated 6C isolates are better equipped to adhere to the nasopharyngeal epithelium, allowing for greater chances of carriage and acquisition of resistance genes.

Pneumococcal pilus genes were also identified in several vaccine serotypes, including 1,

4, 6A, 6B, 9V, 14 and 19F. A study published around the time of PCV-13 introduction in

Portugal illustrated an overall decline in pneumococcal pilus frequency, as many piliated types

were contained in PCV-7 (164). As PCV-13 serotypes 7F and 19A also commonly contain pilus

genes, it is reasonable to assume that the frequency of pili will decrease even more with PCV-13

use, as the prevalence of disease caused by serotypes 7F and 19A should continue to decrease in

the coming years.

6.3.1 The Importance of Other Pneumococcal Virulence Factors

Pneumococcal surface proteins and neuraminidases play a key role in immune evasion

and colonization of epithelium. PspA and PspC have been implicated in the inhibition of

classical and alternate complement pathways, respectively (55, 165). Neuraminidases have been

associated with the promotion of colonization through exposing host cell receptors and

promotion of biofilm formation (58, 166). As whole genomes were available for 192 isolates in

this study, *pspA*, *pspC*, *nanA*, *nanB* and *nanC* were investigated to determine whether these

genes were ubiquitous or variably present in different serotypes or clones.

As described in the results, PspA families were highly associated with specific clonal

types; 41.5% possessed a family 1 variant, while 58.0% possessed a family 2 variant. This ratio

is consistent with a study performed by Hollingshead *et al.* that tested an isolate set from

Australia, France, Spain, Sweden, the United Kingdom, Canada and the United States, which

identified very similar values of 40.6% and 58.5% for family 1 and family 2 PspA variants,

respectively (167). This suggests that the clonal composition of pneumococci in the additional

six countries is relatively similar to that of Canada; in contrast is Japan, where family 2 PspA

variants were found in only 49.4% of isolates, indicating their clonal composition may vary from

that seen in Canada (168). It has been suggested that the ubiquitous nature of PspA would make

it an attractive option for a protein-based pneumococcal vaccine (165, 167). As this study and others have determined PspA to be highly variable, a combination of three PspA antigens (family 1 – clade 2 and family 2 – clades 3 and 4) has been suggested to elicit optimal protection against most *S. pneumoniae* strains (167, 168). While it has not been studied in as much detail as a vaccine target, PspC is also commonly present on the pneumococcal cell surface and has many structural similarities with PspA, including variation (55). The only variant noted in this study was "*pspC7*", present in CC63 isolates and unique due to its two downstream insertion sequences. This variant was originally reported in 2002 and has not been discussed in great detail since (134); it would be interesting to study recent isolates from other countries to determine if this variant is present around the world or solely in Canadian isolates.

Pneumococcal neuraminidases have also been identified as potential targets for protein-based vaccines. Our study identified *nanA* in 100% of isolates tested, as well as *nanB* and *nanC* in 88.5% and 44.0% of isolates, respectively. An ABCs study conducted by Pettigrew *et al*. resulted in similar prevalences of *nanA*, *nanB* and *nanC* at 100%, 96% and 51%, respectively (58). NanA has been found to be essential for colonization and pathogenesis, making it the best possible vaccine target due to its ubiquitous nature (58, 169). As discussed in Section 6.1.5, recent studies have suggested that NanB may have a role in regulating the expression of *nanA*, and thus the ability of a strain to colonize hosts (57). While a number of isolates did not possess *nanB* and therefore may be more proficient at invasion than colonization, ST2613 isolates possessed *nanB* interrupted by an insertion sequence. As the presence of insertion elements renders a gene inactive, ST2613 may be deficient in the essential NanA protein; this may help explain why studies have not recently discussed identification of ST2613 from carriage. Lastly, *nanC* was identified in the lowest proportion of isolates. Pettigrew *et al*. noted that *nanC* was

isolated more frequently in invasive isolates than in carriage isolates, particularly those obtained from cerebrospinal fluid (CSF); the authors speculated that *nanC* contributes to the ability of a *S. pneumoniae* strain to cause meningitis (58). In our study, only 2/85 isolates possessing *nanC* were collected from CSF, with the remaining 83 collected from blood. While it is possible that the 83 cases eventually progressed to meningitis and we were unable to access a specimen, this is unlikely. Alternatively, Pettigrew *et al.* postulated that rather than being associated with tissue-specific virulence, *nanC* is instead associated with serotypes or STs that frequently cause meningitis (58). This theory is more appropriate for the data collected in our study; we identified *nanC* in serotypes 1, 4, and multiple clones of 14 and 23F, all of which have been commonly associated with pneumococcal meningitis (170). Future sequencing projects should include a greater number of CSF isolates and commonly meningitis-associated serotypes to confirm the relationship between *nanC* and pathogenesis.

6.4 Serotypes 22F and 33F: An Improvement to the Conjugate Vaccine Formulation?

The 15-valent pneumococcal conjugate vaccine currently undergoing clinical trials in the United States differs from PCV-13 in that it includes serotypes 22F and 33F. At the time of the initial formulation of this vaccine (2007), serotypes 22F and 33F accounted for approximately 10% of IPD cases in American adults (70). A more recent surveillance study of ABCs-collected isolates determined that these two serotypes accounted for 18% of adult and 21% of child IPD cases in 2013 (171). Similar increases in prevalence of one or both serotypes have been noted in Denmark, Germany, Italy, Japan and Sweden (144, 172–175). In addition to increasing prevalence, serotypes 22F and 33F possess increased invasive capacity compared to other non-PCV-13 serotypes, and have been shown in the United Kingdom to cause the highest loss of quality adjusted life years in pediatric patients (176, 177).

Serotype 22F was the most commonly collected serotype in SAVE 2015 (77), and was identified as being highly clonal. The predominant serotype 22F clone was ST433, a finding that has been noted in molecular studies performed by many other countries, including Japan, Sweden and the United States (173, 178, 179). Roughly one-quarter of serotype 22F isolates in this study were resistant to clarithromycin, mediated by either *mefA* or *ermB*; however, a similar Canadian study performed WGS analyses on a much larger sample of serotype 22F isolates and noted *mefA* to be the most common macrolide resistance determinant in this clone (180). The clonal nature of this serotype bears striking resemblance to serotype 7F; one of the most clonal serotypes described in this study, almost all tested 7F isolates were ST191 or a variant related to PMEN39, and few were resistant to antimicrobials. Once the most common serotype isolated by the SAVE study, use of PCV-13 resulted in a dramatic and significant decrease in prevalence of serotype 7F (77). Studies have estimated the specific PCV-13 vaccine effectiveness for serotype 7F to be over 90% (143); the clonal nature of this serotype, and thus the lack of serotype variability may have contributed to the success of opsonophagocytic killing of serotype 7F in PCV-13. As serotype 22F shares many properties with vaccine-success serotype 7F, it is possible that 22F will react similarly to vaccine use when PCV-15 becomes available.

Though the SAVE study identified serotype 33F less frequently than 22F, serotype 33F demonstrated a statistically significant increase in prevalence from 2011-2015 (77). As previously described, serotype 33F diversified over the course of the study, expressing novel STs and antimicrobial-resistant phenotypes. To date, few studies have done in-depth molecular characterization of this serotype. A 2017 study done in Italy noted that ST717, an antimicrobial-resistant ST seen in our study, had emerged in 2012-2014. This strain commonly demonstrated resistance to macrolides, though the mechanism was unclear (175); in our study, ST717 and

related isolates possessed *ermB* and *tetM*-mediated resistance. A 2016 study of American isolates noted similar resistance mechanisms in their ST717 isolates (38). This study also found that isolates related to ST100 were commonly resistant to clarithromycin, trimethoprim-sulfamethoxazole or both, mediated by *mefA* and *folA/P* alterations. The same phenotype was identified in our study, however the genotypes were not verified. As in our Canadian study, the American publication detailed five new STs identified in serotype 33F; unlike our study, these novel STs demonstrated the antimicrobial resistance characteristics of ST100 - all isolates had resistance patterns mediated by *mefA* and *folA/P* alterations (38). In the current study, our novel serotype 33F STs most commonly possessed *ermB* and *tetM* and demonstrated relatedness to ST717. This suggests that serotype 33F is diversifying from two different antimicrobial resistance lineages, and is currently geographically variable: in the United States, ST100 is diversifying into new STs, while ST717 is diversifying in Canada. The ease of adaptation of serotype 33F and the presence of antimicrobial resistance genes indicates this serotype was a good choice for inclusion in the conjugate vaccine formulation.

6.4.1 Serotypes to Consider in Future Conjugate Vaccine Formulations

Of the serotypes discussed in this study, many merit inclusion in future conjugate vaccines. It would of course be ideal if a vaccine could be effective against all pneumococcal capsular types; however, the practicalities of manufacturing a vaccine make it unlikely that an "omni"-vaccine will be available any time soon. Of the seven serotypes discussed in this study that are not currently (or will soon be included) in PCV-13 or PCV-15, many can be ruled out as future types of inclusion. The presence of serotype 6A in PCV-13 has allowed for a partial cross-protective response against structurally and immunologically related serotype 6C (152, 153), therefore it is likely not a priority for inclusion at this time. The same could be suggested for

PCV-13 serotype 9V cross-protecting against emerging serotype 9N, however this has not been studied in detail in recent years. Serotypes 8 and 11A did not demonstrate any appreciable antimicrobial resistance, and were relatively clonal; these types, along with 9N, are at least currently present in the PPV23 vaccine.

The results of this study suggest that a serotype warranting serious consideration for future inclusion is 35B. Our study, as well as studies in the United States and France, have identified two increasingly prevalent and resistant clones (77, 125, 127); both pose problems for disease burden and treatment. Serotype 35B has been shown to effectively colonize the nasopharynx as well as cause invasive disease (130); in addition, ST558 threatens the use of the most common anti-pneumococcal agent, penicillin, while ST156 poses an old threat, as a MDR clone that has caused problems over many years. There is currently no coverage for this serotype in any pneumococcal vaccine, as it nor any related types are included in PPV23.

Two other serotypes worth noting are 15A and 12F. Serotype 15A makes up the majority of one of the most commonly MDR clonal complexes identified in this study (CC63); however, it has not seen the significant increase in prevalence noted for other prevalent, emerging serotypes. As noted in the introductory SAVE material, serotype 15A has exhibited an overall decreasing trend from 2011-2015; future monitoring of this serotype will be necessary to determine if the decrease continues, or if a sharp increase warrants its inclusion in a conjugate vaccine. Serotype 12F would be of interest for inclusion as it is an outbreak serotype. Outbreaks of IPD cause a strain on the healthcare system, thus the ability to vaccinate at-risk populations against this type would be appealing. However, most reported outbreaks have occurred in adults outside the range of vaccine-eligibility, including those who are generally healthy, and those living in rural or marginalized communities (148–150). As serotype 12F continues to expand as

macrolide-resistant ST218 and MDR ST989, outbreaks could result in increasing morbidity and mortality due to a decreasing number of treatment options. Including this serotype in a conjugate vaccine would at least ensure that future outbreaks would be unlikely to occur in vaccine-eligible children under the age of two and adults over 65.

6.5 Temporal Analysis of Serotypes 3 and 33F: From BESST (2007-2009) to SAVE (2011-2014)

The eight years' worth of isolates examined by WGS in this study did not provide a particularly lengthy period in which to identify temporal trends. As a new vaccine pressure was introduced halfway through the study period, it is possible that we only caught the tail-end of PCV-7 effects to the pneumococcal population, and just the beginning of PCV-13 effects. Despite this, serotypes 3 and 33F appeared to demonstrate trends over the course of the BESST (2007-2009) and SAVE (2011-2014) studies.

Serotype 3 demonstrated a trend towards increasing prevalence of ST180-clade II isolates. WGS analysis identified ST180-clade I isolates from 2008-2012, while all ST180-clade II isolates characterized in this study were obtained from 2011-2014. Examining the data collected from traditional MLST and comparing it to antimicrobial susceptibility results indicated that ST180-clade I isolates likely continued to be present up to and including the SAVE 2015 study year. Additionally, five other putative clade II isolates were identified from SAVE 2013-2015: one ST180 isolate not included in the WGS sampling, and four un-typed isolates that matched the MDR pattern demonstrated by clade II isolates in this study. Including those analyzed in the current study, BESST did not collect any MDR serotype 3 isolates, suggesting that MDR ST180-clade II isolates have become common only in recent years. However, Croucher *et al*. indicated that many historical ST180-clade II isolates demonstrated

little antimicrobial resistance (146), meaning that clade II isolates may have been present in the BESST study, just without a resistance phenotype. This fact aside, it is clear that at minimum the MDR-nature of ST180-clade II isolates in Canada is a recent trend. As previously discussed, serotype 3 has not demonstrated the projected decrease in prevalence noted in other PCV-13 serotypes (131, 144, 145). Although a rapidly diverging ST may be a good explanation for this, any hypotheses that the appearance of ST180-clade II participated in the persistence of serotype 3 appear to be unlikely. Azarian *et al.* performed time scale phylogeny analyses on serotype 3 isolates, determining that the first ST180-clade II isolates were collected long before the introduction of PCVs in any country (147). This also indicates that the emergence of clade II is not likely associated with vaccine effectiveness, meaning that the reason behind the persistence of serotype 3 remains unknown.

From 2007 to 2014, serotype 33F demonstrated a trend towards increasing diversity. As previously discussed, our study identified numerous STs within serotype 33F, including ST100, ST717 and variants, as well as newly identified STs demonstrating antimicrobial resistance. ST100 was collected throughout the BESST and SAVE study periods, however the novel and resistant STs were only identified beginning in 2011. Although only two serotype 33F isolates from the BESST study were fully characterized, overall none exhibited a MDR phenotype (30) and only one demonstrated low-level resistance to clarithromycin (H.J. Adam, unpublished data); this suggests the presence of *mefA,* which has been previously linked to ST100 (38). The lack of resistance in the serotype 33F isolates collected by the BESST study suggests that most were related to antimicrobial susceptible ST673, or an older clone of ST100 that had not yet acquired *mefA*. The ST100 and related isolates tested from both studies demonstrated few areas of predicted recombination, in contrast to the novel STs that had numerous areas. The most recently

collected isolate analyzed by Gubbins also had the largest presence of unique blocks of recombination, suggesting that this serotype will continue to diversify over time.

As MLST and WGS were not performed on the entire cohort of isolates, and the only molecular work performed with BESST isolates was on the 30 included in this study, it is impossible to calculate a true prevalence of the novel types discussed. Acquiring a more historical set of Canadian isolates for molecular characterization (from at least the pre-PCV-7 period, if not earlier) would enable a more accurate description and characterization of trends within *S. pneumoniae* isolates collected in Canada.

6.6 Comparison of Genetic Characterization Methods

This study held the great advantage of being able to utilize a variety of methods to characterize the genetic properties of strains. Over the course of five years, PFGE (once the gold-standard for bacterial strain typing) waned in popularity to be replaced by MLST. Now, the increasing availability and decreasing costs of WGS make it the obvious choice for characterizing bacterial populations.

PFGE has been described as labor intensive and subjective; it also suffers from a lack of a centralized database for *S. pneumoniae* which results in the method generally not being reproducible between laboratories (181). PFGE is less expensive than the sequence-based MLST methodology, however the centralized database for MLST data (http://pubmlst.org/spneumoniae) makes it a highly portable and unambiguous method to type isolates (181). In this study, PFGE and MLST were generally able to discriminate equally well between different clones within the same serotype. This has been previously reported for organisms such as *Enterococcus faecalis* and *Staphylococcus aureus* (182); however, other studies found that PFGE was more discriminatory than MLST for food-borne and outbreak related organisms such as *Salmonella*

enterica serovar Typhimurium and *E. coli* O157:H7 (182, 183). In this study, PFGE provided greater discrimination for serotypes 12F and 3. As serotype 12F is an outbreak-related serotype, this makes sense as PFGE was once an essential tool in outbreak investigation; PulseNet International utilized PFGE for 20 years to differentiate food-borne outbreak cases from geographically- and temporally-associated sporadic cases (184). Serotype 12F isolates were predominantly typed as ST218 by MLST, however this designation could not differentiate between macrolide-resistant strains or detect variations in ST218 collected from different regions of Canada. PFGE successfully clustered the macrolide-resistant ST218 strains together (Canadian outbreak strain), and separated the susceptible ST218 isolates by region of isolation. Like 12F, serotype 3 isolates were mostly typed as ST180 by MLST. We now know from phylogenetic and recombination analyses that ST180 isolates can be members of two very distinct clades; MLST was unable to differentiate between the clades, while PFGE was at least able to cluster these isolates apart, indicating some difference between them. Despite the previously described pitfalls of PFGE, the advantage remains that this methodology uses the total genetic content of an organism; MLST may miss the big picture as it only studies seven slowly adapting genes.

Although PFGE uses the whole genetic content of an organism, it does not distinguish strains at the nucleotide level as WGS does. As described by Gilmour *et al.*, WGS represents the "ultimate epidemiological typing method", as it encompasses the entire genetic content at the single nucleotide level and is universally applicable (73); once utilized only for research purposes, this method is becoming the new gold-standard for public health investigations and clinical microbiology (73, 75). As hypothesized, WGS was clearly the best characterization method for our isolates based on the sheer amount of information generated; the whole genome

of each isolate was able to provide a ST, antimicrobial resistance gene profile and virulence gene database, as well as to infer phylogenetic relationships and recombination patterns. Serotype and antimicrobial susceptibilities were identified separately in this study, however in the very near future it will likely be possible to obtain all necessary information with a single assay. Serotype prediction is already possible using the automated pipeline PneumoCaT, which can accurately predict 89/94 serotypes to the type level and the rest to group level, thus reducing the need for phenotypic testing (185). While WGS offers the potential for *in silico* antimicrobial susceptibility testing, the current evidence in favour of this method is lacking. A recent report by the European Committee on Antimicrobial Susceptibility Testing (EUCAST) described *S. pneumoniae* as an organism that has the least available evidence for predicting resistance phenotype using WGS, stating that more studies are necessary before the gaps in knowledge can even be addressed (186). Compounding this issue is the fact that resistance to many antimicrobial classes in *S. pneumoniae* is mediated by the development of mosaic genes, as opposed to the simple presence or absence of a gene (186). It became clear in our current study that the presence of a resistance gene or mutation did not necessarily correlate to phenotypic resistance, as many isolates possessing *ermB*, *tetM* or mutations in all three PBPs were susceptible to clarithromycin, doxycycline and penicillin, respectively. Conversely, lacking a gene or mutation did not necessarily correlate to susceptibility, as was evident in an isolate that demonstrated full trimethoprim-sulfamethoxazole resistance with neither of the common *fol* alterations. Much more research is necessary to generate evidence in favour of this approach, followed by early, international agreement on the appropriate standardization, harmonization and validation of analytical approaches, interpretive criteria and databases for WGS-based prediction of susceptibility data (75, 186).

Despite the numerous advantages to WGS, there was one pitfall noted in this study that was *S. pneumoniae*-specific: the lack of high-quality, well annotated and complete reference genomes available on NCBI. At the time of this study, 33 complete *S. pneumoniae* genomes were available on NCBI, in comparison to 361 and 168 for *E. coli* and *S. aureus*, respectively. The standard *S. pneumoniae* R6 reference genome, though in use for many years for various purposes, did not represent our isolate set well given its obscure serotype and avirulent nature; additionally, the unfinished genomes deposited on NCBI were often of low quality. This left few useful (appropriate serotype, ST) closed genomes available for use, supplemented with internal references of as high quality as possible. As the SNVPhyl pipeline utilized in this study provided a reference-based discovery of SNVs, it was highly dependent on the strain chosen as a reference. This became noticeable in one particular set of strains: serotypes 18C and 33F. In the initial analysis using *S. pneumoniae* R6 as reference, all serotype 33F isolates clustered together regardless of ST or resistance profile, maintaining close relatedness to serotype 18C isolates. When an internal serotype 33F isolate was utilized for subsequent analysis, resistant serotype 33F isolates were more closely related to the serotype 18C background strains than they were to the other serotype 33F isolates. As the SNVPhyl analysis only uses core genome SNVs present in the reference strain to produce the phylogeny, it is possible that when using *S. pneumoniae* R6 as a reference, the areas with SNVs differentiating the serotype 33F isolates from each other were masked due to their absence in the core genome of the standard reference. Utilizing an internal serotype 33F strain as a reference allowed for additional sequence included as core genome, thus allowing more information to be available to discriminate between isolates. This finding emphasizes the importance of utilizing a suitably related reference genome in reference-based approaches, particularly for an organism as variable as *S. pneumoniae*, as crucial

information may be removed from the analysis if the reference is too dissimilar. An alternative method to determine the genetic relationship between strains involves a gene-by-gene approach to identify allelic changes, as opposed to SNVs (187, 188). Whole genome MLST (wgMLST), an expansion of the traditional MLST scheme, is reproducible, accurate and offers standardization and comparability between laboratories (75, 187, 189). A wgMLST schema encompasses the core genome of a species, as well as the accessory genome that is present in a subset of isolates (189). Despite the plasticity of many species, this method has become popular in many settings (189); for example, wgMLST schema are being validated to strengthen the real-time surveillance capabilities of PulseNet International (184). A reduced version of wgMLST is also available, encompassing only the core genome of an organism (cgMLST). Despite wgMLST and cgMLST being available for 14 and 8 organisms, respectively, there are no functional schema for *S. pneumoniae* (http://www.applied-maths.com/applications/wgmlst, www.cgmlst.org/ncs). With no schema currently even in development for this organism, this creates an attractive option for future research.

6.7 Study Limitations

As this project drew its isolate collection from the SAVE study, it therefore carried with it the inherent limitations associated with SAVE. In particular, the study is limited due to its lack of participation of all Canadian provinces. As no isolates were received from Alberta and British Columbia, any regional distributions may be skewed to underrepresent the Western region of Canada. In addition, the reporting of IPD cases in Canada is voluntary and passive, meaning only a subset of isolates from certain provinces may have been submitted to PHAC-NML and subsequently forwarded to CARA for the SAVE study. The nature of isolate collection also meant that incidence data for the serotypes of interest could not be calculated; coupled with the

lack of information on serotype carriage, this resulted in the inability to calculate our own study-specific invasive capacities for comparison to the literature. The SAVE study also received little demographic information; a more complete picture may have been drawn had antimicrobial usage, comorbidity and patient outcome data been available.

In an ideal situation, all isolates would have been tested using all methodologies; however, this was not possible due to time and cost constraints. Most importantly, the sample size included in the WGS analysis was only a very small portion of *S. pneumoniae* isolates collected by the SAVE study. Inclusion of more isolates of interest and more background strains would allow for better representation of the breadth of genetic diversity in the Canadian pneumococcal population; additionally, as all isolates were not included in the WGS analyses, the rate of capsular switching could not be calculated. The current study also predominantly focused on sequencing IPD isolates collected from adults over the age of 65; as this age group had the largest pool of isolates, it offered the most diverse set of isolates from which to sample. However, this is only one age group that is commonly affected by IPD; children under the age of two are a primary group affected by *S. pneumoniae*, therefore a similar study focusing on pediatric isolates is crucial to understanding the diversity of common serotypes in the conjugate vaccine-eligible population.

7. CONCLUSIONS

This thesis presents data on twelve serotypes of invasive *S. pneumoniae* commonly collected in Canada from 2011 to 2015 (3, 6C, 7F, 8, 9N, 11A, 12F, 15A, 19A, 22F, 33F and 35B). The incredible diversity of *S. pneumoniae* resulted in the data presented only partially supporting many of our hypotheses. We hypothesized that these serotypes would frequently be antimicrobial resistant and/or MDR, genetically similar within the serotype and virulent due to the presence of pneumococcal pilus genes; in many ways, these hypotheses were tied together, with the results being relatively interdependent.

S. pneumoniae isolates tested in this study demonstrated variable susceptibilities to key antimicrobials, with clarithromycin demonstrating the worst overall activity. Aside from penicillin, β-lactam agents retained excellent activity against all isolates tested and antimicrobial agents of last resort demonstrated 100% susceptibility. Antimicrobial resistance in *S. pneumoniae* was due to a variety of acquired and chromosomal resistance determinants. Determinants were generally not unique to a particular serotype, however the dual *mefA/ermB* genotype was only present in serotype 19A isolates, and *cat* was most commonly identified from serotype 3. Though the rate of MDR isolates statistically significantly decreased over the study period, four serotypes accounted for the majority of MDR demonstrated by *S. pneumoniae*, including 15A, 19A, 33F and 6C. These serotypes were also the most highly diverse of the twelve serotypes analyzed in this thesis, demonstrating the highest numbers of STs within the serotype and the most significant amounts of predicted recombination. Other serotypes which demonstrated little recombination were also less diverse and uncommonly MDR, such as 7F and 22F.

The presence of pilus genes was a highly clonal property; however, it was only variably associated with MDR and not necessarily connected to virulence. MDR serotype 19A isolates possessed both PI-1 and PI-2 genes, while MDR serotype 15A isolates did not; only ST62 isolates possessed PI-2 genes within serotype 11A, but only two-thirds of isolates within the complex retained the gene. A clear correlation was not identified between the presence of pilus genes and most serotypes and STs, necessitating further study to determine what effects pili have on pathogenesis. Conversely to pili, presence of virulence genes such as pneumococcal surface proteins and neuraminidases had clear correlations to particular strains and international clones.

Numerous capsular switch variants were identified in this study, emphasizing the impact that recombination has on creating diversity in the pneumococcal population. Capsule switching enabled some variants to escape the action of conjugate vaccines; others obtained resistance determinants and MDR phenotypes not normally associated with the serotype, thus increasing diversity and spreading resistance. Despite the recombinogenic nature of *S. pneumoniae* and its propensity to undergo capsular switches, serotype is still a relevant molecular descriptor for this organism. For as long as pneumococcal vaccines are composed of capsular polysaccharide antigens, serotype will remain an important factor with which to characterize isolates. In general, phylogenetic analysis identified isolates as clustering by serotype. However, as capsular switching plays a significant role in adaptation, propagation and serotype replacement, ST will also be an important molecular descriptor in conjunction with serotype; many properties, such as possession of resistance and virulence determinants, were identified to be clonal properties in this study, not fully related to serotype but instead to a specific clone. The results presented demonstrate the importance of recombination in generating variation and maintaining genetic plasticity amongst the pneumococcal population; it is crucial to continue studying as many

different serotypes as possible, as by no means can one study a single serotype and assume the same characteristics apply to the rest.

This study illustrated the increased ability of WGS to discriminate between closely related isolates, in comparison to PFGE and MLST. Phylogenetic and recombination analysis assisted in the ability to identify and characterize serotype 3-ST180 isolates belonging to two unique clades; MLST was unable to discriminate beyond ST, while PFGE was only able to note that the clades were genetically distinct. We hypothesized that MLST would provide greater discriminatory power than PFGE, however in a few cases, including serotype 3, PFGE was more discriminatory. In particular, serotype 12F-ST218 isolates were clustered by PFGE into outbreak-related isolates (macrolide-resistant) and susceptible isolates, which were further clustered by region; MLST, however, grouped all isolates together as ST218. The work presented in this thesis outlines the importance of being able to include the entire genome in molecular analyses, as opposed to singling out relatively few genes. WGS has a bright future in public health genomics and clinical microbiology processes; with one assay, WGS analyses have the ability to detect outbreaks, type isolates, identify relatedness using SNVs or expanded-MLST schemes, and hopefully in the future predict antimicrobial susceptibilities.

8. FUTURE DIRECTIONS

(A) Continued surveillance of pneumococcal serotypes circulating in Canada

Though this study has focused on twelve specific serotypes, the pneumococcal population is constantly shifting in Canada due to numerous factors, including antimicrobial and vaccine use. It is crucial to continue monitoring IPD-associated serotypes to ensure vaccine types decrease in prevalence, and to ascertain which non-vaccine serotypes will rise in prevalence to fill the vacated niche. It is essential to continue antimicrobial susceptibility testing to determine whether emerging non-vaccine serotypes are resistant to antimicrobials. This study identified serotypes 35B, 15A and 12F as possible additions to a new conjugate vaccine formulation; it is particularly vital to monitor serotype evolution after PCV-15 is introduced in Canada to determine whether these types continue to demonstrate cause for inclusion, or if other serotypes become more concerning.

(B) Implementation of WGS-based surveillance

Although general continuation of surveillance practices is important, implementation of WGS-based surveillance would vastly improve the ability to follow the population dynamics of *S. pneumoniae*. Continuously performing WGS on all isolates collected in the SAVE study would provide a wealth of information, as it would result in the availability of the total genetic content of isolates collected by the study, as opposed to just phenotypes or individual genes. Many phenotypic and molecular tests such as PCR would become unnecessary, as serotyping, gene extraction and eventually antimicrobial susceptibility prediction will be available using bioinformatics. This method of surveillance would vastly increase the ability to predict and evaluate serotype trends, monitor the spread of horizontally transferable resistance mechanisms and identify capsular switch variants.

(C) Creation of a schema for *S. pneumoniae* wgMLST/cgMLST

This thesis discussed the reproducibility, ease of use and portability of extended-MLST methodologies such as wgMLST and cgMLST. No schema for *S. pneumoniae* is currently available, and to our knowledge, none are under development. Creation of a schema for this organism would provide another useful tool to differentiate *S. pneumoniae* strains using the entire genome if SNV-based methodologies are unavailable or not feasible.

(D) Identify high-quality genomes to close and submit to the NCBI genome database

An issue that arose during the data analysis of this thesis was the low number high-quality, well-annotated and closed genomes available on NCBI to use as references. Although it is costly and time-consuming, identifying a high-quality genome of a MDR 15A, outbreak 12F or commonly isolated 22F to close and submit to NCBI would assist in expanding the *S. pneumoniae* genome database.

(E) Further characterization of virulence factors

Many questions pertaining to virulence factors were left unanswered in this study. In particular, a correlation between pilus gene presence, serotype/ST and resistance pattern was not identified. The presence of pilus genes does not necessarily mean they are being expressed; therefore, a study looking at the expression of these genes may help our understanding of pathogenesis, why some MDR isolates possess pilus genes while others do not, and why some STs only variably possess the genes. Additionally, the presence of *nanC* was of interest in this study, as it was postulated to be associated with CSF or meningitis-associated serotypes. We did not find conclusive evidence in this study, as few CSF isolates were included in the WGS analyses. Additional studies on relevant isolates would be useful to truly confirm or deny the role *nanC* has in *S. pneumoniae* pathogenesis.

9. REFERENCES

1.   **Spellerberg B**, **Brandt C**. 2015. Streptococcus, p. 383–402. *In* J. H. Jorgensen, M. A. Pfaller, K. C. Carroll, G. Funke, M. Landry, S. S. Richter, DWW (ed.), Manual of Clinical Microbiology, 11, vol I ed. ASM Press, Washington, D.C.

2.   **Levinson W**. 2010. Review of Medical Microbiology and Immunology, 11th ed. McGraw-Hill Companies, Inc.

3.   **van der Poll T**, **Opal SM**. 2009. Pathogenesis, treatment, and prevention of pneumococcal pneumonia. Lancet **374**:1543–1556.

4.   **Straume D**, **Stamsås GA**, **Håvarstein LS**. 2015. Natural transformation and genome evolution in Streptococcus pneumoniae. Infect Genet Evol **33**:371–380.

5.   **Griffith F**. 1928. The Significance of Pneumococcal Types. J Hyg (Lond) **27**:113–59.

6.   **Avery OT**, **Macleod CM**, **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 Exp Med **79**:137–58.

7.   **Shanker E**, **Federle MJ**. 2017. Quorum sensing regulation of competence and bacteriocins in Streptococcus pneumoniae and mutans. Genes (Basel) **8**:15.

8.   **Johnston C**, **Campo N**, **Bergé MJ**, **Polard P**, **Claverys JP**. 2014. *Streptococcus pneumoniae*, le transformiste. Trends Microbiol **22**:113–119.

9.   **Claverys J-P**, **Prudhomme M**, **Martin B**. 2006. Induction of competence regulons as a general response to stress in Gram-positive bacteria. Annu Rev Microbiol **60**:451–475.

10.   **Henriques-Normark B**, **Tuomanen EI**. 2013. The pneumococcus: Epidemiology, microbiology, and pathogenesis. Cold Spring Harb Perspect Med **3**:a010215–a010215.

11.   **Mehr S**, **Wood N**. 2012. *Streptococcus pneumoniae* - a review of carriage, infection, serotype replacement and vaccination. Paediatr Respir Rev **13**:258–264.

12.   **Lynch JP**, **Zhanel GG**. 2010. *Streptococcus pneumoniae*: epidemiology and risk factors, evolution of antimicrobial resistance, and impact of vaccines. Curr Opin Pulm Med **16**:217–225.

13.   **Bogaert D**, **De Groot R**, **Hermans PWM**. 2004. *Streptococcus pneumoniae* colonisation: The key to pneumococcal disease. Lancet Infect Dis **4**:144–154.

14.   **Lynch J**, **Zhanel G**. 2009. *Streptococcus pneumoniae*: epidemiology, risk factors, and strategies for prevention. Semin Respir Crit Care Med **30**:189–209.

15. **World Health Organization**. 2012. Pneumococcal vaccines. WHO position paper - 2012. Wkly Epidemiol Rec **87**:129–144.

16. **Centers for Disease Control and Prevention (CDC)**. 2015. Epidemiology and prevention of vaccine-preventable diseases, 13th ed. Public Health Foundation, Washington, D.C.

17. **Centers for Disease Control and Prevention (CDC)**. 2014. Active bacterial core surveillance report, Emerging infections program network, *Streptococcus pneumoniae*, 2014. https://www.cdc.gov/abcs/reports-findings/survreports/spneu14.html.

18. **Swartz MN**. 2002. Attacking the pneumococcus -- a hundred years' war. N Engl J Med **346**:722.

19. **Kislak JW**, **Razavi LM**, **Daly AK**, **Finland M**. 1965. Susceptibility of pneumococci to nine antibiotics. Am J Med Sci **250**:261–8.

20. **Hansman D**, **Bullen MM**. 1967. A resistant pneumococcus. Lancet **290**:264–265.

21. **Kim L**, **McGee L**, **Tomczyk S**, **Beall B**. 2016. Biological and epidemiological features of antibiotic-resistant *Streptococcus pneumoniae* in pre- and post-conjugate vaccine eras: A United States perspective. Clin Microbiol Rev **29**:525–552.

22. **Jacobs MR**, **Koornhof HJ**, **Robins-Browne RM**, **Stevenson CM**, **Vermaak ZA**, **Freiman I**, **Miller GB**, **Witcomb MA**, **Isaäcson M**, **Ward JI**, **Austrian R**. 1978. Emergence of multiply resistant pneumococci. N Engl J Med **299**:735–740.

23. **McGee L**, **McDougal L**, **Zhou J**, **Spratt BG**, **Tenover FC**, **George R**, **Hakenbeck R**, **Hryniewicz W**, **Lefévre JC**, **Tomasz A**, **Klugman KP**. 2001. Nomenclature of major antimicrobial-resistant clones of *Streptococcus pneumoniae* defined by the pneumococcal molecular epidemiology network. J Clin Microbiol **39**:2565–2571.

24. **Centers for Disease Control and Prevention (CDC)**. 2013. Antibiotic resistance threats in the United States, 2013. https://www.cdc.gov/drugresistance/threat-report-2013/pdf/ar-threats-2013-508.pdf.

25. **Tunkel AR**, **Hartman BJ**, **Kaplan SL**, **Kaufman BA**, **Roos KL**, **Scheld WM**, **Whitley RJ**. 2004. Practice guidelines for the management of bacterial meningitis. Clin Infect Dis **39**:1267–84.

26. **Tunkel AR**, **Hasbun R**, **Bhimraj A**, **Byers K**, **Kaplan SL**, **Michael Scheld W**, **van de Beek D**, **Bleck TP**, **Garton HJL**, **Zunt JR**. 2017. 2017 Infectious Diseases Society of America's Clinical Practice Guidelines for Healthcare-Associated Ventriculitis and Meningitis. Clin Infect Dis **64**:e34–e65.

27. 2012. The Sanford Guide to Antimicrobial Therapy 2012, 42nd ed. Antimicrobial Therapy, Inc.

28. **Smith AM**, **Klugman KP**. 1998. Alterations in PBP 1a essential for high-level penicillin resistance in *Streptococcus pneumoniae*. Antimicrob Agents Chemother **42**:1329–1333.

29. **Richter SS**, **Diekema DJ**, **Heilmann KP**, **Dohrn CL**, **Riahi F**, **Doern G V.** 2014. Changes in pneumococcal serotypes and antimicrobial resistance after introduction of the 13-valent conjugate vaccine in the United States. Antimicrob Agents Chemother **58**:6484–6489.

30. **Adam HJ**, **Karlowsky JA**, **Nichol KA**, **Gilmour MW**, **Hoban DJ**, **Embree J**, **Zhanel GG**. 2012. Baseline epidemiology of *Streptococcus pneumoniae* serotypes in Canada prior to the introduction of the 13-valent pneumococcal vaccine. Microb Drug Resist **18**:176–82.

31. **Golden AR**, **Rosenthal M**, **Fultz B**, **Nichol KA**, **Adam HJ**, **Gilmour MW**, **Baxter MR**, **Hoban DJ**, **Karlowsky JA**, **Zhanel GG**. 2015. Characterization of MDR and XDR *Streptococcus pneumoniae* in Canada, 2007-13. J Antimicrob Chemother **70**:2199–2202.

32. **Wierzbowski AK**, **Karlowsky JA**, **Adam HJ**, **Nichol KA**, **Hoban DJ**, **Zhanel GG**. 2014. Evolution and molecular characterization of macrolide-resistant *Streptococcus pneumoniae* in Canada between 1998 and 2008. J Antimicrob Chemother **69**:59–66.

33. **Schroeder MR**, **Stephens DS**. 2016. Macrolide resistance in *Streptococcus pneumoniae*. Front Cell Infect Microbiol **6**:98.

34. **Zhanel GG**, **Adam HJ**, **Baxter MR**, **Fuller J**, **Nichol KA**, **Denisuik AJ**, **Lagacé-Wiens PRS**, **Walkty A**, **Karlowsky JA**, **Schweizer F**, **Hoban DJ**, **Canadian Antimicrobial Resistance Alliance**. 2013. Antimicrobial susceptibility of 22746 pathogens from Canadian hospitals: results of the CANWARD 2007-11 study. J Antimicrob Chemother **68 Suppl 1**:i7–i22.

35. **Van Bambeke F**, **Michot JM**, **Van Eldere J**, **Tulkens PM**. 2005. Quinolones in 2005: An update. Clin Microbiol Infect **11**:256–280.

36. **Brueggemann AB**, **Coffman SL**, **Rhomberg P**, **Huynh H**, **Almer L**, **Nilius A**, **Flamm R**, **Doern G V**. 2002. Fluoroquinolone resistance in *Streptococcus pneumoniae* in United States since 1994-1995. Antimicrob Agents Chemother **46**:680–688.

37. **Karlowsky JA**, **Adam HJ**, **Desjardins M**, **Lagacé-Wiens PRS**, **Hoban DJ**, **Zhanel GG**, **Canadian Antimicrobial Resistance Alliance**. 2013. Changes in fluoroquinolone resistance over 5 years (CANWARD 2007-11) in bacterial pathogens isolated in Canadian hospitals. J Antimicrob Chemother **68 Suppl 1**:i39–i46.

38. **Metcalf BJ**, **Gertz RE**, **Gladstone RA**, **Walker H**, **Sherwood LK**, **Jackson D**, **Li Z**, **Law C**, **Hawkins PA**, **Chochua S**, **Sheth M**, **Rayamajhi N**, **Bentley SD**, **Kim L**, **Whitney CG**, **McGee L**, **Beall B**. 2016. Strain features and distributions in pneumococci from children with invasive disease before and after 13-valent conjugate vaccine implementation in the USA. Clin Microbiol Infect **22**:60.e9-60.e29.

39. **Sung H**, **Hee BS**, **Kim MN**, **Lee K**, **Kim EC**, **Song W**, **Seok HJ**, **Lee WG**, **Park YJ**, **Eliopoulos GM**. 2006. Vancomycin-tolerant *Streptococcus pneumoniae* in Korea. J Clin Microbiol **44**:3524–3528.

40. **Henriques Normark B**, **Novak R**, **Ortqvist A**, **Källenius G**, **Tuomanen E**, **Normark S**. 2001. Clinical isolates of *Streptococcus pneumoniae* that exhibit tolerance of vancomycin. Clin Infect Dis **32**:552–8.

41. **Rodriguez CA**, **Atkinson R**, **Bitar W**, **Whitney CG**, **Edwards KM**, **Mitchell L**, **Li J**, **Sublett J**, **Li C**, **Liu T**, **Chesney PJ**, **Tuomanen EI**. 2004. Tolerance to vancomycin in pneumococci: detection with a molecular marker and assessment of clinical impact. J Infect Dis **190**:1481–1487.

42. **Dong W**, **Chochua S**, **McGee L**, **Jackson D**, **Klugman KP**, **Vidal JE**. 2014. Mutations within the *rplD* gene of linezolid-nonsusceptible *Streptococcus pneumoniae* strains isolated in the United States. Antimicrob Agents Chemother **58**:2459–2462.

43. **Geno KA**, **Gilbert GL**, **Song JY**, **Skovsted IC**, **Klugman KP**, **Jones C**, **Konradsen HB**, **Nahm MH**. 2015. Pneumococcal capsules and their types: past, present, and future. Clin Microbiol Rev **28**:871–899.

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

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

46. **Bentley SD**, **Aanensen DM**, **Mavroidi A**, **Saunders D**, **Rabbinowitsch E**, **Collins M**, **Donohoe K**, **Harris D**, **Murphy L**, **Quail MA**, **Samuel G**, **Skovsted IC**, **Kaltoft MS**, **Barrell B**, **Reeves PR**, **Parkhill J**, **Spratt BG**. 2006. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS Genet **2**:262–269.

47. **Weinberger DM**, **Trzcinski K**, **Lu YJ**, **Bogaert D**, **Brandes A**, **Galagan J**, **Anderson PW**, **Malley R**, **Lipsitch M**. 2009. Pneumococcal capsular polysaccharide structure predicts serotype prevalence. PLoS Pathog **5**:e1000476.

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

49. **Chaguza C**, **Andam CP**, **Harris SR**, **Cornick JE**, **Yang M**, **Bricio-Moreno L**, **Kamng'ona AW**, **Parkhill J**, **French N**, **Heyderman RS**, **Kadioglu A**, **Everett DB**, **Bentley SD**, **Hanage WP**. 2016. Recombination in *Streptococcus pneumoniae* lineages increase with carriage duration and size of the polysaccharide capsule. MBio **7**:e01053-16.

50. **Aanensen DM**, **Mavroidi A**, **Bentley SD**, **Reeves PR**, **Spratt BG**. 2007. Predicted functions and linkage specificities of the products of the *Streptococcus pneumoniae* capsular biosynthetic loci. J Bacteriol **189**:7856–7876.

51. **Telford JL**, **Barocchi MA**, **Margarit I**, **Rappuoli R**, **Grandi G**. 2006. Pili in Gram-positive pathogens. Nat Rev Microbiol **4**:509–519.

52. **Bagnoli F**, **Moschioni M**, **Donati C**, **Dimitrovska V**, **Ferlenghi I**, **Facciotti C**, **Muzzi A**, **Giusti F**, **Emolo C**, **Sinisi A**, **Hilleringmann M**, **Pansegrau W**, **Censini S**, **Rappuoli R**, **Covacci A**, **Masignani V**, **Barocchi MA**. 2008. A second pilus type in *Streptococcus pneumoniae* is prevalent in emerging serotypes and mediates adhesion to host cells. J Bacteriol **190**:5480–5492.

53. **Barocchi MA**, **Ries J**, **Zogaj X**, **Hemsley C**, **Albiger B**, **Kanth A**, **Dahlberg S**, **Fernebro J**, **Moschioni M**, **Masignani V**, **Hultenby K**, **Taddei AR**, **Beiter K**, **Wartha F**, **von Euler A**, **Covacci A**, **Holden DW**, **Normark S**, **Rappuoli R**, **Henriques-Normark B**. 2006. A pneumococcal pilus influences virulence and host inflammatory responses. Proc Natl Acad Sci U S A **103**:2857–62.

54. **Zähner D**, **Gudlavalleti A**, **Stephens DS**. 2010. Increase in pilus islet 2-encoded pili among *Streptococcus pneumoniae* isolates, Atlanta, Georgia, USA. Emerg Infect Dis **16**:955–962.

55. **Croney CM**, **Nahm MH**, **Juhn SK**, **Briles DE**, **Crain MJ**. 2013. Invasive and noninvasive *Streptococcus pneumoniae* capsule and surface protein diversity following the use of a conjugate vaccine. Clin Vaccine Immunol **20**:1711–8.

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

57. **Brittan JL**, **Buckeridge TJ**, **Finn A**, **Kadioglu A**, **Jenkinson HF**. 2012. Pneumococcal neuraminidase A: an essential upper airway colonization factor for *Streptococcus pneumoniae*. Mol Oral Microbiol **27**:270–283.

58. **Pettigrew MM**, **Fennie KP**, **York MP**, **Daniels J**, **Ghaffar F**. 2006. Variation in the presence of neuraminidase genes among *Streptococcus pneumoniae* isolates with identical sequence types. Infect Immun **74**:3360–3365.

59.	**Manco S**, **Hernon F**, **Yesilkaya H**, **Paton JC**, **Andrew PW**, **Kadioglu A**. 2006. Pneumococcal neuraminidases A and B both have essential roles during infection of the respiratory tract and sepsis. Infect Immun **74**:4014–4020.

60.	**Durando P**, **Faust SN**, **Fletcher M**, **Krizova P**, **Torres A**, **Welte T**. 2013. Experience with pneumococcal polysaccharide conjugate vaccine (conjugated to CRM197 carrier protein) in children and adults. Clin Microbiol Infect **19**:1–9.

61.	**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. https://www.prevnar.ca/files/pdfs/NACI.pdf.

62.	**Pichichero ME**. 2013. Protein carriers of conjugate vaccines: Characteristics, development, and clinical trials. Hum Vaccines Immunother **9**:2505–2523.

63.	**Bettinger JA**, **Scheifele DW**, **Kellner JD**, **Halperin SA**, **Vaudry W**, **Law B**, **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.

64.	**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 **54**:893–7.

65.	**Pilishvili T**, **Lexau C**, **Farley MM**, **Hadler J**, **Harrison LH**, **Bennett NM**, **Reingold A**, **Thomas A**, **Schaffner W**, **Craig AS**, **Smith PJ**, **Beall BW**, **Whitney CG**, **Moore MR**. 2010. Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. J Infect Dis **201**:32–41.

66.	**Brueggemann AB**, **Pai R**, **Crook DW**, **Beall B**. 2007. Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. PLoS Pathog **3**:1628–1636.

67.	**Wyres KL**, **Lambertsen LM**, **Croucher NJ**, **McGee L**, **Von Gottberg A**, **Liñares J**, **Jacobs MR**, **Kristinsson KG**, **Beall BW**, **Klugman KP**, **Parkhill J**, **Hakenbeck R**, **Bentley SD**, **Brueggemann AB**. 2013. Pneumococcal capsular switching: A historical perspective. J Infect Dis **207**:439–449.

68.	**National Advisory Committee on Immunization (NACI)**. 2010. Update on the use of conjugate pneumococcal vaccines in childhood. Canada Commun Dis Rep **36**:1–21.

69.	**Sobanjo-ter Meulen A**, **Vesikari T**, **Malacaman EA**, **Shapiro SA**, **Dallas MJ**, **Hoover PA**, **McFetridge R**, **Stek JE**, **Marchese RD**, **Hartzel J**, **Watson WJ**, **Musey LK**. 2015. Safety, tolerability and immunogenicity of 15-valent pneumococcal conjugate vaccine in toddlers previously vaccinated with 7-valent pneumococcal conjugate vaccine. Pediatr Infect Dis J **34**:186–94.

70. **McFetridge R**, **Meulen AST**, **Folkerth SD**, **Hoekstra JA**, **Dallas M**, **Hoover PA**, **Marchese RD**, **Zacholski DM**, **Watson WJ**, **Stek JE**, **Hartzel JS**, **Musey LK**. 2015. Safety, tolerability, and immunogenicity of 15-valent pneumococcal conjugate vaccine in healthy adults. Vaccine **33**:2793–2799.

71. **Pichichero ME**, **Khan MN**, **Xu Q**. 2016. Next generation protein based *Streptococcus pneumoniae* vaccines. Hum Vaccines Immunother **12**:194–205.

72. **Moffitt KL**, **Malley R**. 2011. Next generation pneumococcal vaccines. Curr Opin Immunol **23**:407–413.

73. **Gilmour MW**, **Graham M**, **Reimer A**, **Van Domselaar G**. 2013. Public health genomics and the new molecular epidemiology of bacterial pathogens. Public Health Genomics **16**:25–30.

74. **Klugman KP**. 2002. The successful clone: the vector of dissemination of resistance in *Streptococcus pneumoniae*. J Antimicrob Chemother **50**:1–6.

75. **Tagini F**, **Greub G**. 2017. Bacterial genome sequencing in clinical microbiology: a pathogen-oriented review. Eur J Clin Microbiol Infect Dis doi: 10.1007/s10096-017-3024-6.

76. **Drancourt M**, **Roux V**, **Fournier P-E**, **Raoult D**. 2004. *rpoB* gene sequence-based identification of aerobic Gram-positive cocci of the genera Streptococcus, Enterococcus, Gemella, Abiotrophia, and Granulicatella. J Clin Microbiol **42**:497–504.

77. **Adam HJ**, **Golden AR**, **Karlowsky JA**, **Baxter MR**, **Nichol KA**, **Martin I**, **Demczuk WHB**, **Mulvey MR**, **Gilmour MW**, **Hoban DJ**, **Zhanel GG**, **Canadian Antimicrobial Resistance Alliance (CARA)**. 2017. Analysis of multi-drug resistance in the predominant *Streptococcus pneumoniae* serotypes in Canada: the SAVE study, 2011-2015. J Antimicrob Chemother In Press.

78. **Streptococcus and STI Unit, Bacteriology and Enteric Diseases Program NML**. National Laboratory Surveillance of Invasive Streptococcal Disease in Canada - Annual Summary 2014. https://www.canada.ca/en/public-health/services/publications/drugs-health-products/national-laboratory-surveillance-invasive-streptococcal-disease-canada-annual-summary-2014.html.

79. **Gertz RE**, **Li Z**, **Pimenta FC**, **Jackson D**, **Juni BA**, **Lynfield R**, **Jorgensen JH**, **Carvalho M da G**, **Beall BW**. 2010. Increased penicillin nonsusceptibility of nonvaccine-serotype invasive pneumococci other than serotypes 19A and 6A in post-7-valent conjugate vaccine era. J Infect Dis **201**:770–775.

80. **Clinical and Laboratory Standards Institute (CLSI)**. 2015. Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically - 9th Edition, Approved Standard M07-A10. Wayne, PA.

81.   **Clinical and Laboratory Standards Institute (CLSI)**. 2016. Performance Standards for Antimicrobial Susceptibility Testing. 26th Informational Supplement (M100-S26). Wayne, PA.

82.   **McEllistrem MC**, **Stout JE**, **Harrison LH**. 2000. Simplified protocol for pulsed-field gel electrophoresis analysis of *Streptococcus pneumoniae*. J Clin Microbiol **38**:351–3.

83.   **Louie M**, **Louie L**, **Papia G**, **Talbot J**, **Lovgren M**, **Simor AE**. 1999. Molecular analysis of the genetic variation among penicillin-susceptible and penicillin-resistant *Streptococcus pneumoniae* serotypes in Canada. J Infect Dis **179**:892–900.

84.   **Enright MC**, **Spratt BG**. 1998. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: Identification of clones associated with serious invasive disease. Microbiology **144**:3049–3060.

85.   **Centers for Disease Control and Prevention (CDC)**. 2009. Alternative MLST primers for *S. pyogenes* and *S. pneumoniae*. https://www.cdc.gov/streplab/alt-mlst-primers.html.

86.   **Nascimento M**, **Sousa A**, **Ramirez M**, **Francisco AP**, **Carriço JA**, **Vaz C**. 2017. PHYLOViZ 2.0: providing scalable data integration and visualization for multiple phylogenetic inference methods. Bioinformatics **33**:128–129.

87.   **Zhanel GG**, **Adam HJ**, **Low DE**, **Blondeau J**, **DeCorby M**, **Karlowsky JA**, **Weshnoweski B**, **Vashisht R**, **Wierzbowski A**, **Hoban DJ**. 2011. Antimicrobial susceptibility of 15,644 pathogens from Canadian hospitals: Results of the CANWARD 2007-2009 study. Diagn Microbiol Infect Dis **69**:291–306.

88.   **Magoč T**, **Salzberg SL**. 2011. FLASH: Fast length adjustment of short reads to improve genome assemblies. Bioinformatics **27**:2957–2963.

89.   **Bankevich A**, **Nurk S**, **Antipov D**, **Gurevich AA**, **Dvorkin M**, **Kulikov AS**, **Lesin VM**, **Nikolenko SI**, **Pham S**, **Prjibelski AD**, **Pyshkin A V.**, **Sirotkin A V.**, **Vyahhi N**, **Tesler G**, **Alekseyev MA**, **Pevzner PA**. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **19**:455–477.

90.   **Seemann T**. 2014. Prokka: Rapid prokaryotic genome annotation. Bioinformatics **30**:2068–2069.

91.   **Petkau A**, **Mabon P**, **Sieffert C**, **Knox NC**, **Cabral J**, **Iskander M**, **Iskander M**, **Weedmark K**, **Zaheer R**, **Katz LS**, **Nadon C**, **Reimer A**, **Taboada E**, **Beiko RG**, **Hsiao W**, **Brinkman F**, **Graham M**, **Van Domselaar G**. 2017. SNVPhyl: a single nucleotide variant phylogenomics pipeline for microbial genomic epidemiology. Microb Genomics **3**.

92.   **Kurtz S**, **Phillippy A**, **Delcher AL**, **Smoot M**, **Shumway M**, **Antonescu C**, **Salzberg SL**. 2004. Versatile and open software for comparing large genomes. Genome Biol **5**:R12.

93.   **Garrison E**, **Marth G**. 2012. Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907.

94.   **Li H**, **Handsaker B**, **Wysoker A**, **Fennell T**, **Ruan J**, **Homer N**, **Marth G**, **Abecasis G**, **Durbin R**. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**:2078–2079.

95.   **Li H**. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics **27**:2987–2993.

96.   **Guindon S**, **Dufayard JF**, **Lefort V**, **Anisimova M**, **Hordijk W**, **Gascuel O**. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst Biol **59**:307–321.

97.   **Ragonnet-Cronin M**, **Hodcroft E**, **Hué S**, **Fearnhill E**, **Delpech V**, **Brown AJ**, **Lycett S**. 2013. Automated analysis of phylogenetic clusters. BMC Bioinformatics **14**:317.

98.   **Zankari E**, **Hasman H**, **Cosentino S**, **Vestergaard M**, **Rasmussen S**, **Lund O**, **Aarestrup FM**, **Larsen MV**. 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother **67**:2640–2644.

99.   **Hyatt D**, **Chen G-L**, **LoCascio PF**, **Land ML**, **Larimer FW**, **Hauser LJ**. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11**:119.

100.   **Camacho C**, **Coulouris G**, **Avagyan V**, **Ma N**, **Papadopoulos J**, **Bealer K**, **Madden TL**. 2009. BLAST+: architecture and applications. BMC Bioinformatics **10**:421.

101.   **Cock PJA**, **Grüning BA**, **Paszkiewicz K**, **Pritchard L**. 2013. Galaxy tools and workflows for sequence analysis with applications in molecular plant pathology. PeerJ **1**:e167.

102.   **Larkin MA**, **Blackshields G**, **Brown NP**, **Chenna R**, **Mcgettigan PA**, **McWilliam H**, **Valentin F**, **Wallace IM**, **Wilm A**, **Lopez R**, **Thompson JD**, **Gibson TJ**, **Higgins DG**. 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23**:2947–2948.

103.   **Nichol KA**, **Zhanel GG**, **Hoban DJ**. 2002. Penicillin-binding protein 1A, 2B, and 2X alterations in Canadian isolates of penicillin-resistant *Streptococcus pneumoniae*. Antimicrob Agents Chemother **46**:3261–3264.

104.   **Hakenbeck R**, **Brückner R**, **Denapaite D**, **Maurer P**. 2012. Molecular mechanisms of β-lactam resistance in *Streptococcus pneumoniae*. Future Microbiol **7**:395–410.

105.   **Rice P**, **Longden I**, **Bleasby A**. 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet **16**:276–7.

106. **Blumental S**, **Granger-Farbos A**, **Moïsi JC**, **Soullié B**, **Leroy P**, **Njanpop-Lafourcade BM**, **Yaro S**, **Nacro B**, **Hallin M**, **Koeck JL**. 2015. Virulence factors of *Streptococcus pneumoniae*. Comparison between African and French invasive isolates and implication for future vaccines. PLoS One **10**:e0133885.

107. **Inouye M**, **Dashnow H**, **Raven L-A**, **Schultz MB**, **Pope BJ**, **Tomita T**, **Zobel J**, **Holt KE**. 2014. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome Med **6**:90.

108. **Li H**. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303-3997.

109. **Thorvaldsdóttir H**, **Robinson JT**, **Mesirov JP**. 2013. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. Brief Bioinform **14**:178–192.

110. **Robinson JT**, **Thorvaldsdóttir H**, **Winckler W**, **Guttman M**, **Lander ES**, **Getz G**, **Mesirov JP**. 2011. Integrative genomics viewer. Nat Biotechnol **29**:24–26.

111. **Croucher NJ**, **Page AJ**, **Connor TR**, **Delaney AJ**, **Keane JA**, **Bentley SD**, **Parkhill J**, **Harris SR**. 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res **43**:e15.

112. **Hadfield J**, **Croucher NJ**, **Goater RJ**, **Abudahab K**, **Aanensen DM**, **Harris SR**. 2017. Phandango: an interactive viewer for bacterial population genomics. bioRxiv 119545.

113. **Croucher NJ**, **Kagedan L**, **Thompson CM**, **Parkhill J**, **Bentley SD**, **Finkelstein JA**, **Lipsitch M**, **Hanage WP**. 2015. Selective and genetic constraints on pneumococcal serotype switching. PLoS Genet **11**:e1005095.

114. **Moore MR**, **Gertz RE**, **Woodbury RL**, **Barkocy-Gallagher GA**, **Schaffner W**, **Lexau C**, **Gershman K**, **Reingold A**, **Farley M**, **Harrison LH**, **Hadler JL**, **Bennett NM**, **Thomas AR**, **McGee L**, **Pilishvili T**, **Brueggemann AB**, **Whitney CG**, **Jorgensen JH**, **Beall B**. 2008. Population snapshot of emergent *Streptococcus pneumoniae* serotype 19A in the United States, 2005. J Infect Dis **197**:1016–27.

115. **Karlowsky JA**, **Adam HJ**, **Golden AR**, **Baxter MR**, **Nichol KA**, **Martin I**, **Demczuk WHB**, **Mulvey MR**, **Gilmour MW**, **Hoban DJ**, **Zhanel GG**, **Canadian Antimicrobial Resistance Alliance (CARA)**. 2017. Antimicrobial susceptibility testing of invasive isolates of *Streptococcus pneumoniae* from Canadian patients: the SAVE Study, 2011-2015. J Antimicrob Chemother In Press.

116. **Demczuk WHB**, **Martin I**, **Griffith A**, **Lefebvre B**, **McGeer A**, **Lovgren M**, **Tyrrell GJ**, **Desai S**, **Sherrard L**, **Adam H**, **Gilmour M**, **Zhanel GG**, **Network TBD**, **Network CPHL**. 2013. Serotype distribution of invasive *Streptococcus pneumoniae* in Canada after the introduction of the 13-valent pneumococcal conjugate vaccine, 2010–2012. Can J

Microbiol **59**:778–788.

117.   **Croucher NJ**, **Chewapreecha C**, **Hanage WP**, **Harris SR**, **McGee L**, **Van Der Linden M**, **Song JH**, **Ko KS**, **De Lencastre H**, **Turner C**, **Yang F**, **Sa-Leao R**, **Beall B**, **Klugman KP**, **Parkhill J**, **Turner P**, **Bentley SD**. 2014. Evidence for soft selective sweeps in the evolution of pneumococcal multidrug resistance and vaccine escape. Genome Biol Evol **6**:1589–1602.

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

119.   **Metcalf BJ**, **Chochua S**, **Gertz RE**, **Li Z**, **Walker H**, **Tran T**, **Hawkins PA**, **Glennen A**, **Lynfield R**, **Li Y**, **McGee L**, **Beall B**, **Reingold A**, **Brooks S**, **Randel H**, **Miller L**, **White B**, **Aragon D**, **Barnes M**, **Sadlowski J**, **Petit S**, **Cartter M**, **Marquez C**, **Wilson M**, **Farley M**, **Thomas S**, **Tunali A**, **Baughman W**, **Harrison L**, **Benton J**, **Carter T**, **Hollick R**, **Holmes K**, **Riner A**, **Holtzman C**, **Danila R**, **MacInnes K**, **Scherzinger K**, **Angeles K**, **Bareta J**, **Butler L**, **Khanlian S**, **Mansmann R**, **Nichols M**, **Bennett N**, **Zansky S**, **Currenti S**, **McGuire S**, **Thomas A**, **Schmidt M**, **Thompson J**, **Poissant T**, **Schaffner W**, **Barnes B**, **Leib K**, **Dyer K**, **McKnight L**, **Gierke R**, **Almendares O**, **Hudson J**, **McGlone L**, **Langley G**. 2016. Using whole genome sequencing to identify resistance determinants and predict antimicrobial resistance phenotypes for year 2015 invasive pneumococcal disease isolates recovered in the United States. Clin Microbiol Infect **22**:1002.e1-1002.e8.

120.   **Mayanskiy N**, **Savinova T**, **Alyabieva N**, **Ponomarenko O**, **Brzhozovskaya E**, **Lazareva A**, **Katosova L**, **Kozlov R**. 2017. Antimicrobial resistance, penicillin-binding protein sequences, and pilus islet carriage in relation to clonal evolution of *Streptococcus pneumoniae* serotype 19A in Russia, 2002–2013. Epidemiol Infect **145**:1–12.

121.   **Setchanova LP**, **Alexandrova A**, **Dacheva D**, **Mitov I**, **Kaneva R**, **Mitev V**. 2015. Dominance of multidrug-resistant Denmark(14)-32 (ST230) clone among *Streptococcus pneumoniae* serotype 19A isolates causing pneumococcal disease in Bulgaria from 1992 to 2013. Microb Drug Resist **21**:35–42.

122.   **Makarewicz O**, **Lucas M**, **Brandt C**, **Herrmann L**, **Albersmeier A**, **Ruckert C**, **Blom J**, **Goesmann A**, **Van Der Linden M**, **Kalinowski J**, **Pletz MW**. 2017. Whole genome sequencing of 39 invasive *Streptococcus pneumoniae* sequence type 199 isolates revealed switches from serotype 19A to 15B. PLoS One **12**:e0169370.

123.   **Chewapreecha C**, **Harris SR**, **Croucher NJ**, **Turner C**, **Marttinen P**, **Cheng L**, **Pessia A**, **Aanensen DM**, **Mather AE**, **Page AJ**, **Salter SJ**, **Harris D**, **Nosten F**, **Goldblatt D**, **Corander J**, **Parkhill J**, **Turner P**, **Bentley SD**. 2014. Dense genomic sampling identifies highways of pneumococcal recombination. Nat Genet **46**:305–309.

124. **Kyaw MH**, **Lynfield R**, **Schaffner W**, **Craig AS**, **Hadler J**, **Reingold A**, **Thomas AR**, **Harrison LH**, **Bennett NM**, **Farley MM**, **Facklam RR**, **Jorgensen JH**, **Besser J**, **Zell ER**, **Schuchat A**, **Whitney CG**, **Active Bacterial Core Surveillance of the Emerging Infections Program Network**. 2006. Effect of introduction of the pneumococcal conjugate vaccine on drug-resistant *Streptococcus pneumoniae*. N Engl J Med **354**:1455–1463.

125. **Janoir C**, **Lepoutre A**, **Gutmann L**, **Varon E**. 2016. Insight into resistance phenotypes of emergent non 13-valent pneumococcal conjugate vaccine type pneumococci isolated from invasive disease after 13-valent pneumococcal conjugate vaccine implementation in France. Open Forum Infect Dis **3**:ofw020.

126. **Beall B**, **McEllistrem MC**, **Gertz RE**, **Boxrud DJ**, **Besser JM**, **Harrison LH**, **Jorgensen JH**, **Whitney CG**. 2002. Emergence of a novel penicillin-nonsusceptible, invasive serotype 35B clone of *Streptococcus pneumoniae* within the United States. J Infect Dis **186**:118–22.

127. **Chochua S**, **Metcalf BJ**, **Li Z**, **Walker H**, **Tran T**, **McGee L**, **Beall B**. 2017. Invasive serotype 35B pneumococci including an expanding serotype switch lineage, United States, 2015–2016. Emerg Infect Dis **23**:922–930.

128. **Olarte L**, **Kaplan SL**, **Barson WJ**, **Romero JR**, **Lin PL**, **Tan TQ**, **Hoffman JA**, **Bradley JS**, **Givner LB**, **Mason EO**, **Hulten KG**. 2017. Emergence of multidrug-resistant pneumococcal serotype 35B among children in the United States. J Clin Microbiol **55**:724–734.

129. **Navarro-Torné A**, **Dias JG**, **Hruba F**, **Lopalco PL**, **Pastore-Celentano L**, **Gauci AJA**, **Invasive Pneumococcal Disease Study Group**. 2015. Risk factors for death from invasive pneumococcal disease, Europe, 2010. Emerg Infect Dis **21**:417–25.

130. **Domenech M**, **Damián D**, **Ardanuy C**, **Liñares J**, **Fenoll A**, **García E**. 2015. Emerging, non-PCV13 serotypes 11A and 35B of *Streptococcus pneumoniae* show high potential for biofilm formation in vitro. PLoS One **10**:e0125636.

131. **Sheppard C**, **Fry NK**, **Mushtaq S**, **Woodford N**, **Reynolds R**, **Janes R**, **Pike R**, **Hill R**, **Kimuli M**, **Staves P**, **Doumith M**, **Harrison T**, **Livermore DM**. 2016. Rise of multidrug-resistant non-vaccine serotype 15A *Streptococcus pneumoniae* in the United Kingdom, 2001 to 2014. Eurosurveillance **21**:30423.

132. **van der Linden M**, **Perniciaro S**, **Imöhl M**. 2015. Increase of serotypes 15A and 23B in IPD in Germany in the PCV13 vaccination era. BMC Infect Dis **15**:207.

133. **Duvvuri VR**, **Deng X**, **Teatero S**, **Memari N**, **Athey T**, **Fittipaldi N**, **Gubbay JB**. 2016. Population structure and drug resistance patterns of emerging non-PCV-13 *Streptococcus pneumoniae* serotypes 22F, 15A, and 8 isolated from adults in Ontario, Canada. Infect Genet Evol **42**:1–8.

134. **Iannelli F**, **Oggioni MR**, **Pozzi G**. 2002. Allelic variation in the highly polymorphic locus pspC of *Streptococcus pneumoniae*. Gene **284**:63–71.

135. **van der Mee-Marquet N**, **Fourny L**, **Arnault L**, **Domelier A-S**, **Salloum M**, **Lartigue M-F**, **Quentin R**. 2008. Molecular characterization of human-colonizing *Streptococcus agalactiae* strains isolated from throat, skin, anal margin, and genital body sites. J Clin Microbiol **46**:2906–11.

136. **Eickel V**, **Kahl B**, **Reinisch B**, **Dübbers A**, **Küster P**, **Brandt C**, **Spellerberg B**. 2009. Emergence of respiratory *Streptococcus agalactiae* isolates in cystic fibrosis patients. PLoS One **4**:e4650.

137. **Dopazo J**, **Mendoza A**, **Herrero J**, **Caldara F**, **Humbert Y**, **Friedli L**, **Guerrier M**, **Grand-Schenk E**, **Gandin C**, **de Francesco M**, **Polissi A**, **Buell G**, **Feger G**, **Garcia E**, **Peitsch M**, **Garcia-Bustos JF**. 2001. Annotated draft genomic sequence from a *Streptococcus pneumoniae* type 19F clinical isolate. Microb Drug Resist **7**:99–125.

138. **Martin JM**, **Hoberman A**, **Paradise JL**, **Barbadora KA**, **Shaikh N**, **Bhatnagar S**, **Shope T**, **Block SL**, **Haralam MA**, **Kurs-Lasky M**, **Colborn DK**, **Green M**. 2014. Emergence of *Streptococcus pneumoniae* serogroups 15 and 35 in nasopharyngeal cultures from young children with acute otitis media. Pediatr Infect Dis J **33**:e286-90.

139. **Golden AR**, **Adam HJ**, **Karlowsky JA**, **Baxter MR**, **Nichol KA**, **Martin I**, **Demczuk WHB**, **Van Caeseele P**, **Gubbay JB**, **Lefebvre B**, **Horsman G**, **Zahariadis G**, **Haldane D**, **German G**, **Gilmour MW**, **Mulvey MR**, **Hoban DJ**, **Zhanel GG**, **Canadian Antimicrobial Resistance Alliance (CARA)**. 2017. Molecular characterization of predominant *Streptococcus pneumoniae* serotypes causing invasive infections in Canada: the SAVE study, 2011-2015. J Antimicrob Chemother In Press.

140. **Ardanuy C**, **De La Campa AG**, **García E**, **Fenoll A**, **Calatayud L**, **Cercenado E**, **Pérez-Trallero E**, **Bouza E**, **Liñares J**. 2014. Spread of *Streptococcus pneumoniae* serotype 8-ST63 multidrug-resistant recombinant clone, Spain. Emerg Infect Dis **20**:1848–1856.

141. **Sá-Leao R**, **Pinto F**, **Aguiar S**, **Nunes S**, **Carriço JA**, **Frazao N**, **Gonçalves-Sousa N**, **Melo-Cristino J**, **De Lencastre H**, **Ramirez M**. 2011. Analysis of invasiveness of pneumococcal serotypes and clones circulating in Portugal before widespread use of conjugate vaccines reveals heterogeneous behavior of clones expressing the same serotype. J Clin Microbiol **49**:1369–1375.

142. **De Cueto M**, **Rodríguez JM**, **Soriano MJ**, **López-Cerero L**, **Venero J**, **Pascual A**. 2008. Fatal levofloxacin failure in treatment of a bacteremic patient infected with *Streptococcus pneumoniae* with a preexisting *parC* mutation. J Clin Microbiol **46**:1558–1560.

143. **Andrews NJ**, **Waight PA**, **Burbidge P**, **Pearce E**, **Roalfe L**, **Zancolli M**, **Slack M**, **Ladhani SN**, **Miller E**, **Goldblatt D**. 2014. Serotype-specific effectiveness and correlates of protection for the 13-valent pneumococcal conjugate vaccine: A postlicensure indirect cohort study. Lancet Infect Dis **14**:839–846.

144. **Slotved H-C**, **Dalby T**, **Harboe ZB**, **Valentiner-Branth P**, **Casadevante VF de**, **Espenhain L**, **Fuursted K**, **Konradsen HB**. 2016. The incidence of invasive pneumococcal serotype 3 disease in the Danish population is not reduced by PCV-13 vaccination. Heliyon **2**:e00198.

145. **Horacio AN**, **Silva-Costa C**, **Lopes JP**, **Ramirez M**, **Melo-Cristino J**, **Portuguese Group for the Study of Streptococcal Infections**. 2016. Serotype 3 remains the leading cause of invasive pneumococcal disease in adults in Portugal (2012-2014) despite continued reductions in other 13-valent conjugate vaccine serotypes. Front Microbiol **7**:1616.

146. **Croucher NJ**, **Mitchell AM**, **Gould KA**, **Inverarity D**, **Barquist L**, **Feltwell T**, **Fookes MC**, **Harris SR**, **Dordel J**, **Salter SJ**, **Browall S**, **Zemlickova H**, **Parkhill J**, **Normark S**, **Henriques-Normark B**, **Hinds J**, **Mitchell TJ**, **Bentley SD**. 2013. Dominant role of nucleotide substitution in the diversification of serotype 3 pneumococci over decades and during a single infection. PLoS Genet **9**:e1003868.

147. **Azarian T**, **Mitchell P**, **Pollard A**, **von Gottberg A**, **Antonio M**, **Kwambana B**, **Du Plessis M**, **Clarke S**, **Everett D**, **Cornick J**, **Sadowy E**, **Skoczynska A**, **Moïsi J**, **McGee L**, **Beall B**, **Breiman R**, **Ho P**, **Gladstone R**, **Hanage W**, **Bentley S**. 2017. Global emergence and population dynamics of divergent serotype 3 ST180 pneumococci post-pneumococcal conjugate vaccine (PCV). ASM Microbe 2017. New Orleans, Louisiana.

148. **Zulz T**, **Wenger JD**, **Rudolph K**, **Robinson DA**, **Rakov A V.**, **Bruden D**, **Singleton RJ**, **Bruce MG**, **Hennessy TW**. 2013. Molecular characterization of *Streptococcus pneumoniae* serotype 12F isolates associated with rural community outbreaks in Alaska. J Clin Microbiol **51**:1402–1407.

149. **Schillberg E**, **Isaac M**, **Deng X**, **Peirano G**, **Wylie JL**, **Van Caeseele P**, **Pillai DR**, **Sinnock H**, **Mahmud SM**. 2014. Outbreak of invasive *Streptococcus pneumoniae* serotype 12F among a marginalized inner-city population in Winnipeg, Canada, 2009-2011. Clin Infect Dis **59**:651–657.

150. **Deng X**, **Peirano G**, **Schillberg E**, **Mazzulli T**, **Gray-Owen SD**, **Wylie JL**, **Robinson DA**, **Mahmud SM**, **Pillai DR**. 2016. Whole-genome sequencing reveals the origin and rapid evolution of an emerging outbreak strain of *Streptococcus pneumoniae* 12F. Clin Infect Dis **62**:1126–1132.

151. **Gladstone RA**, **Du Plessis M**, **Kwambana B**, **Hawkins P**, **Alaerts M**, **Brooks A**, **Zomer A**, **Cremers AJ**, **Pollard AJ**, **van der Ende A**, **Metcalf BJ**, **Beall B**, **Njanpop-Lafourcade BM**, **Sheppard C**, **Chewapreecha C**, **van de Beek D**, **Jauneikaite E**,

Mobegi F, **Kapatai G**, **Moisi JC**, **de Jonge MI**, **Hibberd ML**, **Fry N**, **Turner P**, **Kandasamy R**, **Burr SE**, **Madhi SA**, **Srifuengfung S**, **Clarke S**, **Nzenze SA**, **Balaji V**, **Contributers CPGSB**, **Everett DB**, **von Gottberg A**, **Antonio M**, **Klugman KP**, **McGee L**, **Breiman R**, **Bentley SD**. 2016. Detection of *Streptococcus pneumoniae* non vaccine serotype 12F clone 989 with multidrug antibiotic resistance, circulating globally. 10th International Symposium on Pneumococci and Pneumococcal Diseases. Glasgow, Scotland.

152. **Cooper D**, **Yu X**, **Sidhu M**, **Nahm MH**, **Fernsten P**, **Jansen KU**. 2011. The 13-valent pneumococcal conjugate vaccine (PCV13) elicits cross-functional opsonophagocytic killing responses in humans to *Streptococcus pneumoniae* serotypes 6C and 7A. Vaccine **29**:7207–11.

153. **McEllistrem MC**, **Nahm MH**. 2012. Novel pneumococcal serotypes 6C and 6D: Anomaly or harbinger. Clin Infect Dis **55**:1379–1386.

154. **Gherardi G**, **D'Ambrosio F**, **Visaggio D**, **Dicuonzo G**, **Del Grosso M**, **Pantosti A**. 2012. Serotype and clonal evolution of penicillin-nonsusceptible invasive *Streptococcus pneumoniae* in the 7-valent pneumococcal conjugate vaccine era in Italy. Antimicrob Agents Chemother **56**:4965–4968.

155. **Simões AS**, **Pereira L**, **Nunes S**, **Brito-Avô A**, **De Lencastre H**, **Sá-Leão R**. 2011. Clonal evolution leading to maintenance of antibiotic resistance rates among colonizing pneumococci in the PCV7 era in Portugal. J Clin Microbiol **49**:2810–2817.

156. **Golden AR**, **Adam HJ**, **Zhanel GG**. 2016. Invasive *Streptococcus pneumoniae* in Canada, 2011-2014: Characterization of new candidate 15-valent pneumococcal conjugate vaccine serotypes 22F and 33F. Vaccine **34**:2527–2530.

157. **Golden AR**, **Adam HJ**, **Gilmour MW**, **Baxter MR**, **Martin I**, **Nichol KA**, **Demczuk WHB**, **Hoban DJ**, **Zhanel GG**. 2015. Assessment of multidrug resistance, clonality and virulence in non-PCV-13 *Streptococcus pneumoniae* serotypes in Canada, 2011-13. J Antimicrob Chemother **70**:1960–1964.

158. **Pantosti A**, **Gherardi G**, **Conte M**, **Faella F**, **Dicuonzo G**, **Beall B**. 2002. A Novel, multiple drug–resistant, serotype 24F strain of *Streptococcus pneumoniae* that caused meningitis in patients in Naples, Italy. Clin Infect Dis **35**:205–208.

159. **Aguinagalde L**, **Corsini B**, **Domenech A**, **Domenech M**, **Cámara J**, **Ardanuy C**, **García E**, **Liñares J**, **Fenoll A**, **Yuste J**. 2015. Emergence of amoxicillin-resistant variants of Spain9V-ST156 pneumococci expressing serotype 11A correlates with their ability to evade the host immune response. PLoS One **10**:e0137565.

160. **Aguiar SI**, **Serrano I**, **Pinto FR**, **Melo-Cristino J**, **Ramirez M**. 2008. The presence of the pilus locus is a clonal property among pneumococcal invasive isolates. BMC Microbiol **8**:41.

161. **Del Grosso M**, **Camilli R**, **D'Ambrosio F**, **Petrucci G**, **Melchiorre S**, **Moschioni M**, **Gherardi G**, **Pantosti A**. 2013. Increase of pneumococcal serotype 19A in Italy is due to expansion of the piliated clone ST416/CC199. J Med Microbiol **62**:1220–1225.

162. **Adam HJ**, **Golden AR**, **Baxter MR**, **Nichol KA**, **Martin I**, **Demczuk WHB**, **Gilmour M**, **Mulvey MR**, **Karlowsky JA**, **Hoban DJ**, **Zhanel GG**, **Canadian Antimicrobial Resistance Alliance (CARA)**. 2017. Comparison of antimicrobial resistance patterns in *Streptococcus pneumoniae* from blood and respiratory isolates in Canadian hospitals from 2007-2016. ASM Microbe 2017. New Orleans, Louisiana.

163. **Sabharwal V**, **Figueira M**, **Pelton SI**, **Pettigrew MM**. 2012. Virulence of *Streptococcus pneumoniae* serotype 6C in experimental otitis media. Microbes Infect **14**:712–718.

164. **Aguiar SI**, **Melo-Cristino J**, **Ramirez M**. 2012. Use of the 13-valent conjugate vaccine has the potential to eliminate pilus carrying isolates as causes of invasive pneumococcal disease. Vaccine **30**:5487–5490.

165. **Rolo D**, **Ardanuy C**, **Fleites A**, **Martín R**, **Liñares J**. 2009. Diversity of pneumococcal surface protein A (PspA) among prevalent clones in Spain. BMC Microbiol **9**:80.

166. **Owen CD**, **Lukacik P**, **Potter JA**, **Sleator O**, **Taylor GL**, **Walsh MA**. 2015. *Streptococcus pneumoniae* NanC: Structural insights into the specificity and mechanism of a sialidase that produces a sialidase inhibitor. J Biol Chem **290**:27736–27748.

167. **Hollingshead SK**, **Baril L**, **Ferro S**, **King J**, **Coan P**, **Briles DE**. 2006. Pneumococcal surface protein A (PspA) family distribution among clinical isolates from adults over 50 years of age collected in seven countries. J Med Microbiol **55**:215–221.

168. **Hotomi M**, **Togawa A**, **Kono M**, **Ikeda Y**, **Takei S**, **Hollingshead SK**, **Briles DE**, **Suzuki K**, **Yamanaka N**. 2013. PspA family distribution, antimicrobial resistance and serotype of *Streptococcus pneumoniae* isolated from upper respiratory tract infections in Japan. PLoS One **8**:e58124.

169. **Xu G**, **Kiefel MJ**, **Wilson JC**, **Andrew PW**, **Oggioni MR**, **Taylor GL**. 2011. Three *Streptococcus pneumoniae* sialidases: three different products. J Am Chem Soc **133**:1718–1721.

170. **Imöhl M**, **Möller J**, **Reinert RR**, **Perniciaro S**, **van der Linden M**, **Aktas O**. 2015. Pneumococcal meningitis and vaccine effects in the era of conjugate vaccination: results of 20 years of nationwide surveillance in Germany. BMC Infect Dis **15**:61.

171. **Moore MR**, **Link-Gelles R**, **Schaffner W**, **Lynfield R**, **Lexau C**, **Bennett NM**, **Petit S**, **Zansky SM**, **Harrison LH**, **Reingold A**, **Miller L**, **Scherzinger K**, **Thomas A**, **Farley MM**, **Zell ER**, **Taylor TH**, **Pondo T**, **Rodgers L**, **McGee L**, **Beall B**, **Jorgensen JH**, **Whitney CG**. 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.

172. **Van Der Linden M**, **Falkenhorst G**, **Perniciaro S**, **Imöhl M**. 2015. Effects of infant pneumococcal conjugate vaccination on serotype distribution in invasive pneumococcal disease among children and adults in Germany. PLoS One **10**:e0131494.

173. **Galanis I**, **Lindstrand A**, **Darenberg J**, **Browall S**, **Nannapaneni P**, **Sjöström K**, **Morfeldt E**, **Naucler P**, **Blennow M**, **Örtqvist A**, **Henriques-Normark B**. 2016. Effects of PCV7 and PCV13 on invasive pneumococcal disease and carriage in Stockholm, Sweden. Eur Respir J **47**:1208–1218.

174. **Furuya Y**, **Yamagishi Y**, **Okade H**, **Kadota T**, **Funatsu T**, **Eto M**, **Nomura N**, **Mitsuyama J**, **Mikamo H**. 2017. Impact of the pneumococcal conjugate vaccine on serotype distribution of adult non-invasive *Streptococcus pneumoniae* isolates in Tokai region, Japan, 2008–2016. J Infect Chemother **23**:394–399.

175. **Camilli R**, **D'Ambrosio F**, **Del Grosso M**, **de Araujo FP**, **Caporali MG**, **Del Manso M**, **Gherardi G**, **D'Ancona F**, **Pantosti A**, **Pneumococcal Surveillance Group**. 2017. Impact of pneumococcal conjugate vaccine (PCV7 and PCV13) on pneumococcal invasive diseases in Italian children and insight into evolution of pneumococcal population structure. Vaccine **35**:4587–4593.

176. **Yildirim I**, **Hanage WP**, **Lipsitch M**, **Shea KM**, **Stevenson A**, **Finkelstein J**, **Huang SS**, **Lee GM**, **Kleinman K**, **Pelton SI**. 2010. Serotype specific invasive capacity and persistent reduction in invasive pneumococcal disease. Vaccine **29**:283–288.

177. **van Hoek AJ**, **Andrews N**, **Waight PA**, **George R**, **Miller E**. 2012. Effect of serotype on focus and mortality of invasive pneumococcal disease: coverage of different vaccines and insight into non-vaccine serotypes. PLoS One **7**:e39150.

178. **Nakano S**, **Fujisawa T**, **Ito Y**, **Chang B**, **Suga S**, **Noguchi T**, **Yamamoto M**, **Matsumura Y**, **Nagao M**, **Takakura S**, **Ohnishi M**, **Ihara T**, **Ichiyama S**. 2016. Serotypes, antimicrobial susceptibility, and molecular epidemiology of invasive and non-invasive *Streptococcus pneumoniae* isolates in paediatric patients after the introduction of 13-valent conjugate vaccine in a nationwide surveillance study conducted in Japan in 2012-2014. Vaccine **34**:67–76.

179. **Chang Q**, **Stevenson AE**, **Croucher NJ**, **Lee GM**, **Pelton SI**, **Lipsitch M**, **Finkelstein JA**, **Hanage WP**. 2015. Stability of the pneumococcal population structure in Massachusetts as PCV13 was introduced. BMC Infect Dis **15**:68.

180. **Demczuk WHB**, **Martin I**, **Hoang L**, **Van Caeseele P**, **Lefebvre B**, **Horsman G**, **Haldane D**, **Gubbay J**, **Ratnam S**, **German G**, **Daley Bernier J**, **Strudwick L**, **McGeer A**, **Zhanel GG**, **Van Domselaar G**, **Graham M**, **Mulvey MR**. 2017. Phylogenetic analysis of emergent *Streptococcus pneumoniae* serotype 22F causing invasive pneumococcal disease using whole genome sequencing. PLoS One **12**:e0178040.

181. **van Cuyck H**, **Pichon B**, **Leroy P**, **Granger-Farbos A**, **Underwood A**, **Soullié B**, **Koeck J-L**. 2012. Multiple-locus variable-number tandem-repeat analysis of *Streptococcus pneumoniae* and comparison with multiple loci sequence typing. BMC Microbiol **12**:241.

182. **Nemoy LL**, **Kotetishvili M**, **Tigno J**, **Keefer-Norris A**, **Harris AD**, **Perencevich EN**, **Johnson JA**, **Torpey D**, **Sulakvelidze A**, **Morris JG**, **Stine OC**. 2005. Multilocus sequence typing versus pulsed-field gel electrophoresis for characterization of extended-spectrum beta-lactamase-producing *Escherichia coli* isolates. J Clin Microbiol **43**:1776–1781.

183. **Fakhr MK**, **Nolan LK**, **Logue CM**. 2005. Multilocus sequence typing lacks the discriminatory ability of pulsed-field gel electrophoresis for typing *Salmonella enterica* serovar Typhimurium. J Clin Microbiol **43**:2215–2219.

184. **Nadon C**, **Van Walle I**, **Gerner-Smidt P**, **Campos J**, **Chinen I**, **Concepcion-Acevedo J**, **Gilpin B**, **Smith AM**, **Kam KM**, **Perez E**, **Trees E**, **Kubota K**, **Takkinen J**, **Nielsen EM**, **Carleton H**, **Pietzka A**, **Trost E**, **Kiil K**, **Damjanova I**, **Michelacci V**, **Mossong J**, **Franz E**, **Van Pelt W**, **Wołkowicz T**, **Borges V**, **Jernberg C**, **Alm E**, **Fisher I**, **Grant K**, **Dallman T**, **Agren J**, **Maiden M**, **Brisse S**, **Carrico JA**, **Cochrane G**, **Guerra B**, **Da Silva Felicio MT**, **Palm D**, **Revez J**, **Struelens M**. 2017. PulseNet International: vision for the implementation of whole genome sequencing (WGS) for global foodborne disease surveillance. Eurosurveillance **22**:30544.

185. **Kapatai G**, **Sheppard CL**, **Al-Shahib A**, **Litt DJ**, **Underwood AP**, **Harrison TG**, **Fry NK**. 2016. Whole genome sequencing of *Streptococcus pneumoniae*: development, evaluation and verification of targets for serogroup and serotype prediction using an automated pipeline. PeerJ **4**:e2477.

186. **Ellington MJ**, **Ekelund O**, **Aarestrup FM**, **Canton R**, **Doumith M**, **Giske C**, **Grundman H**, **Hasman H**, **Holden MTG**, **Hopkins KL**, **Iredell J**, **Kahlmeter G**, **Köser CU**, **MacGowan A**, **Mevius D**, **Mulvey M**, **Naas T**, **Peto T**, **Rolain JM**, **Samuelsen**, **Woodford N**. 2017. The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. Clin Microbiol Infect **23**:2–22.

187. **Mellmann A**, **Andersen PS**, **Bletz S**, **Friedrich AW**, **Kohl TA**, **Lilje B**, **Niemann S**, **Prior K**, **Rossen JW**, **Harmsen D**. 2017. High interlaboratory reproducibility and accuracy of next-generation-sequencing-based bacterial genotyping in a ring trial. J Clin Microbiol **55**:908–913.

188. **Maiden MCJ**, **van Rensburg MJJ**, **Bray JE**, **Earle SG**, **Ford SA**, **Jolley KA**, **McCarthy ND**. 2013. MLST revisited: the gene-by-gene approach to bacterial genomics. Nat Rev Microbiol **11**:728–736.

189. **Pontarotti P**. 2016. Evolutionary Biology: Convergent Evolution, Evolution of Complex Traits, Concepts and Methods. Springer International Publishing.

APPENDIX A

Detailed breakdown of which work described in this thesis was performed by the candidate.

Thesis Section	Test/Procedure	Breakdown of Contributions
1.7	Serotyping and serotype demographics analysis	Performed by the Streptococcus and STI Unit at PHAC-NML as part of routine reference laboratory activities.
4.1	Antimicrobial susceptibility testing	Routinely performed by CARA as part of the ongoing SAVE study; however, testing for the 2012 and 2013 study years was performed by the candidate.
4.2	Polymerase chain reaction detection of pilus genes	Performed by the candidate.
4.3	Pulsed-field gel electrophoresis	Performed by the candidate.
4.4	Multi-locus sequence typing	DNA extraction/amplification and sequence analysis were performed by the candidate. Product purification and sequencing reactions were performed by the Genomics Core Facility at PHAC-NML.
4.5	Whole genome sequencing	DNA extraction and quantification were performed by the candidate. Sequencing reactions were performed by the Genomics Core Facility at PHAC-NML. All bioinformatics analyses were performed by the candidate.

APPENDIX B

Detailed information for 162 SAVE and 30 BESST isolates selected for WGS.

Study	Year	Isolate Number	Isolate Source[A]	Serotype	ST	Phenotype[B]	Genotype[C]	PBP Pattern[D]	Province[E]	Age Group[F]	Gender	Original Reason for Isolate Inclusion[G]
						Antimicrobial Resistance			**Demographic Information**			
BESST	2007	72063/ SC16-2494-P	Blood	19A	13191	CLR, CLD	I100L	0-31-0	ON	≥65	F	B
BESST	2007	72136/ SC10-0657-P	Blood	25F	105	DOX	*tetM, folP*ins	0-0-23	ON	≥65	M	B
BESST	2007	72331/ SC16-2495-P	Blood	14	9	CLR	*mefA*	2-0-64	QC	≥65	F	B
BESST	2007	72813/ SC16-2496-P	Blood	23A	1839	-	-	12-0-0	SK	≥65	M	B
BESST	2007	73508/ SC16-2497-P	Blood	4	3004	-	-	0-0-3	ON	≥65	M	B
BESST	2007	73852/ SC10-0275-P	Blood	38	393	-	-	2-4-0	QC	≥65	F	B
BESST	2007	73861/ SC16-2498-P	Blood	23F	36	-	-	0-0-2	QC	≥65	M	B
BESST	2007	73892/ SC16-2499-P	Blood	22F	433	CLR	*mefA*	1-2-2	QC	≥65	M	B
BESST	2007	75723/ SC16-2500-P	Blood	11A	62	-	S79F	24-6-10	QC	≥65	F	B
BESST	2007	76271/ SC16-2501-P	Blood	10A	585	-	*folP*ins	0-0-0	ON	≥65	F	B
BESST	2007	78532/ SC10-0283-P	Blood	33F	100	-	-	2-0-6	SK	≥65	F	B
BESST	2008	81322/ SC16-2502-P	Blood	3	180	-	-	2-3-2	ON	≥65	M	B
BESST	2008	81641/ SC16-2503-P	Blood	9N	517	-	-	1-0-0	NS	≥65	M	B
BESST	2008	82193/ SC16-2504-P	Blood	15A	63	CLR, CLD, DOX	*ermB, tetM*	24-27-28	QC	≥65	F	B
BESST	2008	82386/ SC16-2505-P	Blood	15B	199	-	-	2-0-2	SK	≥65	M	B
BESST	2008	82669/ SC16-2506-P	Blood	3	180	-	-	2-3-2	ON	≥65	F	B
BESST	2008	82810/ SC16-2507-P	Blood	22F	433	-	-	1-2-2	QC	≥65	F	B

252

Study	Year	ID	Source	Serotype	ST	Resistance	Genes	Number	Region	Age	Sex	Class
BESST	2008	83492/ SC16-2508-P	Blood	19A	63	CLR, CLD, DOX	e*rmB, tetM*	67-27-35	ON	≥65	M	B
BESST	2008	83974/ SC10-0292-P	Blood	6C	3101	CLR	*mefA*	2-0-3	NS	≥65	M	B
BESST	2008	84034/ SC16-2509-P	Blood	9V	1556	SXT	I100L+*folP*ins	15-12-18	NS	≥65	F	B
BESST	2008	84129/ SC10-0293-P	Blood	33F	13186	-	*folP*ins	2-0-128	SK	≥65	F	B
BESST	2008	84264/ SC16-2510-P	Blood	14	124	-	-	23-26-2	QC	≥65	M	B
BESST	2009	85096/ SC10-0180-P	Blood	18C	496	-	-	0-0-0	QC	≥65	M	B
BESST	2009	85139/ SC16-2511-P	Blood	8	404	-	-	3-2-5	ON	≥65	M	B
BESST	2009	85244/ SC10-0184-P	Blood	7F	191	-	-	2-0-2	ON	≥65	F	B
BESST	2009	85267/ SC10-0185-P	Blood	19A	695	-	-	8-0-11	ON	≥65	F	B
BESST	2009	86884/ SC10-0234-P	Blood	3	180	-	-	2-3-2	QC	≥65	F	B
BESST	2009	87356/ SC10-0551-P	Blood	19A	199	-	-	0-29-11	ON	≥65	F	B
BESST	2009	87594/ SC10-0566-P	Blood	4	3004	CLR	-	0-0-3	NB	≥65	F	B
BESST	2009	87701/ SC10-0569-P	Blood	9N	10191	-	-	1-0-0	ON	≥65	F	B
SAVE	2011	SC11-0099-P	Blood	9V	156	CLR, PEN, SXT	*mefA*, I100L+*folP*ins	15-12-18	ON	≥65	F	CS-D
SAVE	2011	SC11-0288-P	Blood	38	393	-	-	2-4-0	MB	≥65	M	B
SAVE	2011	SC11-0858-P	Blood	9N	66	-	-	69-0-0	MB	≥65	M	CS-D
SAVE	2011	SC11-0900-P	Blood	19A	320	DOX, PEN, SXT	*mefA, ermB*, tetM*, I100L+*folP*ins	13-11-16	ON	≥65	F	MDR
SAVE	2011	SC11-0902-P	Blood	19A	81	CHL, CLR, DOX, PEN, SXT	*ermB, tetM, cat*, I100L+*folP*ins	13-12-16	ON	≥65	F	CS-P
SAVE	2011	SC11-1200-P	Blood	19A	320	CLR, CLD, DOX, LEV, PEN, SXT	*mefA, ermB, tetM*, S79F, I100L+*folP*ins	13-80-123	MB	≥65	M	XDR
SAVE	2011	SC11-1406-P	Blood	5	289	-	*folP*ins	23-0-32	MB	6-<18	F	B
SAVE	2011	SC11-1495-P	Blood	14	9	CLR	*mefA*	2-0-64	SK	≥65	M	B
SAVE	2011	SC11-1574-P	Blood	7C	1797	-	-	2-4-2	SK	≥65	F	B
SAVE	2011	SC11-1577-P	Blood	22F	433	-	-	1-2-2	SK	≥65	M	MDR
SAVE	2011	SC11-1746-P	Blood	34	547	-	-	0-0-2	ON	≥65	M	B
SAVE	2011	SC11-1824-P	Blood	35B	1666	-	-	4-7-7	NB	≥65	M	B
SAVE	2011	SC11-1841-P	Blood	35B	9346	CLR, PEN, SXT	*mefA*, I100L+*folP*ins	4-12-7	ON	≥65	F	CS-P
SAVE	2011	SC11-1862-P	Blood	6B	1876	CLR	*mefA*	36-34-44	ON	≥65	M	B

253

SAVE	2011	SC11-1866-P	Blood	31	13187	-	-	12-0-0	ON	≥65	F	B
SAVE	2011	SC11-1882-P[H]	Blood	7F	13188	CLR, DOX, SXT	-	2-0-?	ON	18-<50	M	MDR
SAVE	2011	SC11-1929-P	Blood	19A	320	CLR, CLD, DOX, PEN, SXT	*mefA, ermB, tetM,* I100L+*folP*ins	13-11-16	MB	≥65	F	MDR
SAVE	2011	SC11-2110-P	Blood	22F	433	LEV	S79Y	1-2-2	QC	≥65	F	B
SAVE	2011	SC11-2116-P	Blood	12F	218	-	-	0-0-0	QC	≥65	M	MDR
SAVE	2011	SC11-2316-P	Blood	20	6805	-	-	1-44-2	ON	≥65	F	B
SAVE	2011	SC11-2325-P	Blood	8	NF	-	-	3-6-5	MB	≥65	M	CS-D
SAVE	2011	SC11-2451-P	Blood	16F	1840	-	-	0-10-127	MB	≥65	F	B
SAVE	2011	SC11-2658-P	Blood	15B	83	CHL, CLR, CLD, DOX, PEN, SXT	*tetM, ermB, cat,* I100L+*folP*ins	70-12-18	QC	≥65	M	B
SAVE	2011	SC11-2703-P[H]	Blood	6C	NF	CLR, CLD, DOX	*ermB, tetM*	2-53-77	NB	≥65	M	MDR
SAVE	2011	SC11-2895-P	Blood	19A	9345	CLR, PEN, SXT	*mefA,* I100L+*folP*ins	68-14-36	NB	50-<65	M	MDR
SAVE	2011	SC11-3271-P	Other	19A	320	CLR, DOX, PEN, SXT	*mefA, ermB, tetM,* I100L+*folP*ins	13-14-26	NB	50-<65	F	MDR
SAVE	2011	SC11-3339-P	Blood	1	306	-	-	23-6-5	ON	50-<65	M	B
SAVE	2011	SC11-3593-P	Blood	35B	558	PEN	-	4-7-7	QC	6-<18	-	B
SAVE	2011	SC11-3712-P	Blood	9N	66	-	S81F	1-0-0	SK	≥65	M	CS-P
SAVE	2011	SC11-3889-P	Blood	7F	63	CLR, CLD, DOX	*ermB, tetM*	24-27-28	SK	≥65	M	MDR
SAVE	2011	SC11-3902-P	Blood	8	63	CLR, CLD, DOX	*ermB, tetM*	24-27-28	ON	50-<65	M	CS-P
SAVE	2011	SC11-3903-P	CSF	3	177	-	-	1-0-0	SK	0-<1	M	B
SAVE	2011	SC11-3929-P	Blood	15A	2613	CLR, CLD, DOX, PEN	*ermB, tetM*	34-75-119	ON	50-<65	M	MDR
SAVE	2011	SC11-4141-P	Blood	23A	190	-	-	2-0-6	ON	≥65	-	B
SAVE	2011	SC11-4183-P	Blood	6D	1692	-	-	2-6-0	ON	≥65	M	B
SAVE	2011	SC11-4254-P	Blood	19A	319	CLR, CLD, DOX, SXT	*ermB, tetM,* I100L+*folP*ins	10-9-13	ON	50-<65	M	MDR
SAVE	2011	SC11-4275-P	Blood	15A	63	CLR, CLD, DOX	*ermB, tetM*	24-27-28	QC	≥65	F	MDR
SAVE	2011	SC11-4375-P	Blood	35F	446	-	-	0-0-3	ON	≥65	F	B
SAVE	2011	SC11-4503-P[H]	Blood	19A	2343	CLR, CLD, DOX, LEV	*ermB, tetM,* S79F+S81F	0-0-122	ON	≥65	F	MDR
SAVE	2011	SC11-4510-P	Other	19A	63	CLR, CLD, DOX	*ermB, tetM*	0-27-28	ON	≥65	F	CS-P
SAVE	2011	SC11-4519-P	Blood	4	205	-	-	0-0-3	NLA	-	-	B
SAVE	2011	SC11-4567-P	Blood	38	393	-	-	2-4-0	NS	≥65	F	B
SAVE	2011	SC11-4599-P	Blood	3	180	CHL, CLR, CLD, DOX	*ermB, tetM, cat*	2-0-2	ON	50-<65	F	MDR
SAVE	2011	SC12-0005-P	Blood	23F	36	-	-	0-0-83	NB	≥65	M	B
SAVE	2011	SC12-0023-P	Blood	15A	63	CLR, CLD, DOX, SXT	*ermB, tetM,* I100L+*folP*ins	24-27-28	ON	18-<50	M	MDR
SAVE	2011	SC12-0054-P	Other	19A	320	CLR, CLD, PEN, SXT	*mefA, ermB, tetM,* I100L+*folP*ins	13-11-16	ON	18-<50	F	MDR
SAVE	2011	SC12-0121-P	Blood	6C	1390	-	-	2-0-2	MB	≥65	M	MDR

SAVE	2011	SC12-0347-P	Blood	24F	230	CLR, DOX	*ermB, tetM, folP*ins	17-15-22	QC	≥65	F	B
SAVE	2011	SC12-0350-P	Blood	22F	433	-	-	1-2-2	QC	≥65	F	CS-D
SAVE	2011	SC12-0688-P	Blood	17F	123	-	-	11-0-0	NB	≥65	M	B
SAVE	2011	SC12-0762-P	Blood	15A	63	CLR, DOX	*ermB, tetM*	24-27-8	QC	≥65	F	MDR
SAVE	2012	SC12-0322-P	Blood	19A	320	CLR, DOX, SXT	*mefA, ermB*, tetM,* I100L+*folP*ins	13-11-16	MB	2-<6	M	MDR
SAVE	2012	SC12-0452-P	Blood	22F	9352	CLR, CLD, DOX, LEV	*ermB, tetM,* S79F+S81L	1-27-2	ON	≥65	F	CS-P
SAVE	2012	SC12-0774-P	Blood	22F	433	CLR, CLD, DOX	-	1-2-2	QC	≥65	M	MDR
SAVE	2012	SC12-0793-P	Blood	15C	8325	-	*folP*ins	2-0-2	ON	≥65	F	B
SAVE	2012	SC12-0846-P	Other	3	180	-	-	2-3-2	SK	≥65	F	MDR
SAVE	2012	SC12-1032-P	Blood	17F	123	CLD	*ermB, tetM*	11-0-0	ON	≥65	F	B
SAVE	2012	SC12-1054-P	Blood	7F	191	-	-	2-0-2	NLA	≥65	F	MDR
SAVE	2012	SC12-1088-P	Blood	16F	1840	-	-	0-10-2	ON	≥65	M	B
SAVE	2012	SC12-1091-P	Blood	3	180	CHL, CLR, DOX	*ermB, tetM, cat*	2-0-111	ON	≥65	M	MDR
SAVE	2012	SC12-1166-P	Blood	35B	198	-	-	0-0-2	PEI	≥65	M	MDR
SAVE	2012	SC12-1225-P	Blood	4	205	-	-	0-0-3	SK	≥65	F	B
SAVE	2012	SC12-1227-P	Blood	12F	10198	CLR, CLD, DOX	*ermB, tetM*	37-4-121	SK	50-<65	F	MDR
SAVE	2012	SC12-1272-P	Other	15A	2613	CLR, CLD, DOX, PEN	*ermB, tetM*	34-75-119	QC	≥65	M	MDR
SAVE	2012	SC12-1277-P	Blood	35F	13189	-	-	0-0-3	QC	≥65	F	B
SAVE	2012	SC12-1374-P	Blood	24B	177	-	-	1-0-0	ON	≥65	M	B
SAVE	2012	SC12-1751-P	Blood	35B	558	CLR, SXT	*mefA,* I100L+*folP*ins	4-79-7	QC	≥65	F	B
SAVE	2012	SC12-1810-P	Blood	19A	320	CLR, CLD, DOX, SXT	*mefA, ermB, tetM,* I100L+*folP*ins	13-11-16	MB	-	-	MDR
SAVE	2012	SC12-1898-P	Blood	6A	1876	LEV	*mefA,* S79F+S81F	36-34-44	ON	≥65	M	B
SAVE	2012	SC12-1906-P	Blood	35B	558	CLR, PEN	*mefA*	4-76-7	ON	≥65	F	MDR
SAVE	2012	SC12-1985-P	Blood	15A	63	CLR, CLD, DOX	*ermB, tetM*	24-27-28	QC	≥65	M	CS-R
SAVE	2012	SC12-1997-P	Blood	33F	9350	CLR, CLD, DOX	*ermB, tetM*	2-0-6	QC	≥65	M	MDR
SAVE	2012	SC12-2045-P	Blood	35F	498	-	-	0-0-3	NS	≥65	F	B
SAVE	2012	SC12-2050-P	Blood	19A	667	-	-	0-0-0	NS	≥65	F	CS-D
SAVE	2012	SC12-2051-P	Blood	19A	320	CLR, CLD, DOX, PEN, SXT	*mefA, ermB, tetM,* I100L+*folP*ins	13-11-16	NS	≥65	M	MDR
SAVE	2012	SC12-2054-P	Blood	18C	496	SXT	*folP*ins	0-0-0	NB	≥65	F	B
SAVE	2012	SC12-2118-P	Blood	19A	320	DOX, PEN, SXT	*mefA, ermB*, tetM,* S79F, I100L+*folP*ins	13-11-120	MB	≥65	M	MDR
SAVE	2012	SC12-2839-P	Blood	19A	63	CLR, CLD, DOX	*ermB, tetM*	67-27-35	NLA	≥65	F	MDR
SAVE	2012	SC12-2856-P	Blood	9N	3224	CLR, DOX	*ermB, tetM*	1-0-0	ON	50-<65	M	B
SAVE	2012	SC12-2899-P	Blood	15A	63	CLR, CLD, DOX	*ermB, tetM*	24-27-28	QC	≥65	F	MDR
SAVE	2012	SC12-2902-P	Blood	15A	2613	CLR, DOX, PEN	*ermB, tetM*	65-75-119	QC	≥65	M	MDR
SAVE	2012	SC12-2905-P	Blood	15C	1262	-	*folP*ins	2-0-2	QC	≥65	M	B
SAVE	2012	SC12-3012-P	CSF	19A	10194	CLR, DOX, PEN	*ermB, tetM, folP*ins	17-39-18	QC	-	M	MDR

SAVE	2012	SC12-3017-P	Other	33F	100	-	*folP*ins	2-0-6	SK	≥65	M	MDR
SAVE	2012	SC12-3268-P	Blood	12F	218	CLR	*mefA*	0-0-0	MB	≥65	F	MDR
SAVE	2012	SC12-3310-P	Blood	22F	9352	CLR, CLD, DOX, LEV	*ermB, tetM,* S79F+S81L	1-27-2	ON	≥65	F	CS-P
SAVE	2012	SC12-3334-P	Blood	19A	320	CLR, DOX, PEN, SXT	*mefA, ermB*, tetM,* I100L+*folP*ins	13-11-16	MB	18-<50	F	MDR
SAVE	2012	SC12-3569-P	Blood	3	180	CHL, CLR, CLD, DOX	*ermB, tetM, cat*	2-0-111	ON	18-<50	F	MDR
SAVE	2012	SC12-3606-P	Blood	19F	2631	-	-	2-4-3	NLA	≥65	F	B
SAVE	2012	SC12-3711-P	Blood	23A	42	CLR, CLD	*ermB*	2-0-6	ON	≥65	F	B
SAVE	2012	SC12-3916-P	Blood	15A	410	-	-	0-0-2	SK	≥65	F	MDR
SAVE	2012	SC12-4014-P	Blood	6B	90	CLR, DOX, PEN, SXT	*ermB, tetM,* I100L+*folP*ins	34-81-43	MB	≥65	F	B
SAVE	2012	SC12-4032-P	Blood	23A	42	-	-	2-0-6	ON	≥65	F	B
SAVE	2012	SC12-4037-P	Blood	15C	1262	-	*folP*ins	2-0-2	ON	≥65	M	B
SAVE	2012	SC12-4157-P	Blood	3	180	CLR, CLD, DOX	*ermB, tetM*	2-3-2	ON	≥65	M	MDR
SAVE	2012	SC13-0056-P	Blood	23B	338	-	-	0-1-1	SK	≥65	M	B
SAVE	2012	SC13-0334-P	Blood	15B	1262	-	*folP*ins	2-0-2	NS	≥65	M	B
SAVE	2012	SC13-2370-P	Blood	15A	2613	CLR, DOX, PEN	*ermB, tetM*	34-75-119	QC	2-<6	M	MDR
SAVE	2013	SC13-0142-P	Blood	19A	193	-	-	66-0-2	MB	-	F	MDR
SAVE	2013	SC13-0368-P	Blood	6B	138	CLR, CLD, DOX, SXT	*ermB, tetM,* I100L+*folP*ins	0-0-0	ON	≥65	F	B
SAVE	2013	SC13-0467-P	CSF	8	1480	-	-	3-5-5	ON	50-<65	M	B
SAVE	2013	SC13-0525-P	Blood	15A	2613	CLR, CLD, DOX	*ermB, tetM*	34-78-119	SK	18-<50	M	MDR
SAVE	2013	SC13-0990-P	Blood	6C	1379	CLR, PEN, SXT	*mefA,* I100L+*folP*ins	19-31-8	ON	50-<65	M	MDR
SAVE	2013	SC13-1198-P	Blood	9N	517	-	-	1-0-0	ON	≥65	F	CS-P
SAVE	2013	SC13-1199-P	Blood	21	432	-	-	0-0-2	ON	≥65	M	B
SAVE	2013	SC13-1642-P	Blood	23F	13190	CHL, CLR, CLD, DOX, LEV, PEN, SXT	*ermB, tetM, cat,* S79F+S81F, I100L+*folP*ins	15-12-18	ON	50-<65	F	XDR
SAVE	2013	SC13-1734-P	Blood	10A	3135	DOX, SXT	*tetM,* I100L+*folP*ins	71-1-22	QC	≥65	M	B
SAVE	2013	SC13-1746-P	Blood	33F	9689	CLR, CLD	*ermB, tetM*	2-0-6	QC	≥65	M	MDR
SAVE	2013	SC13-1756-P	Blood	17F	2355	CLR, CLD, DOX	*ermB, tetM*	67-26-126	ON	≥65	F	B
SAVE	2013	SC13-1760-P	Blood	6A	1876	-	*mefA*	36-34-44	ON	≥65	F	B
SAVE	2013	SC13-1815-P	Blood	23F	36	LEV	S79F+S81F	0-3-2	ON	≥65	M	B
SAVE	2013	SC13-2375-P[H]	CSF	14	9	CLR	*mefA*	2-0-64	MB	≥65	M	CS-R
SAVE	2013	SC13-2507-P	Blood	21	432	-	-	0-0-2	QC	1-<2	F	B
SAVE	2013	SC13-2616-P[I]	Other	15A	3811	-	-	1-0-0	SK	18-<50	F	CS-P
SAVE	2013	SC13-2685-P	Blood	19A	193	-	-	0-0-11	NS	≥65	F	CS-D
SAVE	2013	SC13-2686-P	Blood	37	447	-	-	11-4-0	NS	≥65	M	B

SAVE	2013	SC13-2710-P	Other	19A	320	DOX, PEN, SXT	*mefA, ermB*, tetM,* S81F, I100L+*folP*ins	13-11-16	ON	≥65	M	MDR
SAVE	2013	SC13-2731-P	Other	21	193	-	-	2-0-2	MB	≥65	F	B
SAVE	2013	SC13-2733-P	Blood	11A	10196	LEV, SXT	S79Y+S81F, I100L+*folP*ins	12-4-2	MB	≥65	M	B
SAVE	2013	SC13-3015-P	Blood	3	180	CHL, CLR, CLD, DOX	*ermB, tetM, cat*	2-74-111	ON	≥65	M	MDR
SAVE	2013	SC13-3068-P	Blood	23B	1373	-	*folP*ins	0-1-1	ON	≥65	M	B
SAVE	2013	SC13-3151-P	Blood	19A	179	CLR, CLD, DOX, PEN	*ermB, tetM*	34-32-43	QC	50-<65	F	MDR
SAVE	2013	SC13-3212-P	Blood	19A	320	CLR, CLD, DOX, SXT	*mefA, ermB, tetM,* I100L+*folP*ins	13-11-16	SK	18-<50	M	MDR
SAVE	2013	SC13-3603-P	Blood	12F	6945	CLR, CLD, DOX	*ermB, tetM*	37-4-23	NB	≥65	M	MDR
SAVE	2013	SC13-4131-P	Blood	9V	156	PEN, SXT	I100L+*folP*ins	15-12-18	MB	≥65	F	B
SAVE	2013	SC13-4463-P	Blood	9N	10191	-	-	1-0-0	ON	50-<65	M	B
SAVE	2013	SC13-4497-P	Blood	5	4219	SXT	I100L+*folP*ins	23-0-125	SK	50-<65	M	B
SAVE	2013	SC13-4579-P	Blood	19A	320	CLR, CLD, DOX, PEN, SXT	*mefA, ermB, tetM,* I100L+*folP*ins	13-11-16	SK	≥65	M	MDR
SAVE	2013	SC13-4822-P	Blood	11A	62	-	-	2-6-10	NS	≥65	F	MDR
SAVE	2013	SC13-4858-P	Blood	6C	338	-	-	19-1-1	NLA	50-<65	M	B
SAVE	2013	SC14-0090-P I	Blood	14	156	DOX, LEV, PEN	-	15-12-18	ON	≥65	F	B
SAVE	2014	SC14-0291-P	Blood	19F	271	CLR, CLD, DOX, PEN, SXT	*ermB, mefA, tetM,* I100L+*folP*ins	13-11-33	ON	≥65	M	B
SAVE	2014	SC14-0428-P	Blood	19A	320	CLR, CLD, DOX, PEN, SXT	*mefA, ermB, tetM.* I100L+*folP*ins	13-11-16	NB	50-<65	M	XDR
SAVE	2014	SC14-0474-P	Blood	18C	496	-	-	0-0-0	ON	≥65	F	B
SAVE	2014	SC14-0594-P	Blood	20	1257	-	-	1-4-2	QC	≥65	F	B
SAVE	2014	SC14-0911-P	Blood	19A	320	CLR, CLD, DOX, PEN, SXT	*mefA, ermB, tetM,* I100L+*folP*ins	13-11-16	NLA	≥65	M	MDR
SAVE	2014	SC14-0933-P	Blood	11A	62	CLR, CLD	*ermB*	2-6-0	ON	50-<65	M	MDR
SAVE	2014	SC14-0936-P	Blood	3	180	CHL, CLR, CLD, DOX	*ermB, tetM, cat*	2-0-111	ON	≥65	M	MDR
SAVE	2014	SC14-1096-P	Other	22F	433	CLR, CLD	*ermB*	1-2-2	ON	50-<65	F	MDR
SAVE	2014	SC14-1327-P	Blood	29	7486	CLR	*mefA*	4-7-7	ON	≥65	M	B
SAVE	2014	SC14-1380-P	Blood	15A	374	CLR, CLD, DOX	*ermB, tetM*	24-27-119	NLA	≥65	M	MDR
SAVE	2014	SC14-1411-P	Blood	16F	570	-	-	0-6-2	ON	≥65	F	B
SAVE	2014	SC14-1542-P	Blood	19A	320	CLR, CLD, DOX, SXT	*ermB, mefA, tetM,* I100L+*folP*ins	13-11-16	NB	≥65	M	MDR
SAVE	2014	SC14-1569-P	Blood	13	574	-	-	11-0-0	ON	≥65	M	B
SAVE	2014	SC14-1578-P	Blood	33F	9348	CLR, CLD	*ermB, tetM*	2-0-6	QC	≥65	F	MDR
SAVE	2014	SC14-1599-P	Blood	34	5854	-	-	0-0-0	NLA	≥65	F	B
SAVE	2014	SC14-1645-P	Blood	10A	97	-	-	0-0-2	MB	≥65	F	B
SAVE	2014	SC14-1752-P	Blood	24F	72	-	-	2-0-0	ON	≥65	F	B
SAVE	2014	SC14-2419-P	Blood	18C	496	-	-	0-0-0	QC	≥65	M	B

257

SAVE	2014	SC14-3062-P	Blood	35B	558	CLR, LEV	*mefA*, S81F	4-7-7	ON	50-<65	M	B
SAVE	2014	SC14-3643-P	Blood	31	1766	-	-	12-0-0	ON	≥65	M	B
SAVE	2014	SC14-3649-P	Blood	9V	156	CLR, PEN, SXT	*mefA*, I100L+*folP*ins	15-12-18	ON	18-<50	-	B
SAVE	2014	SC14-3781-P	Blood	19A	320	CLR, DOX, PEN, SXT	*mefA, ermB*, tetM*, I100L+*folP*ins	13-11-16	SK	≥65	F	MDR
SAVE	2014	SC14-3783-P	Blood	15A	2613	CLR, CLD, DOX, PEN	*ermB, tetM*	34-75-119	SK	≥65	F	MDR
SAVE	2014	SC14-3841-P	Blood	11A	99	CLR, CLD, DOX, SXT	*ermB, tetM*, I100L+*folP*ins	23-77-0	QC	≥65	F	MDR
SAVE	2014	SC14-3964-P	Blood	31	433	-	-	1-2-0	ON	≥65	M	B
SAVE	2014	SC14-4061-P	Blood	11B	10552	-	-	1-4-0	QC	≥65	M	B
SAVE	2014	SC14-4115-P	Blood	15B	199	CLR	*mefA*	2-0-124	SK	≥65	M	B
SAVE	2014	SC14-4407-P	Blood	6A	490	LEV	*mefA*, S79Y+S81F, I100L+*folP*ins	1-4-2	QC	≥65	M	B
SAVE	2014	SC15-0236-P	Blood	1	306	-	-	23-6-5	NS	18-<50	F	B
SAVE	2014	SC15-0238-P	Blood	15A	9355	CLR, CLD, DOX	*ermB, tetM*	24-27-28	NS	≥65	F	MDR
SAVE	2014	SC15-0672-P	Blood	34	5854	-	-	0-0-0	NLA	≥65	F	B

[A] CSF, cerebrospinal fluid; Other, isolate was collected from a sterile site other than blood/CSF.

[B] CHL, chloramphenicol; CLD, clindamycin; CLR, clarithromycin; DOX, doxycycline; LEV, levofloxacin; PEN, penicillin; SXT, trimethoprim-sulfamethoxazole.

[C] I100L = *folA* mutation; *folP*ins = one or two codon insert in *folP*, see Table 5.5; S79 mutations in *parC*, S81 mutations in *gyrA*.

[D] PBP 1A-2B-2X.

[E] MB, Manitoba; NB, New Brunswick; NLA, Newfoundland and Labrador; NS, Nova Scotia; ON, Ontario; PEI, Prince Edward Island; QC, Quebec; SK, Saskatchewan.

[F] Age in years.

[G] B, background; CS-D, capsule switch donor; CS-P, capsule switch progeny; CS-R, capsule switch recipient; MDR, multidrug resistant; XDR, extensively drug resistant.

[H] Isolate mapped poorly to the reference in phylogenetic analysis and was removed from further analysis of that type.

[I] Would not grow for repeat MIC testing, were removed from the analysis of resistance genes.

* Truncated *ermB*.

APPENDIX C

Transpeptidase domain sequences identified in 196 invasive *S. pneumoniae* isolates collected in Canada from 2007-2014. Sequences identified in this study are labeled as such, with all other sequences previously described by Metcalf *et al*. (38).

PBP1A

>0

TMKPITDYAPALEYGVYDSTATIVHDEPYNYPGTNTPVYNWDRGYFGNITLQYALQQS
RNVPAVETLNKVGLNRAKTFLNGLGIDYPSIHYSNAISSNTTESDKKYGASSEKMAAAY
AAFANGGTYYKPMYIHKVVFSDGSEKEFSNVGTRAMKETTAYMMTDMMKTVLTYGT
GRNAYLAWLPQAGKTGTSNYTDEEIENHIKTSQFVAPDELFAGYTRKYSMAVWTGYSN
RLTPLVGNGLTVAAKVYRSMMTYLSEGSNPEDWNIPEGLYRNGEFV

>1

TMKPITDYAPALEYGIYDSTATIVHDEPYNYPGTNTPVYNWDRGYFGNITLQYALQQSR
NVPAVETLNKVGLNRAKTFLNGLGIDYPSIHYSNAISSNTTESDKKYGASSEKMAAAYA
AFANGGTYYKPMYIHKVVFSDGSEKEFSNVGTRAMKETTAYMMTDMMKTVLSYGTG
RNAYLAWLPQAGKTGTSNYTDEEIENHIKTSQFVAPDELFAGYTRKYSMAVWTGYSNR
LTPLVGNGLTVAAKVYRSMMTYLSEGSNPEDWNIPEGLYRNGEFV

>2

TMKPITDYAPALEYGVYDSTATIVHDEPYNYPGTNTPVYNWDRGYFGNITLQYALQQS
RNVPAVETLNKVGLNRAKTFLNGLGIDYPSIHYSNAISSNTTESDKKYGASSEKMAAAY
AAFANGGTYYKPMYIHKVVFSDGSEKEFSNVGTRAMKETTAYMMTDMMKTVLSYGT
GRNAYLAWLPQAGKTGTSNYTDEEIENHIKTSQFVAPDELFAGYTRKYSMAVWTGYSN
RLTPLVGNGLTVAAKVYRSMMTYLSEGSNPEDWNIPEGLYRNGEFV

>3

TMKPITDYAPALEYGVYDSTAAIVHDEPYNYPGTNTPVYNWDRGYFGNITLQYALQQS

RNVPAVETLNKVGLNRAKTFLNGLGIDYPSIHYSNAISSNTTESDKKYGASSEKMAAAY

AAFANGGTYYKPMYIHKVVFSDGSEKEFSNVGTRAMKETTAYMMTDMMKTVLTYGT

GRNAYLAWLPQAGKTGTSNYTDEEIENHIKTSQFVAPDELFAGYTRKYSMAVWTGYSN

RLTPLVGNGLTVAAKVYRSMMTYLSEGSNPEDWNIPEGLYRNGEFV

>4

SMKPITDYAPALEYGVYDSTATMVNDIPYNYPGTSTPVYNWDRAYFGNISLQYALQQS

RNVPAVETLNKVGLDRAKTFLNGLGIDYPSIHYSNAISSNTTESSKQYGASSEKMAAAY

AAFANGGIYHKPMYINKVVFSDGSEKEFSDAGTRAMKETTAYMMTEMMKTVLTYGTG

RGAYLPWLPQAGKTGTSNYTDDEIEKYIKNTGYVAPDEMFVGYTRKYSMAVWTGYSN

RLTPIIGDGFLVAAKVYRSMISYLSEDDHPGDWTMPEGVYRSGEFV

>8

TMKPITDYAPALEYGVYDSTATIVHDEPYNYPGTDTPVYNWDRGYFGNITLQYALQQS

RNVPAVETLNKVGLNRAKTFLNGLGIDYPSLHYSNAISSNTTESDKKYGASSEKMAAAY

AAFANGGTYYKPMYIHKVVFSDGSEKEFSNVGTRAMKETTAYMMTDMMKTVLVYGI

GRGAYLPWLPQAGKTGTSNYTDEEIEKYIKNTGYVAPDEMFVGYTRKYAMAVWTGYS

NRLTPLVGDGLTVAAKVYRSMMTYLSEGSNPEDWNIPEGLYRNGEFV

>10

TMKSITDYAPALEYGVYDSTATIVHDVPYNYPGTDTPVYNWDKGYFGNITIQYALQQSR

NVTAVETLNKVGLDRAKTFLNGLGIDYPSIHYANAISSNTTESNKQYGASSEKMAAAYA

AFANGGIYHKPMYINKVVFSDGSEKEFSDAGTRAMKETTAYMMTEMMKTVLAYGTGR

GAYLPWLAQAGKTGTSNYTDDEIEKHIKNTGYVAPDETFVGYTRKYSMAVWTGYSNR

LTPIVGDGFLVAAKVYRSMITYLSEDTHPEDWTMPDGLFRNGEFV

>11

TMKPITDYAPALEYGVYDSTATIVHDEPYNYPGTNTPVYNWDRGYFGNITLQYALQQS
RNVPAVETLNKVGLNRAKTFLNGLGIDYPSIHYSNAISSNTTESDKKYGASSEKMAAAY
AAFANGGTYYKPMYIHKVVFSDGSEKEFSNVGTRAMKETTAYMMTDMMKTVLTYGT
GRNAYLAWLPQAGKTGTSNYTDEEIENHIKTSQFVAPDELFAGYTRKYSMAVWTGYSN
RLTPLVGNGLTVAAKVYRSMMTYLSEGRNPEDWNIPEGLYRNGEFV

>12

TMKPITDYAPALEYGVYDSTATIVHDEPYNYPGTNTPVYNWDRGYFGNITLQYALQQS
RNVPAVETLNKVGLNRAKTFLNGLGIDYPSIHYSNAISSNTTESDKKYGASSEKMAAAY
AAFANGGTYYKPMYIHKVVFSDGSEKEFSNVGTRAMKETTAYMMTEMMKTVLSYGT
GRNAYLAWLPQAGKTGTSNYTDEEIENHIKTSQFVAPDELFAGYTRKYSMAVWTGYSN
RLTPLVGNGLTVAAKVYRSMMTYLSEGSNPEDWNIPEGLYRNGEFV

>13

SMKPITDYAPALEYGVYDSTASIVHDVPYNYPGTDTPLYNWDHVYFGNITIQYALQQSR
NVTAVETLNKVGLDRAKTFLNGLGIDYPSMHYANAISSNTTESNKKYGASSEKMAAAY
AAFANGGIYHKPMYINKIVFSDGSEKEFSDAGTRAMKETTAYMMTEMMKTVLTYGTG
RGAYLPWLPQAGKTGTSNYTDEEIEKYIKNTGYVAPDEMFVGYTRKYSMAVWTGYSN
RLTPIVGDGFLVAAKVYRSMMTYLSEGSNPEDWNIPEGLYRNGEFV

>15

AMKPITDYAPAIEYGVYDSTATMVNDIPYNYPGTSTPVYNWDRAYFGNITLQYALQQS
RNVTAVETLNKVGLDRAKTFLNGLGIDYPSMHYANAISSNTTESNKQYGASSEKMAAA
YAAFANGGTYYKPMYIHKVVFSDGSKKEFSNVGTRAMKETTAYMMTDMMKTVLTYG

TGRGAYLPWLPQAGKTGTSNYTDEEVENHIKNTGYVAPDEMFVGYTRKYSMAVWTGY
SNRLTPIVGDGFLVAAKVYRSMITYLSEDTHPEDWTMPDGLFRNGEFV

>17

AMKPITDYAPALEYDIYDSTASIVHDVPYNYPGTDTPLYNWDKVYFGNITIQYALQQSR
NVTAVETLNKVGLDRAKTFLNGLGIDYPSMHYANAISSNTTESNKKYGASSEKMAAAY
AAFANGGIYHKPMYINKIVFSDGSSKEYADPGTRAMKETTAYMMTEMMKTVLAYGTG
RGAYLPWLPQAGKTGTSNYTDDEIENYIKNTGYVAPDEMFVGYTRKYSMAVWTGYSN
RLTPIVGDGFYVAAKVYRSMMTYLSEDNNPGDWTMPEGLYRSGEFV

>19

TMKPITDYAPALEYGVYDSTATIVHDEPYNYPGTDTPVYNWDRGYFGNITLQYALQQS
RNVPAVETLNKVGLNRAKTFLNGLGIDYPSLHYSNAISSNTTESDKKYGASSEKMAAAY
AAFANGGTYYKPMYIHKVVFSDGSEKEFSNVGTRAMKETTAYMMTDMMKTVLVYGI
GRGAYLPWLPQAGKTGTSNYTDEEIEKYIKNTGYVAPDEMFVGYTRKYAMAVWTGYS
NRLTPLVGDGLTVAAKVYRSMMTYLSEGSNPEDWNMPEGLYRNGEFV

>23

TMKPITDYAPALEYGIYDSTATIVHDEPYNYPGTNTPVYNWDRGYFGNITLQYALQQSR
NVPAVETLNKVGLNRAKTFLNGLGIDYPSIHYSNAISSNTTESDKKYGASSEKMAAAYA
AFANGGTYYKPMYIHKVVFSDGSEKEFSNVGTRAMKETTAYMMTDMMKTVLTYGTG
RNAYLAWLPQAGKTGTSNYTDEEIENHIKTSQFVAPDELFAGYTRKYSMAVWTGYSNR
LTPLVGNGLTVAAKVYRSMMTYLSEGSNPEDWNIPEGLYRNGEFV

>24

TMKPITDYAPAIEYGVYDSTATMVNDIPYNYPGTSTPVYNWDRAYFGNITLQYALQQSR
NVPAVETLNKVGLDRAKNFLNGLGIDYPDMHYSNAISSNTTESNKQYGASSEKMAAAF

AAFANGGIYHKPMYINKIVFSDGSEKEFSDAGTRAMKETTAYMMTEMMKTVLSYGTG

RNAYLAWLPQAGKTGTSNYTDEEIENHIKTSQFVAPDELFAGYTRKYSMAVWTGYSNR

LTPLVGNGLTVAAKVYRSMMTYLSEGSNPEDWNIPEGLYRNGEFV

>34

AMKPITDYAPAIEYGVYDSTATMVNDIPYNYPGTSTPVYNWDRAYFGNITLQYALQQS

RNVTAVETLNKVGLDRAKTFLNGLGIDYPSMHYANAISSNTTESNKQYGASSEKMAAA

YAAFANGGIYHKPMYINKVVFSDGSKKEFSDVGTRAMKETTAYMMTEMMKTVLAYG

TGRGAYLPWLAQAGKTGTSNYTDDEIEKHIKNTGYVAPDEMFVGYTRKYSMAVWTGY

SNRLTPIVGDGFLVAAKVYRSMITYLSEDTHPEDWTMPDGLFRNGEFV

>36

TMKPITDYAPALEYGVYDSTATIVHDEPYNYPGTNTPVYNWDRGYFGNITLQYALQQS

RNVPAVETLNKVGLNRAKTFLNGLGIDYPSIHYSNAISSNTTESDKKYGASSEKMAAAY

AAFANGGTYYKPMYIHKVVFSDGSEKEFSNVGTRAMKETTAYMMTDMMKTVLVYGI

GRGAYLPWLPQAGKTGTSNYTDEEIEKYIKNTGYVAPDEMFVGYTRKYAMAVWTGYS

NRLTPLVGDGLTVAAKVYRSMMTYLSEGSNPEDWNIPEGLYRNGEFV

>37

TMKPITDYAPALEYGVYDSTATIVHDEPYNYPGTNTPVYNWDRGYFGNITLQYALQQS

RNVPAVETLNKVGLNRAKTFLNGLGIDYPSIHYSNAISSNTTESDKKYGASSEKMAAAY

AAFANGGTYYKPMYIHKVVFSDGSEKEFSNVGTRAMKETTAYMMTDMMKTVLSYGT

GRNAYLAWLPQAGKTGTSNYTDEEIENHIKTSQFVAPDELFAGYTRKYSMAVWTGYSN

RLTPLVGNGLTVAAKVYRSMMTYLSEGRNPEDWNIPEGLYRNGEFV

>65 (this study)

AMKPITDYAPAIEYGVYDSTATMVNDIPYNYPGTNTPVYNWDRAYFGNITLQYALQQS

RNVTAVETLNKVGLDRAKTFLNGLGIDYPSMHYANAISSNTTESNKQYGASSEKMAAA

YAAFANGGIYHKPMYINKVVFSDGSKKEFSDVGTRAMKETTAYMMTEMMKTVLAYG

TGRGAYLPWLAQAGKTGTSNYTDDEIEKHIKNTGYVAPDEMFVGYTRKYSMAVWTGY

SNRLTPIVGDGFLVAAKVYRSMITYLSEDTHPEDWTMPDGLFRNGEFV

>66 (this study)

TMKPITDYAPALEYGVYDSTATIVHDEPYNYPGTNTPVYNWDRGYFGNITLQYALQQS

RNVPAVETLNKVGLNRAKTFLNGLGIDYPSIHYSNAISSNTTESDKKYGASSEKMAAAY

AAFANGGTYYKPMYIHKVVFSDGSEKEFSNVGTRAMKETTAYMMTDMMKTVLSYGT

GRTAYLAWLPQAGKTGTSNYTDEEIENHIKTSQFVAPDELFAGYTRKYSMAVWTGYSN

RLTPLVGNGLTVAAKVYRSMMTYLSEGSNPEDWNIPEGLYRNGEFV

>67 (this study)

TMKPITDYAPALEYGVYDSTATIVHDEPYNYPGTNTPVYNWDRGYFGNITLQYALQQS

RNVPAVETLNKVGLNRAKTFLNGLGIDYPSIHYSNAISSNTTESDKKYGASSEKMAAAY

AAFANGGTYYKPMYIHKVVFSDGSEKEFSNVGTRAMKETTAYMMTEMMKTVLSYGT

GRNAYLAWLPQAGKTGTSNYTDEEIEKHIKNTGYVAPDETFVGYTRKYSMAVWTGYT

NRLTPIMGDGLTVAAKVYRSMMTYLSEGSNPEDWNIPEGLYRNGEFV

>68 (this study)

AMKPITDYAPAIEYGVYDSTATMVNDIPYNYPGTSTPVYNWDRAYFGNITLQYALQQS

RNVTAVETLNKVGLDRAKTFLNGLGIDYPSMYYANAISSNTTESNKQYGASSEKMAVA

YAAFANGGIYHKPMYINKIVFSDGSEKEFSDAGTRAMKETTAYMMTEMMKTVLVYGI

GRGAYLPWLPQAGKTGTSNYTDEEIEKYVKNTGYVAPDESFVGYTPKYSMAVWTGYS

NRFTPIVGDGFLVAAKVYRSMISYLSEDEQPEDWTMPEGLYRNGEFV

>69 (this study)

TMKPITDYAPALEYGVYDSTATIVHDEPYNYPGTNTPVYNWDRGYFGNITLQYALQQS

RNVPAVETLNKVGLNRAKTFLNGLGIDYPSIHYSNAISSNTTESDKKYGANSEKMAAAY

AAFANGGTYYKPMYIHKVVFSDGSEKEFSNVGTRAMKETTAYMMTDMMKTVLSYGT

GRNAYLAWLPQAGKTGTSNYTDEEIENHIKTSQFVAPDELFAGYTRKYSMAVWTGYSN

RLTPLVGNGLTVAAKVYRSMMTYLSEGSNPEDWNIPEGLYRNGEFV

>70 (this study)

AMKPITDYAPAIEYGVYDSTATMVNDIPYNYPGTSTPVYNWDRAYFGNITLQYALQQS

RNVTAVETLNKVGLDRAKTFLNSLGIDYPSMHYANAISSNTTESNKQYGASSEKMAAA

YAAFANGGTYYKPMYIHKVVFSDGSKKEFSNVGTRAMKETTAYMMTDMMKTVLTYG

TGRGAYLPWLPQAGKTGTSNYTDEEVENHIKNTGYVAPDEMFVGYTRKYSMAVWTGY

SNRLTPIVGDGFLVAAKVYRSMITYLSEDTHPEDWTMPDGLFRNGEFV

>71 (this study)

TMKPITDYAPALEYGVYDSTATIVHDEPYNYPGTDTPVYNWDRGYFGNITLQYALQQS

RNVPAVETLNKVGLNRAKTFLNGLGIDYPSIHYSNAISSNTTESDKKYGASSEKMAAAY

AAFANGGTYYKPMYIHKVVFSDGSEKEFSNVGTRAMKETTAYMMTDMMKTVLTYGA

GRNAYLAWLPQAGKTGTSNYTDEEIENHIKTSQYVAPDELFAGYTRKYSMAVWTGYSN

RLTPLVGDGLTVAAKVYRSMMTYLSEGSNPEDWNIPEGLYRNGEFV

PBP2B

>0

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQSIVFQGSAPINSWYTQAYGSFPITAVQ

ALEYSSNTYMVQTALGLMGQTYQPNMFVGTSNLESAMEKLRSTFGEYGLGTATGIDLP

DESTGFVPKEYSFANYITNAFGQFDNYTPMQLAQYVATIANNGVRVAPRIVEGIYGNND

KGGLGDLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSGLTTGRAFSNGALVSISG

KTGTAESYVADGQQATNTNAVAYAPSDNPQIAVAVVFPHNTN

>1

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQPIVFQGSAPINSWYTQAYGSFPITAVE

ALEYSSNAYMVQTALGIMGQTYQPNMFVLTNNLESAMGKLRSTFAEYGLGASTGIDLP

DESTGFIPKEYNFANYITNAFGQFDNYTPMQLAQYVGTIANNGVRIAPHIVEGIYGNNEQ

GGLGNLIQSVESKEMNKINISESDVSILQQGFYQVSHGGSALTTGRAFSNGASVSISGKTG

TAESYVEGGQEANNTNAVAYAPSDNPQIAVAVVFPHNTN

>2

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQSIVFQGSAPINSWYTQAYGSFPITAVQ

ALEYSSNTYMVQTALGLMGQTYQPNMFVGTSNLESAMEKLRSTFGEYGLGTATGIDLP

DESTGFVPKEYSFANYITNAFGQFDNYTPMQLAQYVATIANDGVRVAPRIVEGIYGNND

KGGLGDLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSELTTGRAFSNGALVSISG

KTGTAESYVADGQQATNTNAVAYAPSDNPQIAVAVVFPHNTN

>3

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQSIVFQGSAPINSWYTQAYGSFPITAVQ

ALEYSSNTYMVQTALGLMGQTYQPNMFVGTSNLESAMEKLRSTFGEYGLGTATGIDLP

DESTGFVPKEYSFANYITNAFGQFDNYTPMQLAQYVATIANNGVRVAPRIVEGIYGNND

KGGLGDLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSGLTTGRAFSNGALVSISG

KTGTAESYVADGQQATNTNAVAYAPYDNPQIAVAVVFPHNTN

>4

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQSIVFQGSAPINSWYTQAYGSFPITAVQ

ALEYSSNTYMVQTALGLMGQTYQPNMFVGTSNLESAMEKLRSTFGEYGLGTATGIDLP

DESTGFVPKEYSFANYITNAFGQFDNYTPMQLAQYVATIANNGVRVAPRIVEGIYGNND

KGGLGDLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSELTTGRAFSNGALVSISG

KTGTAESYVADGQQATNTNAVAYAPSDNPQIAVAVVFPHNTN

>5

TNVFAPGSVVKAATISSGWENGVLSGNQTLTDQSIVFQGSAPINSWYTQAYGSFPITAVQ

ALEYSSNTYMVQTALGLMGQTYQPNMFVGTSNLEFAMEKLRSTFGEYGLGTATGIDLP

DESTGFVPKEYSFANYITNAFGQFDNYTPMQLAQYVATIANDGVRVAPRIVEGIYGNND

KGGLGDLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSELTTGRAFSNGALVSISG

KTGTAESYVADGQQATNTNAVAYAPSDNPQIAVAVVFPHNTN

>6

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQSIVFQGSAPINSWYTQAYGSFPITAVQ

ALEYSSNTYMVQTALGLMGQTYQPNMFVGTSNLESAMEKLRSTFGEYGLGTATGIDLP

DESTGFVPKEYSFANYITNAFGQFDNYTPMQLAQYVATIANNGVRVAPRIVEGIYGNND

KGGLGDLIQQLQPTEINKVNISDSDMSILHQGFYQVAHGTSGLTTGRAFSNGALVSISGK

TGTAESYVADGQQATNTNAVAYAPSDNPQIAVAVVFPHNTN

>7

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQPIVFQGSAPIYSWYKLAYGSFPITAVE

ALEYSSNAYMVQTALGIMGQTYQPNMFVGTSNLETAMGKLRATFGEYGLGAATGIDLP

DESTGFVPKEYSFANYITNAFGQFDNYTPMQLAQYVATIANDGVRVAPRIVEGIYGNND

KGGLGDLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSGLTTGRAFSNGALVSISG

KTGTAESYVAGGQEANNTNAVAYAPSDNPQIAVAVVFPHNTN

>9

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQPIVFQGSAPINSWYTQAYYGSFPITAV

EALEYSSNAYMVQTALGIMGQTYQPNMFVLTNNLESAMGKLRSTFAEYGLGASTGIDL

267

PDESTGFIPKEYNFANYITNAFGQFDNYTPMQLAQYVATIANNGVRVAPRIVEGIYGNN

DKGGLGNLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSGLTTGRAFSNGALVSIS

GKTGTAESYVADGQQATNTNAVAYAPSDNPQIAVAVVFPHNTN

>10

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQSIVFQGSAPINSWYTQAYGSFPITAVQ

ALEYSSNTYMVQTALGLMGQTYQPNMFVGTSNLESAMEKLRSTFGEYGLGTETGIDLP

DESTGFVPKEYSFANYITNAFGQFDNYTPMQLAQYVATIANNGVRVAPRIVEGIYGNND

KGGLGDLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSGLTTGRAFSNGALVSISG

KTGTAESYVADGQQATNTNAVAYAPSDNPQIAVAVVFPHNTN

>11

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQPIVFQGSAPIYSWYKLAYGSFPITAVE

ALEYSSNAYMVQTALGIMGQTYQPNMFVGTSNLETAMGKLRATFGEYGLGAATGIDLP

DESTGFVPKEYSFANYITNAFGQFDNYTPMQLAQYVATIANDGVRVAPRIVEGIYGNND

KGGLGELIQAIDTKEINKVNISESDMAILHQGFYQVSHGTSPLTTGRAFSDGATVSISGKT

GTGESYVAGGQEANNTNAVAYAPTENPQIAVAVVFPHNTN

>12

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQPIVFQGSAPIYSWYKLAYGSFPITAVE

ALEYSSNAYMVQTALGIMGQTYQPNMFVGTSNLETAMGKLRATFGEYGLGAATGIDLP

DESTGFVPKEYSFANYITNSFGQFDNYTPMQLAQYVATIANNGVRVAPRIVEGIYGNND

KGGLGDLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSGLTTGRAFSNGALVSISG

KTGTAESYVADGQQATNTNAVAYAPSDNPQIAVAVVFPHNTN

>14

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQPIVFQGSAPIYSWYKLAYGSFPITAVE

ALEYSSNAYVVQTALGIMGQTYQPNMFVGTSNLESAMGKLRSTFGEYGLGSATGIDLP

DESTGLVPKEYNFANFITNAFGQFDNYTPMQLAQYVATIANNGVRLAPHIVEGIYDNND

KGGLGELIQAIDTKEINKVNISESDMAILHQGFYQVSHGTSPLTTGRAFSDGATVSISGKT

GTGESYVAGGQEANNTNAVAYAPTENPQIAVAVVFPHNTN

>15

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQSIVFQGSAPINSWYTAFSRPMPITAVQ

ALEYSSNAYMVQTALGLMGQTYQPNMFVGTSNLESAMGKLRSTFGEYGLGSATGIDLP

DESTGFIPKEYSFANYITNAFGQFDNYTPMQLAQYVATIANDGVRVAPRIVEGIYGNND

KGGLGDLIQQLQPTEMNKVNISDSDMSVLHQGFYQVAHGTSGLTTGRAFSNGALVSISG

KTGTAESYVADGQQATNTNAVAYAPSDNPQIAVAVVFPHNTN

>26

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDRSIVFQGSAPINSWYTQAYGSFPITAVQ

ALEYSSNTYMVQTALGLMGQTYQPNMFVGTSNLESAMEKLRSTFGEYGLGTATGIDLP

DESTGFVPKEYSFANYITNAFGQFDNYTPMQLAQYVATIANNGVRVAPRIVEGIYGNND

KGGLGDLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSGLTTGRAFSNGALVSISG

KTGTAESYVADGQQATNTNAVAYAPSDNPQIAVAVVFPHNTN

>27

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQSIVFQGSAPINSWYTQAYGSFPITAVQ

ALEYSSNAYMVQTALGLMGQTYQPNMFVGTSNLESAMGKLRSTFGEYGLGSATGIDLP

DESTGFVPKDYSFANYITNAFGQFDNYTPMQLAQYVATIANDGVRVAPRIVEGIYGNND

KGGLGDLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSGLTTGRAFSNGAAVSISG

KTGTAESYVEGGQEANNTNAVAYAPSDNPQIAVAVVFPHNTN

>29

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQSIVFQGSAPINSWYTQAYGSFPITAVQ

ALEYSSNAYMVQTALGLMGQTYQPNMFVGTSKLESAMGKLRSTFGEYGLGSATGIDLP

DESTGFVPKDYSFANYITNAFGQFDNYTPMQLAQYVATIANDGVRVAPRIVEGIYGNND

KGGLGDLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSGLTTGRAFSNGALVSISG

KTGTAESYVADGQQATNTNAVAYAPSDNPQIAVAVVFPHNTN

>31

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQPIVFQGSAPIYSWYKLAYGSFPITAVE

ALEYSSNAYMVQTALGIMGQTYQPNMFVGTSNLETAMGKLRATFGEYGLGAATGIDLP

DESTGFVPKEYSFANYITNAFGQFDNYTPMQLAQYVATIANNGVRVAPRIVEGIYGNND

KGGLGDLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSGLTTGRAFSNGALVSISG

KTGTAESYVADGQQATNTNAVAYAPSDNPQIAVAVVFPHNTN

>32

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQPIVFQGSAPIYSWYKLAYGSFPITAVE

ALEYSSNAYMVQTALGIMGQTYQPNMFVGTSNLETAMGKLRATFGEYGLGAATGIDLP

DESTGFVPKEYSFANYITNSFGQFDNYTPMQLAQYVATIANDGVRVAPRIVEGIYGNND

KGGLGDLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSGLTTGRAFSNGAAVSISG

KTGTAESYVAGGQEANNTNAVAYAPSDNPQIAVAVVFPHNTN

>34

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQSIVFQGSAPINSWYTQAYGSFPITAVQ

ALEYSSNAYMVQTALGLMGQTYQPNMFVGTSNLESAMGKLRSTFGEYGLGAATGIDLP

DESTGFVPKDYNFANFITNAFGQFDNYTPMQLAQYVATIANDGVRVAPRIVEGIYGNND

KGGLGDLIQQLQPTEMNKVNISDSDMSVLHQGFYQVAHGTSGLTTGRAFSNGALVSISG

KTGTAESYVADGQEATNTNAVAYAPSDNPQIAVAVVFPHNTN

>39

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQPIVFQGSAPIYSWYKLAYGSFPITAVE

ALEYSSNAYMVQTALGIMGQTYQPNMFVGTSNLETAMGKLRATFGEYGLGAATGIDLP

DESTGFVPKEYSFANYITNAFGQFDNYTPMQLAQYVATIANDGVRVAPRIVEGIYGNND

KGGLGDLIQQLQPTEMNKVNISDSDMSVLHQGFYQVAHGTSGLTTGRAFSNGALVSISG

KTGTAESYVADGQQATNTNAVAYAPSDNPQIAVAVVFPHNTN

>44

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQSIVFQGSAPINSWYTQAYGSFPITAVQ

ALEYSSNTYMVQTALGLMGQTYQPNMFVGTSNLESAMEKLRSTFGEYGLGTATGIDLP

DESTGFVPKEYSFANYITNAFGQFDNYTPMQLAQYVATIANNGVRVAPRIVEGIYGNND

KGGLGNLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSGLTTGRAFSNGALVSISG

KTGTAESYVADGQQATNTNAVAYAPSDNPQIAVAVVFPHNTN

>53

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQSIVFQGSAPINSWYTAFSVPMPITAVQ

ALEYSSNAYMVQTALGLMGQTYQPNMFVGTSNLESAMGKLRSTFGEYGLGSATGIDLP

DESTGFIPKEYSFANYITNAFGQFDNYTPMQLAQYVATIANDGVRVAPRIVEGIYGNND

KGGLGDLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSGLTTGRAFSNGALVSISG

KTGTAESYVAGGQEANNTNAVAYAPSDNPQIAVAVVFPHNTN

>74 (this study)

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQSIVFQGSAPINSWYTQAYGSFPITAVQ

ALEYSSNTYMVQTALGLMGQTYQPNMFVGTSNLESAMEKLRSTFGEYGLGTATGIDLP

DESTGFVPKEYSFANYITNAFGQFDNYTPMQLAQYVATIANNGVRVAPRIVEGIYGNND

KGGLGDLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSGLTTGRAFSNGAWVSISG

KTGTAESYVADGQQATNTNAVAYAPSDNPQIAVAVVFPHNTN

>75 (this study)

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQPIVFQGSAPIYSWYKLAYGSFPITAVE

ALEYSSNAYMVQTALGIMGQTYQPNMFVGTSNLETAMGKLRATFGEYGLGAATGIDLP

DESTGFVPKEYSFANYITNAFGQFDNYTPMQLAQYVATIANDGVRVAPRIVEGIYGNND

KGGLGDLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSGLTTGRAFSNGAAVSISG

KTGTAESYVAGGQEANNTNAVAYAPSDNPQIAVAVVFPHNTN

>76 (this study)

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQPIVFQGSAPIYSWYKLAYGSFPITAVE

ALEYSSNAYMVQTALGIMGQTYQPNMFVGTSNLETAMGKLRATFGEYGLGAATGIDLP

DESTGFVPKEYSFANYITNAFGQFDNYTPMQLAQYVATIANDGVRVAPRIVEGIYGNND

KGGLGDLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSGLTTGRAFSNGALVSISG

NTGTAESYVAGGQEANNTNAVAYAPSDNPQIAVAVVFPHNTN

>77 (this study)

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQSIVFQGSAPINSWYTQAYGSFPITAVQ

ALEYSSNTYMVQTALGLMGQTYQPNMFVDTSNLESAMEKLRSTFGEYGLGTATGIDLP

DESTGFVPKEYSFANYITNAFGQFDNYTPMQLAQYVATIANNGVRVAPRIVEGIYGNND

KGGLGDLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSELTTGRAFSNGALVSISG

KTGTAESYVADGQQATNTNAVAYAPSDNPQIAVAVVFPHNTN

>78 (this study)

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQPIVFQGSAPIYSWYKLAYGSFPITAVE

ALEYSSNAYMVQTALGIMGQTYQPNMFVGTSNLETAMGKLRATFGEYGLGAATGIDLP

DESTGFVPKEYSFANYITNAFGQFDNYTPMQLAQYVATIANDGVRVAPRIVEGIYENND

KGGLGDLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSGLTTGRAFSNGAAVSISG

KTGTAESYVAGGQEANNTNAVAYAPSDNPQIAVAVVFPHNTN

>79 (this study)

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQSIVFQGSAPINSWYTQAYGSFPITAVQ

ALEYSSNTYMVQTALGLMGQTYQPNMFVGTSNLETAMGKLRATFGEYGLGAATGIDL

PDESTGFVPKEYSFANYITNAFGQFDNYTPMQLAQYVATIANDGVRVAPRIVEGIYGNN

DKGGLGDLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSGLTTGRAFSNGALVSIS

GKTGTAESYVAGGQEANNTNAVAYAPSDNPQIAVAVVFPHNTN

>80 (this study)

TNVFIPGSVVKAATISSGWENGVLSGNQTLTDQPIVFQGSAPIYSWYKLAYGSFPITAVE

ALEYSSNAYMVQTALGIMGQTYQPNMFVGTSNLETAMGKLRATFGEYGLGVATGIDLP

DESTGFVPKEYSFANYITNAFGQFDNYTPMQLAQYVATIANDGVRVAPRIVEGIYGNND

KGGLGELIQAIDTKEINKVNISESDMAILHQGFYQVSHGTSPLTTGRAFSDGATVSISGKT

GTGESYVAGGQEANNTNAVAYAPTENPQIAVAVVFPHNTN

>81 (this study)

TNVFVPGSVVKAATISSGWENGVLSGNQTLTDQPIVFQGSAPIYSWYKLAYGSFPITAVE

ALEYSSNAYMVQTALGIMGQTYQPNMFVGTSNLETAMGKLRATFGEYGLGAATGIDLP

DESTGFVPKDYSFANYITNAFGQFDNYTPMQLAQYVATIANDGVRVAPRIVEGIYGNND

KGGLGDLIQQLQPTEMNKVNISDSDMSILHQGFYQVAHGTSGLTTGRAFSNGAAVSISG

KTGTAESYVAGGQEANNTNAVAYAPSDNPQIAVAVVFPHNTN

PBP2X

>0

GTDGIITYEKDRLGNIVPGTEQVSQRTMDGKDVYTTISSPLQSFMETQMDAFQEKVKGK

YMTATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTMKVMMLA

AAIDNNTFPGGEVFNSSELKIADATIRDWDVNEGLTGGRMMTFSQGFAHSSNVGMTLLE

QKMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMIR

AFTAIANDGVMLEPKFISAIYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMVLVGTDP

VYGTMYNHSTGKPTVTVPGQNVALKSGTAQIADEKNGGYLVGVTDYIFSAVSMSPAEN

PDFILYV

>1

GKDGIITYEKDRLGNIVPGTEQVSQQTVDGKDVYTTISSTLQSFMETQMNAFQEKVKGK

YMTATLVSAKTGEILATTQRPTFDADTKEGLTKDFVWRDILYQSNYEPGSTMKVMTLA

AAIDNNTFPGGEVFNSSELKVADATIRDWDVNEGLTGGGMMTFSQGFAHSSNVGMTLL

EQKMGDATWLDYLNRFKFGIPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMI

RAFTAIANDGVMLEPKFITALYDPNNQSVRKSQKEIVGNPVSKDAASQTRTHMVLVGT

DPRYGTMYNHSTGKATVNVPGQNVALKSGTAEIADEKNGGYLVGSTNNIFSVVAMNP

AENPDFILYV

>2

GTDGIITYEKDRLGNIVPGTEQVSQRTMDGKDVYTTISSPLQSFMETQMDAFQEKVKGK

YMTATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTMKVMMLA

AAIDNNTFPGGEVFNSSELKIADATIRDWDVNEGLTGGRMMTFSQGFAHSSNVGMTLLE

QKMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMIR

AFTAIANDGVMLEPKFISAIYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMVLVGTDP

VYGTMYNHSTGKPTVTVPGQNVALKSGTAQIADEKNGGYLVGLTDYIFSAVSMSPAEN

PDFILYV

>3

GTDGIITYEKDRLGNIVPGTEQVSQRTMDGKDVYTTISSPLQSFMETQMDAFQEKVKGK

YMTATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTMKVMMLA

AAIDNNTFPGGEVFNSSELKIADATIRDWDVNEGLTGGRTMTFSQGFAHSSNVGMTLLE

QKMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMIR

AFTAIANDGVMLEPKFISAIYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMVLVGTDP

VYGTMYNHSTGKPTVTVPGQNVALKSGTAQIADEKNGGYLVGLTDYIFSAVSMSPAEN

PDFILYV

>5

GTDGIITYEKDRLGNIVPGTEQVSQRTMDGKDVYTTISSPLQSFMETQMDSFQEKVKGK

YMTATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTMKVMMLA

AAIDNNTFPGGEVFNSSELKIADATIRDWDVNEGLTGGRMMTFSQGFAHSSNVGMTLLE

QKMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMIR

AFTAIANDGVMLEPKFISAVYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMVLVGTDP

VYGTMYNHSTGKPTVTVPGQNVALKSGTAQIADEKNGGYLVGLTDYIFSAVSMSPAEN

PDFILYV

>6

GTDGIITYEKDRLGNIVPGTEQVSQRTMDGKDVYTTISSPLQSFMETQMDAFQEKVKGK

YMTATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTMKVMMLA

AAIDNNTFPGGEVFNSSELKIADATIRDWDVNEGLTGGRTMTFSQGFAHSSNVGMTLLE

QKMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMIR

AFTAIANDGVMLEPKFISAIYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMVLVGTDP

VYGTMYNHSTGKPTVTVPGQNVALKSGTAQIADEKNGGYLVGVTDYIFSAVSMSPAEN

PDFILYV

>7

GTDGIITYEKDRVGNIVPGTELVSQQTVDGKDVYTTLSSPLQSFMETQMDAFLEKVKGK
YMTATLVSAKTGEILATTQRPTFNADTKEGITEDFVWRDILYQSNYEPGSAMKVMTLAS
SIDNNTFPSGEYFNSSEFKIADATTRDWDVNEGLTTGGMMTFLQGFAHSSNVGMSLLEQ
KMGDATWLDYLKRFKFGVPTRFGLTDEYAGQLPADNIVSIAQSSFGQGISVTQTQMLRA
FTAIANDGVMLEPKFISAIYDTNNQSVRKSQKEIVGNPVSKEAASTTRNHMILVGTDPLY
GTMYNHYTGKPIITVPGQNVAVKSGTAQIADEKNGGYLVGSTNYIFSVVTMNPAENPDF
ILYV

>8

GKDGIITYEKDRLGNIVPGTEQVSQQTVDGKDVYTTLSSPLQSFMETQMDAFLEKVKGK
YMTATLVSAKTGEILATTQRPTFNADTKEGITEDFVWRDILYQSNYEPGSAMKVMTLAS
SIDNNTFPSGEYFNSSEFKIADATTRDWDVNEGLTTGGMMTFLQGFAHSSNVGMSLLEQ
KMGDATWLDYLKRFKFGVPTRFGLTDEYAGQLPADNIVSIAQSSFGQGISVTQTQMLRA
FTAIANDGVMLEPKFISAIYDTNNQSVRKSQKEIVGNPVSKEAASTTRNHMILVGTDPLY
GTMYNHYTGKPIITVPGQNVAVKSGTAQIADEKNGGYLVGSTNYIFSAVTMNPAENPDF
ILYV

>10

GTDGIITYEKDRLGNIVPGTEQVSQRTMDGKDVYTTISSPLQSFMETQMDAFQEKVKGK
YMTATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTMKVMMLA
AAIDNNTFPGGEVFNSSELKIADATIRDWDVNEGLTGGRMMTFSQGFAHSSNVGMTLLE
QKMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMIR
AFTAIANDGVMLEPKFLSAIYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMVLVGTDP
VYGTMYNHSTGKPTVTVPGQNVALKSGTAQIADEKNGGYLVGVTDYIFSAVSMSPAEN

PDFILYV

>11

GTDGIITYEKDRLGNIVPGTEQISQQTVDGKDVYTTISSPLQSFMETQMDAFQEKVKGKY
MTATLVSAKTGEILATTQRPTFNADTKEGITKDFVWRDILYQSNYEPGSAMKVMTLAAS
IDNNTFPGGEYFNSSELKIADVTIRDWDVNEGLTGGGMMTFSQGFAHSSNVGMTLLEQ
KMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMIRA
FTAIANDGVMLEPKFISAIYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMVLVGTDPV
YGTMYNHSTGKPTVTVPGQNVALKSGTAQIADEKNGGYLVGLTNYIFSAVSMSPAENP
DFILYV

GTDGIITYEKDRLGNIVPGTEQVSQQTVDGKDVYTTISSPLQSFMETQMDAFQEKVKGK
YMTATLVSAKTGEILATTQRPTFNADTKEGITKDFVWRDILYQSNYEPGSTMKVMTLAS
SIDNNTFPSGEYFNSSELKIADVTIRDWDVNDGLTTGRMMTFLQGFALSSNVGMTLLEQ
KMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMIRA
FTAIANDGVMLEPKFISAIYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMVLVGTDPV
YGTMYNHSTGKPIITVPGQNVAVKSGTAQIADEKNGGYLVGSTNYIFSVVTMNPAENPD
FILYV

>16

GTDGIITYEKDRVGNIVPGTELVSQQTVDGKDVYTTLSSPLQSFMETQMDAFLEKVKGK
YMTATLVSAKTGEILATTQRPTFNADTKEGITEDFVWRDILYQSNYEPGSAMKVMTLAS
SIDNNTFPSGEYFNSSEFKIADATTRDWDVNEGLTTGGMMTFLQGFAHSSNVGMSLLEQ
KMGDATWLDYLKRFKFGVPTRFGLTDEYAGQLPADNIVSIAQSSFGQGISVTQTQMLRA
FTAIANDGVMLEPKFISAIYDTNNQSVRKSQKEIVGKPVSEDTASLTRTNMILVGTDPLY
GTMYNHYTGKPIITVPGQNVAVKSGTAQIADEKNGGYLVGSTNYIFSAVTMNPAENPDF

ILYV

>18

GTDGIITYEKDRVGNIVPGTELVSQQTVDGKDVYTTLSSPLQSFMETQMDAFLEKVKGK
YMTATLVSAKTGEILATTQRPTFNADTKEGITEDFVWRDILYQSNYEPGSAMKVMTLAS
SIDNNTFPSGEYFNSSEFKIADATTRDWDVNDGLTTGGMMTFLQGFAHSSNVGMSLLEQ
KMGDATWLDYLKRFKFGVPTRFGLTDEYAGQLPADNIVSIAQSSFGQGISVTQTQMLRA
FTAIANDGVMLEPKFISAIYDTNNQSVRKSQKEIVGNPVSKEAASTTRNHMILVGTDPLY
GTMYNHYTGKPIITVPGQNVAVKSGTAQIADEKNGGYLVGSTNYIFSVVTMNPAENPDF
ILYV

>22

GTDGIITYEKDRLGNIVPGTEQVSQRTMDGKDVYTTISSPLQSFMETQMDAFQEKVKGK
YMTATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTMKVMMLA
AAIDNNTFPSGEYFNSSELKIADATIRDWDVNDGLTTGGMMTFLQGFAHSSNVGMSLLE
QKMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAMSAFGQGISVTQTQML
RAFTAIANDGVMLEPKFISALYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMILVGTD
PVYGTMYNHSTGKPTVTVPGQNVALKSGTAEIADEKNGGYLVGSTNNIFSVVAMNPAE
NPDFILYV

>23

GTDGIITYEKDRLGNIVPGTEQVSQRTMDGKDVYTTISSPLQSFMETQMDAFQEKVKGK
YMTATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTMKVMMLA
AAIDNNTFPGGEVFNSSELKIADATIRDWDVNEGLTGGRMMTFSQGFAHSSNVGMTLLE
QKMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMIR
AFTAIANDGVMLEPKFISAIYDPNNQTARKSQKEIVGNPVSKDAASLTRTNMVLVGTDP

VYGTMYNHSTGKPTVTVPGQNVALKSGTAQIADEKNGGYLVGVTDYIFSAVSMSPAEN

PDFILYV

>26

GTDGIITYEKDRVGNIVPGTELVSQQTVDGKDVYTTLSSPLQSFMETQMDAFLEKVKGK

YMTATLVSAKTGEILATTQRPTFNADTKEGITEDFVWRDILYQSNYEPGSAFKVMMLAS

SIDNNTFPSGEYFNSSEFKIADATTRDWDVNAGLTTGGMMTFLQGFAHSSNVGTSLLEQ

KMGDATWLDYLKRFKFGVPTRFGLTDEYAGQLPADNIVSIAQSSFGQGISVTQTQMLRA

FTAIANDGVMLEPKFISAIYDTNNQSVRKSQKEIVGNPVSKEAASTTRNHMILVGTDPLY

GTMYNHYTGKPIITVPGQNVAVKSGTAQIADEKNGGYLVGSTNYIFSVVTMNPAENPDF

ILYV

>28

GTDGIITYEKDRLGNIVPGTEQVSQQTVDGKDVYTTISSPLQSFMETQMDAFQEKVKGK

YMTATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTMKVMMLA

AAIDNNTFPGGEVFNSSELKIADATIRDWDVNEGLTGGRMMTFSQGFAHSSNVGMTLLE

QKMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAMSAFGQGISVTQTQML

RAFTAIANDGVMLEPKFISALYDPNDQSVRKSQKEIVGNPVSKEAASVTRDHMVMVGT

DPTYGTMYNHSTGKATVNVPGQNVALKSGTAEIADEKNGGYLTGSTNNIFSVVSMHPA

ENPDFILYV

>32

GTDGIITYEKDRLGNIVPGTEQVSQQTVDGKDVYTTISSPLQSFMETQMDAFQEKVKGK

YMTATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTMKVMMLA

AAIDNNTFPGGEVFNSSELKIADATIRDWDVNEGLTGGRMMTFSQGFAHSSNVGMTLLE

QKMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMIR

AFTAIVNDGVMLEPKFISAIYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMILVGTDPV

YGTMYNHSTGKPTVTVPGQNVALKSGTAEIADEKNGGYLVGSTNNIFSVVAMNPAENP

DFILYV

>33

GKDGIITYEKDRLGNIVPGTEQVSQQTVDGKDVYTTLSSPLQSFMETQMDAFLEKVKGK

YMTATLVSAKTGEILATTQRPTFNADTKEGITEDFVWRDILYQSNYEPGSAFKVMMLAS

SIDNNTFPSGEYFNSSEFKIADATTRDWDVNAGLTTGGMMTFLQGFAHSSNVGTSLLEQ

KMGDATWLDYLKRFKFGVPTRFGLTDEYAGQLPADNIVSIAQSSFGQGISVTQTQMLRA

FTAIANDGVMLEPKFISAIYDTNNQSVRKSQKEIVGNPVSKEAASTTRNHMILVGTDPLY

GTMYNHYTGKPIITVPGQNVAVKSGTAQIADEKNGGYLVGSTNYIFSVVTMNPAENPDF

ILYV

>35

GTDGIITYEKDRLGNIVPGTEQVSQQTVDGKDVYTTISSPLQSFMETQMDAFQEKVKGK

YMTATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTMKVMMLA

AAIDNNTFPSGEYFNSSELKIADVTIRDWDVNDGLTTGRMMTFLQGFALSSNVGMSLLE

QKMGDTTWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMIR

AFTAIANDGVMLEPKFISAIYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMILVGTDPV

YGTMYNHSTGKPTVTVPGQNVALKSGTAQIADEKNGGYLVGLTNYIFSAVSMNPAENP

DFILYV

>36

GTDGIITYEKDRVGNIVPGTELVSQQTVDGKDVYTTLSSPLQSFMETQMDAFLEKVKGK

YMTATLVSAKTGEILATTQRPTFNADTKEGITEDFVWRDILYQSNYEPGSAMKVMTLAS

SIDNNTFPSGEYFNSSEFKIADATTRDWDVNEGLTTGGMMTFLQGFAHSSNVGMSLLEQ

KMGDATWLDYLKRFKFGVPTRFGLTDEYAGQLPADNIVSIAQSSFGQGISVTQTQMLRA

FTAIANDGVMLEPKFISAIYDTNNQSVRKSQKEIVGNPVSKEAASTTRNHMILVGTDPLY

GTMYNHYTGKPIITVPGQNVAVKSGTAQIADEKNGGYLVGSTNYIFSAVTMNPAENPDF

ILYV

>43

GTDGIITYEKDRLGNIVPGTEQVSQQTVDGKDVYTTISSTLQSFMETQMDAFLEKVKGK

YMTATLVSAKTGEILATTQRPTFNADTKEGITEDFVWRDILYQSNYEPGSAMKVMTLAS

SIDNNTFPSGEYFNSSEFKIADATTRDWDVNEGLTTGGMMTFLQGFAHSSNVGMSLLEQ

KMGDATWLDYLKRFKFGVPTRFGLTDEYAGQLPADNIVSIAQSSFGQGISVTQTQMLRA

FTAIANDGVMLEPKFISAIYDTNNQSVRKSQKEIVGNPVSKEAASTTRNHMILVGTDPLY

GTMYNHYTGKPIITVPGQNVAVKSGTAQIADEKNGGYLVGSTNYIFSAVTMNPAENPDF

ILYV

>44

GTDGIITYEKDRLGNIVPGTEQVSQQTVDGKDVYTTISSTLQSFMETQMNAFQEKVKGK

YMTATLVSAKTGEILATTQRPTFDADTKEGLTKDFVWRDILYQSNYEPGSTMKVMTLA

AAIDNNTFPGGEVFNSSELKIADATIRDWDVNDGLTTGGMMTFLQGFAHSSNVGMSLL

EQKMGDATWLDYLNRFKFGVPTRFGLTDEYSGQLPADNIVNIAMSAFGQGISVTQTQMI

RAFTAIANDGVMLEPKFISAIYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMILVGTDP

VYGTMYNHSTGKPTVTVPGQNVALKSGTAQIADEKNGGYLVGLTNYIFSAVSMNPAEN

PDFILYV

>64

GTDGIITYEKDRLGNIVPGTEQVSQRTMDGKDVYTTISSPLQSFMETQMDAFQEKVKGK

YMTATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTMKVMTLAA

AIDNNTFPGGEVFNSSELKIADATIRDWDVNEGLTGGRMMTFSQGFAHSSNVGMTLLEQ

KMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMIRA

FTAIANDGVMLEPKFISAIYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMVLVGTDPV

YGTMYNHSTGKPTVTVPGQNVALKSGTAQIADEKNGGYLVGLTDYIFSAVSMSPAENP

DFILYV

>77

GTDGIITYEKDRLGNIVPGTEQVFQQTVDGKDVYTTISSTLQSFMETQMNAFQEKVKGK

YMTATLVSAKTGEILATTQRPTFDADTKEGLTKDFVWRDILYQSNYEPGSAMKVMTLA

AAIDNNTFPGGEVFNSSELKVADVTIRDWDVNEGLTGGGMMTFSQGFAHSSNVGMTLL

EQKMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMI

RAFTAIANDGVMLEPKFISAIYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMVLVGTD

PVYGTMYNHSTGKPTVTVPGQNVALKSGTAQIADEKNGGYLVGLTNYIFSAVSMSPAE

NPDFILYV

>83

GTDGIITYEKDRLGNIVPGTEQVSQRTMDGKDVYTTISSPLQSFMEPQMDAFQEKVKGK

YMTATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTMKVMMLA

AAIDNNTFPGGEVFNSSELKIADATIRDWDVNEGLTGGRMMTFSQGFAHSSNVGMTLLE

QKMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMIR

AFTAIANDGVMLEPKFISAIYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMVLVGTDP

VYGTMYNHSTGKPTVTVPGQNVALKSGTAQIADEKNGGYLVGLTDYIFSAVSMSPAEN

PDFILYV

>111

GTDGIITYEKDRLGNIVPGTEQASQHTVDGKDVYTTLSSPLQSFMETQMDAFQEKVKGK

YMTATLVSAKTGEILATTQRPTFNADTKDGITKDFVWRDILYQSNYEPGSAMKVMTLA
SAIDNNTFPGGEYFNSSELKIADATIRDWDVNDGLTTGGMMTFSQGFAHSSNVGMSLLE
QKMGDATWLDYLNRFKFGVPTRFGLTDEYTGQLPADNIVNIAMSAFGQGISVTQTQML
RAFTAIANDGVMLEPKFISALYDPNDQSVRKSQKEIVGNPVSKAAASSTREHMVMVGT
DPVYGTMYNHSTGKPNVNVPGQNVALKSGTAQIADEKNGGYLTGETNYIFSVVSMHPA
ENPDFILYV

>119 (this study)

GTDGIITYEKDRLGNIVPGTEQVSQQTVDGKDVYTTISSTLQSFMETQMDAFLEKVKGK
YMTATLVSAKTGEILATTQRPTFNADTKEGITEDFVWRDILYQSNYEPGSAMKVMTLAS
SIDNNTFPSGEYFNSSEFKIADATTRDWDVNEGLTTGGMMTFLQGFAHSSNVGMSLLEQ
KMGDATWLDYLKRFKFGVPTRFGLTDEYAGQLPADNIVSIAQSSFGQGISVTQTQMLRA
FTAIANDGVMLEPKFISAIYDTNNQSVRKSQKEIVGNPVSKEAASTTRNHMILVGTDPLY
GTMYNHYTGKPIITVPGQNVAVKSGTAQIADEKNGGYLVGSTNYIFSVVTMNPAENPDF
ILYV

>120 (this study)

GTDGIITYEKDRVGNIVPGTELVSQQTVDGKDVYTTLSSPLQSFMETQMDAFLEKVKGK
YMTATLVSAKTGEILATTQRPTFNADTKEGITEDFVWRDILYQSNYEPGSAMKVMTLAS
SIDNNTFPSGEYFNSSEFKIADATTRDWDVNEGLTTGGMMTFLQGFAHSSNVGMSLLEQ
KMGDATWLDYLKRFKFGVPTRFGLTDEYAGQLPADNIVSIAQSSFGQGISVTQTQMLRA
FTAIANDGVMLEPKFISAIYDTNNQSVRKSQKEIVGKPVSEDAASLTRTNMILVGTDPLY
GTMYNHYTGKPIITVPGQNVAVKSGTAQIADEKNGGYLVGSTNYIFSAVTMNPAENPDF
ILYV

>121 (this study)

GTDGIITYEKDRLGNIVPGTEQVSQRTMDGKDVYTTISSPLQSFMETQMDAFQEKVKGK

YMTATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTMKVMMLA

AAIDNNTFPGGEVFNSSELKIADATIRDWDVNEGLTGGRMMTFSQGFAHSSNVGMTLLE

QKMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMIR

AFTAIANDGVMLEPKFISAIYDPNNQTARKSQKEIVGNPVSKDAASLTRTNMVLVGTDP

VYGTMYNHSTGKPTVTVPGQNVAIKSGTAQIADEKNGGYLVGVTDYIFSAVSMSPAEN

PDFILYV

>122 (this study)

GTDGIITYEKDRLGNIVPGTEQVSQRTMDGKDVYTTISSPLQSFMETQMDAFQEKVKGK

YMTATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTMKVMMLA

AAIDNNTFPGGEVFNSSELKIADATIRDWDVNEGLTGGRMMTFSQGFAHSSNVGMTLLE

QKMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNLAQSSFGQGISVTQTQMI

RAFTAIANDGVMLEPKFISAIYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMVLVGTD

PVYGTMYNHSTGKPTVTVPGQNVALKSGTAQIADEKNGGYLVGVTDYIFSAVSMSPAE

NPDFILYV

>123 (this study)

GTDGIITYEKDRVGNIVPGTELVSQQTVDGKDVYTTLSSPLQSFMETQMDAFLEKVKGK

YMTATLVSAKTGEILATTQRPTFNADTKEGITEDFVWRDILYQSNYEPGSAMKVITLASS

IDNNTFPSGEYFNSSEFKIADATTRDWDVNEGLTTGGMMTFLQGFAHSSNVGMSLLEQK

MGDATWLDYLKRFKFGVPTRFGLTDEYAGQLPADNIVSIAQSSFGQGISVTQTQMLRAF

TAIANDGVMLEPKFISAIYDTNNQSVRKSQKEIVGKPVSEDTASLTRTNMILVGTDPLYG

TMYNHYTGKPIITVPGQNVAVKSGTAQIADEKNGGYLVGSTNYIFSAVTMNPAENPDFI

LYV

>124 (this study)

GTDGIITYEKDRLGNIVPGTEQVSQRTMDGKDVYTTISSPLQSFMETQMDAFQEKVKGK
YMTATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTMKVMMLA
AAIDNNTFPGGEVFNSSELKIADATIRDWDVNEGLTGGRMMTFSQGFAHSSNVGMTLLE
QKMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQML
RAFTAIANDGVMLEPKFISAIYDTNNQSVRKSQKEIVGNPVSKETASTTRNHMILVGTDP
LYGTMYNHYTGKPIITVPGQNVAVKSGTAQIADEKNGGYLVGSTNYIFSVVTMNPAENP
DFILYV

>125 (this study)

GKDGIITYEKDRLGNIVPGTEQVSQQTVDGKDVYTTISSTLQSFMETQMNVFQEKVKGK
YMTATLVSAKTGEILATTQRPTFDADTKEGLTKDFVWRDILYQSNYEPGSTMKVMTLA
AAIDNNTFPSGEYFNSSELKIADVTIRDWDVNDGLTTGRMMTFLQGFALSSNVGMTLLE
QKMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMIR
AFTAIANDGVMLEPKFISAIYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMVLVGTDP
VYGTMYNHSTGKPTVTVPGQNVALKSGTAQIADEKNGGYLVGLTNYIFSAVSMSPAEN
PDFILYV

>126 (this study)

GTDGIITYEKDRLGNIVPGTEQVSQRTMDGKDVYTTISSPLQSFMETQMDAFQEKVKGK
YMTATLVSAKTGEILATTQRPTFDADTKEGLTKDFVWRDILYQSNYEPGSTMKVMTLA
AAIDNNTFPGGEVFNSSELKIADATIRDWDVNEGLTGGRMMTFSQGFAHSSNVGMTLLE
QKMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAMSAFGQGISVTQTQML
RAFTAIANDGVMLEPKFISALYDPNDQSARKSQKEVVGNPVSKEAASVTRDQMVMVGT
DPTYGTMYNHSTGKATVNVPGQNVALKSGTAEIADEKNGGYLVGSTNNIFSVVAMNPA

ENPDFILYV

>127 (this study)

GTDGIITYEKDRLGNIVPGTEQVSQRTMDGKDVYTTISSPLQSFMETQMDAFQEKVKGK

YMTATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTMKVMMLA

AAIDNNTFPGGEVFNSSELKIADATIRDWDVNEGLTGGRMMTFSQGFAHSSNVGMTLLE

QKMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMIR

AFTAIANDGVMLEPKFISAIYDTNNQSVRKSQKEIVGNPVSKEAASTTRNHMILVGTDPL

YGTMYNHYTGKPIITVPGQNVAVKSGTAQIADEKNGGYLVGSTNYIFSAVTMNPAENP

DFILYV

>128 (this study)

GTDGIITYEKDRLGNIVPGTEQVSQRTMDGKDVYTTISSPLQSFMETQMDAFQEKVKGK

YMTATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTMKVMMLA

AAIDNNTFPGGEVFNSSELKIADATIRDWDVNEGLTGGRTMTFSQGFTHSSNVGMTLLE

QKMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMIR

AFTAIANDGVMLEPKFISAIYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMVLVGTDP

VYGTMYNHSTGKPTVTVPGQNVALKSGTAQIADEKNGGYLVGVTDYIFSAVSMSPAEN

PDFILYV

>129 (PMEN21, ERR1022105)

GTDGIITYEKDRLGNIVPGTEQVSQRTMDGKDVYTTISSPLQSFMETQMDAFQEKVKGK

YMTATLVSAKTGEILATTQRPTFDADTKEGITKDFVWRDILYQSNYEPGSTMKVMMLA

AAIDNNTFPGGEVFNSSELKIADATIRDWDVNEGLTGGGMMTFSQGFAHSSNVGMTLL

EQKMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMI

RAFTAIANDGVMLEPKFISAIYDPNDQTARKSQKEVVGNPVSKDAASLTRTNMILVGTD

PVYGTMYNHSTGKPTVNVPGQNVALKSGTAEIADEKNGGYLVGSTNNIFSVVAMNPAE

DPDFILYV

>130 (ERR057775)

GTDGIITYEKDRLGNIVPGTEQVFQRTMDGKDVYTTISSPLQSFMETQMDAFQEKVKGK

YMTATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTMKVMMLA

AAIDNNTFPGGEVFNSSELKIADATIRDWDVNEGLTGGRMMTFSQGFAHSSNVGMTLLE

QKMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMIR

AFTAIANDGVMLEPKFISAIYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMVLVGTDP

VYGTMYNHSTGKPTVTVPGQNVALKSGTAQIADEKNGGYLVGLTDYIFSAVSMSPAEN

PDFILYV