University of the Witwatersrand

Faculty of Health Sciences

Molecular characterisation of nontypeable *Streptococcus pneumoniae* in South Africa

Submitted by

Thabo Mohale

(1006116)

In fulfilment of the requirements for the degree of Master of Science in Medicine

28 August 2018

Supervisor: Dr Nicole Wolter

Co-supervisors: Dr Mushal Allam and Professor Anne von Gottberg

DECLARATION

I, Thabo Mohale, declare that this dissertation is my own work. Experiments described here were conducted under the supervision of Dr Nicole Wolter, Dr Mushal Allam and Professor Anne von Gottberg at the Centre for Respiratory Diseases and Meningitis, National Institute for Communicable Diseases, a division of the National Health Laboratory Service, Johannesburg. It is being submitted for the degree of Master of Science in Medicine to the Faculty of Health Sciences at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination to this or any other university. The results of this dissertation have been published (**Chapter 3),** and **Chapter 4** is in preparation for submission (see under **publications**). All co-authors have agreed (by signing a declaration which is submitted to the University) to the use of the results as part of this dissertation. I declare that I contributed adequately towards research findings in the two manuscripts as stated below:

- Assisted in conceiving the study
- Drafted the proposal which was critically reviewed and approved by my supervisors
- Performed all the laboratory work
- Performed whole genome analysis under supervision
- Interpreted the data
- Drafted the two manuscripts of which I am the first author

_____

Thabo Mohale

_____

Date

For my family

PRESENTATIONS

Mohale T, Wolter N, Allam M, Nzenze S.A, Madhi S.A, du Plessis M and von Gottberg A. Characterization of nontypeable carriage pneumococci in South Africa in the conjugate vaccine era, 2009-2012. *Wits Faculty of Health Sciences Biennial Research Day and Postgraduate Expo,* 1 September 2016, Johannesburg, South Africa. **(Oral)**

Mohale T, Wolter N, Allam M, Nzenze S.A, Madhi S.A, du Plessis M and von Gottberg A. Characterization of nontypeable carriage pneumococci in South Africa in the conjugate vaccine era, 2009-2012. *10th International Symposium on Pneumococci and Pneumococcal Diseases*, Glasgow, Scotland, United Kingdom, 26th -30th June 2016. **(Oral)**

Mohale T, Wolter N, Allam M, Ndlangisa K, Crowther-Gibson P, du Plessis M, von Gottberg A. Genotypic characterization of nontypeable pneumococci in South Africa, 2003-2013. *National Institute for Communicable Diseases* (NICD) *Research Forum*, 30 March 2016, NICD, Sandringham, Johannesburg, South Africa. **(Oral)**

Mohale T. Nontypeable *Streptococcus pneumoniae* - How important are they? *Group for Enteric, Respiratory, and Meningeal Disease Surveillance in South Africa Principal Investigators meeting*, 14-15 October 2014, NICD, Sandringham, Johannesburg, South Africa. **(Oral)**

Mohale T, Wolter N, Allam M, du Plessis M, von Gottberg A. Molecular characterisation of nontypeable *Streptococcus pneumoniae* causing invasive disease in South Africa, 2003-2013.*UNIPATH*, 19-21 September 2014, Pretoria, South Africa. **(Poster)**.

Mohale T, Wolter N, Allam M, du Plessis M, Ismail A, von Gottberg A. Molecular basis of the nontypeability of *Streptococcus pneumoniae* isolates causing meningitis in South Africa, 2003 – 2013. *Wits Faculty of Health Sciences Biennial Research Day and Postgraduate Expo*, 17 September 2014, Johannesburg South Africa. **(Oral)**.

PUBLICATIONS

Mohale T, Wolter N, Allam M, Ndlangisa K, Crowther-Gibson P, du Plessis M, von Gottberg A. Genomic analysis of nontypeable pneumococci causing invasive pneumococcal disease in South Africa, 2003-2013. *BMC Genomics* 2016, 17:470. doi: 10.1186/s12864-016-2808-x.:470-2808.

Mohale T, Wolter N, Allam M, Nzenze S.A, Madhi S.A, du Plessis M and von Gottberg A. Characterisation of carriage nontypeable pneumococci in South Africa and comparison with invasive nontypeable isolates. (In preparation)

ABSTRACT

Streptococcus pneumoniae colonises the nasopharynx and is also a significant pathogen causing diseases such as meningitis and pneumonia. The capsular polysaccharide is an important virulence determinant and a target for current pneumococcal vaccines. However, pneumococci showing no serological evidence of capsule expression (nontypeable pneumococci [NTPn]) have been identified, mainly in carriage studies and rarely in invasive disease. NTPn are not targeted by current vaccines and are believed to play an important role in the evolution of *Streptococcus* species. Limited data exist which describe NTPn and their population structure, especially from the African continent. Our aim was to describe, characterise and compare invasive and carriage NTPn isolates from South Africa.

Invasive NTPn were detected through national, laboratory-based surveillance for invasive pneumococcal disease in South Africa between 2003 and 2013. Carriage NTPn were obtained from cross-sectional studies assessing the impact of pneumococcal conjugate vaccine (PCV) on nasopharyngeal colonisation in South Africa between 2009 and 2012. Isolates were characterised by capsular locus sequence analysis, multilocus sequence typing and whole genome phylogenetic analysis. Antimicrobial non-susceptibility patterns and potentially inactivating capsule mutations were determined. We also performed comparative genomic analysis of invasive and carriage NTPn isolates.

The prevalence of invasive and carriage NTPn was 0.1% (39/32,824) and 3.7% (137/3721) (P <0.01), respectively. Carriage NTPn increased from 3.0% (60/2015) to 4.5% (77/1706) (P = 0.02), post-PCV7. Five percent (2/39) and 24.0% (33/137) (P <0.01) of invasive and carriage NTPn were co-infected and co-colonised with encapsulated pneumococci, respectively. Non-susceptibility to cotrimoxazole [84.0% (112/133) vs. 44.0% (17/39)], penicillin [77.0%

(102/133) vs. 36.0% (14/39)], erythromycin [53.0% (70/133) vs. 31.0% (12/39)] and clindamycin [36.0% (48/133) vs. 18.0% (7/39)] were significantly higher ($P \leq 0.03$) among carriage than invasive NTPn.

Fifty-six percent (22/39) of invasive and 9.0% (13/137) of carriage NTPn isolates ($P < 0.01$) had at least partial capsule genes (Group I). The remaining invasive (17/39, 44.0%) and carriage (124/137, 91.0%) ($P < 0.01$) NTPn had complete deletion of the capsular locus (Group II). We identified a variety of potentially inactivating capsule mutations in Group I invasive and carriage NTPn. Invasive NTPn were more diverse (Simpson's diversity index [D] = 0.97 [95.0% CI: 0.947 - 0.996]) than carriage NTPn (D = 0.92 [95.0% CI: 0.899 - 0.945]). Seventy-nine percent (31/39) of invasive NTPn belonged to a lineage that was related to encapsulated pneumococci compared carriage NTPn, where 67.0% (92/137) belonged to a lineage exclusive to NTPn strains. We identified 275 genes that were significantly associated with invasive NTPn compared to carriage NTPn.

In conclusion, carriage NTPn were frequently isolated and had higher rates of antimicrobial non-susceptibility than invasive NTPn. Carriage NTPn were also less diverse, with the majority belonging to classical lineage and lacking capsule genes, whereas invasive NTPn were more diverse and the majority belonged to a lineage related to encapsulated pneumococci. Comparative genomic analysis revealed genes that may be responsible for capsule-independent virulence mechanisms of invasive NTPn.

ACKNOWLEDGEMENTS

I would like to express my greatest, deepest appreciation to my supervisor, Dr Nicole Wolter. Thank you for your guidance, support and for providing me with opportunities to share this work. I could not have asked for a better supervisor.

To my co-supervisors, Professor Anne von Gottberg and Dr Mushal Allam, thank you for your invaluable assistance and for sharing your time and knowledge with me. This project could not have been possible without your high levels of expertise. To Dr Mignon du Plessis, thank you for your useful comments.

To Professor Shabir Madhi and Dr Suzan Nzenze, thank you for providing me with carriage NTPn isolates. To Prabha Naidoo, Noluthando Duma, Happy Skosana and Dineo Mogale, thank you for assisting with the phenotypic work. Thanks to my colleagues, Karistha Ganesh, Fahima Moosa, Maimuna Carrim and Kedibone Ndlangisa for stimulating and useful discussions. A huge thank you to my friend and colleague, Malefu Moleleki, for being there for me. A special thanks to all co-authors who agreed to the use of manuscripts as part of this dissertation

Last but not least, I would like to thank my family (my mother, Fikile Malakoane, my brother, Edward Mohale and my sister, Makamohelo Malakoane) for unconditional support and encouragement during this difficult time. You were my motivation for completing this project.

This research would not have been possible without financial support from National Health Laboratory Service Research Trust development grant (PI: Dr Nicole Wolter), Medical Research Council of South Africa and Centre for Respiratory Diseases and Meningitis.

CONTENTS

FIGURES

TABLES

APPENDICES

NOMENCLATURE

<	-	Less than
CPS	-	Capsular polysaccharide
%	-	Percentage
CO_2	-	Carbon dioxide
α	-	Alpha
CDS	-	Coding sequences
bp	-	Base pairs
IPD	-	Invasive pneumococcal disease
e.g.	-	Example
°C	-	Degrees Celsius
RT-PCR	-	Real-time polymerase chain reaction
HIV	-	Human immunodeficiency virus
>	-	Greater than
β	-	Beta
PCV	-	Polysaccharide-protein conjugate vaccine
nm	-	Nanometres
cps	-	Capsular polysaccharide synthesis locus
dexB	-	Glucan 1,6-α-glucosidase gene
aliA	-	Oligopeptide ABC transporter gene
NTPn	-	Nontypeable pneumococci
EcPn	-	Encapsulated pneumococci
DNA	-	Deoxyribonucleic acid
NCC	-	Null capsule clade
MLST	-	Multilocus sequence typing

ST	-	Sequence type
mg	-	Milligram
ml	-	Millilitre
mm	-	Millimetre
µg	-	Microgram
µl	-	Microlitre
MIC	-	Minimum inhibitory concentrations
O_2	-	Oxygen
rpm	-	Revolutions per minute
®	-	Registered sign
ng	-	Nanogram
NaOH	-	Sodium hydroxide
PI	-	Principal investigator
D	-	Simpson’s diversity index

1 Introduction

1.1 Background

Streptococcus pneumoniae, or the pneumococcus, is an important human pathogen causing invasive diseases such as meningitis and bacteraemia, as well as non-invasive diseases such as pneumonia, otitis media and sinusitis (1;2). Pneumococcal disease occurs in all age groups, with children <5 years of age, the elderly and immunocompromised individuals most at risk. Pneumococcus is a human commensal that is carried in the nasopharynx, with high colonisation rates in children <5 years of age (3). The capsular polysaccharide (CPS) is the principal virulence factor (4), and based on the structure and antigenicity of the CPS, at least 98 serotypes have been identified (5-10). The CPS is also the target of all currently available pneumococcal vaccines (11). Despite the importance of the CPS as a virulence factor, pathogenic *S. pneumoniae* isolates lacking a capsule have been described (12;13).

1.2 History of the pneumococcus

The bacterium was simultaneously and independently discovered in 1881 by Louis Pasteur in France and George Sternberg in the United States (3;14). Pasteur named it *Microbe septicemique du salive* and Sternberg called it *Micrococcus pasteuri*. Due to its recognition as the most common cause of lobar pneumonia, the bacterium was called pneumococcus in the late 1880s. In 1926, the bacterium was assigned the name *Diplococcus pneumoniae* because of its appearance in Gram-stained sputum, and in 1974, the name was again changed to *Streptococcus pneumoniae* because of its morphology during growth in liquid medium.

1.3 Microbiology of the pneumococcus

S. pneumoniae is a Gram-positive coccus and bacterial cells are usually seen as pairs of cocci. The bacterium is a non-motile, facultative anaerobe that does not form spores. It is catalase negative and ferments glucose to lactic acid. Pneumococcus is a fastidious bacterium, growing in 5.0% CO_2, and is usually cultured in a medium that contains blood (as a source of catalase). The bacterium is α-haemolytic due to the production of pneumolysin, which breaks down haemoglobin and results in a green zone around colonies when grown on blood agar. As the colonies age, they collapse due to the production of autolysin, which disrupt and disintegrate cells, giving the characteristic draughtsman appearance (**Figure 1**). Pneumococcus is inhibited by ethyl hydrocupreine (optochin) and undergoes lysis in the presence of bile salts. Almost all clinical isolates of *S. pneumoniae* contain CPS that covers the peptidoglycan cell wall. The assignment of capsular serotype is traditionally done by using the Quellung reaction (15), a gold standard, phenotypic method of serotyping pneumococcal isolates using anti-sera against serotype-specific CPS (**Figure 2**).

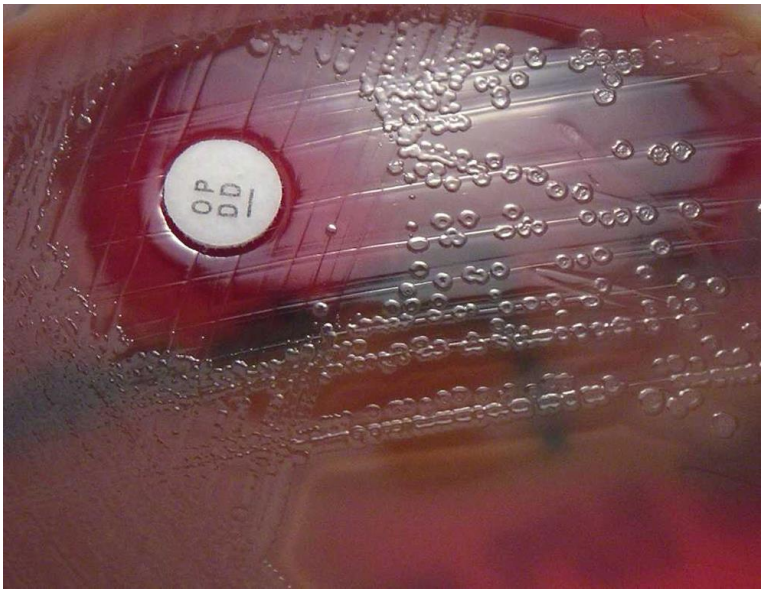


Figure 1. Growth of *Streptococcus pneumoniae* on blood agar showing draughtsman colonies and inhibition by optochin after overnight incubation

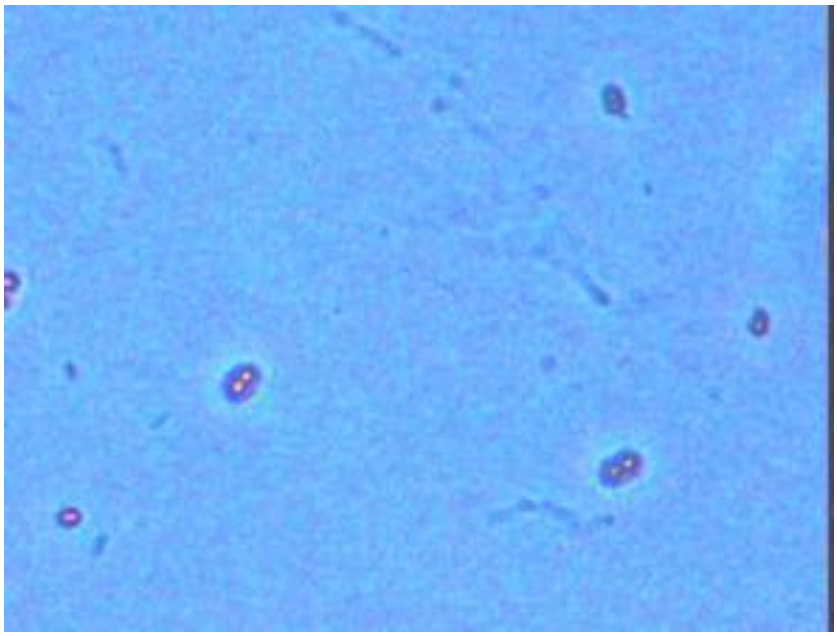


Figure 2. A phase-contrast microscopy of pneumococcal cells indicating swelling of the capsule in the presence of paired anti-sera

1.4 Genome of the pneumococcus

The pneumococcus contains a single circular genome of approximately 2 000 000 bp in size (16;17). It contains approximately 2000 predicted coding sequences (CDS) and about 5.0% of the genome is composed of insertion sequences. The guanine-cytosine content (39.0% - 40.0%) is lower compared to other bacterial pathogens. The genome of pneumococcus is highly variable, with only about 74.0% of the genome conserved (18). On average, the core genome (CDS conserved in all isolates) of pneumococcus consists of 1647 predicted CDS and contains genes essential for survival. The accessory genome (CDS that are not conserved in all isolates) plays an important role in bacterial pathogen evolution (19). The pneumococcal pan-genome is open,

meaning it is always increasing as new isolates are being continuously sequenced and adds novel genes to the current pool. The pan-genome, which includes core and accessory genomes, is defined as the full complement of CDS in a species. The genomic variations seen in the pneumococcus are mainly introduced by transformation, conjugation and transduction.

Pneumococci are competent and naturally transformable, meaning they are able to take up exogenous DNA from other pneumococci or bacterial species, and incorporate it into their genome by homologous recombination (20;21). The magnitude of transformation frequencies varies significantly between pneumococcal strains (22). Unencapsulated pneumococci exhibit a higher rate of genetic recombination compared to encapsulated pneumococci (23), this is because the capsule has been hypothesised to hinder transformation. Transformation plays an important role in the development of antibiotic resistance and can also result in serotype/capsular switching (24), whereby genes encoding one serotype are replaced by genes encoding another serotype. Serotype or capsule switching has significant clinical impact, because of the potential for vaccine-serotype strains to escape the vaccine by switching their capsules.

1.5 Pathogenesis and pneumococcal disease

S. pneumoniae spreads from person-to-person via respiratory droplets. Following acquisition, the pneumococcus colonises the nasopharynx by adhering to epithelial cells through a variety of mechanisms involving bacterial surface proteins and host epithelial cell receptors. Colonisation is predominantly asymptomatic, and this asymptomatic carrier state can persist for as long as six weeks in children <5 years of age. In adults, the duration is shorter, usually about three weeks (25). Once nasopharyngeal colonisation is established, the organism may penetrate the mucosal barriers

and, under normal circumstances, is usually rapidly cleared by the host primary defense mechanisms. However, in some cases, including preceding viral infections, exposure to smoking, reduced immune competence in the human host, and/or pathogen virulence factors (listed in **Table 1**), the host defense mechanisms fail to clear the organism, resulting in invasive or non-invasive disease.

This transition of pneumococcus from colonisation to disease may result in a variety of different clinical syndromes, depending on the route of dissemination. A common dissemination route in children is spread of the organism via the Eustachian tube to the middle ear causing acute otitis media. Symptoms include fever, pain and hearing loss. The organism may also spread to the sinus cavities, causing sinusitis. Pneumococcus is one of the most prevalent bacterial pathogens causing sinusitis and acute otitis media (26;27). Another common route of dissemination is the invasion of the alveoli, resulting in pneumonia. Cough, fever, sweats and shortness of breath are the most frequent symptoms of pneumonia. In 2015, pneumonia was the leading cause of death in children <5 years of age globally (28). Pneumococcus may also invade the bloodstream causing bacteraemia. Meningitis (inflammation of the meninges) may result from direct spread from middle ear, sinuses or from the bloodstream. Symptoms of meningitis include headache, stiff neck, photophobia, fever and confusion. Pneumococcus was the leading cause of bacterial meningitis in the United States of America from 1998 - 2007 (29).

1.6 Diagnosis

A definitive laboratory diagnosis of invasive pneumococcal disease (IPD) is based on isolation of *S. pneumoniae* from normally sterile body sites (e.g. cerebrospinal fluid, blood, pleural fluid).

Usually, specimens are cultured on 5.0% sheep blood agar and incubated overnight at 37°C in the presence of 5.0% CO_2. During microscopic evaluation, pneumococcus will appear as lancet-shaped, Gram-positive diplococci. On sheep blood agar, pneumococcal colonies will be surrounded by a greenish zone of α-haemolysis, and are typically optochin susceptible and bile soluble (3;30). For rapid identification, commercial latex agglutination tests such as BinaxNOW are used directly on specimens (31-33). However, these tests have been shown to have low sensitivity and specificity. In cases where culture of the specimen does not yield growth of pneumococci, possibly due to antibiotic therapy prior to specimen taking, nucleic acid amplification tests such as real-time polymerase chain reaction (RT-PCR) have proven to be an important diagnostic tool. For pneumococcus, a RT-PCR that detects the *lytA* (autolysin) gene is commonly used (34).

1.7 Burden of disease

Pneumococcal disease continues to result in significant morbidity and mortality worldwide, particularly in children <5 years of age. Approximately half a million children <5 years of age die every year due to pneumococcal infections, with most deaths occurring in developing countries (28). An estimated 5.9 million children <5 years of age died globally in 2015, of which 2.8 million were in Africa. Pneumonia, accounting for 13.0% of all cases, was the leading cause of death, killing 767,000 children. In South Africa, 42,000 children <5 years of age died in 2015. The HIV epidemic plays a major role in the burden of pneumococcal disease. HIV-infected individuals have a higher risk of developing IPD than HIV-uninfected individuals. A study in South Africa found that there was a 41.7 - fold increase in severe IPD in HIV-infected children <12 years of age compared with HIV-uninfected children (35).

1.8 Treatment and prevention of disease

1.8.1 Treatment

Pneumococcal diseases are managed through treatment with antimicrobial agents. Treatment usually depends on the clinical syndrome. The American Academy of Paediatrics and Family Physicians recommends amoxicillin as the first line of treatment for otitis media in all age groups (36). If amoxicillin fails, amoxicillin-clavulanic acid, fluoroquinolone or ceftriaxone are used and in the case of penicillin allergy, a macrolide is recommended. The South African Department of Health also gives similar recommendations (37). Sinusitis in all age groups is treated with amoxicillin only or together with clavulanic acid (37-39). Pneumococcal meningitis is usually treated with penicillin G or alternatively with ceftriaxone. If the patient has allergies or there is evidence of resistance to either of these antimicrobial agents, vancomycin plus a carbapenem are used (40). For treatment of community-acquired pneumonia in outpatient settings, the Paediatric Infectious Diseases Society and the Infectious Diseases Society of America (41), and the South African Department of Health (37)(37)(37)(32) recommends using amoxicillin for infants, children >3 months and adults with mild to moderate community-acquired pneumonia. For inpatients, ampicillin or penicillin G is used and if resistance is high in that particular setting or the patient has other complications, a third-generation cephalosporin (ceftriaxone or cefotaxime) is used instead (37;41). If a patient is allergic to penicillin, moxifloxacin is recommended (37;37;41).

1.8.2 Antimicrobial non-susceptibility

Antimicrobial non-susceptibility complicates the management of pneumococcal infections and may lead to treatment failure. Pneumococci were susceptible to almost all relevant antimicrobial

agents until the emergence of penicillin-resistant isolates in the mid-1970s (42;43). The prevalence of pneumococcal antimicrobial non-susceptibility has been increasing globally, including in South Africa, not only to penicillin, but also to non-β-lactam antimicrobials (44;45). In the pre-vaccine era in the United States (between 1995 and 1998); there were increases in the proportions of isolates that were non-susceptible to penicillin (from 21 to 25.0%), cefotaxime (10 to 14.0%), meropenem (10 to 16.0%), erythromycin (11 to 15.0%) and trimethoprim-sulfamethoxazole (25 to 29.0%). In South Africa, between 2003 to 2008, non-susceptibility to the following antibiotics increased significantly ($P < 0.001$): penicillin (23 to 38.0%), tetracycline (15 to 20.0%), erythromycin (13 to 15.0%), clindamycin (10 to 11.0%) and chloramphenicol (3 to 5.0%) (44). However, the introduction and continued use of polysaccharide-protein conjugate vaccines (PCV) in South Africa has greatly reduced the prevalence of non-susceptible vaccine serotype isolates (46). From pre-vaccine (2005 – 2008) to post-vaccine (2011 – 2012) periods, the prevalence of penicillin non-susceptible isolates declined by 82.0%, ceftriaxone by 85.0% and multidrug-resistant isolates by 84.0% among children less than 2 years of age.

1.8.3 Prevention

The majority of disease caused by *S. pneumoniae* is preventable through the use of pneumococcal vaccines. Currently two types of vaccines are available: polysaccharide and polysaccharide-protein conjugate vaccines. The 23-valent polysaccharide vaccine (Pneumovax®, Merck & CO, Whitehouse station, NJ) was the first licensed pneumococcal vaccine and provides protection against 23 serotypes (1, 2, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19A, 19F, 20, 22F, 23F and 33F). This vaccine contains purified capsular polysaccharides, induces a T-cell independent immune response which is poorly immunogenic in children <2 years of age and does

not induce immune memory. Since it is not effective in children <2 years, it is only recommended for adults and immunocompromised children >2 years of age (47).

The 7-valent polysaccharide-protein conjugate vaccine (PCV7, Pfizer Inc., Philadelphia, United States) was the first licensed PCV and contained purified capsular polysaccharides conjugated to a carrier protein. This vaccine provided protection against 7 serotypes (4, 6B, 9V, 14, 18C, 19F and 23F). PCV7 induces immune memory and a T-cell dependent immune response which results in protective immunity in children aged <2 years (48). In addition, PCV7 also confers mucosal immunity, and thereby reduces nasopharyngeal colonisation of serotypes included in the vaccine. PCV7 reduced pneumococcal disease in many parts of the world, including South Africa, and also showed herd effect by reducing disease in unvaccinated adults (46;49-52). Since the widespread use of PCV7, several studies reported an increase of non-vaccine serotypes in carriage (53) and IPD (54-56) (a process called serotype replacement). However, these increases remained low relative to decreases in vaccine serotypes.

Newer vaccine formulations, PCV10 [PHiD-CV (Synflorix™, GlaxoSmithKline Biologicals, Rixensart, Belgium)] and PCV13 (Pfizer Inc., Philadelphia) contain PCV7 serotypes plus additional serotypes (1, 5, 7F and 1, 3, 5, 6A, 7F, 19A, respectively), some of which are those that increased in prevalence after PCV7 rollout. These vaccines provide greater serotype coverage and have further reduced the incidence of pneumococcal disease (57-60). In South Africa, PCV7 was introduced into the Expanded Programme on Immunisation in April 2009 and was replaced with PCV13 in July 2011. Although PCVs are highly effective at preventing IPD caused by serotypes included in the vaccines, they do not offer protection against all disease-causing pneumococcal serotypes. New generation vaccines containing conserved proteins in S. pneumoniae have the

potential to provide serotype-independent protection, thereby preventing serotype replacement. Such vaccines are currently being developed and assessed and have shown promising results in phase 2, randomized, controlled clinical trials (61-63).

1.9 Virulence factors

S. pneumoniae produces a range of virulence factors that play a role in facilitating the disease process. The major virulence determinants include the anti-phagocytic properties of the CPS, adherence, invasion and evasion factors, iron and other heavy-metal transporters, pneumolysin and bacteriocin production, quorum sensing and biofilm formation (64). A summary of known pneumococcal virulence factors are listed in **Table 1**.

1.9.1 Capsular polysaccharide

The CPS is one of the major virulence factors as it forms a shield to protect pneumococci from host cell-mediated phagocytosis or killing by neutrophils, and reduces the total load of complement, mainly C3b, on the bacterial surface (4). It is 200-400 nm thick, and is made up of repeating oligosaccharides that are covalently bound to the outer surface of the peptidoglycan cell wall (3). A close relationship between the thickness of encapsulation and the virulence of pneumococcal strains has been observed.

With the exception of serotypes 3 and 37, the CPS is encoded by genes located at the capsular polysaccharide synthesis (*cps*) locus between the glucan 1,6-α-glucosidase (*dexB*) and oligopeptide

ABC transporter (*aliA*) genes (5;65). **Figure 3** shows serotypes 1 and 9A as examples. The *cps* locus ranges in size from 10,337 (serotype 3) to 30,298 bp (serotype 38) with an average of 20,714 bp (5). The first four genes of the *cps* locus at the 5' end, *wzg*, *wzh*, *wzd* and *wze* (also known as *cpsA, B, C, and D*), are conserved in all serotypes with high sequence identity and are always in this order (5). These genes play a role in the regulation and processing of the *cps* (66;67). In most *cps* loci, the fifth gene encodes *wchA* (*cpsE*), which is responsible for the first step in CPS synthesis. Downstream from the *cps* locus, the polysaccharide polymerase (*wzy*) and flippase (*wzx*) genes are always present together with serotype-specific genes. Genes encoding insertion-sequence transposases are present in the regions between *cps* genes and the flanking *dexB* and *aliA* genes in most serotypes. In addition, four serotypes (19F, 25F, 25A and 38) have group-II introns within the *cps* locus (5). Variations between *cps* loci can range from one bp to many genes. Based on their genetic relatedness, the *cps* loci from 88 serotypes have been grouped into eight clusters and 21 sub-clusters (68). Serotypes within the same serogroup were grouped into the same cluster except for serotypes within serogroups 7, 16, 17, 33, 35 and 47 which were split into two or three different subclusters.

In all but two serotypes (3 and 37), CPS synthesis occurs via the wzx/wzy-dependent pathway (**Figure 4**) (5). In this pathway, a transferase (encoded by *wchA* gene in most serotypes) initiates the transfer of the first sugar to a lipid receptor. This is followed by several transferases which generate the polysaccharide repeat unit which is transported and polymerised by *wzx* and *wzy* genes, respectively. The wzd/wze complex then translocates the mature CPS to the cell surface. Serotypes 3 and 37 CPSs are synthesised via the synthase pathway (69;70).

Table 1*. Virulence factors of *Streptococcus pneumoniae* and their role in disease.

Virulence factor	Mechanism of action	Level of evidence[a]
Anti-phagocytic and invasion factors		
Capsular polysaccharide	Prevents mucosal clearance, antiphagocytic, sterically inhibits complement and immunoglobulin-binding to host receptors	4+
Pneumococcal surface protein A (*pspA*)	Blocks complement component 3b binding to factor B, binds to epithelial membranes	2+
Pneumococcal surface protein C (*pspC*)	Binds factor H and blocks C3b fi xation, binds to human polymeric immunoglobulin receptor during invasion	2+
Pili	Pili on cell surface inhibits phagocytosis, promotes invasion	1+
Metal-binding transporters		
Pneumococcal surface antigen A (*psaA*)	Mediates metal ion uptake (zinc and manganese), protects against oxidant stress, binds to N-acetylglucosamine-beta-(1-4)-galactose disaccharide	1-2+
Pneumococcal iron acquisition A and iron uptake A (*piaA*/*piuA*)	ABC transporters that acquire iron for bacterial growth and virulence	1+
Toxins and enzymes		
Pneumolysin (*ply*)	Cytolytic, toll-like receptor ligand, induces ciliostasis, impairs respiratory burst, activates complement, cytokine, chemokine production	3+
Bacteriocin	Inhibits bacterial competition, might have other cytotoxic actions	2+
IgA protease	Degrades human IgA1	1-2+
Autolysin (*lytA*)	Releases peptidoglycan, teichoic acid, pneumolysin and other intracellular contents	2+
Hyaluronate lyase (*hyl*)	Degrades hyaluronan in the extracellular matrix	1+
Surface adherence factors		
Neuraminidase (*nanA*)	Contributes to adherence, removes sialic acids on host glycopeptides and mucin to expose binding sites	1+
Enolase (*eno*)	Binds to fi bronectin in host tissues	1+
Pneumococcal adhesion and virulence (*pavA*)	Binds to plasminogen within host tissues	1+
Pneumococcal serine-rich repeat protein (*psrP*)	Surface adhesion, binds to platelet surface and promotes tissue invasion	2+
Other factors		
Cell wall polysaccharide	Activates complement, pro-infl ammatory	1+
Lipoteichoic acid	Binds to PAFR, TLR2 ligand, proinfl ammatory	2+
Sortase A (*strA*)	Links surface proteins to cell wall	1+
Biofilm and competence	Biofilm organisms upregulate competence-stimulating peptide; competence genes contribute to pneumonia and meningitis	2+
Phosphorylcholine	Binds to platelet-activating factor receptor. on human epithelial cells	2+
Pneumococcal choline binding protein A	Low manganese-induced protein, unknown function but important in lungs and blood	1-2+

*Taken from reference (64).

[a]; 1+: one or few animal studies, 2+: some evidence from several animal studies, 3+: confirmed in several animal studies and 4+: confirmed in people.

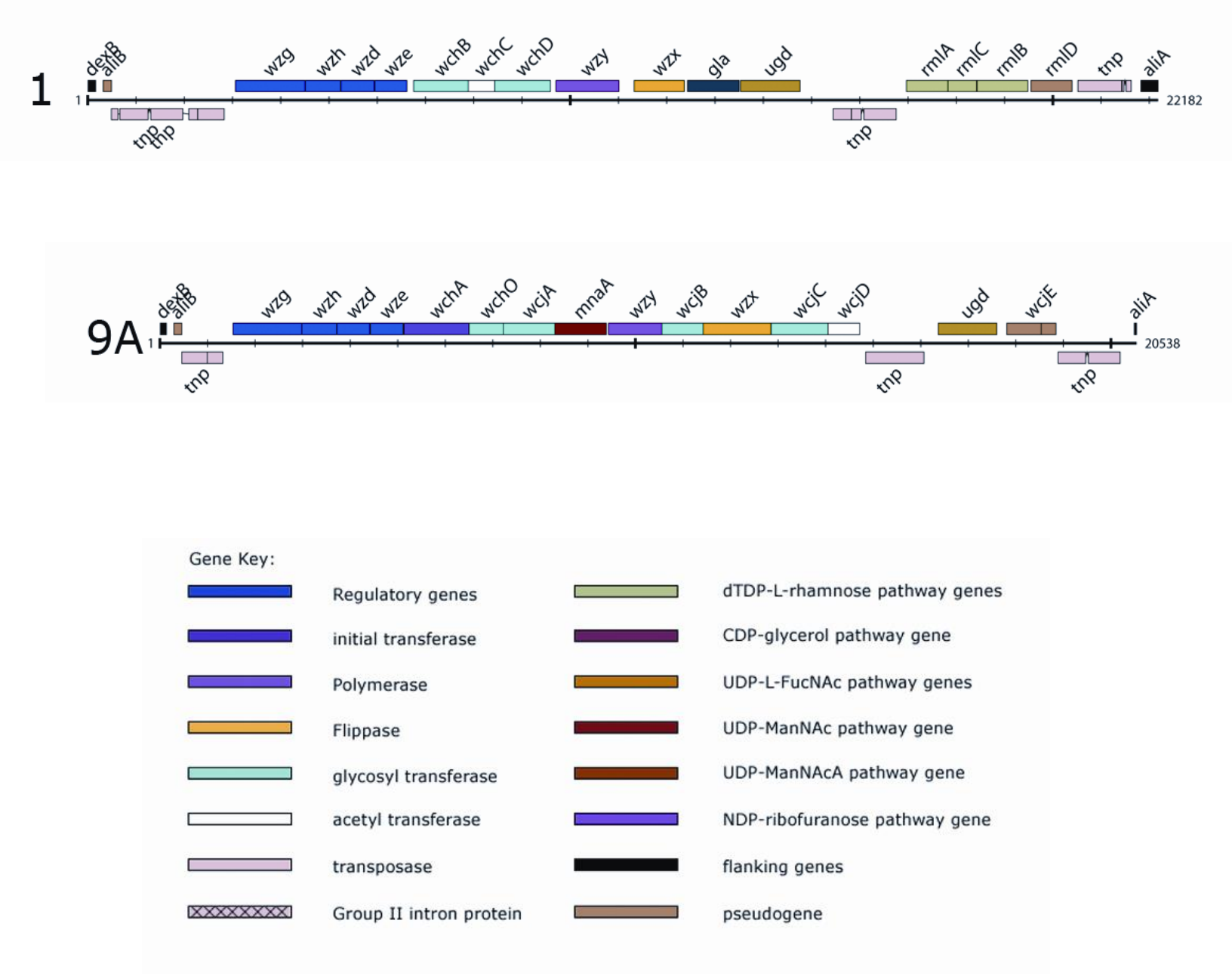


Figure 3. Capsular polysaccharide biosynthesis genes of *Streptococcus pneumoniae* serotypes 1 and 9A. Taken from reference (5).

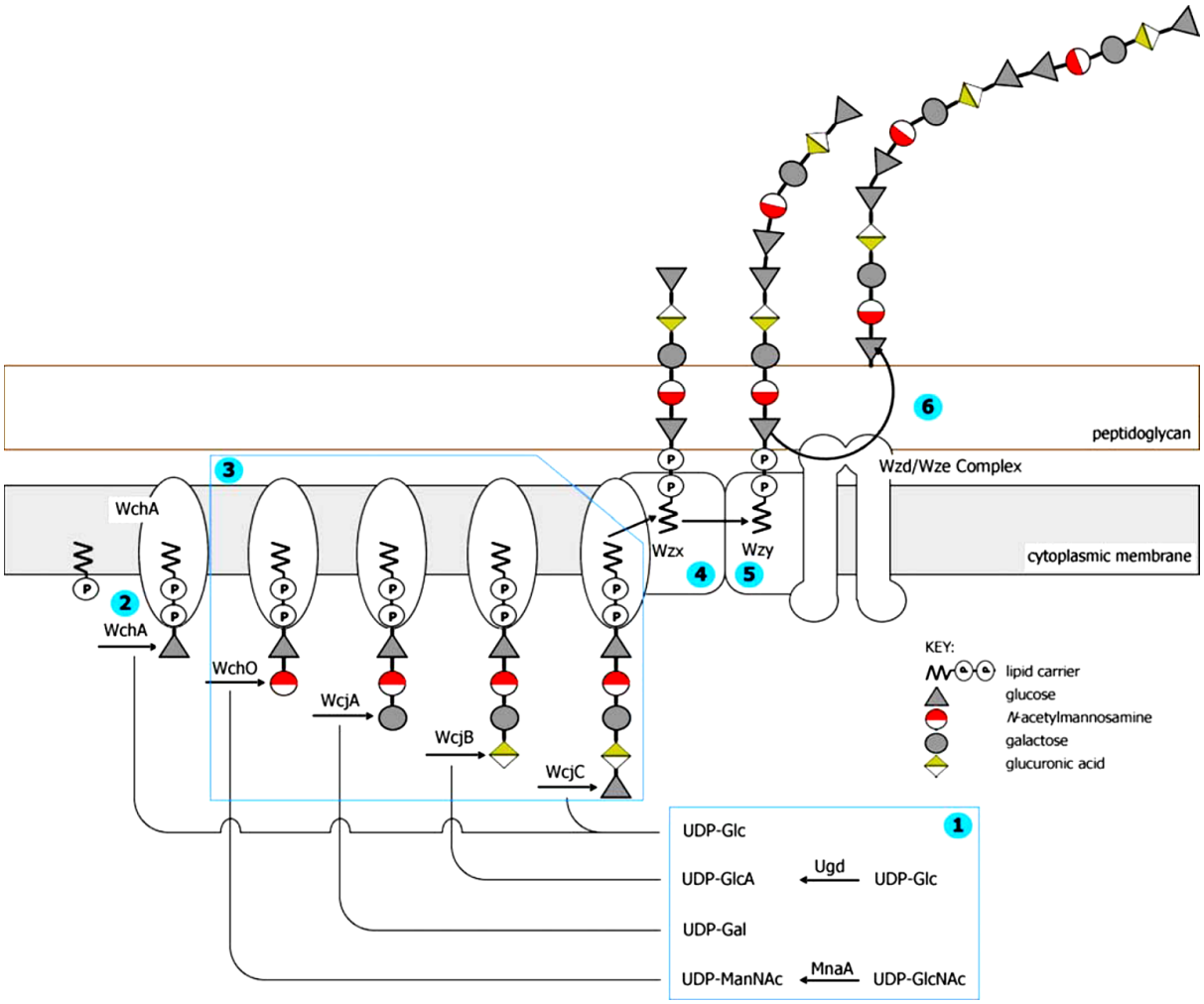


Figure 4. Representation of the wzx/wzy-dependent pathway for biosynthesis of serotype 9A capsular polysaccharide. Taken from reference (5): Pictured is a hypothetical model for capsule biosynthesis in *S. pneumoniae* based on a mixture of experimental evidence and speculation.
1) Non-housekeeping nucleotide sugar biosynthesis

2) The initial transferase (*wchA*) in this case links the initial sugar as a sugar phosphate (Glc-P) to a membrane-associated lipid carrier (widely assumed to be undecaprenyl phosphate).

3) Glycosyl transferases sequentially link further sugars to generate repeat unit.

4) *wzx* flippase transports the repeat unit across the cytoplasmic membrane.

5) *wzy* polymerase links individual repeat units to form lipid-linked CPS.

6) Wzd/Wze complex translocates mature CPS to the cell surface and may be responsible for the attachment to peptidoglycan.

1.10 Nontypeable *Streptococcus pneumoniae*

Some pneumococcal strains, defined as nontypeable pneumococci (NTPn), show no serological evidence of capsule expression when serotyped with the Quellung reaction (15). The inability to do so may be due to several factors namely, low-level CPS expression below Quellung limit of detection, novel serotypes that are not detectable by serotype-specific anti-sera or genetic modifications in the cps locus such as specific single nucleotide polymorphisms (71-74), partial deletions of cps genes (13;74;75), gene duplications (76) and complete cps genes deletions and/or replacement with other non-*cps* genes (23;73-75;77;78). NTPn are predominantly detected in carriage (79) and rarely cause IPD (73;74;77;80;81). From 3.9% to 19.0% of *S. pneumoniae* isolates from carriage and 0.1% to 3.4% of IPD cases are estimated to be due to NTPn. However, this may be an underestimate because NTPn form smaller, drier colonies on blood agar, resembling viridians group streptococci and may therefore be misidentified. In addition, NTPn may also cause non-invasive disease such as acute otitis media (82;83) and are a common cause of bacterial conjunctivitis, causing either outbreaks (84;85) or sporadic cases (86).

Although lacking the principal virulence factor, the lack of CPS can be advantageous to NTPn. These isolates show increased adherence to epithelial cells when colonising the nasopharynx (87-89) and are more easily transformed and therefore able to take up foreign DNA, compared to encapsulated pneumococcal isolates (EcPn) (90;91). In addition, NTPn have been shown to have a growth advantage over EcPn (72;92). One study showed that this is due to the fact that EcPn use more energy to synthesise the CPS compared to NTPn (92). Thus, the loss of CPS increases the fitness of the pneumococcal strains.

NTPn are believed to play a role in pneumococcal ecology and disease, although their true role is not known. Because of their high prevalence in carriage combined with high transformability, they may act as vectors of genetic material, such as antimicrobial and virulence genes for EcPn, highlighted by the fact that vast majority of colonising NTPn are multidrug-resistant (93-95). In addition, NTPn are found to co-colonise the nasopharynx with EcPn (93;94;96) and this provides an opportunity for horizontal gene transfer. One study showed that a globally disseminated *S. pneumoniae* serotype 19F clone had evolved from low-level to high-level penicillin resistance by acquisition of resistance gene fragments from NTPn isolates (97). Another study in Thailand (2007-2010) among infants and their mothers, showed that the highest rates of receipt and donation of DNA occurred in NTPn compared to EcPn, and the most frequently exchanged genes were those associated with antimicrobial resistance and immune interactions (23).

NTPn have been categorised into two groups based on the contents of the *cps* locus (77). Group I NTPn have at least partial *cps* genes, whereas in Group II the *cps* genes are replaced with genes encoding novel proteins or the *cps* genes are completely deleted. Group II is further subdivided into four null capsule clades (NCC) (**Figure 5**) (12;73). NCC1 has the *psk* gene (also called *nspA*), NCC2 has *aliC* (*aliB* ORF1) and *aliD* (*aliB* ORF2) genes either with (NCC2a) or without (NCC2b) a putative toxin-antitoxin system (encoded by *ntaAB* genes), NCC3 has the *aliD* gene, and NCC4 contains only transposable elements in the *cps* locus. The *psk* gene has been shown to play a role in adherence to epithelial cells and colonisation (98), while *aliC* and *aliD* genes facilitate upregulation of competence for genetic transformation and mediate colonisation, respectively (99). Group II NTPn are frequently isolated from carriage compared to Group I, which predominantly cause IPD (23;73). On the basis of multilocus sequence typing (MLST) and whole genome sequencing, two distinct lineages have been identified, namely, a lineage of exclusively NTPn

strains (classic lineage) and a lineage related to EcPn strains (sporadic lineage) (23;100;101). Group II NTPn are typical of the classical lineage but can also be found in sporadic lineage whereas Group I NTPn are thought to be derived from EcPn that have lost the ability to express the CPS and fall within the sporadic lineage

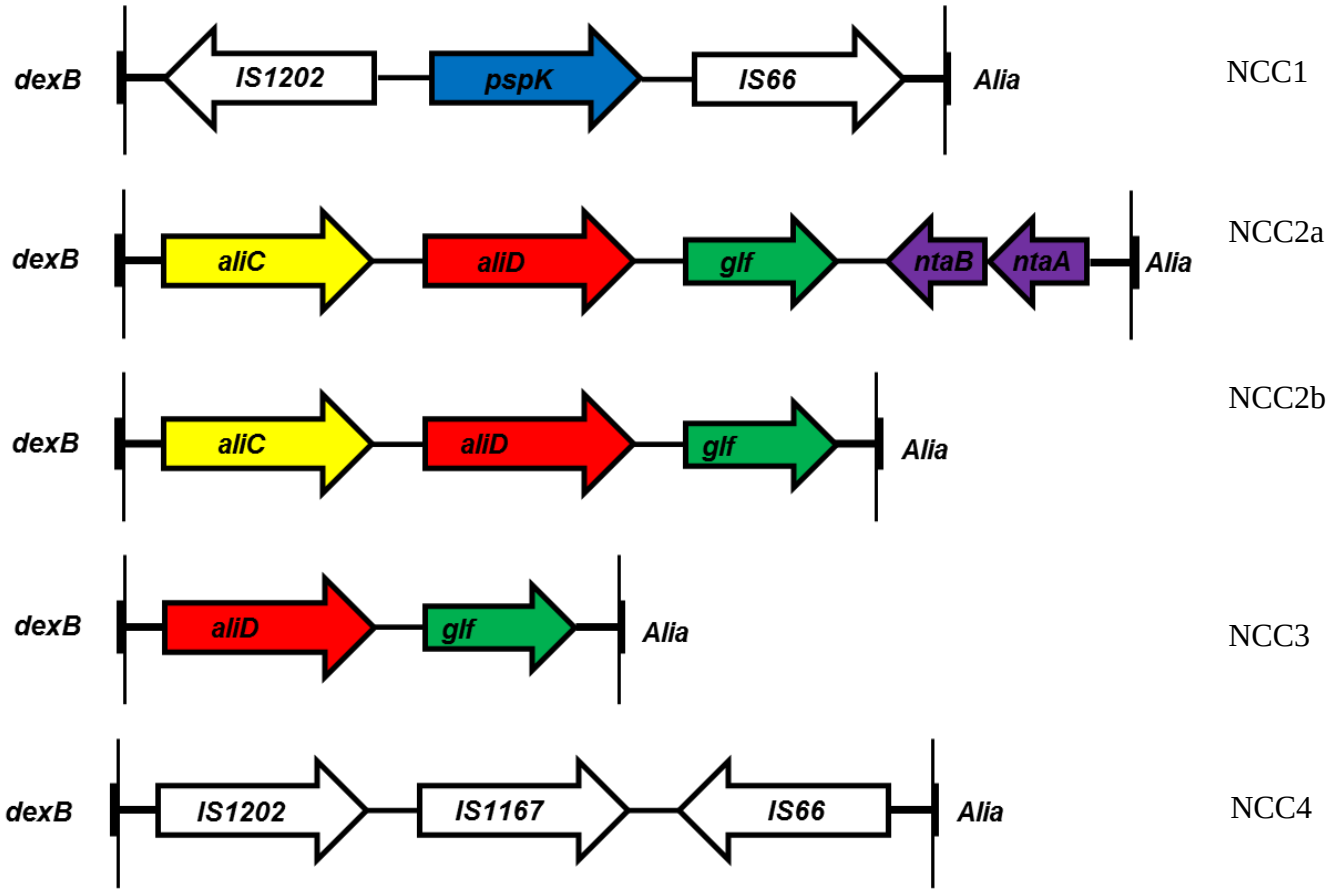


Figure 5. Schematic representation of nontypeable *Streptococcus pneumoniae* Group II capsule loci. Genes are depicted with arrows. NCC: Null capsule clade. Adapted from reference (12).

Most studies describing NTPn have been from conjunctivitis outbreaks. A few studies have characterised NTPn from carriage (12;23;75;78;93-95) and only three studies from IPD (13;73;77). The largest collection of invasive NTPn characterised (N=88) was from the Centers for Disease Control and Prevention (CDC) Active Bacterial Core surveillance (2006-2009), where they found that the vast majority of isolates (89.8%) were Group I (73). Of these, 25.0% were derived from ancestral serotype 8 (serotype from which NTPn may have originated) that had lost the ability to express a capsule due to mutations in *wchA* gene. Sixty-seven percent of Group II isolates were NCC2 and related to an established NTPn lineage. Another study in the United States among Native American communities (all ages; years 1994-2007) found almost all of invasive NTPn (96.4%) to be Group I, genetically diverse although ancestral serotype 1 (18.0%) and 7F (14.0%) were more prevalent (13). This study reported no increase in invasive NTPn post-PCV7. Antimicrobial non-susceptibilities were not reported for these two studies. In Switzerland, 57.0% of invasive NTPn were found to be Group I and all isolates were susceptible to penicillin (77).

In contrast to invasive NTPn, the vast majority of carriage NTPn are Group II (23;78). Ninety-one percent of carriage NTPn (2007-2010) in Thailand among infants and their mothers were Group II, with majority (57.0%) having a NCC1 (23). About 83.0% of these isolates were predicted to be resistant to cotrimoxazole and β-lactam antibiotics based on the detection of known resistance genes. A subset of a collection of carriage NTPn (1997-2007) from children in Lisbon, Portugal (representative of lineages detected over the study period) were found to be all Group II-NCC2 (78). Seventy-seven percent and 85.0% of isolates were resistant to penicillin and cotrimoxazole, respectively, and 38.0% were multidrug resistant (resistant to at least three antimicrobial classes). In another study conducted in Lisbon and Oeiras (Portugal), carriage NTPn increased from 1.5% in 2001 to 5.1% in 2006 (post-PCV7) among day-care attendees (102). A study in the Gambia

(2011-2012) among mother-infant pairs found 88.0% of their carriage NTPn to be Group I and reported an increase in NTPn from 0.3% in 2011 to 6.0% in 2012 (P <0.001) post-PCV13 (103). Because the majority of isolates (88.0%) were in fact EcPn that had lost the ability to express the capsule, the authors concluded that the increase is likely due to PCV13 selection pressure.

The mechanisms by which NTPn are able to cause IPD without the principal virulence factor (CPS) are not known. The immune suppression of the host and/or other lesser known or novel virulence factors expressed by NTPn might possibly explain the success of these strains in causing IPD (71). In all studies describing invasive NTPn, the populations from whom NTPn were isolated are often not well described (e.g. basic demographics and immunological status) (13;73;77). Comparative genomic studies of invasive and carriage NTPn could provide insights into the pathogenicity of invasive NTPn isolates. One study conducted among Native American communities (all ages; 1994-2007) compared invasive and carriage NTPn, but using only MLST and microarray analysis of *cps* locus (only 12.0% of carriage NTPn were analysed by microarray) (13). In that study, almost all of invasive NTPn (96.4%) had Group I CPS, sequence types (ST) found among EcPn and were genetically diverse, whereas carriage NTPn were clonal and had STs typically found among carriage NTPn strains. The authors concluded that invasive NTPn fundamentally differed from the majority of carriage NTPn.

1.11 Multilocus sequence typing

A number of methods have been used to study evolutionary relationships and measure genetic diversity of *S. pneumoniae* (104-107). This includes older methods such as multilocus enzyme electrophoresis, which uses electrophoretic movement of metabolic enzymes (encoded by conserved

housekeeping loci) as a measure of allelic variation among strains within a population, and pulsed-field gel electrophoresis, which is based on the whole genome fingerprint patterns which are compared to determine genetic relatedness between isolates. These methods are either subjective, poorly reproducible, difficult to compare amongst laboratories or define variation within a small proportion of the genome. Recently, nucleotide-based methods such as MLST have been developed to study evolutionary relationships and measure genetic diversity of *S. pneumoniae*.

Multilocus sequence typing is based upon variation in the nucleotide sequences of approximately 450 bp fragments of seven housekeeping genes. Sequence data for each allele are submitted to a centralised global database and assigned an allele number. The unique combination of seven allele numbers is submitted to the same database and a ST is assigned. Isolates sharing at least six of the seven alleles are grouped into clusters called a clonal complex (CC) (a less stringent definition where isolates share five out of seven alleles is also sometimes used). Members of a CC are assumed to have a recent common ancestor/founder. Sequence types that vary from that of a founder at a single locus are called single-locus variants, at two loci (double locus variants), at three loci (triple locus variants) and so on. Using the MLST database, some serotypes have been associated with certain STs.

MLST is easily accessible via the internet, generate unambiguous and reproducible data and is standardised, therefore easily comparable between laboratories. However, MLST define variation within a small proportion of the genome. Some isolates may exhibit high levels of variation in other parts of the genome. Nonetheless, MLST is still widely used and regarded as a valuable tool for bacterial typing.

1.12 Whole genome sequencing

In order to understand the genomic diversity of NTPn, robust methods with high resolution are required. Whole genome sequencing is a powerful method that is being widely used to study evolutionary relationships and measure genetic diversity in many bacterial species including *S. pneumoniae*. This is because the introduction of next generation sequencing technologies has significantly reduced the cost and time required for whole genome sequencing. This method gives high resolution on the genetic structure and the relationship between strains as it takes into account the entire genome of the organism.

There are a number of next generation sequencing platforms which use different sequencing technologies. **Table 2** shows a summary of their performance metrics. The first next generation sequencing platform was the Roche/454 FLX (Roche 454 Life Sciences, Connecticut, United States) which uses pyrosequencing, where incorporation of a nucleotide by DNA polymerase results in the release of pyrophosphate and this initiates reactions that produce light by the luciferase enzyme. The light emitted is then used to infer the sequence. The Illumina Genome Analyzer (Illumina, Inc., San Diego, United States) uses a sequencing-by-synthesis method. DNA strands are immobilised on the surface of a flow cell. During each cycle, a single labelled deoxynucleoside triphosphate is added. The nucleotide label serves as a terminator for polymerisation, so after each deoxynucleoside triphosphate incorporation, the fluorescent dye is imaged to identify the base and then enzymatically cleaved to allow incorporation of the next nucleotide. The SOLiD™ System (Applied Biosystems, Santa Clara, CA, USA) uses a sequencing-by-ligation method. In this method, a set of four fluorescently-labelled probes (8 bp in length, first 2 bp are specific and the rest are degenerate) compete for ligation to the sequencing primer. Each

ligation step is followed by fluorescence detection and cleavage of the fluorescent probe. Ion Torrent (Thermofisher Scientific Inc, Waltham, Massachusetts, USA) uses semiconductor sequencing. This method works on the principle that during incorporation of deoxynucleoside triphosphate, hydrogen ion is released. The released hydrogen ion changes the pH of the solution which is detected by the ion sensor and converted to nucleotide. PacBio RS II (Pacific Biosciences, Menlo park, CA, USA), considered the next evolution of sequencing, works on single molecule real-time sequencing. This technology utilises zero-mode waveguide, which creates illuminated observation volume to observe only a single nucleotide of DNA being incorporated. A single DNA polymerase is attached to the bottom of zero-mode waveguide with a single molecule of DNA. Fluorescently labelled deoxynucleoside triphosphates are added and only the complimentary nucleotides will be incorporated. When a nucleotide is incorporated by DNA polymerase, the fluorescent molecule leaves the zero-mode waveguide. A detector at the bottom of zero-mode waveguide detects fluorescence and calls the base according to the corresponding fluorescence of the dye.

Table 2. Comparison of next-generation sequencing platforms

Instrument	**Roche/454 FLX**	**Illumina MiSeq**	**SOLiD 5500W**	**Ion Torrent PGM**	**PacBio RS II**
Sequencing technology	Pyrosequencing	Sequencing by synthesis	Sequencing by ligation	Semiconductor sequencing	Single-molecule real-time sequencing
Read length	750 (average) up to 1000	2 X 250	75 for fragments 2 X 250 for mate pairs	35 to 400	5000 (average) up to 20000
Accuracy (%)	99.9	98.0	99.9	98.0	99.9
Time per run	24 hours	24 hours	7 to 12 days	2 hours	30 to 120 minutes
Cost (US$)[a]	10	0.05 to 0.15	0.13	1	2
Advantages	Long reads	High throughput	High accuracy Low cost per base	Low capital outlay Fast	No amplification, fast Longest reads
Disadvantages	Expensive sequencing Homopolymer errors		Expensive capital Short reads	Homopolymer errors	Instrument cost Lower yield at high accuracy

[a]Per 1-million bases

1.13 Study rationale

Despite evidence of their increase in prevalence post PCV introduction, and important role they play in the evolution and adaptation of the species (the absence of a capsule enhances transformability thereby increasing the uptake and recombination of DNA), NTPn are poorly understood. As such, limited data exist which describe these organisms in terms of their epidemiology and genetic diversity. Although typically associated with colonisation, NTPn have also been associated with outbreaks of bacterial conjunctivitis and occasionally cause otitis media and IPD. Thus studying the epidemiology, molecular structure and genomic diversity of NTPn is important, especially because the current pneumococcal vaccines are not effective against NTPn.

1.13.1 Aim

To investigate the epidemiology, population structure and genomic diversity of NTPn in South Africa.

1.13.2 Objectives

1. To describe the epidemiology of invasive (2003 – 2013) and carriage (2009 – 2012) NTPn in South Africa
2. To determine and compare the antimicrobial susceptibility profiles of invasive and carriage NTPn
3. To perform whole genome sequencing on invasive and carriage NTPn in order to:

- Determine and compare the genetic basis for their nontypeability
- Describe their genetic relatedness
- Compare the genomes to identify potential virulence factors in invasive isolates that are absent in carriage isolates

2 Materials and methods

2.1 Bacterial isolates

2.1.1 Invasive NTPn isolates

Isolates were obtained through the GERMS-SA (Group for Enteric, Respiratory and Meningeal Disease Surveillance in South Africa) network, which conducts active, national, laboratory-based surveillance for IPD. Surveillance began in 1999 and was enhanced in 2003 in 24 sites in which dedicated surveillance officers collected additional clinical information (46;108). Over 200 diagnostic laboratories across South Africa routinely submit clinical isolates and basic patient demographic data to the network. A case of IPD was defined as the detection of *S. pneumoniae* from a normally sterile site (e.g., cerebrospinal fluid, blood) from January 2003 through December 2013. Isolates were cultured on 5.0% horse blood agar (Diagnostic Media Products, Johannesburg, South Africa) and incubated at 37°C in 5.0% CO_2 for 18-24 hours. Optochin sensitivity and bile solubility assays were performed to confirm *S. pneumoniae* identification. In addition, for invasive NTPn isolates, real-time PCR detecting the *lytA* gene was performed (34).

2.1.2 Carriage NTPn isolates

Isolates were obtained from community cross-sectional studies assessing the impact of PCV7 on pneumococcal nasopharyngeal colonisation in South Africa. Cross-sectional carriage studies were conducted among randomly selected households with at least one child <2 years of age in rural Agincourt, Mpumalanga Province in the months of May to October during 2009 and 2011 (92), and among HIV-infected and HIV-uninfected mother-child pairs in urban Soweto, Johannesburg,

Gauteng Province, between May and September during 2010 and 2012 (93). Nasopharyngeal swabs (Medical Wire and Equipment Co, Ltd) were collected, and placed in skim milk, tryptose, glycerol and glucose broth transport media (STGG) and stored at -70°C until they were processed. Nasopharyngeal swabs were cultured on 5.0% horse blood agar plus 5 mg/ml gentamicin sulphate (Diagnostic Media Products) and incubated at 37°C in 5.0% CO_2 for 48 hours. Pneumococcal isolates were identified by susceptibility (≥14mm) to optochin (5 µg; Becton Dickinson Microbiology Systems, Cockeysville, MD, USA), bile solubility as well as RT-PCR detecting *lytA* gene (34)

2.2 Phenotypic serotyping

Serotyping was performed by the Quellung reaction using serotype-specific antisera (Statens Serum Institut, Copenhagen, Denmark). Bacterial cells from overnight pneumococcal cultures were suspended in 1 ml of saline (Diagnostic Media Products) to a turbidity equivalent to that of <0.5 McFarland standard (Diagnostic Media Products). A smear of the bacterial suspension was made on a glass slide and air dried. A drop (approximately 5µl) of methylene blue solution was mixed with a drop (approximately 5µl) of omni anti-serum (Statens Serum Institut) on a coverslip and placed over the air dried suspension. A drop of microscope oil (Merck, Darmstadt, Germany) was placed on top of the coverslip and this was viewed under the phase contrast microscope (Nikon Instech, Kanagawa, Japan) using an oil immersion lens at 100X magnification. A positive reaction was defined by the isolate appearing swollen. *S. pneumoniae* ATCC 49619 and R6 were used as positive and negative controls, respectively. Co-infection and co-colonisation were defined as simultaneous isolation of at least two different serotypes (including NTPn isolates) from the same specimen.

2.3 Antimicrobial susceptibility testing

Minimum inhibitory concentrations (MIC) for invasive and carriage NTPn were determined by broth microdilution using commercially customised TREK panels (Trek Diagnostics Inc., Ohio, United States) following manufacturer's instructions. Each Sensititre® HPB susceptibility plate is coated with erythromycin, penicillin, clindamycin, linezolid, gentamicin, meropenem, cotrimoxazole, tetracycline, ceftriaxone, amoxicillin, ceftaroline, quinupristin/dalfopristin, levofloxacin, rifampicin, chloramphenicol and vancomycin. A 0.5 McFarland standard Mueller-Hinton broth (with TES buffer) suspension of isolates was prepared. 100µl of the suspension was transferred into Mueller-Hinton broth with TES/lysed horse blood and mixed. This was poured into a sterile trough and 100µl transferred into each well of Sensititre® susceptibility plates. The susceptibility plates were sealed and incubated at 37°C for 18-24 hours in the O_2 incubator. The lysed horse blood suspensions were inoculated onto 5.0% blood agar plates (to check for purity) and incubated at 37°C for 18-24 hours in the CO_2 incubator. MICs for EcPn isolates were determined by agar dilution or Etest for the years 2003-2008 and this was changed to broth microdilution from 2009 (46;109). MIC breakpoints were interpreted using Clinical and Laboratory Standards Institute guidelines (110). For penicillin, isolates with MICs of ≥0.12mg/L were defined as non-susceptible. Isolates that were intermediately resistant or resistant to any of the antibiotics were regarded as non-susceptible. Multidrug resistance was defined as non-susceptibility to three or more classes of antibiotics.

2.4 Preparation of DNA extracts for whole genome sequencing

2.4.1 Pre-lysis

Pure overnight cultures were inoculated into 3 ml of Brain Heart Infusion broth (Diagnostic Media Products) and incubated for 18-24 hours at 37°C in 5.0% CO_2. 1.2 ml of the suspension was transferred into a sterile 1.5 ml eppendorf tube and centrifuged for 3 minutes at 13,000 rpm to pellet the cells. The bacterial pellets were suspended in 200µl of 10 mg/ml of lysozyme (Sigma-Aldrich, St. Louis, MO, USA) and incubated at 37°C for 1 hour.

2.4.2 DNA extraction

Genomic DNA was extracted using the QIAamp® DNA Mini kit (QIAGEN, Hilden, Germany) following manufacturer's instructions. 50 µl of AL lysis buffer and 20 µl of Proteinase K were added to the resuspended pellets and this was incubated at 56°C for 1 hour. 150 µl of lysis buffer AL was added and this was incubated at room temperature for 10 minutes. 250 µl of 100.0% molecular grade ethanol was added. 700 µl of the samples were transferred to Qiagen columns and centrifuged at 8000 rpm for 1 minute. Columns were transferred to new collection tubes and 500 µl of buffer AW1 was added to each column and this was centrifuged at 8000 rpm for 1 minute. Columns were transferred to new collection tubes and 500 µl of buffer AW2 was added. This was centrifuged at 14,000 rpm for 3 minutes. Columns were transferred to 1.5ml sterile eppendorf tubes, 60µl of pre-warmed elution buffer was added to each column and this was incubated for 5 minutes at room temperature, followed by centrifugation at 12,000rpm for 2 minutes. The columns were discarded and the 1.5ml sterile eppendorf tubes with 60 µl of eluted DNA were retained.

2.4.3 DNA concentration quantification

DNA was quantified using the Qubit® 2.0 fluorometer (Invitrogen, Carlsbad, CA, USA) and the Qubit™ dsDNA BR assay kit (Invitrogen) according to manufacturer's instructions. A working solution for each sample was prepared by mixing 199 µl BR buffer with 1 µl BR reagent. 195 µl of the working solution was mixed with 5 µl of extracted DNA and this was incubated at room temperature for 10 minutes. The samples were read on the Qubit® 2.0 fluorometer (Invitrogen). DNA samples with a concentration of at least 15 ng/µl (in a total volume of 55 µl) were sent for whole genome sequencing, whereas DNA samples that yielded < 15 ng/µl were re-extracted and quantified again.

2.5 Genome sequencing, assembly and annotation

Extracted genomic DNA was subjected to library preparation with Nextera XT DNA sample preparation kit (Illumina, San Diego, CA, USA) following manufacturer's instructions. Briefly, DNA samples were quantified with Qubit® Fluorometer using Qubit® dsDNA High Sensitivity Assay Kit (Invitrogen) and diluted to a starting concentration of 1 ng/µl. Each DNA sample was tagmented (fragmented and tagged) at 55°C for 5 minutes to generate adapter-ligated DNA fragments. Tagmented DNA fragments were amplified by a limited-cycle PCR (15 cycles) to add Illumina sequencing adapters and to create a dual indexed library for each sample. The resulting indexed libraries were purified and size-selected using Agencourt AMPure® XP beads (Beckman Coulter Inc., Brea, California, USA). Purified DNA libraries were quantified and assessed for quality on an Agilent Bioanalyser (Agilent Technologies, Santa Clara, California, USA). Libraries

were normalized, pooled at equimolar concentrations and denatured using NaOH prior to loading onto the reagent cartridge. Genome sequencing was carried out on Illumina MiSeq instrument using a 600-cycle MiSeq Reagent v3 Kit (Illumina) to generate paired-end reads (2 x 300 bp). The resulting paired-end reads were quality trimmed using Trim sequences module in CLC Genomics Workbench v8 (CLC bio, Aarhus, Denmark) with the limit set at 0.05 (e.g. sequences with PHRED score of < 30 were trimmed) and ambiguous nucleotides were also trimmed. The paired-end reads were mapped to the reference genome of *S. pneumoniae* ATCC 700669 (GenBank accession no. FM211187, serotype 23F, ST81) using CLC Genomics Workbench (CLC bio). *De novo* assembly was performed for all isolates using CLC Genomics Workbench (CLC bio) and ordered relative to *S. pneumoniae* ATCC 700669 using Mauve (111). The ordered contigs were concatenated and annotated using Prokka v1.11 (112).

2.6 Data analysis

2.6.1 *cps* locus and sequence types

The sequences of the *cps* loci were manually extracted from whole genome sequence data using CLC Genomics Workbench (CLC bio). The *cps* loci were used to categorize NTPn into different groups and clades (12;73). The seven MLST alleles were derived from whole genome sequence data using the Bio-MLST-Check module (http://search.cpan.org/dist/Bio-MLST-Check/lib/Bio/MLST/Check.pm) This module also assigned the MLST allele numbers and ST. New alleles/allelic profiles determined by the module were submitted to the MLST database (https://pubmlst.org/bigsdb?db=pubmlst_spneumoniae_seqdef&page=submit) for assignment.

The eBURST v3 algorithm (http://eburst.mlst.net) was used to determine genetic relationships between invasive and carriage NTPn and grouped isolates into CC. A CC was defined as a group of isolates sharing at least five of the seven alleles with any other isolate in the group.

2.6.2 *In silico* prediction of ancestral serotypes for Group I NTPn isolates

To predict ancestral serotypes of Group I NTPn isolates (serotypes from which NTPn may have originated), three methods were used, namely serotype-ST associations, BLAST of the *cps* loci and phylogenetic analysis. Using the MLST database, we checked for serotypes that are most commonly associated with the ST of our Group I NTPn and those serotypes were assumed to be the likely ancestral serotype. The *cps* loci of Group I NTPn were blasted against reference *cps* loci of 94 serotypes using CLC Genomics Workbench (CLC bio) (GenBank accession numbers CR93162 – CR931722, EF538714, HM171374, GU074953 and JQ653094). The serotype that gave the highest BLAST bit score was assigned as the likely ancestral serotype. Phylogenetic analysis (described in section **2.6.4**) of NTPn from this study together with invasive EcPn with different serotypes from our setting (some matching the predicted ancestral serotypes) was used to confirm predicted ancestral serotypes. If our Group I NTPn clustered closely with a particular serotype in the phylogenetic tree, that serotype was assumed to be the likely ancestral serotype. If there were conflicting ancestral serotypes identified by the different methods, the final result was based on phylogenetic analysis as this method takes into account the entire genome.

2.6.3 Capsule mutations in Group I NTPn

To determine the genetic variations within the *cps* locus of Group I NTPn isolates that may be responsible for their phenotypic nontypeability, we used the alignment module in CLC Genomics Workbench to align the *cps* locus of each isolate together with a number of its predicted ancestral serotype *cps* loci (**Appendix A**).

2.6.4 Phylogenomic tree construction

A rapid large-scale prokaryote pan genome analysis pipeline (113) and randomized accelerated maximum likelihood (RAxML) (114), were used to construct maximum likelihood phylogenetic trees based on core genome single nucleotide polymorphisms. The core genome alignment module in the rapid large-scale prokaryote pan genome analysis pipeline was used to extract predicted coding regions from annotated assemblies and convert them to protein sequences. All protein sequences were compared with each other using BLASTP. Proteins that had alignment similarity of ≥70.0% and were present in at least 90.0% of the isolates were defined as the core genome. RAxML was used to create a bootstrapped maximum likelihood phylogenetic tree from the resulting core genome alignment and this was visualized in Figtree v1.4.2. Two phylogenetic trees were constructed. The first tree included invasive NTPn [from this study (n=39)], EcPn (n=42) from South Africa and carriage NTPn from other countries (n=6) (115;116). This tree was constructed to examine the relationships among our invasive NTPn and also with EcPn. The clustering of NTPn with EcPn was also used confirm the predicted ancestral serotypes. The second tree included invasive (n=39) and carriage (n=137) NTPn from this study, a previously described

global collection of carriage NTPn (n=131) (115) and invasive EcPn from South Africa (n=48) (**Appendix B**). In this tree, we wanted to assess the relationship of our carriage NTPn isolates with our invasive NTPn, EcPn and carriage NTPn from other settings. The tree was also used to confirm predicted ancestral serotypes of carriage isolates.

2.6.5 Genomic comparison of invasive and carriage NTPn

To compare the gene content of invasive and carriage NTPn isolates, a rapid large-scale prokaryote pan genome analysis pipeline was used to generate a pan-genome of the isolates (113). From the pan-genome, the pipeline determines the presence/absence of these genes among the isolates. A gene was classified as unique if it was present in at least one invasive isolate and absent in all carriage isolates. Gene differences between invasive and carriage isolates were analysed and the statistical significance determined. Genes that were significantly associated with invasive NTPn and present in at least 70.0% of invasive NTPn were further analysed using WEGO software to perform gene ontology functional classification (117).

2.6.6 Statistical analyses

Proportional differences were analysed using the chi-square test or Fisher's exact test where appropriate. Statistical analyses were performed with GraphPad InStat 3 (GraphPad software Inc., California, USA) and Microsoft Excel. Sequence type diversity was calculated using Simpson's diversity index (D) [http://darwin.phyloviz.net/dokuwiki/doku.php (under comparing partitions

tab]. D ranges from 0 to 1, and values closer to 1 indicate higher diversity. Statistical significance was assessed at $P < 0.05$.

2.6.7 Ethics

Ethical approval for national-laboratory based surveillance (protocol number: M081117), the two carriage studies (protocol numbers: M090114 and M090115) and this project (protocol number: M140450) (**Appendix C**) was obtained from the Human Research Ethics Committee of the University of the Witwatersrand, Johannesburg, South Africa.

3 Genomic analysis of nontypeable pneumococci causing invasive pneumococcal disease in South Africa, 2003–2013

3.1 Results

3.1.1 Invasive pneumococcal disease surveillance in South Africa, 2003-2013

From 2003 through 2013, 46,485 cases of IPD were reported, of which 32,824 (71.0%) had viable isolates. Thirty-nine (0.1%) were nontypeable by the Quellung reaction and were thus classified as NTPn. Case characteristics are shown in **Table 3**. Forty-four percent (17/38) of cases were in children <5 years. Two patients with known vaccination status had received one dose of PCV13 and two doses of PCV7, respectively. Thirteen percent (2/15) of patients with known immune status were immunocomprised. Of ten patients with known HIV status, 50.0% (5/10) were seropositive. Twenty-two percent (4/18) of patients with known outcome data died.

3.1.2 Classification of invasive NTPn

Fifty-six percent (22/39) of invasive NTPn had at least partial *cps* genes and were thus categorized as Group I (**Table 4**). The remaining isolates (17/39, 44.0%) had complete deletion of *cps* genes (n=17) and replacement by *pspK* (n=1), *aliC*, *aliD* and *ntaAB* (n=12) or *aliC* and *aliD* genes only (n=2) and were thus categorized as Group II. Based on the classification proposed by Park *et al* (12;73), these isolates were defined as NCC1 (n=1), NCC2a (n=12), NCC2b (n=2) and NCC4 (n=2).

3.1.3 Antimicrobial resistance patterns

Non-susceptibility among invasive NTPn was highest for cotrimoxazole (17/39, 44.0%), penicillin (14/39, 36.0%), tetracycline (13/39, 33.0%) and erythromycin (12/39, 31.0%) (**Figure 6**). Prevalence of non-susceptibility was higher among invasive NTPn than EcPn for most antimicrobials, however this difference was only significant for tetracycline [33.0% (13/39) versus 18.0%, (5902/32785), $P = 0.02$] and erythromycin [31.0% (12/39) versus 14.0% (4589/32785), P = 0.005]. In addition, 31.0% (12/39) of invasive NTPn were multidrug resistant [of which 21.0% (8/39) included penicillin] compared to 19.0% (6230/32785) of EcPn, however this difference was not statistically significant ($P = 0.09$). With the exception of cotrimoxazole and chloramphenicol, Group II invasive NTPn had a higher prevalence of non-susceptibility than Group I, although this difference was only statistically significant for penicillin [59.0% (10/17) versus 18.0% (4/22), $P = 0.02$] and erythromycin [53.0% (9/17) versus 14.0% (3/22), $P = 0.01$]. Multidrug resistance was also higher in Group II invasive NTPn (47.0%, 8/17) compared to Group I (18.0%, 4/22), although not statistically significant ($P = 0.08$).

3.1.4 Multi-locus sequence typing

A total of 28 STs were identified (**Table 4**) with a Simpson's diversity index of 0.97 (95.0% CI: 0.947 - 0.996). Among Group I isolates, 18 STs were identified: 6 were novel STs and 12 were STs usually associated with EcPn. Among Group II isolates, 10 STs were identified: 7 were novel STs and 3 (NCC2a) were ST344 which is usually associated with carriage NTPn, and ST105 and ST218 (NCC4) which are associated with EcPn. Two CCs were identified among Group I isolates; CC217 (n=6) and CC53 (n=3) (**Figure 7**). The remainder (n=12) of Group I isolates were not

assigned to any CC, except for ST5604 which is a single-locus variant of a ST105 isolate from Group II. CC344 (n=6) was identified among Group II isolates. The remainder (n=10) of Group II isolates were not assigned to any CC.

Table 3*. Characteristics of patients with invasive disease due to nontypeable and encapsulated *Streptococcus pneumoniae*, South Africa, 2003 – 2013

Characteristic		Nontypeable *S. pneumoniae* N=39	Encapsulated *S. pneumoniae* N=32785	Total N=32824
		n (%)	n (%)	n (%)
Year	2003	2 (5)	2885 (9)	2887 (9)
	2004	4 (10)	3465 (11)	3469 (11)
	2005	5 (13)	3643 (11)	3648 (11)
	2006	5 (13)	3411 (10)	3416 (10)
	2007	2 (5)	3318 (10)	3320 (10)
	2008	6 (15)	3319 (10)	3325 (10)
	2009	7 (18)	3377 (10)	3384 (10)
	2010	3 (8)	2869 (9)	2872 (9)
	2011	0	2409 (7)	2409 (7)
	2012	3 (8)	2158 (7)	2161 (7)
	2013	2 (5)	1931 (6)	1933 (6)
Gender	Male	23 (59)	15842 (48)	15865 (48)
	Female	16 (41)	16291 (50)	16307 (50)
	Unknown	0	652 (2)	652 (2)
Age group (years)	<5	17 (44)	9446 (29)	9463 (29)
	5-14	3 (8)	2960 (9)	2963 (9)
	15-24	2 (5)	1918 (6)	1920 (6)
	25-44	10 (26)	11714 (36)	11724 (36)
	45-64	4 (10)	4355 (13)	4359 (13)
	>64	2 (5)	1137 (3)	1139 (3)
	Unknown	1 (3)	1255 (4)	1256 (4)
Specimen type	Cerebrospinal fluid	11 (28)	11486 (35)	11497 (35)
	Blood	22 (56)	19284 (59)	19306 (59)
	Other[a]	6 (16)	2015 (6)	2021 (6)
PCV vaccination[b]	Yes[c]	2 (100)	580 (77)	582 (77)
	No	0	131 (17)	131 (17)
	Unknown	0	46 (6)	46 (6)
PCV13 serotypes[d]	PCV13 serotypes	13[e] (59)	23619 (72)	23632 (72)
	Non PCV13 serotypes	9[e] (41)	9166 (28)	9175 (28)
HIV status	Positive	5 (13)	7718 (23)	7723 (23)
	Negative	5 (13)	2489 (8)	2494 (8)
	Unknown	29 (74)	22578 (69)	22607 (69)
Immunocompromising conditions[f]	Yes	2 (5)	3114 (10)	3116 (10)
	No	13 (33)	6974 (21)	6987 (21)
	Unknown	24 (62)	22697 (69)	22721 (69)
Comorbid conditions[g]	Yes	2 (5)	1496 (5)	1498 (5)
	No	13 (33)	8592 (26)	8605 (26)
	Unknown	24 (62)	22697 (69)	22721 (69)

Outcome	Died	4 (10)	4024 (12)	4028 (12)
	Recovered	14 (36)	9887 (30)	9901 (30)
	Unknown	21 (54)	18874 (58)	18895 (58)

*Taken from reference (74)

[a]Other specimen types: pleural, joint and vitreous fluid.

[b]Only children born after February 2009 were included.

[c]Received at least 1 dose of the vaccine.

[d]PCV13 serotypes: 4, 6B, 9V, 14, 18C, 19F, 23F, 1, 5, 7F, 3, 6A and 19A.

[e]Nontypeable predicted ancestral serotypes for Group I isolates.

[f]Immunocompromising conditions were defined as medical record-documented pre-existing history of head injury, connective tissue disease, asplenia, pregnancy, premature birth, malignancy, burns, gastric acid suppression, aplastic anemia, organ transplant, primary immunodeficiency conditions, chromosomal conditions, protein energy malnutrition, alcohol dependency, current smoking, or immunosuppressive therapy.

[g]Comorbid conditions were defined as medical record-documented pre-existing history of pulmonary disease, renal disease, cerebrovascular accident, hepatic disease, cardiac disease, or diabetes mellitus.

Table 4*. Classification, predicted ancestral serotypes and *cps* mutations of invasive nontypeable *Streptococcus pneumoniae* in South Africa, 2003-2013 (N=39).

Group	Isolate number	Sequence type	Sequence type associated serotype(s)/serogroup[a]	*cps* locus BLAST	Ancestral serotype	Mutations in the *cps* locus when compared to ancestral serotypes	Comments
I	NT224	217	1	1	**1**	*rmlA, rmlB, rmlC, rmlD* and transposons present, all other *cps* genes deleted	
I	NT12	217	1	1	**1**		
I	NT17	217	1	1	**1**		
I	NT18	217	1	1	**1**		
I	NT45	8314	1	1	**1**		SLV of ST217
I	NT11	612	1	1	**1**		SLV of ST217
I	NT5	5604	25F	25F	25F	*glf, wzd, wze* and transposons present. All other *cps* genes deleted	SLV of ST105
I	NT1	53	8	8	8	Insertion in *wzx* and 13bp deletion in *wchA*	
I	NT3	53	8	8	8	SNP in *wchA*	
I	NT36	3847	8	8	8	Deletion of *wzg*, partial deletion of *wzh* and *wciS*. SNPs in *wciS*.	SLV of ST53
I	NT225	9817	None	18C	**18C**	SNP in *wze* and insertion in *wchA*	Novel ST is SLV of different STs mostly associated with 18C
I	NT6	9813	None	35B	35B	SNP in *wciB*	Novel ST has no SLV or DLV
I	NT40	989	12	12F	12F	2 bp deletion in *wciJ*	
I	NT34	2416	12F & others[b]	12F	12F	Partial deletions *of wciJ, mnaA,wzx* and insertion in *fnlA*	ST also associated with 7F, 19A
I	NT13	9818	None	19A	**19A**	SNPs in *wchA, wchO, wchQ, wzy* and *mnaA*. Partial deletion in *wchQ, wzy* and *wzx*	Novel ST SLV of ST4870,4873 (associated with serotype 19A)
I	NT10	6373	6A	6A	**6A**	SNPs in *wchA* and *wciP*. Partial deletion in *wciN*	
I	NT20	9814	None	6A	**6A**	SNPs in *wze, wchA, HG262, wciP,wzx.* Partial deletion of *wzx*	Novel ST SLV of ST8718 (associated with serotype 6A)
I	NT48	2208	24F	24F	24F	Deletion in *wcxK, rbsF* and *glf*. SNPs in *rbsF* and *wzx.*	
I	NT14	2652	14	14	**14**	Deletion of *wciY*. SNPs in *wchA, wch*J, *wchK, wchL, wchM*, and *wchN* and *Irp*.	
I	NT32	230	14 & others[b]	14	**14**	SNP in *wze.* Deletion in *wchK, wzy, wciY* and *Irp.*	ST also associated with 19A, 19F, 20, 23F, 24 & 24F
I	NT49	9816	None	23F	**23F**	Insertion in *wzg, wzy, wzx, gtp3, rmlB* and *wchF*	Novel ST DLV of ST353 (associated with serotype 23F)
I	NT30	9815	None	16F	16F	SNP in *wzy*. SNPs and deletions in *wzg*	Novel ST has no SLV or DLV
II-NCC 1	NT38	9812	None	ND	N/A	Absence of known *cps* genes. Only *pspK* gene	Novel ST SLV of ST9519 (associated with serotype 3)
II-NCC 2a	NT27	344	NT	ND	N/A	Absence of known *cps* genes (except *glf*)[c]. Only *aliC, aliD, ntaA* & 2 *ntaB*	
II-NCC 2a	NT2	344	NT	ND	N/A	Absence of known *cps* genes (except *glf*)[c]. Only *aliC, aliD, ntaA* & *ntaB* genes	

II-NCC 2a	NT39	344	NT	ND	N/A	Absence of known *cps* genes (except *glf*)[c]. Only *aliC, aliD, ntaA & ntaB* genes	Novel STs SLV of ST344 (associated with nonencapsulated *S. pneumoniae* strains)
II-NCC 2a	NT16	9809	None	ND	N/A		
II-NCC 2a	NT44	10239	None	ND	N/A		
II-NCC2b	NT47	10241	None	ND	N/A	Absence of known *cps* genes (except *glf*)[c]**.** Only *aliC* and *aliD* genes	
II-NCC 2a	NT46	9810	None	ND	N/A	Absence of known *cps* genes (except *glf*)[c]. Only *aliC, aliD, ntaA* & *ntaB* genes	Novel ST SLV of ST448 associated with nonencapsulated *S. pneumoniae* strains)
II-NCC 2a	NT50	9810	None	ND	N/A		
II-NCC 2b	NT4	10240	None	ND	N/A	Absence of known *cps* genes (except *glf*)[c]**.** Only *aliC* and *aliD* genes	
II-NCC 2a	NT15	9811	None	ND	N/A	Absence of known *cps* genes(except *glf*)[c]. Only *aliC, aliD, ntaA* & *ntaB* genes	Novel STs DLV of ST7252 (associated with serotype 38)
II-NCC 2a	NT31	9811	None	ND	N/A		
II-NCC 2a	NT42	9811	None	ND	N/A		
II-NCC 2a	NT43	9811	None	ND	N/A		
II-NCC 2a	NT29	9811	None	ND	N/A		
II-NCC 4	NT28	218	12F,7F	ND	NP	Deletion of all *cps* genes	
	NT33	105	25,38,NT	ND	NP	Deletion of all *cps* genes	

*Taken from reference (74)

[a]As per *Streptococcus pneumoniae* multilocus sequence typing database (http://pubmlst.org/spneumoniae/, accessed March 2015)

[b]Other serotype/s associated with that ST. See comments.

SLV = single-locus variant, differs at one of the seven alleles. DLV= Double-locus variant, differs at two of the seven alleles.

None = Novel ST and therefore no serotype/ST association exists.

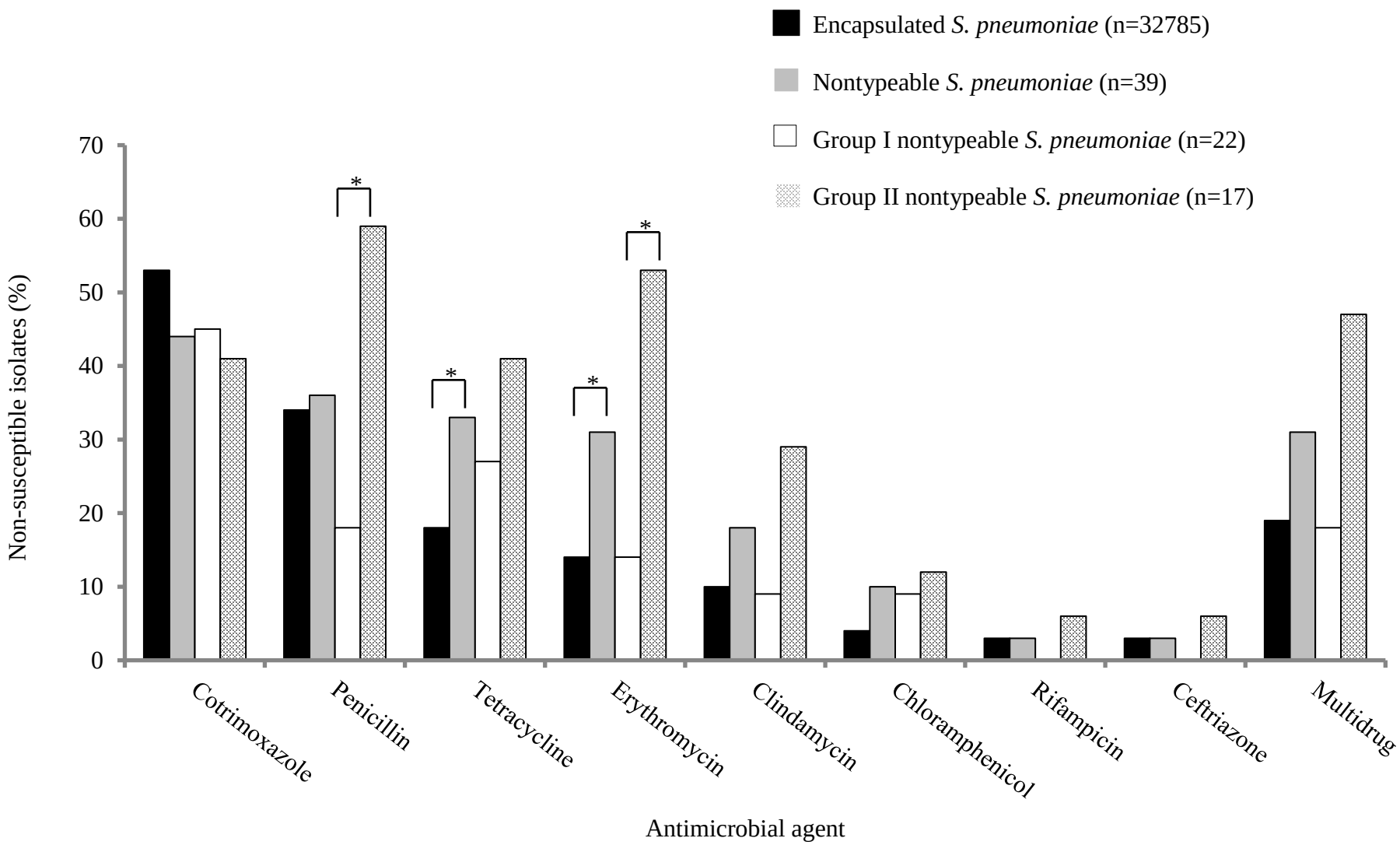
ND = not done, isolates have no *cps* genes.

N/A = Not applicable, resembles nontypeable pneumococci from carriage (no *cps* genes, but instead have typical nontypeable genes).

NP = not predicted. Isolates have no *cps* genes and/or are associated with more than 1 serotype.

Bold type represents serotypes included in PCV13.

[c]*glf* pseudo gene present within the *cps* locus.



* denotes statistically significant differences (P <0.05)

Figure 6. Antimicrobial non-susceptibility of *Streptococcus pneumoniae* isolates causing invasive disease in South Africa, 2003-2013. Taken from reference (74)

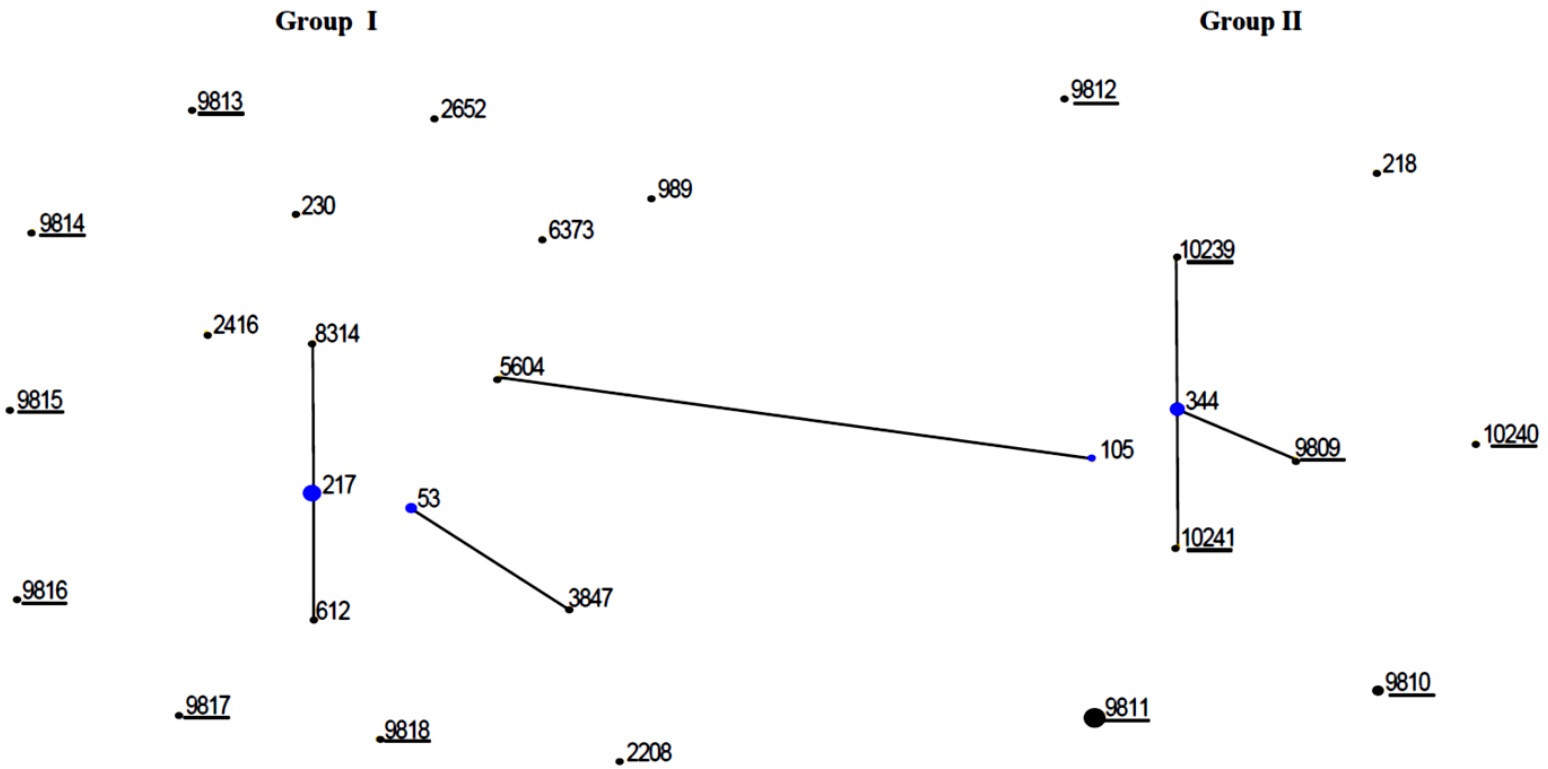


Figure 7. eBURST population snapshot showing relationships among sequence types (ST) identified in Group I (n=22) and Group II (n=17) invasive nontypeable *Streptococcus pneumoniae* in South Africa (2003-2013). The size of each circle corresponds with the number of isolates. Clusters of linked isolates correspond to clonal complexes (isolates sharing six of seven alleles) and blue indicates the founding genotype. New STs are underlined. Taken from reference (74)

3.1.5 Predicted ancestral serotypes of Group I isolates

The most common predicted ancestral serotypes were serotype 1 (n=6), 8 (n=3), 12F (n=2), 6A (n=2) and 14 (n=2). Fifty-nine percent (13/22) of the predicted ancestral serotypes were PCV13 serotypes 1 (n=6), 18C, 19A, 6A (n=2), 14 (n=2) and 23F (**Table 4**). Two of the invasive NTPn isolates (NT224 and NT225) were detected in two patients that were co-infected with a serotype 1 and serotype 18C isolate, respectively.

3.1.6 *cps* mutations in Group I isolates

Comparisons between Group I isolates and their ancestral serotype *cps* regions revealed a variety of mutations within the *cps* region of Group I isolates. These mutations are summarized in **Table 4**, described in detail in **Appendix D** and sketched in **Appendix E**. Briefly, predicted ancestral serotype 1 isolates had identical partial deletions of *cps* genes (NT11, NT12, NT17, NT18, NT45 and NT224). Isolate NT5 also had partial deletions. The remaining isolates had almost all *cps* genes present, as in their ancestral serotypes; however, we identified a variety of mutations (single nucleotide polymorphisms, insertions and deletions) in some of the *cps* genes. Most isolates had a combination of mutations. The vast majority of single nucleotide polymorphisms resulted in amino acid changes and a small number in stop codons. Base deletions and insertions were not common as single nucleotide polymorphisms.

3.1.7 Phylogenomic analysis

Ancestral serotype prediction was further confirmed by phylogenomic clustering of Group I isolates with representative genomes of their predicted ancestral serotypes (**Figure 8**). An exception was NT30 which was predicted to be derived from serotype 16F but was more closely related to serotype 11A. For some isolates, there was more divergence between the invasive NTPn and their ancestral serotypes as indicated by the long branches. Group II NCC1 (NT38), NCC4 (NT28 and NT33) and two carriage NTPn (MNZ37 and MNZ11b) from other studies (116) also clustered with invasive EcPn and appear to be related to serotypes 3, 7F, 25A and 15A, respectively. Group II NCC2 invasive NTPn belonging to ST9811 (n=5) and ST10240 formed a distinct clade separate from other Group II NCC2 isolates and seem to be more closely related to invasive EcPn. The remaining Group II NCC2 invasive NTPn (n=8) formed a distinct clade with carriage NTPn from other studies (n=4) (115;116). Overall, 31/39 (79.0%) of invasive NTPn from this study [Group I (n=22), Group II NCC1 (n=1), Group II NCC4 (n=2) and Group II NCC2 (n=6) isolates belonging to ST9811 (n=5) and ST10240] appear to have been derived from EcPn.

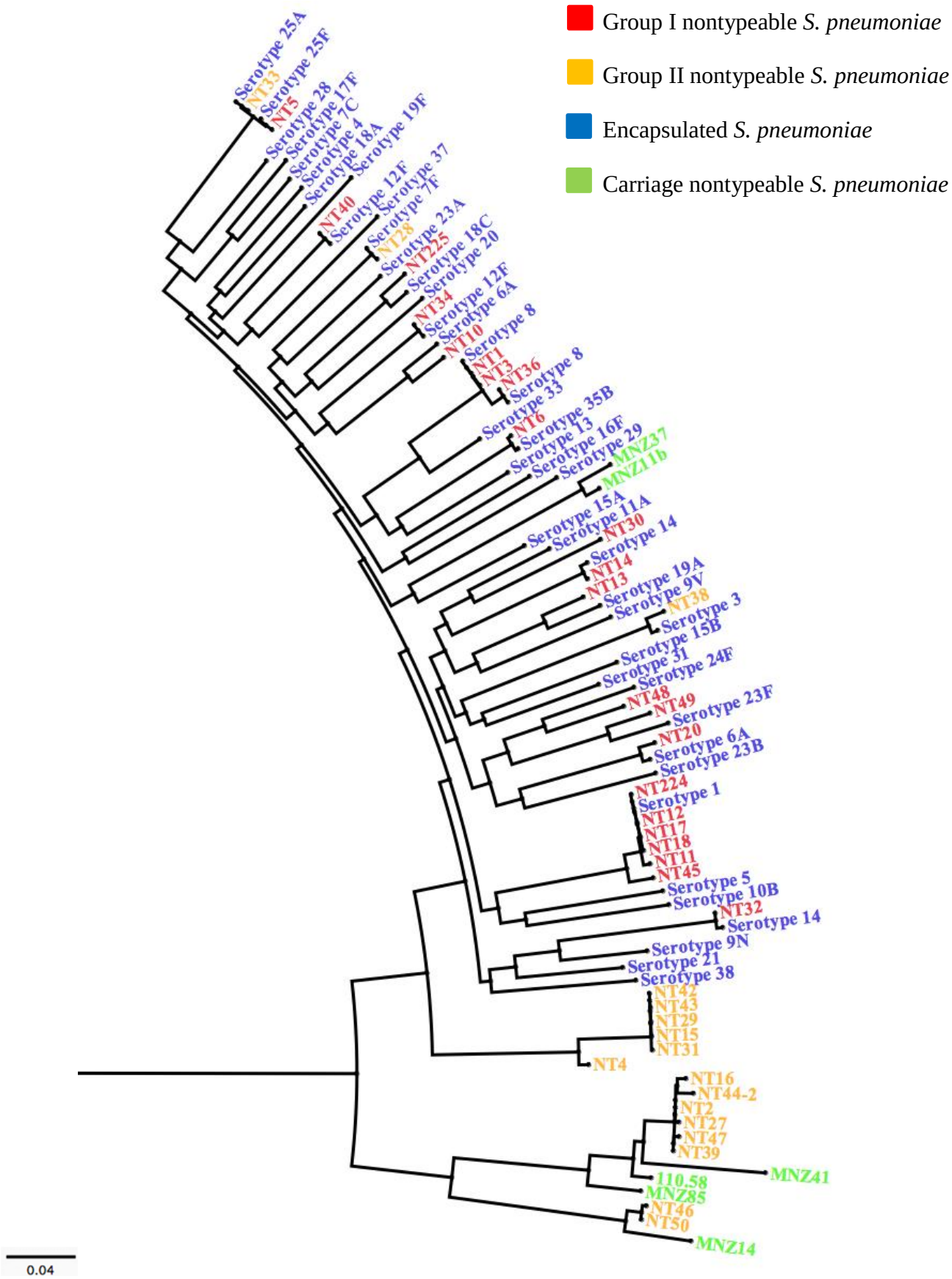

Group I nontypeable S. pneumoniae
Group II nontypeable S. pneumoniae
Encapsulated S. pneumoniae
Carriage nontypeable S. pneumoniae
Serotype 25A
NT33
Serotype 25F
NT5
Serotype 28
Serotype 17F
Serotype 7C
Serotype 4
Serotype 18A
Serotype 19F
NT40
Serotype 12F
Serotype 37
Serotype 7F
NT28
Serotype 23A
NT225
Serotype 18C
Serotype 20
NT34
Serotype 12F
Serotype 6A
NT10
Serotype 8
NT1
NT3
NT36
Serotype 8
Serotype 33
NT6
Serotype 35B
Serotype 13
Serotype 16F
Serotype 29
MNZ37
MNZ11b
Serotype 15A
Serotype 11A
NT30
Serotype 14
NT14
NT13
Serotype 19A
Serotype 9V
NT38
Serotype 3
Serotype 15B
Serotype 31
Serotype 24F
NT48
NT49
Serotype 23F
NT20
Serotype 6A
Serotype 23B
NT224
Serotype 1
NT12
NT17
NT18
NT11
NT45
Serotype 5
Serotype 10B
NT32
Serotype 14
Serotype 9N
Serotype 21
Serotype 38
NT42
NT43
NT29
NT15
NT31
NT4
NT16
NT44-2
NT2
NT27
NT47
NT39
MNZ41
110.58
MNZ85
NT46
NT50
MNZ14
0.04

Figure 8. Maximum likelihood phylogenetic tree based on core genome SNPs of invasive nontypeable *Streptococcus pneumoniae* (n=39), shown in red and yellow from South Africa (2003-2013). Encapsulated *S. pneumoniae* (representative of predicted ancestral serotypes) from South Africa (n=42, shown in blue) and publically available nonencapsulated *S. pneumoniae* genomes ((115;116) (n=6, indicated green) were included for comparison. The scale bar represents 0.04 nucleotide changes per 100 base pairs. Taken from reference (74)

3.2 Discussion

We characterised NTPn causing IPD in South Africa, a middle-income, high HIV prevalence country with PCV introduction in 2009. NTPn represented 0.1% of IPD cases, much lower than studies in the US, which have reported prevalence of between 0.6% and 3.0% (13;73). Comparison of NTPn prevalence is challenging due to the refinement of and greater sensitivity in Quellung over the years and, NTPn morphology that contrasts with EcPn and therefore making identification difficult. In our study and the studies from the US, NTPn isolates were retyped with the refined Quellung antisera. The high prevalence's in the US could represent true differences between the two countries, or may represent different lab practices resulting in different levels of sensitivity to detect NTPn. Our study showed that among patients with NTPn infection and for whom data were available, 44.0% were children < 5 years, proportionately more than EcPn (29.0%), 50.0% were HIV positive versus 76.0% for EcPn and the case-fatality ratio (22.0%) was similar to cases with EcPn (29.0%) (P = 0.71). Fifty-six percent of our invasive NTPn were Group I, for which the most common predicted ancestral serotypes were 1 (27.0%) and 8 (14.0%). Fifty-nine percent of the predicted ancestral serotypes were PCV13 serotypes. The *cps* loci of the Group I isolates harboured a variety of mutations. Non-susceptibility to tetracycline and erythromycin was significantly higher in invasive NTPn than EcPn. Non-susceptibility to penicillin and erythromycin was also significantly higher in Group II NTPn than Group I.

The mechanisms by which NTPn are able to cause IPD without the capsule are not clear. Immune suppression of the host and/or other lesser known or novel virulence factors might possibly explain the success of NTPn in causing IPD. In this study, the majority of patients with NTPn infection (87.0%, 13/15) were not immunocompromised and this was similar to patients with EcPn strains

(69.0%), however our study was not sufficiently powered to exclude the role of immune suppression.

We observed a higher prevalence (44.0%) of Group II isolates amongst our invasive NTPn that has not been described in other studies, where 4.0% to 10.0% of invasive NTPn were Group II isolates (13;73). The majority of Group II isolates (82.0%) were NCC2 and no isolates belonging to NCC3 were identified, a finding similar to invasive NTPn collected through the Active Bacterial Core surveillance program in the United States during 2006-2009 (73). In addition, two isolates were classified as NCC4, a newly defined clade in Group II, characterized by deletion of all *cps* genes (73). Similar to what has been shown in the US Active Bacterial Core surveillance, NCC4 isolates in this study also appeared to be derived from EcPn (73).

We were able to predict ancestral serotypes for all Group I isolates because they had partial *cps* genes, suggesting that these isolates were derived from EcPn which, at some point, lost their capsule through mutations in the *cps* locus. Indeed, we found a diverse range of mutations in the *cps* locus of Group I isolates, some of which may be responsible for the nontypeable phenotype. For predicted ancestral serotype 1 isolates, all *cps* mutations were identical. The deleted region is flanked by identical IS1167 repeats and this, together with the fact that the predicted ancestral serotype 1 isolates were genotypically related, indicates that the *cps* mutations probably occurred as a result of a single deletion event in a particular strain rather than multiple independent deletions in different strains. The same mutations have been described for serotype 1-derived invasive NTPn in other studies (13;75).

Serotype predictions were confirmed by the phylogenomic analyses as these isolates clustered with encapsulated strains expressing the same serotype as the predicted ancestral serotypes. One isolate (NT30), however, was more closely related to serotype 11A than its predicted serotype 16F. This isolate has a novel ST, which shares three of seven alleles with the 11A isolate, suggesting that NT30 may have been derived from a common ancestor with serotype 11A at some point in time. For some isolates, there was more divergence between the invasive NTPn and their ancestral serotypes. This could be that certain serotypes are inherently diverse while others are more stable.

In this study, NTPn that had a EcPn background were more likely to be invasive than NTPn that were from a classic NTPn lineage (79% versus 21%), therefore indicating that in the absence of the capsule, which has been shown to play a key role in the invasiveness of pneumococcal isolates, the genetic background appears to also play a role. This maybe due to either invasive NTPn with EcPn background producing low-level CPS that is not detectable by Quellung or they may be switching CPS production on/off. However, our data show in some of these isolates, there were complete deletions of *cps* genes making expression or switching on/off of a capsule impossible. The genetics factors responsible for this increased ability to invade should be further investigated.

It is difficult to say whether Group I isolates invaded as encapsulated or unencapsulated. One hypothesis that have been put forward is that Group I isolates invaded as encapsulated, and during infection as the host immune system respond by targeting the CPS, then some encapsulated pneumococci, but not all, lose the CPS. The encapsulated pneumococci would be rapidly opsonised, reducing complement levels, therefore allowing transient survival of unencapsulated pneumococci (73). This could explain why sometimes NTPn are detected simultaneously with encapsulated pneumococci in invasive disease (73;118). Alternatively, it could be that Group I

isolates lost the capsule in the nasopharynx, in which they would have high transformation efficiency in this state, therefore easily able to take up exogenous DNA (including virulence and resistance) that enables them to survive in invasive disease without the CPS. This could explain why there was more resistance among invasive NTPn compared to EcPn in our study.

Predicted ancestral serotypes 1 (27.0%) and 8 (14.0%) were most prevalent among our invasive NTPn isolates. Analysis of invasive NTPn isolates from Native American communities collected from 1994-2007 showed that predicted ancestral serotype 1 (19.0%) and 7F (15.0%) were most prevalent (13). In addition, serotype 8 (25.0%) was most prevalent among invasive NTPn from the US Active Bacterial Core surveillance (73). These data suggest that serotypes 1 and 8 may be more prone to mutations in their *cps* region than other serotypes.

Two of the invasive NTPn isolates (NT224 and NT225) were detected in mixed infections in two patients with encapsulated serotype 1 and 18C isolates, respectively. For the serotype 1 and NT224 mixed infection, *S. pneumoniae* was identified from the same culture, with the NT isolate representing approximately 98.0% of the culture (using the Quellung reaction). After several attempts to separate the two variants, only the invasive NTPn isolate could be obtained, however, RT-PCR confirmed the presence of the serotype 1 in the mixed culture. NT224 was confirmed to be ST217 which is associated with serotype 1. For the serotype 18C and NT225 mixed infection, both isolates were available and NT225 had the same *cps* region, with the exception of a single nucleotide polymorphism, compared to its encapsulated counterpart (serotype 18C), the same ST and their core genomes were identical (118).

Pneumococcal serotypes targeted by PCV have the potential to switch from vaccine serotype to non-vaccine serotype or to nontypeable, thereby providing a mechanism whereby they may be able to evade vaccine pressure. However, no evidence of adaptation to PCV occurred among invasive NTPn from Native American communities, where 45.0% of NTPn isolates had ancestral PCV7 serotypes pre-vaccine and none of their NTPn isolates were ancestral PCV7 serotypes post-vaccine (13). Although 59.0% of the predicted ancestral serotypes of invasive NTPn in our study were PCV13 serotypes, there is no evidence to suggest that serotype switching occurred as a result of PCV pressure as the vast majority (12/13, 92.0%) of invasive NTPn that are derived from PCV13 serotypes were already present prior to PCV13 introduction.

Antimicrobial non-susceptibility data for invasive NTPn are limited. In our study, there was a significantly higher prevalence of non-susceptibility among invasive NTPn than EcPn for erythromycin and tetracycline. These higher non-susceptibility rates coupled with higher transformation rates of NTPn (90) enable these strains to serve as a reservoir of antibiotic resistance genes. A serotype 19F clone from Switzerland became increasingly resistant to penicillin by acquisition of a *pbp*2x gene from NTPn (97). Recently, a comparative genome analysis of over 3000 carriage pneumococci from a refugee camp in Thailand showed that the highest rates of receipt and donation of recombinant of DNA occurred in carriage NTPn and that the most commonly exchanged genes were those associated with antibiotic resistance and immune interactions (23).

3.3 Conclusion

NTPn are currently a rare cause of IPD in South Africa and are genetically diverse but have a higher prevalence of antimicrobial resistance than EcPn. The majority of invasive NTPn were derived from EcPn, of which a significant proportion were PCV13 serotypes. Further studies which explore capsule-independent virulence mechanisms of invasive NTPn should be considered.

4 Characterisation of carriage nontypeable pneumococci in South Africa and comparison with invasive nontypeable isolates

4.1 Results

4.1.1 Prevalence of NTPn among carriage isolates

The overall prevalence of *S. pneumoniae* colonisation in the two study settings was 32.0% (3,721/11,739), of which 3.7% (137/3,721) were classified as NTPn. Amongst participants colonised with NTPn and for which demographic data was available, 61.0% (63/104) were female, 81.0% (110/135) were children aged <5 years and 44.0% (41/94) were HIV positive. The prevalence of NTPn among carriage isolates increased from 3.0% (60/2015) prior to routine PCV introduction to 4.5.0% (77/1706) (P = 0.02) post-PCV. Twenty-four percent (33/137) of carriage NTPn isolates were co-carried with EcPn. The most common co-colonising serotypes were serotype 19F (5/33, 15.0%), 6B (4/33, 12.0%), 6A (3/33, 9.0%), 23F (3/33, 9.0%) and 34 (3/33, 9.0%).

4.1.2 Carriage NTPn *cps* loci

Nine percent (13/137) of carriage isolates had conventional *cps* loci and were thus classified as Group I NTPn (**Table 5**). For 77.0% (10/13) Group I carriage isolates, all methods used to predict ancestral serotypes were in agreement. The most common predicted ancestral serotypes were serotypes 3 (4/13, 31.0%) and 13 (2/13, 15.0%). Compared to their ancestral serotypes, Group I isolates had mutations (mostly single nucleotide polymorphisms and deletions) within their *cps*

loci (**Table 5**). The remaining carriage isolates (124/137, 91.0%) were classified as Group II and had *pspK* (6/124, 5.0%), *aliC*, *aliD* and *ntaAB* (104/124, 84.0%), *aliC* and *aliD* (5/124, 5.0%), *aliD* (3/124, 2.0%), transposable elements (3/124, 2.0%) and *aliC* (3/124, 2.0%) in place of the *cps* locus. Based on the classification proposed by Park *et al* (12;73), these isolates were defined as NCC1 (6/124, 5.0%), NCC2a (104/124, 84.0%), NCC2b (5/124, 5.0%), NCC3 (3/124, 2.0%) and NCC4 (3/124, 2.0%). Three isolates (3/124, 2.0%) contained only *aliC* and did not belong to any clade using this classification system.

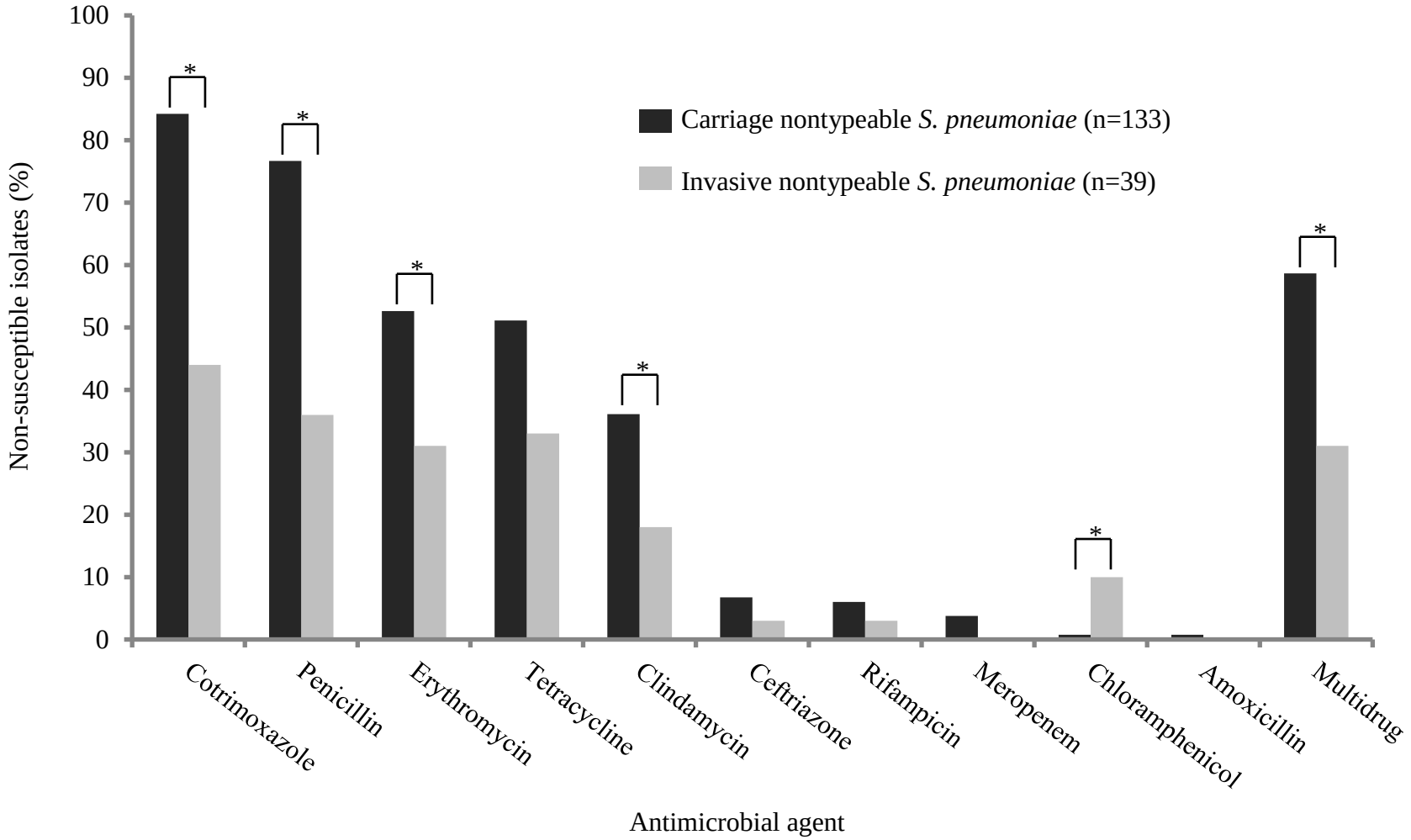
4.1.3 Antimicrobial non-susceptibility

Among carriage NTPn isolates, 84.0% (112/133) were non-susceptible to cotrimoxazole, 77.0% (102/133) to penicillin, 53.0% (70/133) to erythromycin, 51.0% (68/133) to tetracycline and 36.0% (48/133) to clindamycin (**Figure 9**). Antimicrobial susceptibilities could not be performed for four isolates due to lost viability during storage (however, genome sequencing had been done before this isolates lost viability). Overall, with the exception of chloramphenicol non-susceptibility, which was significantly higher in invasive NTPn isolates compared to carriage NTPn isolates [10.0% (4/39) versus 1.0% (1/133), $P = 0.01$], non-susceptibility was higher among carriage isolates than invasive NTPn for cotrimoxazole [84.0% (112/133) versus 44.0% (17/39), $P < 0.01$], penicillin [77.0% (102/133) versus 36.0% (14/39), $P < 0.01$], erythromycin [53.0% (70/133) versus 31.0% (12/39), $P = 0.02$] and clindamycin [36.0% (48/133) versus 18.0% (7/39), $P = 0.03$]. Multidrug non-susceptibility was also higher in carriage isolates compared to invasive NTPn isolates [59.0% (78/133) versus 31.0% (12/39), $P < 0.01$].

Table 5. Sequence types, capsular loci and ancestral serotypes of Group I carriage nontypeable *Streptococcus pneumoniae* (n=13) in South Africa, 2009-2012

Isolate number	Sequence type (ST)	ST associated serotype	Capsule locus similarity	Phylogenetic clustering	Predicted ancestral serotype	Mutations in the *cps* locus relative to ancestral serotypes	Comments
NT185	700	3	3	3	3	Deletion of *aliB*, *tnp*, *wzg*, *wzh*. Partial deletion of *wzd* & *wze*. SNPs in *wze*, *ugd*, *wchE* & *galU*. Insertion in *wze*.	
NT69	458	3	3	3	3	Partial deletion of *aliB* & *pgm*. SNPs in *wzh* & *aliB*.	
NT89	458	3	3	3	3	Partial deletion of *aliB*, *pgm* & *tnp*.	
NT205	458	3	3	3	3	Partial deletion of *pgm*.	
NT191	4872	19A	19A	19A	19A	Deletion of *IS1670*. Partial deletion of *IS1202*. SNPs in *IS1202*, *wzg*, *wze*, *wzx*, *rmlD*. Insertion in *IS1202*, *wzx* & *rmlD*.	
NT103	5410	4	19A	4	4	Deletion of *IS1202*. Partial deletion of *IS1670* & *rmlD*. SNPs in *wzg* & *IS1670*. Insertion in *rmlD*.	Co-carried with 19A. *Capsule compared to 19A
NT244	6277	19F	19F	19F	19F	Deletion of *aliB* & introns. Partial deletion & SNPs in *tnp*. Insertion in *wzx*.	Co-carried with 19F
NT134	217	1	1	1	1	Deletion of *aliB*. Partial deletion of *tnp* & *wzx*. SNPs in *tnp*, *wzd*, *wze* & *gla*.	
NT207	2285	6A	6A	6A	6A	SNPs in *wzd*. Partial deletion of *wzx*.	
NT171	5647	13	13	13	13	Deletion of *tnp*. Partial deletion of *whaG*, & *wzy*. SNPs in *wciF*.	
NT177	393	38	38	38	38	SNP in *wcil*. Partial deletion of *wzy*, *wcyD* & *gla*. Insertion in *wzy*.	
NT206	447	37	33F	37	37	SNPs in *wzg*, *wzh*, *wzd*, *wze*, *wchA*, *wciB*, *wzy*, *wciG*, *glf*, & *wcjE*	*Capsule compared to 33F
NT66	10938	None	10A	13	13	SNP in *aliB*. Partial deletion of *tnp*.	*Capsule compared to 10A

*Isolates had capsular genes not matching their predicted ancestral serotypes, maybe due to serotype switching (this is evident for isolate NT103, which was co-carried with serotype 19A and therefore likely took up the 19A capsule), hence the capsule locus was compared to the switched capsule and not the ancestral serotype locus.



* denotes statistically significant differences (P <0.05)

Figure 9. Antimicrobial non-susceptibility of carriage and invasive nontypeable *S. pneumoniae* from South Africa

4.1.4 Sequence type analysis of carriage NTPn and comparison to invasive NTPn

Overall, among the 137 carriage NTPn isolates, 44 STs were identified (of which 24 were novel STs) with a Simpson's diversity index of 0.92 (95.0% CI: 0.899 - 0.945). Group I isolates were mostly unrelated as only three isolates belonged to the same ST (ST458). (**Figure 10** and **Table 5**). Among Group II isolates, 33 STs were identified of which 23 were novel STs. Sequence types 9810 (25/124, 20.0%), 344 (21/124, 17.0%), 9809 (13/124, 10.0%), 9811 (13/124, 10.0%), 10972 (7/124, 6.0%), 9519 (6/124, 5.0%) and 7604 (5/124, 4.0%) were most common and represented 73.0% (90/124) of Group II isolates. Four CCs namely CC344 (49/124, 40.0%), CC9810 (33/24, 27.0%), CC10972 (21/124, 17.0%) and CC9519 (8/124, 6.0%) were identified and accounted for 90.0% (111/124) of Group II isolates. Sixty-six percent (82/124) of Group II isolates belonged to or were related (sharing at least 5/7 alleles) to international PMEN clones Norway^{NT}-ST344 and USA^{NT}-ST448 (http://www.pneumogen.net/pmen/). Six STs (ST217, ST344, ST9809, ST10241, ST9810 and ST9811) were identified in both carriage and invasive isolates. Four of these STs were identical (ST344) or related [ST9809 and ST10241 are single-locus variants (SLV) of ST344, and ST9810 is a SLV of ST448] to STs associated with established carriage NTPn lineages (94;101).

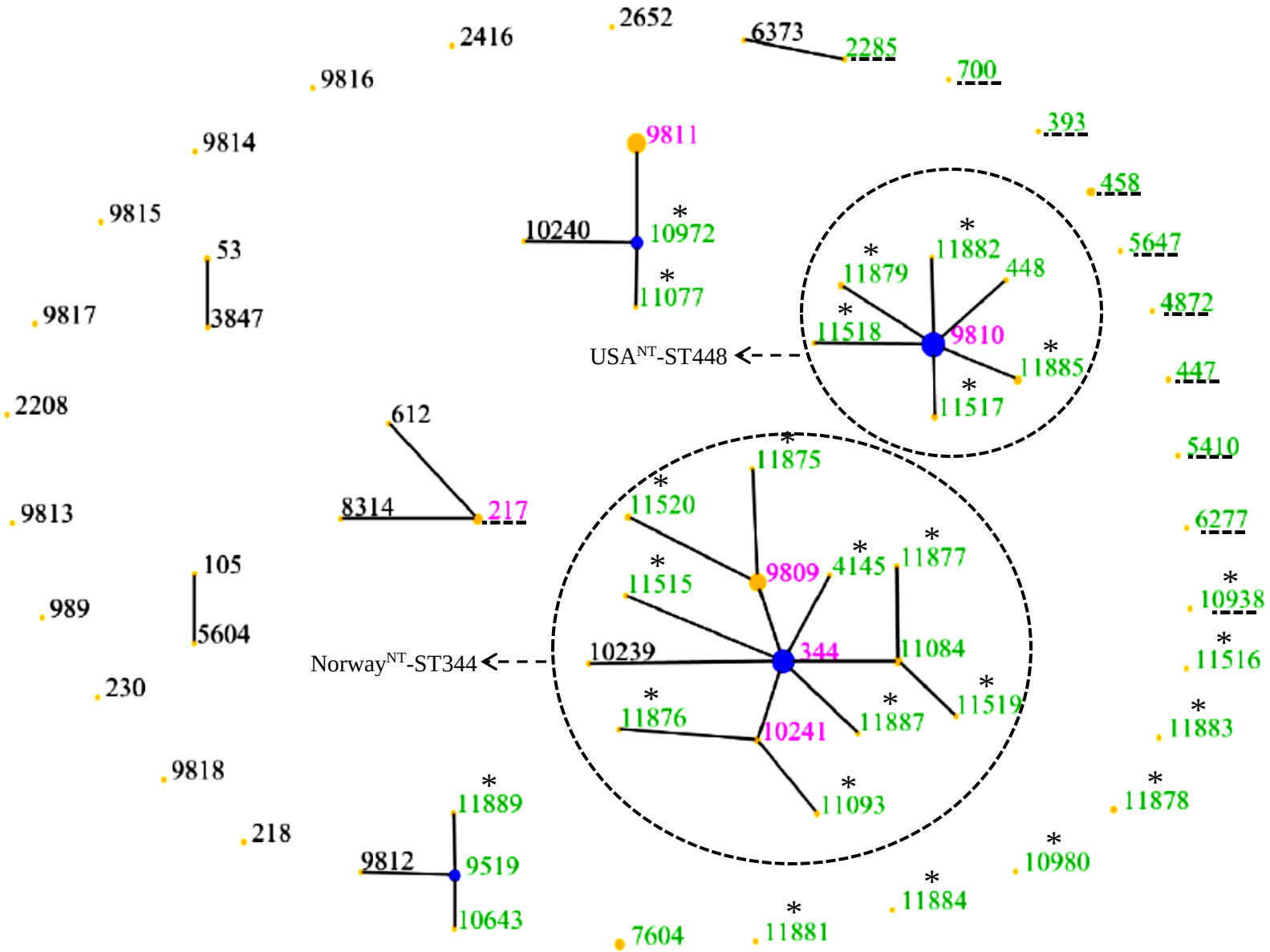

2652
2416
6373
2285
9816
700
393
9814
9811
458
9815
10240
10972
5647
53
11882
11879
448
11077
4872
9817
3847
11518
9810
USA^NT-ST448
11885
447
2208
11517
612
5410
11875
8314
217
11520
9813
6277
11877
105
9809
4145
11515
10938
989
5604
10239
344
11084
11516
Norway^NT-ST344
230
11519
11876
11887
11883
10241
9818
11889
11878
218
11093
10980
9812
9519
11884
10643
7604
11881

Figure 10. eBURST comparison of nontypeable *Streptococcus pneumoniae* (n=176) showing relationships between sequence types found in invasive (n=39) and carriage (n=137) isolates. Black sequence types were only found in invasive, green in carriage and pink in both. The size of each circle corresponds with the number of isolates. Clusters of linked isolates correspond to clonal complexes and blue indicates the founding genotype. * denotes new sequence types. Group I carriage isolates are underlined with a dashline.

4.1.5 Phylogenomic analysis of carriage and invasive NTPn

Invasive and carriage NTPn isolates clustered by ST and lineage (**Figure 11**). Within the classic lineage, two CCs, namely; CC344 and CC9810 were identified. These CCs comprised mainly of ST344, ST9809, ST448 and ST9810. Within CC344, ST344 from our setting clustered more closely with ST9809 South African isolates than ST344 isolates from other countries. Overall, 67.0% (92/137) of carriage and 23.0% (9/39) ($P < 0.01$) of invasive NTPn isolates from our setting belonged to the classic lineage. The remaining carriage (45/137, 33.0%) and invasive (30/39, 77.0%) NTPn isolates clustered with EcPn isolates in the sporadic lineage. Group I carriage isolates were closely related to their predicted ancestral serotypes (**Table 5**).

4.1.6 Putative virulence determinants of invasive NTPn isolates

We identified a total of 11,959 genes in the pan genome of invasive and carriage NTPn isolates, of which 4,777 encoded hypothetical proteins. The core genome for all NTPn isolates was 1,102 coding sequences, which translated to 55.0% of the total genome. Comparison of the gene content of invasive and carriage NTPn isolates revealed 1,298 genes that were present only in invasive isolates. Of these, 42.0% (541/1,298) encoded hypothetical proteins. These unique genes among invasive isolates ranged from 6 to 463 genes in one isolate, with an average of 42 genes per isolate. Two hundred and seventy-five genes were significantly associated with invasive NTPn isolates compared to carriage NTPn isolates (**Appendix F**), of which 59 genes were present in at least 70.0% of invasive NTPn. Of the 59 genes, 52 genes (88.0%) had BLAST matches to known proteins and were assigned to three gene ontology classes with 260 functional terms (**Figure 12**).

Assignments to biological process, molecular function and cellular component were 49.0% (127/260), 34.0% (89/260) and 17.0% (44/260), respectively. Within biological process, metabolic process was highly represented (27.0%, 14/52) and within molecular function, catalytic activity (39.0%, 20/52) and binding (27.0%, 14/52) were highly represented.

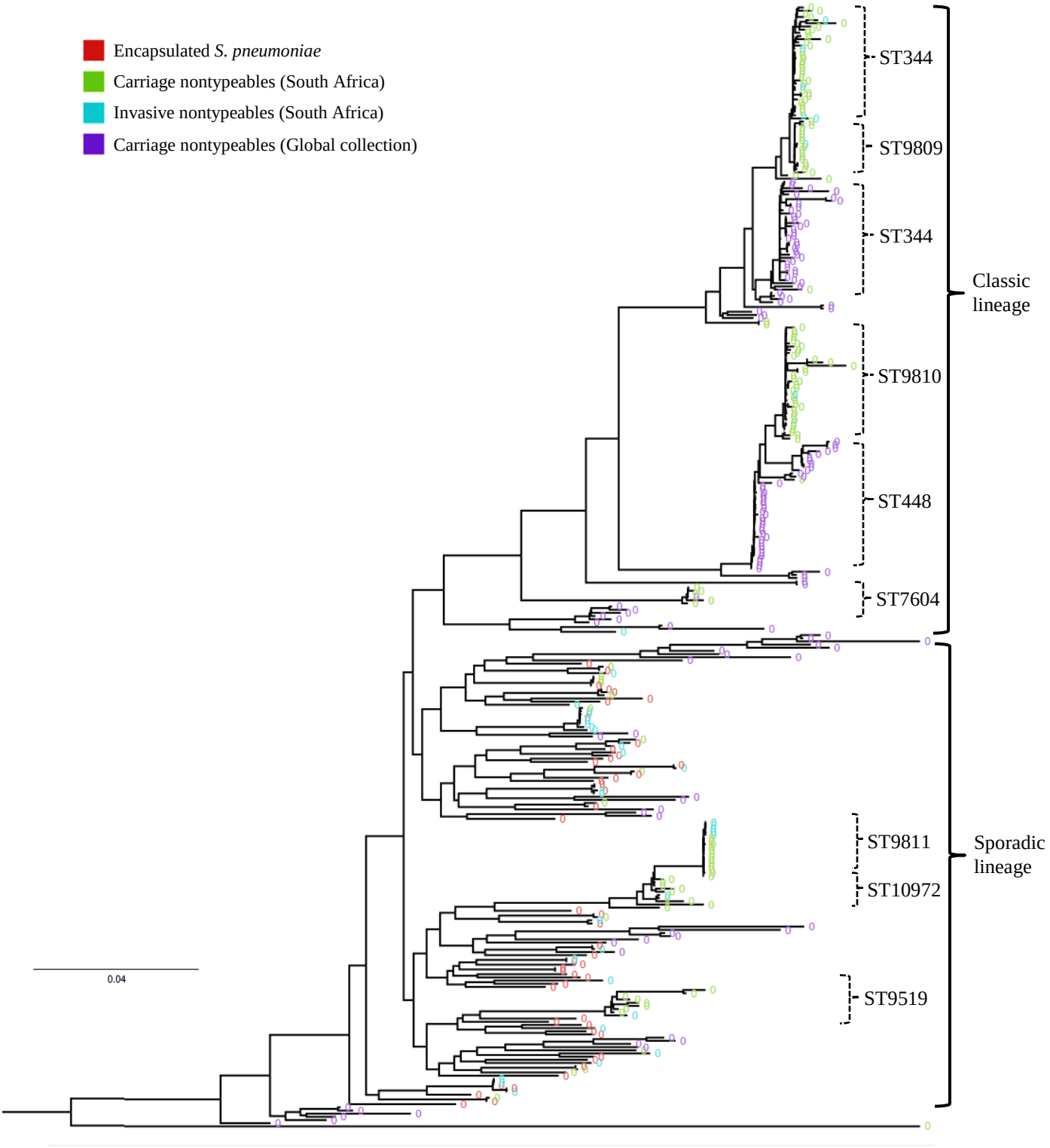

Encapsulated S. pneumoniae
Carriage nontypeables (South Africa)
Invasive nontypeables (South Africa)
Carriage nontypeables (Global collection)
ST344
ST9809
ST344
ST9810
ST448
ST7604
Classic lineage
ST9811
ST10972
ST9519
Sporadic lineage
0.04

Figure 11. Maximum likelihood phylogenetic tree of carriage (n=137) and invasive (n=39) nontypeable *S. pneumoniae* from South Africa and a global collection of carriage nontypeable isolates (n=131) representing 17 countries. Encapsulated *S. pneumoniae* isolates representative of different serotypes from South Africa (n=47) were included for comparison. The tree is based on core genome single nucleotide polymorphisms of the isolates. The scale bar represents 0.04 nucleotide changes per 100 base pairs

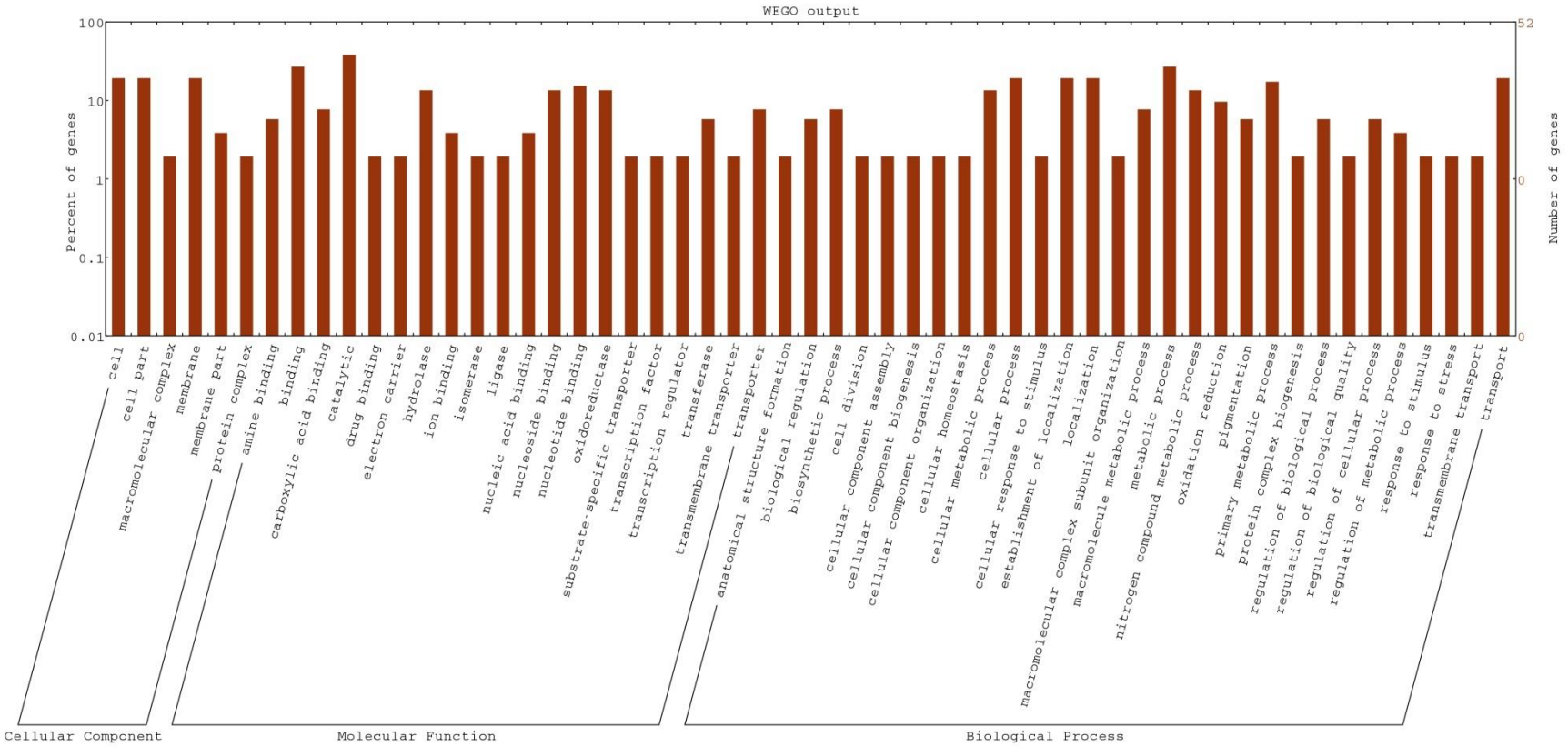


Figure 12. Gene ontology classification of 52 genes significantly associated with invasive nontypeable *Streptococcus pneumoniae* compared to carriage nontypeable strains. The genes were classified into three classes (biological process, molecular function and cellular component) using WEGO software**.**

4.2 Discussion

We identified substantial differences among carriage and invasive NTPn isolates, related to prevalence, antimicrobial resistance, *cps* composition, population structure and genetic composition. Non-susceptibility to cotrimoxazole, penicillin, erythromycin and clindamycin was significantly higher in carriage NTPn compared to invasive NTPn. Ninety-one percent of our carriage NTPn were Group II, of which 84.0% were NCC2a, whereas only 44.0% of invasive NTPn were Group II. Two-thirds of carriage NTPn were related to globally disseminated nontypeable multidrug-resistant clones Norway^{NT}-ST344 and USA^{NT}-ST448. The majority of carriage NTPn (67.0%) belonged to the classic lineage versus 23.0% of invasive NTPn. We identified genes that were exclusively found in invasive NTPn.

As expected, NTPn represented a higher proportion (3.7%) of carriage pneumococcal isolates than invasive isolates (0.1%) (P <0.01). Without the major virulence factor (capsule), invasive NTPn isolates are more likely to be cleared quickly by the host immune system (79), hence their prevalence tends to be lower than NTPn among carriage isolates. Other pneumococcal carriage studies have reported the prevalence of NTPn ranging from 3.9% to 19.0% (79). The prevalence of carriage NTPn isolates increased from 3.0% pre-PCV to 4.5% (P = 0.02) post-PCV in our setting. The increase is unlikely due to PCV selection pressure as the majority of our carriage NTPn isolates (67.0%) were from the classic lineage and therefore not derived from EcPn, but rather the availability of a niche in the nasopharynx left by elimination of PCV serotypes. In the Gambia, carriage NTPn prevalence increased from 0.3% to 6.0% (P <0.01) post-PCV and most isolates (88.0%) were derived from EcPn (Group I) (103). The authors concluded that the increase may be due to PCV selection pressure. In Portugal, carriage NTPn increased from 1.5% to 5.1% post-PCV (102).

There were significantly higher rates of antimicrobial non-susceptibility among carriage NTPn isolates than invasive NTPn for cotrimoxazole, penicillin, erythromycin and clindamycin. Because the majority of carriage NTPn isolates (91.0%) lacked CPS, they were more likely to be transformed with resistance genes compared to invasive NTPn isolates, where 56% were at some point encapsulated (less transformation efficiency in this state) and had lost the ability to express the capsule sporadically. This high resistance in carriage NTPn could also be driven by global multidrug-resistant nontypeable clones Norway^{NT}-ST344 and USA^{NT}-ST448 as two-thirds of our carriage NTPn were related to these clones. These high rates of antimicrobial non-susceptibility are common among carriage NTPn isolates (23;94;95). As co-colonisation of NTPn with EcPn occurs frequently and provides an opportunity for exchange of genetic material (93-95), NTPn can serve as reservoir of resistance genes. Evidence for this was a study in Switzerland that found a serotype 19F clone became resistant to penicillin by acquisition of a gene from NTPn (97). Recently, using whole genome sequencing, a study in Thailand showed that higher rates of DNA exchange occurred in NTPn compared to EcPn and the most commonly exchanged genes were those associated with antimicrobial resistance and immune interaction (23).

A small proportion (9.0%) of our carriage NTPn isolates had partial *cps* genes (for which we predicted their ancestral serotype) compared to our invasive NTPn isolates (56.0%). This proportion of carriage NTPn isolates that had partial *cps* genes is similar to a pneumococcal carriage study conducted in Thailand from 2007 – 2010, where only 9.0% (46/501) of carriage NTPn isolates had *cps* genes (23). The predicted ancestral serotypes among carriage isolates seems to be different to ancestral serotypes among invasive isolates, although ancestral serotypes 19A, 1 and 6A were identified in both carriage and invasive isolates. Three carriage NTPn isolates from our collection (NT103, NT206 and NT66, ancestral serotypes 4, 37 and

13, respectively) had capsular genes (19A, 33F and 10A, respectively) not matching their predicted ancestral serotypes. We hypothesise that there was a serotype switch and that during this event the isolates lost the ability to express the capsule. This is evident for NT103, which was co-carried with serotype 19A and therefore likely took up the 19A capsule.

We observed a high prevalence (91.0%) of Group II carriage isolates compared to our invasive NTPn isolates (44.0%). This high prevalence of Group II carriage isolates was similar to a study in Thailand, where Group II represented 91.0% (455/501) of carriage NTPn (23). However, NCC2a was the most prevalent (84.0%) among our Group II isolates compared to 35.0% (158/455) for Thailand isolates, where NCC1 (57.0%, 258/455) was most prevalent. In Portugal, Group II-NCC2a (61.0%, 8/13) was also most prevalent among representative carriage NTPn isolated from children between 1997 and 2007 (78). Group II-NCC2a has *aliC* and *aliD* genes which facilitate genetic transformation and mediate colonisation, respectively, compared to NCC1 isolates which has the *psk* gene and plays a role in adherence and colonisation. This might possibly explain the success of NCC2a, at least in our setting, but other factors may also play a role. Three of our Group II carriage isolates (2.0%) contained only *aliC* in the *cps* region and thus did not belong to any clade using the current classification (12;73). We propose that CPS composed of *aliC* be classified NCC5.

Compared to our invasive NTPn isolates which were more heterogeneous (D = 0.97), MLST analysis of carriage NTPn isolates showed that this population is slightly less diverse (D = 0.92). This may be due to the fact that majority (two-thirds) of carriage NTPn isolates were of the same clones (NorwayNT-ST344 and USANT-ST448. Six STs (ST217, ST344, ST9809, ST10241, ST9810 and ST9811), accounting for 41.0% of invasive and 54.0% of carriage isolates were shared between invasive and carriage NTPn isolates. Similarly, the population

structure of NTPn isolates from carriage studies conducted in Portugal in children aged <5 years between 2001 and 2007 was also clonal, with 83.0% (20/24) of NTPn isolates belonging to global clones Norway^{NT}-ST344 and USA^{NT}-ST448 (94).

Immune deficiencies of the host and/or other known or novel virulence factors have been proposed as explanations for the ability of NTPn to cause IPD without the capsule (71). However, in our study, the majority of patients with invasive NTPn infection (87.0%) were not immunocomprised, although the numbers are small. We identified genes that were only present in invasive NTPn, of which 42.0% encoded hypothetical proteins. Genes significantly associated with invasive NTPn (and present in at least 70.0% of isolates) were identified and half (49.0%) were assigned to biological process of which approximately one-third were involved in metabolic processes.

Because of significant differences in population structure of invasive and carriage NTPn, genes that are significantly associated with invasive NTPn could be genes that are significantly associated with the sporadic lineage versus the classical lineage, although this was not tested in this study. These unique genes may play an important role in aiding NTPn to cause IPD and should be further investigated. In addition, if these genes play an important role in virulence of NTPn, they are likely to also play a role in virulence in EcPn and therefore would be important genes that could be targeted for vaccine development.

Twenty-four percent of our NTPn isolates were co-carried with EcPn. The EcPn found in co-colonisation with NTPn were diverse, but serotypes 19F, 6B, 6A, 23F and 34 were most common. Consistent with our findings, other studies have often found NTPn as a co-coloniser (94;119). In a study from Portugal, serotypes 19A (14.0%), 14 and 19F (10.0% each) were

common co-colonisers with NTPn (94). One Group I carriage NTPn isolate (NT244), with a predicted 19F ancestral serotype co-colonised with a serotype 19F isolate. It is possible that NT244 is a variant of this 19F isolate, however we did not assess the relationship between co-colonizing pairs. Schaffner *et* al. described a serotype 18C carriage isolate that lost its capsule resulting in co-existence of encapsulated and unencapsulated variants of the same strain (72). We have also previously described two cases of IPD caused by co-infecting encapsulated and unencapsulated variants of serotype 1 and 18C (118).

4.3 Conclusion

We observed a higher prevalence and rate of antimicrobial non-susceptibility among carriage NTPn than invasive NTPn isolates. In contrast to our invasive NTPn isolates, the majority of our carriage NTPn isolates had genetically similar backgrounds belonging to a lineage exclusive to nontypeable strains and lacked conventional *cps* genes. We identified several genes which may play an important role in capsule-independent survival mechanisms of invasive NTPn isolates.

5 Overall conclusion

In this study, we characterised and compared invasive and carriage NTPn circulating in South Africa. NTPn isolates causing IPD differed from those detected in carriage. NTPn caused IPD at a low prevalence (0.1%), but were frequently isolated in carriage (3.7%). Rate of antimicrobial non-susceptibility was significantly higher in carriage NTPn than invasive NTPn. Invasive NTPn were predominantly Group I isolates (56%) and were related to EcPn (77%), whereas carriage NTPn were predominantly Group II isolates (91%) and were related to unencapsulated pneumococcal lineages (67%). The *cps* loci of invasive (56%) and carriage (9%) NTPn that had partial *cps* genes harboured a variety of mutations. Invasive NTPn isolates were more diverse (D = 0.97), whereas carriage NTPn isolates were slightly less diverse (D = 0.92). We identified genes that were exclusively found in invasive NTPn compared to carriage NTPn.

NTPn caused much lower IPD cases (0.1%) in South Africa compared to other settings (0.6% and 3%) (13;73). Similar to another study (13), we found no evidence of pneumococcal adaptation to PCV selection pressure, by capsule deletion, among invasive NTPn, although our study was limited in power. Almost half (44%) of invasive NTPn disease was in children <5 years and half were HIV positive, proportionately more than EcPn (29.0% and 76.0%, respectively). Only a small proportion (13%) of patients were immunocomprised, indicating that NTPn can cause invasive disease in immune competent individuals, although our study was not sufficiently powered to examine the association between NTPn infection and immunosuppression. The case-fatality ratio (22.0%) of invasive NTPn was similar to cases with EcPn (29.0%) (P = 0.71).

Similar to other pneumococcal carriage studies (79), NTPn was frequently detected in carriage (3.7%) and often (24%) as a co-coloniser with EcPn. The prevalence of carriage NTPn increased significantly post-PCV [3.0 to 4.5%, P = 0.02], but this was not due vaccine derived Group I NTPn which were rare in this collection of isolates, unlike other settings where carriage NTPn were predominantly vaccine derived Group I (103). The observed increase in our setting may reflect an indirect effect of PCV, due to elimination of PCV serotypes in the nasopharynx. Majority of participants colonised with NTPn were children aged <5 years (81%), female (61%) and almost half (44%) were HIV positive.

We identified similar trends of antimicrobial non-susceptibility among invasive and carriage NTPn. However, we found significantly higher rate of antimicrobial non-susceptibility among carriage NTPn than invasive NTPn isolates. This maybe due to the fact that some invasive NTPn (56%) were previously encapsulated and haven't had as many opportunities for transformation as the carriage NTPn isolates, where vast majority (91%) were unencapsulated. Because of the high rate of transformation of NTPn (90;91) and co-colonisation of the nasopharynx with EcPn (93;94;96), this provides an opportunity for horizontal gene transfer and may enable these NTPn to serve as a reservoir of antimicrobial resistance genes as previously described (23).

We predicted ancestral serotypes for invasive (56%) and carriage (9%) NTPn isolates that had partial *cps* genes, suggesting that these isolates were derived from EcPn which, at some point, lost their capsule through mutations in the *cps* locus. Indeed, we found a diverse range of mutations in the *cps* locus of these isolates, some of which may be responsible for the nontypeable phenotype. The predicted ancestral serotypes appear to be different among invasive and carriage NTPn isolates.

The population structure of carriage NTPn isolates were slightly less diverse (D = 0.92), with two-thirds of isolates representative of nontypeable multidrug-resistant global clones Norway^{NT}-ST344 and USA^{NT}-ST448 compared invasive NTPn (D = 0.97). Whole genome analysis revealed that the majority of invasive NTPn (77.0%) belonged to a lineage related to encapsulated strains that sporadically lost its ability to express a capsule, of which one-third were derived from PCV13 serotypes. In contrast, majority of carriage isolates (67%) belonged to a lineage exclusive to nontypeable strains. We compared the genomes of invasive and carriage NTPn isolates and identified 1,298 genes that were exclusively present in invasive NTPn, and 275 genes that were significantly associated with invasive NTPn. These unique genes may play an important role in aiding NTPn to cause IPD.

It is possible that our NTPn isolates may have lost the ability to express capsule during sub-culturing in the laboratory, but we believe this was not the case as we would expect NTPn to be detected more frequently. We did not perform *in vitro* experiments to confirm whether the mutations identified in the *cps* locus of our isolates were responsible for the lack of capsule expression. However, in some of these isolates, there were complete deletions of *cps* genes making expression of a capsule impossible. For EcPn isolates, different methods [agar dilution or Etest (2003-2008) or broth microdilution (2009-2013)] were used for MIC testing compared to invasive NTPn (broth microdilution). This may have overestimated the prevalence of β-lactam resistance in invasive NTPn for the 2003-2008 period as broth microdilution detects more resistance than agar dilution as was previously shown (109). We did not investigate the genes that were exclusively present in invasive NTPn compared to carriage NTPn to see if any of them may play a role in virulence. Co-colonisation and co-infection of NTPn with EcPn may have been underestimated, because the Quellung reaction requires serotyping of multiple colonies to detect co-colonisation/co-infection.

This study adds to growing evidence showing that capsule is not a prerequisite for colonisation and virulence and that pneumococcal strains lacking the capsule can frequently colonise the nasopharynx and occasionally cause IPD. Our study shows substantial differences among invasive and carriage NTPn related to epidemiology, antimicrobial resistance, capsule composition, population structure and genome content. Continued surveillance of NTPn is important since these isolates represent a reservoir of antibiotic resistance genes for EcPn and because the current pneumococcal vaccines are not effective against them. NTPn strains should be routinely characterised beyond serotyping to detect potential vaccine escapees. The genes identified as significantly associated with invasive NTPn in this study should be further investigated and their role in virulence determined.

6 Reference List

(1) O'Brien KL, Wolfson LJ, Watt JP, Henkle E, Deloria-Knoll M, McCall N, et al. Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. Lancet 2009 Sep 12;374(9693):893-902.

(2) Otsuka T, Kitami O, Kondo K, Ota H, Oshima S, Tsuchiya A, et al. Incidence survey of acute otitis media in children in Sado Island, Japan--Sado Otitis Media Study (SADOMS). PLoS One 2013 Jul 2;8(7):e68711.

(3) Musher DM. *Streptococcus pneumoniae*. Mandell, Douglas and Bennett's Principles and Practice of Infectious. 7th ed. Philadelphia: Elsevier Churchill Livingstone; 2010. p. 2623-42.

(4) Hyams C, Camberlein E, Cohen JM, Bax K, Brown JS. The *Streptococcus pneumoniae* capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. Infect Immun 2010 Feb;78(2):704-15.

(5) Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabbinowitsch E, Collins M, et al. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS Genet 2006 Mar;2(3):e31.

(6) Calix JJ, Porambo RJ, Brady AM, Larson TR, Yother J, Abeygunwardana C, et al. Biochemical, genetic, and serological characterization of two capsule subtypes among *Streptococcus pneumoniae* Serotype 20 strains: discovery of a new pneumococcal serotype. J Biol Chem 2012 Aug 10;287(33):27885-94.

(7) Calix JJ, Nahm MH. A new pneumococcal serotype, 11E, has a variably inactivated *wcjE* gene. J Infect Dis 2010 Jul 1;202(1):29-38.

(8) Park IH, Pritchard DG, Cartee R, Brandao A, Brandileone MC, Nahm MH. Discovery of a new capsular serotype (6C) within serogroup 6 of *Streptococcus pneumoniae*. J Clin Microbiol 2007 Apr;45(4):1225-33.

(9) Jin P, Kong F, Xiao M, Oftadeh S, Zhou F, Liu C, et al. First report of putative *Streptococcus pneumoniae* serotype 6D among nasopharyngeal isolates from Fijian children. J Infect Dis 2009 Nov 1;200(9):1375-80.

(10) Geno KA, Saad JS, Nahm MH. Discovery of Novel Pneumococcal Serotype 35D, a Natural *WciG*-Deficient Variant of Serotype 35B. J Clin Microbiol 2017 May;55(5):1416-25.

(11) Pneumococcal vaccines WHO position paper--2012. Wkly Epidemiol Rec 2012 Apr 6;87(14):129-44.

(12) Park IH, Kim KH, Andrade AL, Briles DE, McDaniel LS, Nahm MH. Nontypeable pneumococci can be divided into multiple *cps* types, including one type expressing the novel gene *pspK*. MBio 2012 Apr 24;3(3):e00035-12.

(13) Scott JR, Hinds J, Gould KA, Millar EV, Reid R, Santosham M, et al. Nontypeable pneumococcal isolates among navajo and white mountain apache communities: are these really a cause of invasive disease? J Infect Dis 2012 Jul 1;206(1):73-80.

(14) Smeall JT. A short history of the pneumococcus with special reference to lobar pneumonia. Edinb Med J 1948 Mar;55(3):129-41.

(15) Austrian R. The quellung reaction, a neglected microbiologic technique. Mt Sinai J Med 1976 Nov;43(6):699-709.

(16) Tettelin H, Nelson KE, Paulsen IT, Eisen JA, Read TD, Peterson S, et al. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. Science 2001 Jul;%20;293(5529):498-506.

(17) Hoskins J, Alborn WE, Jr., Arnold J, Blaszczak LC, Burgett S, DeHoff BS, et al. Genome of the bacterium *Streptococcus pneumoniae* strain R6. J Bacteriol 2001 Oct;183(19):5709-17.

(18) Donati C, Hiller NL, Tettelin H, Muzzi A, Croucher NJ, Angiuoli SV, et al. Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species. Genome Biol 2010;11(10):R107-R111.

(19) Jackson RW, Vinatzer B, Arnold DL, Dorus S, Murillo J. The influence of the accessory genome on bacterial pathogen evolution. Mob Genet Elements 2011 May;1(1):55-65.

(20) Griffith F. The Significance of Pneumococcal Types. J Hyg (Lond) 1928 J an;27(2):113-59.

(21) Avery OT, MacLeod CM, McCarty M. 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. 1944. Mol Med 1995 May;1(4):344-65.

(22) Evans BA, Rozen DE. Significant variation in transformation frequency in *Streptococcus pneumoniae*. ISME J 2013 Apr;7(4):791-9.

(23) Chewapreecha C, Harris SR, Croucher NJ, Turner C, Marttinen P, Cheng L, et al. Dense genomic sampling identifies highways of pneumococcal recombination. Nat Genet 2014 Mar;46(3):305-9.

(24) Wyres KL, Lambertsen LM, Croucher NJ, McGee L, von GA, Linares J, et al. Pneumococcal capsular switching: a historical perspective. J Infect Dis 2013 Feb 1;207(3):439-49.

(25) Hogberg L, Geli P, Ringberg H, Melander E, Lipsitch M, Ekdahl K. Age- and serogroup-related differences in observed durations of nasopharyngeal carriage of penicillin-resistant pneumococci. J Clin Microbiol 2007 Mar;45(3):948-52.

(26) Casey JR, Adlowitz DG, Pichichero ME. New patterns in the otopathogens causing acute otitis media six to eight years after introduction of pneumococcal conjugate vaccine. Pediatr Infect Dis J 2010 Apr;29(4):304-9.

(27) Gwaltney JM, Jr., Scheld WM, Sande MA, Sydnor A. The microbial etiology and antimicrobial therapy of adults with acute community-acquired sinusitis: a fifteen-year experience at the University of Virginia and review of other selected studies. J Allergy Clin Immunol 1992 Sep;90(3 Pt 2):457-61.

(28) World Health Organisation. World Health Organisation: Global Health Observatory data, child mortality and causes of death in 2015.

(29) Thigpen MC, Whitney CG, Messonnier NE, Zell ER, Lynfield R, Hadler JL, et al. Bacterial meningitis in the United States, 1998-2007. N Engl J Med 2011 May 26;364(21):2016-25.

(30) Werno AM, Murdoch DR. Medical microbiology: laboratory diagnosis of invasive pneumococcal disease. Clin Infect Dis 2008 Mar 15;46(6):926-32.

(31) Smith MD, Derrington P, Evans R, Creek M, Morris R, Dance DA, et al. Rapid diagnosis of bacteremic pneumococcal infections in adults by using the Binax NOW *Streptococcus pneumoniae* urinary antigen test: a prospective, controlled clinical evaluation. J Clin Microbiol 2003 Jul;41(7):2810-3.

(32) Murdoch DR, Laing RT, Mills GD, Karalus NC, Town GI, Mirrett S, et al. Evaluation of a rapid immunochromatographic test for detection of *Streptococcus pneumoniae* antigen in urine samples from adults with community-acquired pneumonia. J Clin Microbiol 2001 Oct;39(10):3495-8.

(33) Gutierrez F, Masia M, Rodriguez JC, Ayelo A, Soldan B, Cebrian L, et al. Evaluation of the immunochromatographic Binax NOW assay for detection of *Streptococcus pneumoniae* urinary antigen in a prospective study of community-acquired pneumonia in Spain. Clin Infect Dis 2003 Feb 1;36(3):286-92.

(34) Carvalho MG, Tondella ML, McCaustland K, Weidlich L, McGee L, Mayer LW, et al. Evaluation and improvement of real-time PCR assays targeting *lytA*, *ply*, and *psaA* genes for detection of pneumococcal DNA. J Clin Microbiol 2007 Aug;45(8):2460-6.

(35) Madhi SA, Petersen K, Madhi A, Wasas A, Klugman KP. Impact of human immunodeficiency virus type 1 on the disease spectrum of *Streptococcus pneumoniae* in South African children. Pediatr Infect Dis J 2000 Dec;19(12):1141-7.

(36) Diagnosis and management of acute otitis media. Pediatrics 2004 May;113(5):1451-65.

(37) Republic of South Africa. Essential Drugs Programme. Primary Healthcare Standard Treatment Guidelines and Essential Medicines List. 5th ed. RSA:NDOH;2014.

(38) Anon JB, Jacobs MR, Poole MD, Ambrose PG, Benninger MS, Hadley JA, et al. Antimicrobial treatment guidelines for acute bacterial rhinosinusitis. Otolaryngol Head Neck Surg 2004 Jan;130(1 Suppl):1-45.

(39) Rosenfeld RM, Piccirillo JF, Chandrasekhar SS, Brook I, Kumar KA, Kramper M, et al. Clinical practice guideline (update): Adult Sinusitis Executive Summary. Otolaryngol Head Neck Surg 2015 Apr;152(4):598-609.

(40) Tunkel AR, Hartman BJ, Kaplan SL, Kaufman BA, Roos KL, Scheld WM, et al. Practice guidelines for the management of bacterial meningitis. Clin Infect Dis 2004 Nov 1;39(9):1267-84.

(41) Bradley JS, Byington CL, Shah SS, Alverson B, Carter ER, Harrison C, et al. Executive summary: the management of community-acquired pneumonia in infants and children older than 3 months of age: clinical practice guidelines by the Pediatric Infectious Diseases Society and the Infectious Diseases Society of America. Clin Infect Dis 2011 Oct;53(7):617-30.

(42) Hansman D, Devitt L, Miles H, Riley I. Pneumococci relatively insensitive to penicillin in Australia and New Guinea. Med J Aust 1974 Sep 7;2(10):353-6.

(43) Jacobs MR, Koornhof HJ, Robins-Browne RM, Stevenson CM, Vermaak ZA, Freiman I, et al. Emergence of multiply resistant pneumococci. N Engl J Med 1978 Oct 5;299(14):735-40.

(44) Crowther-Gibson P, Cohen C, Klugman KP, de Gouveia L., von Gottberg A. Risk factors for multidrug-resistant invasive pneumococcal disease in South Africa, a setting with high HIV prevalence, in the prevaccine era from 2003 to 2008. Antimicrob Agents Chemother 2012 Oct;56(10):5088-95.

(45) Whitney CG, Farley MM, Hadler J, Harrison LH, Lexau C, Reingold A, et al. Increasing prevalence of multidrug-resistant *Streptococcus pneumoniae* in the United States. N Engl J Med 2000 Dec 28;343(26):1917-24.

(46) von Gottberg A., de Gouveia L., Tempia S., Quan V., Meiring S, von Mollendorf C., et al. Effects of vaccination on invasive pneumococcal disease in South Africa. N Engl J Med 2014 Nov 13;371(20):1889-99.

(47) 23-valent pneumococcal polysaccharide vaccine. WHO position paper. Wkly Epidemiol Rec 2008 Oct 17;83(42):373-84.

(48) Black S, Shinefield H, Fireman B, Lewis E, Ray P, Hansen JR, et al. Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children. Northern California Kaiser Permanente Vaccine Study Center Group. Pediatr Infect Dis J 2000 Mar;19(3):187-95.

(49) Harboe ZB, Valentiner-Branth P, Benfield TL, Christensen JJ, Andersen PH, Howitz M, et al. Early effectiveness of heptavalent conjugate pneumococcal vaccination on invasive pneumococcal disease after the introduction in the Danish Childhood Immunization Programme. Vaccine 2010 Mar;%19;28(14):2642-7.

(50) Pilishvili T, Lexau C, Farley MM, Hadler J, Harrison LH, Bennett NM, et al. Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. J Infect Dis 2010 Jan 1;201(1):32-41.

(51) Ruckinger S, van der Linden M, Reinert RR, von KR, Burckhardt F, Siedler A. Reduction in the incidence of invasive pneumococcal disease after general vaccination with 7-valent pneumococcal conjugate vaccine in Germany. Vaccine 2009 Jun 24;27(31):4136-41.

(52) Whitney CG, Farley MM, Hadler J, Harrison LH, Bennett NM, Lynfield R, et al. Decline in invasive pneumococcal disease after the introduction of protein-polysaccharide conjugate vaccine. N Engl J Med 2003 May 1;348(18):1737-46.

(53) Cohen R, Levy C, Bonnet E, Grondin S, Desvignes V, Lecuyer A, et al. Dynamic of pneumococcal nasopharyngeal carriage in children with acute otitis media following PCV7 introduction in France. Vaccine 2010 Aug 23;28(37):6114-21.

(54) Bettinger JA, Scheifele DW, Kellner JD, Halperin SA, Vaudry W, Law B, et al. The effect of routine vaccination on invasive pneumococcal infections in Canadian children, Immunization Monitoring Program, Active 2000-2007. Vaccine 2010 Feb 25;28(9):2130-6.

(55) Pilishvili T, Lexau C, Farley MM, Hadler J, Harrison LH, Bennett NM, et al. Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. J Infect Dis 2010 Jan 1;201(1):32-41.

(56) Hicks LA, Harrison LH, Flannery B, Hadler JL, Schaffner W, Craig AS, et al. Incidence of pneumococcal disease due to non-pneumococcal conjugate vaccine (PCV7) serotypes in the United States during the era of widespread PCV7 vaccination, 1998-2004. J Infect Dis 2007 Nov 1;196(9):1346-54.

(57) Lepoutre A, Varon E, Georges S, Dorleans F, Janoir C, Gutmann L, et al. Impact of the pneumococcal conjugate vaccines on invasive pneumococcal disease in France, 2001-2012. Vaccine 2015 Jan 3;33(2):359-66.

(58) Mackenzie GA, Hill PC, Sahito SM, Jeffries DJ, Hossain I, Bottomley C, et al. Impact of the introduction of pneumococcal conjugate vaccination on pneumonia in The Gambia: population-based surveillance and case-control studies. Lancet Infect Dis 2017 Sep;17(9):965-73.

(59) Bruce MG, Singleton R, Bulkow L, Rudolph K, Zulz T, Gounder P, et al. Impact of the 13-valent pneumococcal conjugate vaccine (pcv13) on invasive pneumococcal disease and carriage in Alaska. Vaccine 2015 Sep 11;33(38):4813-9.

(60) Kaplan SL, Barson WJ, Lin PL, Romero JR, Bradley JS, Tan TQ, et al. Early trends for invasive pneumococcal infections in children after the introduction of the 13-valent pneumococcal conjugate vaccine. Pediatr Infect Dis J 2013 Mar;32(3):203-7.

(61) Odutola A, Ota MO, Ogundare EO, Antonio M, Owiafe P, Worwui A, et al. Reactogenicity, safety and immunogenicity of a protein-based pneumococcal vaccine in Gambian children aged 2-4 years: A phase II randomized study. Hum Vaccin Immunother 2016;12(2):393-402.

(62) Odutola A, Ota MOC, Antonio M, Ogundare EO, Saidu Y, Foster-Nyarko E, et al. Efficacy of a novel, protein-based pneumococcal vaccine against nasopharyngeal carriage of Streptococcus pneumoniae in infants: A phase 2, randomized, controlled, observer-blind study. Vaccine 2017 May 2;35(19):2531-42.

(63) Prymula R, Pazdiora P, Traskine M, Ruggeberg JU, Borys D. Safety and immunogenicity of an investigational vaccine containing two common pneumococcal proteins in toddlers: a phase II randomized clinical trial. Vaccine 2014 May 23;32(25):3025-34.

(64) van der Poll T, Opal SM. Pathogenesis, treatment, and prevention of pneumococcal pneumonia. Lancet 2009 Oct 31;374(9700):1543-56.

(65) Kolkman MA, van der Zeijst BA, Nuijten PJ. Diversity of capsular polysaccharide synthesis gene clusters in *Streptococcus pneumoniae*. J Biochem 1998 May;123(5):937-45.

(66) Morona JK, Paton JC, Miller DC, Morona R. Tyrosine phosphorylation of *CpsD* negatively regulates capsular polysaccharide biosynthesis in streptococcus pneumoniae. Mol Microbiol 2000 Mar;35(6):1431-42.

(67) Morona JK, Miller DC, Morona R, Paton JC. The effect that mutations in the conserved capsular polysaccharide biosynthesis genes *cpsA*, *cpsB*, and *cpsD* have on virulence of *Streptococcus pneumoniae*. J Infect Dis 2004 May 15;189(10):1905-13.

(68) Mavroidi A, Aanensen DM, Godoy D, Skovsted IC, Kaltoft MS, Reeves PR, et al. Genetic relatedness of the *Streptococcus pneumoniae* capsular biosynthetic loci. J Bacteriol 2007 Nov;189(21):7841-55.

(69) Llull D, Munoz R, Lopez R, Garcia E. A single gene (*tts*) located outside the cap locus directs the formation of *Streptococcus pneumoniae* type 37 capsular polysaccharide. Type 37 pneumococci are natural, genetically binary strains. J Exp Med 1999 Jul;%19;190(2):241-51.

(70) Arrecubieta C, Lopez R, Garcia E. Type 3-specific synthase of *Streptococcus pneumoniae* (*Cap3B*) directs type 3 polysaccharide biosynthesis in *Escherichia coli* and in pneumococcal strains of different serotypes. J Exp Med 1996 Aug 1;184(2):449-55.

(71) Melchiorre S, Camilli R, Pietrantoni A, Moschioni M, Berti F, Del Grosso M, et al. Point mutations in *wchA* are responsible for the non-typability of two invasive *Streptococcus pneumoniae* isolates. Microbiology 2012 Feb;158(Pt 2):338-44.

(72) Schaffner TO, Hinds J, Gould KA, Wuthrich D, Bruggmann R, Kuffer M, et al. A point mutation in *cpsE* renders *Streptococcus pneumoniae* nonencapsulated and enhances its growth, adherence and competence. BMC Microbiol 2014 Aug 28;14:210. doi: 10.1186/s12866-014-0210-x.:210-0210.

(73) Park IH, Geno KA, Sherwood LK, Nahm MH, Beall B. Population-based analysis of invasive nontypeable pneumococci reveals that most have defective capsule synthesis genes. PLoS One 2014 May 15;9(5):e97825.

(74) Mohale T, Wolter N, Allam M, Ndlangisa K, Crowther-Gibson P, du Plessis M, et al. Genomic analysis of nontypeable pneumococci causing invasive pneumococcal disease in South Africa, 2003-2013. BMC Genomics 2016 Jun 22;17:470. doi: 10.1186/s12864-016-2808-x.:470-2808.

(75) Salter SJ, Hinds J, Gould KA, Lambertsen L, Hanage WP, Antonio M, et al. Variation at the capsule locus, *cps*, of mistyped and non-typable *Streptococcus pneumoniae* isolates. Microbiology 2012 Jun;158(Pt 6):1560-9.

(76) Waite RD, Penfold DW, Struthers JK, Dowson CG. Spontaneous sequence duplications within capsule genes *cap8E* and tts control phase variation in *Streptococcus pneumoniae* serotypes 8 and 37. Microbiology 2003 Feb;149(Pt 2):497-504.

(77) Hathaway LJ, Stutzmann MP, Battig P, Aebi S, Muhlemann K. A homologue of *aliB* is found in the capsule region of nonencapsulated *Streptococcus pneumoniae*. J Bacteriol 2004 Jun;186(12):3721-9.

(78) Tavares DA, Simoes AS, Bootsma HJ, Hermans PW, de Lencastre H., Sa-Leao R. Non-typeable pneumococci circulating in Portugal are of *cps* type NCC2 and have genomic features typical of encapsulated isolates. BMC Genomics 2014 Oct 6;15:863. doi: 10.1186/1471-2164-15-863.:863-15.

(79) Keller LE, Robinson DA, McDaniel LS. Nonencapsulated *Streptococcus pneumoniae*: Emergence and Pathogenesis. MBio 2016 Mar 22;7(2):e01792-15.

(80) Lagos R, Munoz A, San MO, Maldonado A, Hormazabal JC, Blackwelder WC, et al. Age- and serotype-specific pediatric invasive pneumococcal disease: insights from systematic surveillance in Santiago, Chile, 1994--2007. J Infect Dis 2008 Dec 15;198(12):1809-17.

(81) Scott JR, Millar EV, Lipsitch M, Moulton LH, Weatherholtz R, Perilla MJ, et al. Impact of more than a decade of pneumococcal conjugate vaccine use on carriage and invasive potential in Native American communities. J Infect Dis 2012 Jan 15;205(2):280-8.

(82) Onwubiko C, Shires C, Quin LR, Swiatlo E, McDaniel LS. Characterization of *Streptococcus pneumoniae* isolated from children with otitis media. FEMS Immunol Med Microbiol 2007 Jun;50(1):119-25.

(83) Xu Q, Kaur R, Casey JR, Sabharwal V, Pelton S, Pichichero ME. Nontypeable *Streptococcus pneumoniae* as an otopathogen. Diagn Microbiol Infect Dis 2011 Feb;69(2):200-4.

(84) Crum NF, Barrozo CP, Chapman FA, Ryan MA, Russell KL. An outbreak of conjunctivitis due to a novel unencapsulated *Streptococcus pneumoniae* among military trainees. Clin Infect Dis 2004 Oct 15;39(8):1148-54.

(85) Martin M, Turco JH, Zegans ME, Facklam RR, Sodha S, Elliott JA, et al. An outbreak of conjunctivitis due to atypical *Streptococcus pneumoniae*. N Engl J Med 2003 20;348(12):1112-21.

(86) Barker JH, Musher DM, Silberman R, Phan HM, Watson DA. Genetic relatedness among nontypeable pneumococci implicated in sporadic cases of conjunctivitis. J Clin Microbiol 1999 Dec;37(12):4039-41.

(87) Adamou JE, Wizemann TM, Barren P, Langermann S. Adherence of *Streptococcus pneumoniae* to human bronchial epithelial cells (BEAS-2B). Infect Immun 1998 Feb;66(2):820-2.

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

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

(90) Pearce BJ, Iannelli F, Pozzi G. Construction of new unencapsulated (rough) strains of *Streptococcus pneumoniae*. Res Microbiol 2002 May;153(4):243-7.

(91) Weiser JN, Kapoor M. Effect of intrastrain variation in the amount of capsular polysaccharide on genetic transformation of *Streptococcus pneumoniae*: implications for virulence studies of encapsulated strains. Infect Immun 1999 Jul;67(7):3690-2.

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

(93) Sa-Leao R, Simoes AS, Nunes S, Sousa NG, Frazao N, de Lencastre H. Identification, prevalence and population structure of non-typable *Streptococcus pneumoniae* in carriage samples isolated from preschoolers attending day-care centres. Microbiology 2006 Feb;152(Pt 2):367-76.

(94) Simoes AS, Valente C, de Lencastre H., Sa-Leao R. Rapid identification of noncapsulated *Streptococcus pneumoniae* in nasopharyngeal samples allowing detection of co-colonization and reevaluation of prevalence. Diagn Microbiol Infect Dis 2011 Nov;71(3):208-16.

(95) Andrade AL, Franco CM, Lamaro-Cardoso J, Andre MC, Oliveira LL, Kipnis A, et al. Non-typeable *Streptococcus pneumoniae* carriage isolates genetically similar to invasive and carriage isolates expressing capsular type 14 in Brazilian infants. J Infect 2010 Oct;61(4):314-22.

(96) Marsh R, Smith-Vaughan H, Hare KM, Binks M, Kong F, Warning J, et al. The nonserotypeable pneumococcus: phenotypic dynamics in the era of anticapsular vaccines. J Clin Microbiol 2010 Mar;48(3):831-5.

(97) Hauser C, Aebi S, Muhlemann K. An internationally spread clone of *Streptococcus pneumoniae* evolves from low-level to higher-level penicillin resistance by uptake of penicillin-binding protein gene fragments from nonencapsulated pneumococci. Antimicrob Agents Chemother 2004 Sep;48(9):3563-6.

(98) Keller LE, Jones CV, Thornton JA, Sanders ME, Swiatlo E, Nahm MH, et al. *PspK* of *Streptococcus pneumoniae* increases adherence to epithelial cells and enhances nasopharyngeal colonization. Infect Immun 2013 Jan;81(1):173-81.

(99) Hathaway LJ, Battig P, Reber S, Rotzetter JU, Aebi S, Hauser C, et al. *Streptococcus pneumoniae* detects and responds to foreign bacterial peptide fragments in its environment. Open Biol 2014 Apr 9;4:130224. doi: 10.1098/rsob.130224.:130224.

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

(101) Hanage WP, Kaijalainen T, Saukkoriipi A, Rickcord JL, Spratt BG. A successful, diverse disease-associated lineage of nontypeable pneumococci that has lost the capsular biosynthesis locus. J Clin Microbiol 2006 Mar;44(3):743-9.

(102) Sa-Leao R, Nunes S, Brito-Avo A, Frazao N, Simoes AS, Crisostomo MI, et al. Changes in pneumococcal serotypes and antibiotypes carried by vaccinated and unvaccinated day-care centre attendees in Portugal, a country with widespread use of the seven-valent pneumococcal conjugate vaccine. Clin Microbiol Infect 2009 Nov;15(11):1002-7.

(103) Roca A, Bojang A, Bottomley C, Gladstone RA, Adetifa JU, Egere U, et al. Effect on nasopharyngeal pneumococcal carriage of replacing PCV7 with PCV13 in the Expanded Programme of Immunization in The Gambia. Vaccine 2015 Dec 16;33(51):7144-51.

(104) Elberse KE, Nunes S, Sa-Leao R, van der Heide HG, Schouls LM. Multiple-locus variable number tandem repeat analysis for *Streptococcus pneumoniae*: comparison with PFGE and MLST. PLoS One 2011;6(5):e19668.

(105) Hall LM, Whiley RA, Duke B, George RC, Efstratiou A. Genetic relatedness within and between serotypes of *Streptococcus pneumoniae* from the United Kingdom: analysis of multilocus enzyme electrophoresis, pulsed-field gel electrophoresis, and antimicrobial resistance patterns. J Clin Microbiol 1996 Apr;34(4):853-9.

(106) Lefevre JC, Faucon G, Sicard AM, Gasc AM. DNA fingerprinting of *Streptococcus pneumoniae* strains by pulsed-field gel electrophoresis. J Clin Microbiol 1993 Oct;31(10):2724-8.

(107) van Cuyck H., Pichon B, Leroy P, Granger-Farbos A, Underwood A, Soullie B, et al. Multiple-locus variable-number tandem-repeat analysis of *Streptococcus pneumoniae* and comparison with multiple loci sequence typing. BMC Microbiol 2012 Oct 22;12:241. doi: 10.1186/1471-2180-12-241.:241-12.

(108) Huebner RE, Klugman KP, Matai U, Eggers R, Hussey G. Laboratory surveillance for *Haemophilus influenzae* type B meningococcal, and pneumococcal disease. Haemophilus Surveillance Working Group. S Afr Med J 1999 Sep;89(9):924-5.

(109) von Mollendorf C., Cohen C, de Gouveia L., Quan V, Meiring S, Feldman C, et al. Factors associated with ceftriaxone nonsusceptibility of *Streptococcus pneumoniae*: analysis of South African national surveillance data, 2003 to 2010. Antimicrob Agents Chemother 2014 Jun;58(6):3293-305.

(110) Clinical and Laboratory Standards Institute. Performance standards for antimicrobial susceptibility testing; approved standard M100-S24. 34. 2014. Clinical and Laboratory Standards Institute, Wayne, PA.

(111) Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 2004 Jul;14(7):1394-403.

(112) Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014 Jul 15;30(14):2068-9.

(113) Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, et al. Roary: Rapid large-scale prokaryote pan genome analysis. Bioinformatics 2015 Jul;20. pii: btv421.:btv421.

(114) Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 2006 Nov 1;22(21):2688-90.

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

(116) Keller LE, Thomas JC, Luo X, Nahm MH, McDaniel LS, Robinson DA. Draft Genome Sequences of Five Multilocus Sequence Types of Nonencapsulated *Streptococcus pneumoniae*. Genome Announc 2013 Jul 25;1(4):e00520-13.

(117) Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, et al. WEGO: a web tool for plotting GO annotations. Nucleic Acids Res 2006 Jul 1;34(Web Server issue):W293-W297.

(118) Ndlangisa KM, du Plessis M, Allam M, Wolter N, Mohale T, de Gouveia L, et al. Two cases of serotypeable and non-serotypeable variants of *Streptococcus pneumoniae* detected simultaneously during invasive disease. BMC Microbiol 2016 Jun 24;16(1):126-0745.

(119) Brugger SD, Frey P, Aebi S, Hinds J, Muhlemann K. Multiple colonization with *S. pneumoniae* before and after introduction of the seven-valent conjugated pneumococcal polysaccharide vaccine. PLoS One 2010 Jul 16;5(7):e11638.

7 Appendices

Appendix A. European nucleotide archive accession numbers for isolates of different serotypes used for capsule locus alignment

Serotype	Number of isolates	ENA accession numbers
18C	8	ERR600457, ERR646541, ERR646533, ERR714477 ERR714561, ERR773976, ERR774348, ERR1066295
35B	7	ERR774045, ERR774142, ERR1065850, ERR1065894 ERR1066252, ERR1066328, ERR1066289
12F	8	ERR773980, ERR774114, ERR774330, ERR774535 ERR1065797, ERR1065863, ERR1065868, ERR1065902
19A	112	ERR490718, ERR490719, ERR490735, ERR490736 ERR490751, ERR490757, ERR490760, ERR490779 ERR490797, ERR505914, ERR505928, ERR505940 ERR505897, ERR505953, ERR505959, ERR505977 ERR568443, ERR568453, ERR568461, ERR568466 ERR568514, ERR568578, ERR568638, ERR568643 ERR568658, ERR568666, ERR568672, ERR568700 ERR570422, ERR570444, ERR632880, ERR632891 ERR632912, ERR632922, ERR632932, ERR632933 ERR632960, ERR600327, ERR600317, ERR600374 ERR600430, ERR600468, ERR600511, ERR600542 ERR600554, ERR600579, ERR600603, ERR600618 ERR632980, ERR633024, ERR633034, ERR633045 ERR633085, ERR633100, ERR633117, ERR646549 ERR646551, ERR646565, ERR646581, ERR646613 ERR646615, ERR701797, ERR701816, ERR701859 ERR708300, ERR714479, ERR714449, ERR714586 ERR714672, ERR714689, ERR714638, ERR714699 ERR714703, ERR714722, ERR730550, ERR730563 ERR730581, ERR730629, ERR730675, ERR730681 ERR730685, ERR730704, ERR730721, ERR730672 ERR730814, ERR730816, ERR730817, ERR730853 ERR730871, ERR730887, ERR730892, ERR773861 ERR773899, ERR773909, ERR773927, ERR773932 ERR773964, ERR773977, ERR774055, ERR774102 ERR774158, ERR774179, ERR774248, ERR774300 ERR774343, ERR774522, ERR774534, ERR1065402 ERR1065453, ERR1065820, ERR1066290, ERR1066364
6A	42	ERR490755, ERR490784, ERR505912, ERR505913 ERR505916, ERR505934, ERR505935, ERR505956 ERR505960, ERR568436, ERR568490, ERR568498 ERR568499, ERR568507, ERR568547, ERR568549 ERR568564, ERR568629, ERR568677, ERR568746 ERR570376, ERR586461, ERR586472, ERR586464 ERR632886, ERR632926, ERR632941, ERR632962 ERR600490, ERR600557, ERR600564, ERR600597 ERR600627, ERR633017, ERR633078, ERR633091 ERR633031, ERR633097, ERR633111, ERR633115

		ERR646555, ERR646587, ERR701850, ERR708298 ERR714453, ERR714489, ERR714506, ERR714683 ERR714705, ERR730523, ERR730579, ERR730599 ERR730677, ERR730722, ERR730761, ERR730776 ERR730908, ERR730915, ERR773919, ERR773946 ERR773914, ERR773974, ERR774137, ERR1066299
24F	3	ERR714525, ERR909131, ERR913509
14	23	ERR568455, ERR568470, ERR570415, ERR600414 ERR646567, ERR646584, ERR646618, ERR701802 ERR701822, ERR701857, ERR714702, ERR730582 ERR730589, ERR730697, ERR730706, ERR730729 ERR730790, ERR730855, ERR730898, ERR730921 ERR773863, ERR773979, ERR774275
23F	83	ERR490721, ERR490708, ERR505894, ERR505923 ERR505895, ERR505942, ERR505986, ERR568460 ERR568486, ERR568491, ERR568495, ERR568528 ERR568532, ERR568580, ERR568525, ERR568599 ERR568607, ERR568669, ERR568676, ERR568687 ERR568697, ERR568701, ERR568714, ERR568724 ERR570400, ERR632872, ERR632918, ERR632874 ERR600337, ERR600370, ERR600432, ERR600450 ERR600467, ERR600500, ERR600545, ERR600549 ERR600612, ERR632994, ERR633020, ERR633026 ERR633054, ERR633066, ERR633075, ERR633079 ERR633083, ERR646534, ERR646597, ERR646601 ERR646612, ERR646616, ERR701799, ERR701800 ERR701803, ERR714452, ERR714465, ERR714472 ERR714534, ERR714555, ERR714562, ERR714565 ERR714583, ERR714614, ERR714657, ERR730561 ERR730698, ERR730674, ERR730801, ERR730818 ERR730822, ERR730945, ERR773816, ERR773866 ERR773893, ERR773936, ERR773972, ERR773918 ERR774245, ERR774277, ERR774356, ERR909326 ERR1065798, ERR1066238, ERR1066326
16F	26	ERR714655, ERR714688, ERR714720, ERR730755 ERR730902, ERR730863, ERR774012, ERR774021 ERR774052, ERR774154, ERR774185, ERR774282 ERR774298, ERR774324, ERR774350, ERR774368 ERR774541, ERR774567, ERR1065394, ERR1065829 ERR1065830, ERR1066206, ERR1066247, ERR1066301 ERR1066348, ERR1066350
8	52	ERR714655, ERR714688, ERR714720, ERR730755 ERR730902, ERR730863, ERR774012, ERR774021 ERR774052, ERR774154, ERR774185, ERR774282 ERR774298, ERR774324, ERR774350, ERR774368 ERR774541, ERR774567, ERR1065394, ERR1065829 ERR1065830, ERR1066206, ERR1066247, ERR1066301 ERR1066348, ERR1066350
3	4	ERR505918, ERR490741, ERR490761, ERR490710
19F	4	ERR490782, ERR490790, ERR505904, ERR490707
1	2	ERR568481, ERR568438
13	7	ERR568639, ERR568640, ERR570403, ERR600559 ERR633089, ERR730895, ERR773826
38	1	CR931710
33F	1	CR931702

10A	1	CR931649

Appendix B. Genbank and European nucleotide archive accession numbers for encapsulated *Streptococcus pneumoniae* isolates from South Africa used to construct phylogenetic trees

Serotype	**Sequence type**	**ENA/Genbank accession number**	**Serotype**	**Sequence type**	**ENA/Genbank accession number**
3	458	ERR490725	25	105	ERR646610
35B	9813	ERR490742	19A	4872	ERR646615
18C	1016	ERR490746	22F	445	ERR701804
4	5410	ERR490775	25F	5604	ERR714714
19F	6277	ERR490793	17F	8835	ERR730770
6C	9907	ERR505911	10A	10824	ERR730768
14	230	ERR505919	8	53	ERR773858
14	63	ERR505900	5	5659	ERR773933
23F	353	ERR568460	3	700	ERR773950
38	393	ERR568482	8	3847	ERR773978
10A	10685	ERR568501	6C	2185	ERR774124
12F	2416	ERR568565	7F	218	ERR774166
31	3548	ERR568674	23F	5647	ERR774234
9N	66	ERR570428	29	5483	ERR774552
18A	1232	ERR632904	6B	4084	ERR900706
23A	2319	ERR600334	6A	2285	ERR908907
11A	8817	ERR600395	24	5077	ERR909131
21	193	ERR600402	10B	7055	ERR909157
28	494	ERR600446	34	7067	ERR909213
6A	2285	ERR600502	16A	5072	ERR909284
20	8824	ERR600540	9V	4881	ERR913060
1	217	ERR600580	23B	7687	ERR913484
6C	2185	ERR633103	37	447	ERR925131
4	5410	ERR646545	7A	7053	ERR925258

Appendix C. Project ethics clearance certificate

R14/49 Mr Thabo Mohale et al

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M140450

NAME: (Principal Investigator)	Mr Thabo Mohale et al
DEPARTMENT:	Clinical Microbiology and Infectious Diseases National Institute for Communicable Diseases National Health Laboratory Service, Sandringham Centre for Respiratory Diseases and Meningitis
PROJECT TITLE:	Molecular Characterisation of Nontypeable Streptococcus Pneumoniae in South Africa
DATE CONSIDERED:	25/04/2014
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Dr Nicole Wolter
APPROVED BY:	Professor PE Cleaton-Jones, Chairperson, HREC (Medical)
DATE OF APPROVAL:	29/04/2014

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Secretary in Room 10004, 10th floor, Senate House, University.
I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. **I agree to submit a yearly progress report**.

Principal Investigator Signature Date

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Appendix D. Description of the capsular loci mutations of invasive nontypeable *Streptococcus pneumoniae* isolates (n=39) from South Africa, 2003 - 2013

Isolate	Gene	Mutation	Position of mutation	Nucleotide change	Amino acid change	Amino acid position	Comments
NT11	Not applicable	Partial deletion of *cps* genes	Not applicable	Not applicable	Not applicable	Not applicable	Only rmlABCD genes present
NT12	Not applicable	Partial deletion of *cps* genes	Not applicable	Not applicable	Not applicable	Not applicable	Only rmlABCD genes present
NT17	Not applicable	Partial deletion of *cps* genes	Not applicable	Not applicable	Not applicable	Not applicable	Only rmlABCD genes present.
NT18	Not applicable	Partial deletion of *cps* genes	Not applicable	Not applicable	Not applicable	Not applicable	Only rmlABCD genes present
NT45	Not applicable	Partial deletion of *cps* genes	Not applicable	Not applicable	Not applicable	Not applicable	Only rmlABCD genes present
NT224	wze	SNP	149	C→A	T→K	50	
	wchA	Insertion	472-473	Not applicable	Not applicable	Not applicable	Results in stop codon
NT5	Not applicable	Partial deletion of cps genes	Not applicable	Not applicable	Not applicable	Not applicable	Only *glf*, *wzd* and *wze* present
NT1	*wchA*	13bp deletion	405-417	Not applicable	Not applicable	Not applicable	
	wzx	9bp Insertion	1377-1385	Not applicable	Not applicable	Not applicable	
NT3	*wchA*	SNP	1186	C→T	R→C	395	
NT36	*wzg*	Deletion	whole gene	Not applicable	Not applicable	Not applicable	
	wzh	Deletion	32143	Not applicable	Not applicable	Not applicable	
	wciS	Deletion	353-379	Not applicable	Not applicable	Not applicable	
	wciS	Deletion	392-393	Not applicable	Not applicable	Not applicable	
	wciS	SNP	382	G→A	E→K	127	
	wciS	SNP	390	G→A	None	130	
	wciS	SNP	35	A→G	Stop codon	Not applicable	Results in stop codon
	wciS	SNP	396	T→A	V→L	132	
	wciS	SNP	397	G→T	V→L	132	
	wciS	SNP	399	G→A	V→L	133	
	wciS	SNP	401	T→G	I→S	134	
	wciT	SNP	632	G→A	R→K	211	

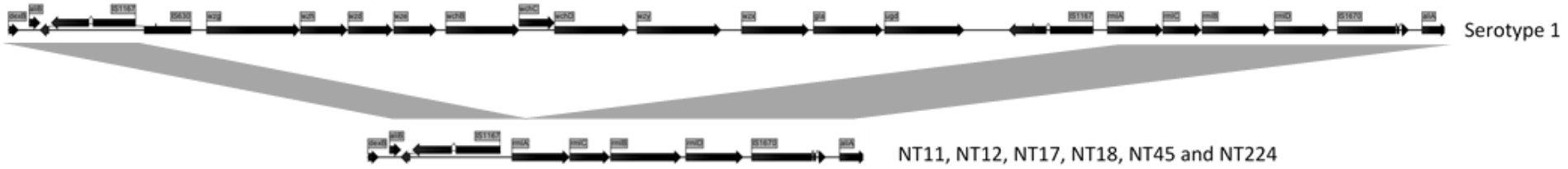
NT225	*wze*	SNP	149	C→A	T→K	50	
	wchA	Insertion	472-473	Not applicable	Stop codon	Not applicable	Results in stop codon
NT6	*wciB*	SNP	274	G→T	G→W	91	
NT40	*wciJ*	Deletion	183-184	Not applicable	Deletion	Not applicable	Results in deletion of Y
NT34	*wciJ*	Deletion	209	Not applicable	Not applicable	Not applicable	
	mnaA	Deletion	424-492	Not applicable	Not applicable	Not applicable	
	wzx	Deletion	522-755	Not applicable	Not applicable	Not applicable	
	fnlA	Insertion	175-196	Not applicable	Not applicable	Not applicable	
NT13	*wchA*	SNP	1216	C→T	Stop codon	Not applicable	Results in stop codon
	wchO	SNP	352	A→C	M→L	117	
	wchQ	SNP	338	A→G	F-G	113	
	wchQ	Deletion	797-879	Not applicable	Not applicable	Not applicable	
	wzy	Deletion	1-350	Not applicable	Not applicable	Not applicable	
	wzy	SNP	1284	C→A	L→I	428	
	wzy	SNP	1304	G→A	Stop codon	Not applicable	Results in stop codon
	wzy	SNP	1357	T→G	I→S	452	
	wzx	Deletion	1-123	Not applicable	Not applicable	Not applicable	
	wzx	SNP	194	G→T	G→V	65	
	wzx	SNP	1212	A→G	None	Not applicable	No amino acid change
	mnaA	SNP	151	T→G	F→V	50	
	mnaA	SNP	153	T→A	F→V	50	
	mnaA	SNP	215	C→A	T→K	72	
NT10	*wchA*	SNP	1202	C→G	P→R	401	
	wciN	Deletion	721-7711	Not applicable	Not applicable	Not applicable	
	wciP	SNP	226	T→C	F→L	75	
	wciP	SNP	781	T→A	Y→N	260	
	wciP	SNP	902	G→A	R→Q	301	
NT20	*wze*	SNP	651	T→C	Y→H	325	
	wchA	SNP	1130	G→T	G→V	565	
	HG262	SNP	90	T→C	None	Not applicable	No amino acid change

	wciP	SNP	392	C→G	A→G	131	
	wciP	SNP	402	T→G	N→K	134	
	wciP	SNP	411	T→A	D→E	137	
	wzx	Deletion	302-338	Not applicable	Not applicable	Not applicable	
	wzx	Deletion	1354-1413	Not applicable	Not applicable	Not applicable	
	wzx	SNP	298	T→G	F→V	99	
	wzx	SNP	341	A→G	Y→C	114	
	wzx	SNP	377	C→A	T→K	126	
	wzx	SNP	416	C→A	T→K	139	
	wzx	SNP	443	T→A	I→N	148	
	wzx	SNP	491	T→A	I→N	164	
NT48	*wcxK*	Deletion	294-299	Not applicable	Not applicable	Not applicable	
	rbsF	SNP	371	A→T	None	Not applicable	No amino acid change
	rbsF	Deletion	988	410-1397	Not applicable	Not applicable	
	rbsF	SNP	1975	G→T	A→S	658	
	rbsF	SNP	2081	G→T	G→W	694	
	rbsF	SNP	1982	G→T	G→V	660	
	rbsF	SNP	2064	C→A	Q→K	688	
	rbsF	SNP	2068	T→A	I→K	689	
	wzx	SNP	3	G→T	M→I	1	
	wzx	SNP	5	G→T	S→I	2	
	wzx	SNP	6	C→T	S→I	2	
	glf	Deletion	208-1154	Not applicable	Not applicable	Not applicable	
NT14	*wchA*	Deletion	277-278	Not applicable	Not applicable	Not applicable	
	wchA	SNP	405	A→C	L→F	135	
	wchA	SNP	968	T→A	V→A	323	
	wchJ	Deletion	217-450	Not applicable	Not applicable	Not applicable	
	wchK	Deletion	127	Not applicable	Not applicable	Not applicable	
	wchK	SNP	154	T→G	L→V	51	
	wchL	SNP	815	A→C	Y→S	272	
	wchL	SNP	829	A→C	K→Q	277	
	wchL	SNP	833	G→C	W→S	278	
	wchL	SNP	834	G→A	W→S	278	
	wchL	SNP	844	G→T	A→S	281	
	wchL	SNP	895	G→T	Stop codon	Not applicable	Results in stop codon
	wchM	SNP	198	A→C	E→D	66	

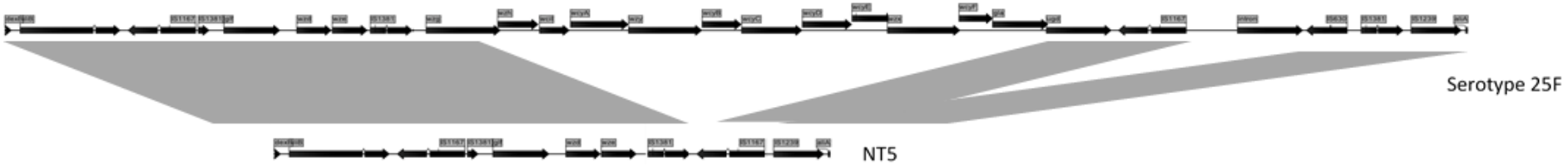
	wciY	Deletion	whole gene	Not applicable	Not applicable	Not applicable	
	wchN	SNP	413	G→T	G→V	138	
	wchN	SNP	433	T→G	L→V	144	
	wchN	SNP	457	T→G	L→V	152	
	wzx	Deletion	1364-14164	Not applicable	Not applicable	Not applicable	
	Irp	SNPs and deletions		Not applicable	Not specified	Not applicable	Many SNPs
NT32	*wze*	SNP	430	A→C	K→Q	143	
	wchK	Deletion	404-504	Not applicable	Not applicable	Not applicable	
	wzy	Deletion	1-209	Not applicable	Not applicable	Not applicable	
	wciY	Deletion	983-1040	Not applicable	Not applicable	Not applicable	
	Irp	SNPs and deletions	Not applicable	Not applicable	Not specified	Not applicable	Many SNPs
NT49	*wzg*	Insertion	999-1018	Not applicable	Not applicable	Not applicable	
	wzy	Insertion	995-1015	Not applicable	Not applicable	Not applicable	
	wzx	Insertion	1261-1264	Not applicable	Not applicable	Not applicable	
	gtp3	Insertion	171-172	Not applicable	Not applicable	Not applicable	
	gtp4	Insertion	202-216	Not applicable	Not applicable	Not applicable	
	rmlB	Insertion	47392	Not applicable	Not applicable	Not applicable	
	wchF	SNP	Not applicable	Not applicable	Not specified	Not applicable	Many SNPs
NT30	*wzy*	SNP	42	A→G	D→G	14	
	wzg	SNPs and deletions	Not applicable	Not applicable	Not specified	Not applicable	Many SNPs

Appendix E. Diagrammatic representation of the capsular locus of invasive nontypeable *Streptococcus pneumoniae* (n=39) from South Africa, 2003-2013. Nucleotide identity is indicated by grey boxes; single nucleotide polymorphisms by horizontal lines; insertions by down arrows and deletions by white boxes with dashed lines. A detailed description of the mutations (position, nucleotide and amino acid changes) is shown in additional file 3. (a) Isolates NT11, NT12, NT17, NT18, NT45 and NT224 (bottom) were compared to their predicted ancestral serotype 1 (top) (b) NT5 (bottom) was compared to its predicted ancestral serotype 25F (top) (c) NT1 (d) NT3 (e) NT36 (bottom) was compared to its ancestral serotype 8 (top) (f) NT225 (g) NT6 (h) NT40 (i) NT34 (j) NT13 (k) NT10 (l) NT20 (m) NT48 (n) NT14 (o) NT32 (p) NT49 (q) NT30 (r) NT38 (s) NT2, NT15, NT16, NT29, NT31, NT39, NT42, NT43, NT44, NT46 and NT50 (t) NT4 and NT47 (u) NT27 (v) NT28 (w) NT33

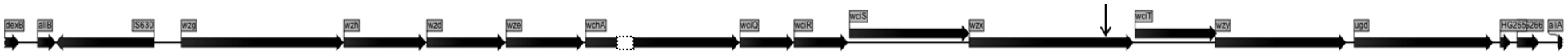
(a) Serotype 1 reference strain and NT11, NT12, NT17, NT18, NT45 NT224



(b) Serotype 25F reference strain and NT5

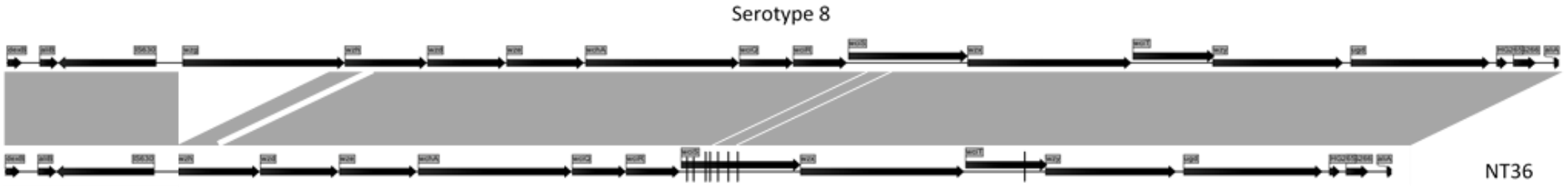


(c) NT1



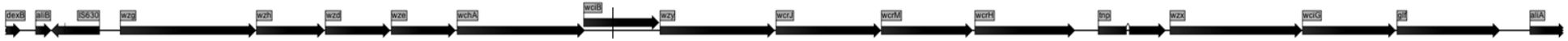
(d) NT3

(e) Serotype 8 reference strain and NT36



(f) NT225

(g) NT6



(h) NT40

(i) NT34

(j) NT13

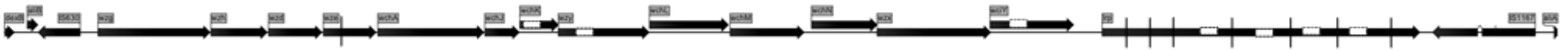
(k) NT10

(l) NT20

(m) NT48

(n) NT14

(o) NT32



(p) NT49

(q) NT30

(r) NT38

(s) NT2, NT15, NT16, NT29, NT31, NT39, NT42, NT43, NT44, NT46 and NT50

(t) NT4, NT47

(u) NT27

(v) NT28

(w) NT33

Appendix F. Genes that showed significant differences in prevalence between invasive and carriage nontypeable pneumococci isolates

Gene	Annotation	Invasive isolates (N=39)		Carriage isolates (N=137)		*P*-value
		n	%	n	%	
potB	Spermidine/putrescine transport system permease protein	39	100	87	63.5	0.01
xpsE	Type II secretion system protein E	39	100	86	62.77	0.01
yesO_1	Putative ABC transporter substrate-binding protein	39	100	75	54.74	<0.001
group_9231	Phosphate transporter permease subunit PtsA	39	100	43	31.39	<0.001
group_7116	DNA mismatch repair protein MutL	39	100	38	27.74	<0.001
group_5092	ABC-type transporter ATP-binding protein EcsA	39	100	33	24.09	<0.001
group_6543	Shikimate dehydrogenase	39	100	19	13.87	<0.001
group_8529	Isopentenyl-diphosphate delta-isomerase	39	100	5	3.65	<0.001
coaE	Dephospho-CoA kinase	38	97.44	88	64.23	0.04
tehB	S-adenosyl-L-methionine-dependent methyltransferase TehB	38	97.44	88	64.23	0.04
gspF	type IV pilin biogenesis protein	38	97.44	86	62.77	0.03
argR_3	Arginine regulator; Arginine repressor	38	97.44	78	56.93	<0.001
group_7920	Phosphate transport system permease protein PstC	38	97.44	42	30.66	<0.001
group_9162	Spermidine/putrescine-binding periplasmic protein precursor	38	97.44	14	10.22	<0.001
group_2848	GDSL-like Lipase/Acylhydrolase; hypothetical protein; Beta-lactamase	37	94.87	81	59.12	0.02
group_5908	GTP-sensing transcriptional pleiotropic repressor CodY	37	94.87	42	30.66	<0.001
pepN	Aminopeptidase N	36	92.31	79	57.66	0.04
nfrA2	FMN reductase [NAD(P)H]; nitroreductase A	36	92.31	57	41.61	<0.001
ftsW	Lipid II flippase FtsW; hypothetical protein	35	89.74	67	48.91	0.01
licT_2	Transcription antiterminator LicT	35	89.74	65	47.45	<0.001
group_5126	Undecaprenyl-diphosphatase BcrC; PAP2 superfamily protein	35	89.74	42	30.66	<0.001
group_4672	putative ABC transporter ATP-binding protein	35	89.74	35	25.55	<0.001
group_6971	Sucrose phosphorylase	35	89.74	10	7.3	<0.001
wfgD	Lipid A export ATP-binding/permease protein Msb	34	87.18	68	49.64	0.02
nrdH	Glutaredoxin-like protein NrdH	34	87.18	63	45.99	0.01
nrdE2	Ribonucleoside-diphosphate reductase subunit alpha	34	87.18	58	42.34	<0.001
group_5439	hypothetical protein	33	84.62	49	35.77	<0.001
group_3491	Penicillin-binding protein 2B	33	84.62	7	5.11	<0.001
group_7136	Bacterial lipoprotein; hypothetical protein	33	84.62	2	1.46	<0.001
pyrK	Dihydroorotate dehydrogenase B (NAD(+))	32	82.05	63	45.99	0.04

group_5380	ABC-2 family transporter protein	32	82.05	60	43.8	0.02
group_8697	Aspartate carbamoyltransferase	32	82.05	6	4.38	<0.001
group_8465	Putative dipeptidase	31	79.49	47	34.31	<0.001
group_6654	SPFH domain / Band 7 family protein; hypothetical protein	31	79.49	45	32.85	<0.001
arcB	Ornithine carbamoyltransferase%2C catabolic	31	79.49	44	32.12	<0.001
group_8437	hypothetical protein	31	79.49	36	26.28	<0.001
group_8114	hypothetical protein	30	76.92	48	35.04	0.01
efeU	hypothetical protein; Ferrous iron permease EfeU precursor	30	76.92	46	33.58	<0.001
adhB	Alcohol dehydrogenase 2	30	76.92	45	32.85	<0.001
yxdL_3	ABC transporter ATP-binding protein YxdL	30	76.92	44	32.12	<0.001
group_4195	CTP synthase	30	76.92	43	31.39	<0.001
group_4669	hypothetical protein	30	76.92	42	30.66	<0.001
group_8924	Membrane lipoprotein TmpC precursor	29	74.36	53	38.69	0.05
group_6802	Aspartokinase 3	29	74.36	51	37.23	0.03
bag_2	IgA FC receptor precursor; Autolysin;	29	74.36	49	35.77	0.02
fepD_1	Ferric enterobactin transport system permease protein FepD	29	74.36	47	34.31	0.01
tcyB	L-cystine transport system permease protein TcyB	29	74.36	47	34.31	0.01
mro	Aldose 1-epimerase precursor	29	74.36	47	34.31	0.01
group_5527	Cysteine desulfurase	29	74.36	43	31.39	<0.001
group_699	Limonene 1%2C2-monooxygenase	29	74.36	43	31.39	<0.001
rbsR	Ribose operon repressor	29	74.36	43	31.39	<0.001
group_8670	hypothetical protein	29	74.36	41	29.93	<0.001
group_2787	ComEC family competence protein; hypothetical protein	29	74.36	33	24.09	<0.001
group_5713	hypothetical protein	29	74.36	27	19.71	<0.001
glnM_2	putative glutamine ABC transporter permease protein GlnM	28	71.79	50	36.5	0.05
group_1397	PD-(D/E)XK nuclease family transposase	28	71.79	48	35.04	0.03
group_3025	Maltose/maltodextrin import ATP-binding protein MalK	28	71.79	40	29.2	<0.001
group_5529	Putative esterase; hypothetical protein	28	71.79	40	29.2	<0.001
group_4519	hypothetical protein	28	71.79	38	27.74	<0.001
group_7048	DNA alkylation repair enzyme; hypothetical protein	27	69.23	37	27.01	<0.001
group_4093	Manganese transport system membrane protein MntB	27	69.23	25	18.25	<0.001
group_3862	Multiple sugar-binding protein precursor; Alpha-galactosidase	27	69.23	17	12.41	<0.001
group_5009	hypothetical protein	26	66.67	42	30.66	0.03
group_3923	VIT family protein; Ribonucleoside-diphosphate reductase subunit beta	26	66.67	41	29.93	0.03
group_5753	Catechol-2%2C3-dioxygenase	26	66.67	41	29.93	0.03

csoR	Copper-sensing transcriptional repressor CsoR	26	66.67	40	29.2	0.02
licB_1	Lichenan-specific phosphotransferase enzyme IIB component	26	66.67	40	29.2	0.02
group_1364	hypothetical protein; Transposase	26	66.67	39	28.47	0.01
bglA_2	6-phospho-beta-glucosidase BglA	26	66.67	39	28.47	0.01
group_5728	hypothetical protein	26	66.67	38	27.74	0.01
group_4876	PTS system trehalose-specific EIIBC component	26	66.67	35	25.55	<0.001
group_4344	hypothetical protein	26	66.67	34	24.82	<0.001
group_4355	putative dual-specificity RNA methyltransferase RlmN	26	66.67	32	23.36	<0.001
srlR_2	Glucitol operon repressor; Lactose phosphotransferase system repressor	26	66.67	20	14.6	<0.001
group_8020	hypothetical protein	26	66.67	1	0.73	<0.001
group_487	hypothetical protein	25	64.1	38	27.74	0.03
group_2898	hypothetical protein	25	64.1	37	27.01	0.02
arcA	Arginine deiminase	25	64.1	34	24.82	0.01
group_10146	hypothetical protein	25	64.1	32	23.36	<0.001
pkn1	Serine/threonine-protein kinase pkn1	25	64.1	32	23.36	<0.001
group_9683	Lichenan-specific phosphotransferase enzyme IIB component	25	64.1	32	23.36	<0.001
group_1936	5-methylcytosine-specific restriction enzyme subunit McrC	25	64.1	26	18.98	<0.001
glnH_1	ABC transporter glutamine-binding protein GlnH precursor	25	64.1	24	17.52	<0.001
group_8140	Chaperone protein DnaK	25	64.1	17	12.41	<0.001
licA_2	Lichenan-specific phosphotransferase enzyme IIA component	25	64.1	7	5.11	<0.001
group_7120	Riboflavin biosynthesis protein RibBA	25	64.1	2	1.46	<0.001
group_9325	Sulfite exporter TauE/SafE	24	61.54	31	22.63	0.01
group_461	Streptococcal histidine triad protein	24	61.54	24	17.52	<0.001
group_4635	putative cysteine desulfurase	24	61.54	11	8.03	<0.001
group_7123	Riboflavin biosynthesis protein RibBA	24	61.54	0	0	<0.001
group_99	Nucleoside diphosphate kinase	23	58.97	34	24.82	0.04
group_647	hypothetical protein; Transposase	23	58.97	31	22.63	0.02
group_2890	spermidine/putrescine ABC transporter membrane protein	23	58.97	29	21.17	0.01
group_7462	ISXO2-like transposase domain protein; hypothetical protein	23	58.97	26	18.98	<0.001
group_5332	Putative esterase	23	58.97	23	16.79	<0.001
group_1732	Immunoglobulin A1 protease precursor	22	56.41	31	22.63	0.04
group_4508	hypothetical protein	22	56.41	30	21.9	0.03
group_7470	hypothetical protein	22	56.41	30	21.9	0.03
group_7699	Glycerol-3-phosphate acyltransferase	22	56.41	30	21.9	0.03
group_3848	hypothetical protein	22	56.41	26	18.98	0.01

group_3379	Histidine--tRNA ligase	22	56.41	22	16.06	<0.001
group_8908	Lichenan permease IIC component	22	56.41	4	2.92	<0.001
group_7705	PTS system fructose-specific EIIABC component	21	53.85	25	18.25	0.01
nudC	hypothetical protein; NADH pyrophosphatase; NUDIX domain protein	21	53.85	23	16.79	<0.001
group_3856	Pyrrolidone-carboxylate peptidase	21	53.85	23	16.79	<0.001
group_2625	hypothetical protein	21	53.85	21	15.33	<0.001
group_64	hypothetical protein	21	53.85	19	13.87	<0.001
group_4011	Putative DNA-invertase from lambdoid prophage Rac	21	53.85	13	9.49	<0.001
hlyB	Alpha-hemolysin translocation ATP-binding protein HlyB	20	51.28	27	19.71	0.05
group_2201	Helix-turn-helix domain protein	20	51.28	26	18.98	0.04
group_5965	hypothetical protein	20	51.28	26	18.98	0.04
group_87	hypothetical protein	20	51.28	24	17.52	0.02
group_8434	Thiosulfate sulfurtransferase GlpE	20	51.28	23	16.79	0.01
group_5603	Sucrose-6-phosphate hydrolase	20	51.28	22	16.06	0.01
group_3290	transcriptional regulator PhoU; Na+/Pi-cotransporter	20	51.28	18	13.14	<0.001
group_2060	hypothetical protein	20	51.28	8	5.84	<0.001
group_4491	putative multidrug resistance ABC transporter ATP-binding	20	51.28	8	5.84	<0.001
lsrC	Branched-chain amino acid transport system	19	48.72	21	15.33	0.01
group_7984	Macrolide export ATP-binding/permease protein MacB	19	48.72	21	15.33	0.01
group_5946	ABC transporter substrate binding protein	19	48.72	20	14.6	0.01
group_6779	hypothetical protein	19	48.72	20	14.6	0.01
cmpC	ABC transporter ATP-binding protein YxdL	19	48.72	19	13.87	0.01
group_5249	hypothetical protein	19	48.72	18	13.14	<0.001
group_5302	Glutamine synthetase	19	48.72	18	13.14	<0.001
group_196	hypothetical protein; Transposase DDE domain protein	19	48.72	11	8.03	<0.001
group_2580	Bacitracin export permease protein BceB	19	48.72	8	5.84	<0.001
group_1201	2'-N-acetylparomamine deacetylase	18	46.15	19	13.87	0.01
pezA	Antitoxin PezA	18	46.15	18	13.14	0.01
group_9296	Bacteriocin (Lactococcin_972)	18	46.15	18	13.14	0.01
group_6056	bifunctional RNase H/acid phosphatase	18	46.15	17	12.41	0.01
group_7420	Tyrosine recombinase XerC; hypothetical protein	18	46.15	11	8.03	<0.001
group_8448	Energy-coupling factor transporter transmembrane protein EcfT	18	46.15	9	6.57	<0.001
group_1643	Plasmin and fibronectin-binding protein A precursor	18	46.15	8	5.84	<0.001
pezT	Toxin PezT; hypothetical protein	17	43.59	20	14.6	0.05
group_6698	hypothetical protein	17	43.59	19	13.87	0.04

group_6499	Methionine import ATP-binding protein MetN 2	17	43.59	18	13.14	0.02
group_3521	Lactococcin A secretion protein LcnD	17	43.59	16	11.68	0.01
group_8946	Transcriptional repressor AdcR	17	43.59	11	8.03	<0.001
group_8353	Alpha-glycerophosphate oxidase	17	43.59	9	6.57	<0.001
group_3543	hypothetical protein	17	43.59	5	3.65	<0.001
group_7990	RNA polymerase-associated protein RapA	16	41.03	16	11.68	0.02
group_10087	Lactose-specific phosphotransferase enzyme IIA component	16	41.03	15	10.95	0.02
group_2132	hypothetical protein; Toxin PezT	16	41.03	12	8.76	<0.001
group_3675	Fe(3+)-citrate-binding protein YfmC precursor	16	41.03	11	8.03	<0.001
group_2229	Xaa-Pro dipeptidyl-peptidase	16	41.03	9	6.57	<0.001
group_616	hypothetical protein	16	41.03	9	6.57	<0.001
group_469	Streptococcal histidine triad protein	16	41.03	7	5.11	<0.001
group_5250	hypothetical protein	16	41.03	6	4.38	<0.001
group_1107	hypothetical protein; Transposase DDE domain protein	15	38.46	15	10.95	0.04
group_2955	PTS system galactitol-specific transporter subunit IIA	15	38.46	15	10.95	0.04
group_5125	Branched-chain amino acid transport system	15	38.46	15	10.95	0.04
group_5948	ABC transporter substrate binding protein	15	38.46	15	10.95	0.04
cpsD	Tyrosine-protein kinase CpsD	15	38.46	13	9.49	0.01
group_5817	GTP-binding protein TypA/BipA; hypothetical protein	15	38.46	13	9.49	0.01
group_5924	hypothetical protein	15	38.46	13	9.49	0.01
cap8A	Capsular polysaccharide type 8 biosynthesis protein cap8A	15	38.46	12	8.76	0.01
group_4252	hypothetical protein	15	38.46	10	7.3	<0.001
group_6061	Phosphoribosylformylglycinamidine synthase	15	38.46	10	7.3	<0.001
soxS	Regulatory protein SoxS	15	38.46	9	6.57	<0.001
group_8511	hypothetical protein	15	38.46	9	6.57	<0.001
group_2045	hypothetical protein	15	38.46	8	5.84	<0.001
group_4162	Glycerophosphoryl diester phosphodiesterase	15	38.46	6	4.38	<0.001
group_4069	Branched-chain-amino-acid aminotransferase	15	38.46	3	2.19	<0.001
group_4560	hypothetical protein	15	38.46	1	0.73	<0.001
group_805	EcoKI restriction-modification system protein HsdS	14	35.9	13	9.49	0.04
bbmA_2	Intracellular maltogenic amylase; Neopullulanase	14	35.9	12	8.76	0.02
group_4902	Serine protease Do-like HtrA	14	35.9	12	8.76	0.02
group_5235	Putative HMP/thiamine import ATP-binding protein YkoD	14	35.9	9	6.57	<0.001
group_8447	hypothetical protein	14	35.9	9	6.57	<0.001
group_7869	Peptidase family M50	14	35.9	4	2.92	<0.001

group_10390	Phosphatidylglycerophosphatase A	14	35.9	2	1.46	<0.001
nanM	N-acetylneuraminate epimerase precursor	14	35.9	2	1.46	<0.001
group_9405	hypothetical protein	14	35.9	2	1.46	<0.001
oppC_2	Oligopeptide transport system permease protein OppC	14	35.9	2	1.46	<0.001
group_2691	hypothetical protein; Internalin-J precursor	14	35.9	1	0.73	<0.001
bglA	hypothetical protein; 6-phospho-beta-glucosidase BglA	13	33.33	11	8.03	0.04
lacG_1	6-phospho-beta-galactosidase	13	33.33	11	8.03	0.04
group_1633	Fibronectin-binding repeat protein; hypothetical protein	13	33.33	10	7.3	0.02
group_4305	Endo-alpha-N-acetylgalactosaminidase precursor	13	33.33	10	7.3	0.02
lacE_1	PTS system lactose-specific EIICB component	13	33.33	8	5.84	0.01
group_5686	hypothetical protein	13	33.33	7	5.11	<0.001
rmlB	dTDP-glucose 4.0%2C6-dehydratase	13	33.33	7	5.11	<0.001
rmlA1	Glucose-1-phosphate thymidylyltransferase 1	13	33.33	6	4.38	<0.001
rfbC	putative dTDP-4-dehydrorhamnose 3.0%2C5-epimerase	13	33.33	6	4.38	<0.001
rmlD	dTDP-4-dehydrorhamnose reductase	13	33.33	5	3.65	<0.001
gsiB	Glutathione-binding protein GsiB precursor	13	33.33	2	1.46	<0.001
gsiC_2	Glutathione transport system permease protein GsiC	13	33.33	2	1.46	<0.001
group_2652	Transposase IS66 family protein	12	30.77	10	7.3	0.05
group_1122	hypothetical protein; Transposase DDE domain protein	12	30.77	7	5.11	0.01
irtA	Iron import ATP-binding/permease protein IrtA	12	30.77	6	4.38	<0.001
yesO_3	Putative ABC transporter substrate-binding protein YesO	12	30.77	4	2.92	<0.001
group_9767	L-arabinose transport system permease protein AraQ	12	30.77	4	2.92	<0.001
group_4046	hypothetical protein; LexA repressor	12	30.77	2	1.46	<0.001
group_490	hypothetical protein	12	30.77	2	1.46	<0.001
gsiA	Glutathione import ATP-binding protein GsiA	12	30.77	2	1.46	<0.001
cdd_3	Cytidine deaminase; cytidine deaminase	12	30.77	1	0.73	<0.001
manP	PTS system mannose-specific EIIBCA component	11	28.21	6	4.38	0.01
hrsA	Heat-responsive suppressor HrsA	11	28.21	6	4.38	0.01
ptsN	Nitrogen regulatory protein	11	28.21	6	4.38	0.01
licR_2	putative licABCH operon regulator	11	28.21	6	4.38	0.01
alsE	D-allulose-6-phosphate 3-epimerase	11	28.21	6	4.38	0.01
bbmA_1	Intracellular maltogenic amylase; Neopullulanase	11	28.21	5	3.65	<0.001
group_5225	hypothetical protein	11	28.21	5	3.65	<0.001
group_5063	Glucose-1-phosphate adenylyltransferase	11	28.21	4	2.92	<0.001
lacF_4	Lactose transport system permease protein LacF	11	28.21	4	2.92	<0.001

group_1037	hypothetical protein; Asparagine synthase	11	28.21	3	2.19	<0.001
group_2687	Internalin-J precursor; hypothetical protein	11	28.21	1	0.73	<0.001
group_168	hypothetical protein; Transposase DDE domain protein	10	25.64	6	4.38	0.03
group_2856	hypothetical protein	10	25.64	6	4.38	0.03
group_787	ApaLI-like restriction endonuclease	10	25.64	6	4.38	0.03
fruA_2	PTS system fructose-specific EIIBC component	10	25.64	6	4.38	0.03
ndvA	putative ABC transporter ATP-binding protein	10	25.64	5	3.65	0.02
group_5053	Aminodeoxychorismate synthase component 1	10	25.64	5	3.65	0.02
group_788	DNA adenine methyltransferase YhdJ	10	25.64	5	3.65	0.02
group_2228	Xaa-Pro dipeptidyl-peptidase	10	25.64	4	2.92	0.01
group_4957	hypothetical protein	10	25.64	4	2.92	0.01
group_4302	Endo-alpha-N-acetylgalactosaminidase precursor	10	25.64	2	1.46	<0.001
group_6203	Glycerol facilitator-aquaporin gla	10	25.64	2	1.46	<0.001
group_453	High-affinity zinc uptake system binding-protein ZnuA precursor	10	25.64	1	0.73	<0.001
group_4813	UDP-N-acetylmuramoylalanine--D-glutamate ligase	10	25.64	1	0.73	<0.001
group_174	Transposase DDE domain protein; hypothetical protein	9	23.08	5	3.65	0.05
group_2621	hypothetical protein	9	23.08	4	2.92	0.02
epsL	putative sugar transferase EpsL; hypothetical protein	9	23.08	4	2.92	0.02
group_5779	Sensor histidine kinase LiaS; Sensor protein VraS	9	23.08	4	2.92	0.02
group_2128	hypothetical protein	9	23.08	3	2.19	0.01
licC_5	Lichenan permease IIC component	9	23.08	3	2.19	0.01
group_6636	Phosphopantetheine adenylyltransferase	9	23.08	2	1.46	<0.001
group_6962	Oligoendopeptidase F.0%2C plasmid; hypothetical protein	9	23.08	2	1.46	<0.001
group_9407	enterobactin exporter EntS	9	23.08	1	0.73	<0.001
group_9772	enterobactin exporter EntS; hypothetical protein	9	23.08	1	0.73	<0.001
group_2710	Transposase IS66 family protein; hypothetical protein	8	20.51	3	2.19	0.03
group_8024	hypothetical protein	8	20.51	3	2.19	0.03
group_3676	Fe(3+)-citrate-binding protein YfmC precursor	8	20.51	2	1.46	0.01
mcrB_2	hypothetical protein	8	20.51	1	0.73	<0.001
hldD	hypothetical protein	8	20.51	1	0.73	<0.001
group_7513	hypothetical protein	8	20.51	1	0.73	<0.001
group_7088	UDP-glucose 4-epimerase	8	20.51	0	0	<0.001
group_7128	Riboflavin biosynthesis protein RibBA	8	20.51	0	0	<0.001
group_7344	Single-stranded DNA-binding protein ssb	8	20.51	0	0	<0.001
group_9486	hypothetical protein	8	20.51	0	0	<0.001

group_10362	hypothetical protein	7	17.95	2	1.46	0.03
group_10364	hypothetical protein	7	17.95	2	1.46	0.03
group_10380	hypothetical protein	7	17.95	2	1.46	0.03
group_7081	hypothetical protein	7	17.95	2	1.46	0.03
group_9384	hypothetical protein; Transposase	7	17.95	2	1.46	0.03
group_5725	hypothetical protein	7	17.95	1	0.73	0.01
group_6639	hypothetical protein	7	17.95	1	0.73	0.01
group_7421	Tyrosine recombinase XerC	7	17.95	1	0.73	0.01
group_8471	hypothetical protein	7	17.95	1	0.73	0.01
group_3492	Penicillin-binding protein 2B	7	17.95	0	0	<0.001
group_10388	hypothetical protein	6	15.38	1	0.73	0.03
group_1578	Major Facilitator Superfamily protein	6	15.38	1	0.73	0.03
group_3273	Inner membrane protein YgaZ	6	15.38	1	0.73	0.03
xerC	hypothetical protein	6	15.38	1	0.73	0.03
group_5191	Type-1 restriction enzyme R protein; hypothetical protein	6	15.38	1	0.73	0.03
group_5241	carbamoyl phosphate synthase-like protein; hypothetical protein	6	15.38	1	0.73	0.03
nagC_2	N-acetylglucosamine repressor; hypothetical protein	6	15.38	1	0.73	0.03
group_8073	hypothetical protein	6	15.38	1	0.73	0.03
group_8249	Glucosamine-6-phosphate deaminase; hypothetical protein	6	15.38	1	0.73	0.03
glxR	NAD binding domain of 6-phosphogluconate dehydrogenase	6	15.38	1	0.73	0.03
iphP	Tyrosine-protein phosphatase precursor	6	15.38	1	0.73	0.03
group_9768	hypothetical protein	6	15.38	1	0.73	0.03
group_9770	hypothetical protein	6	15.38	1	0.73	0.03
group_2993	hypothetical protein	6	15.38	0	0	0.01
group_4047	hypothetical protein	6	15.38	0	0	0.01
group_5187	hypothetical protein	6	15.38	0	0	0.01
group_8475	hypothetical protein	6	15.38	0	0	0.01
group_3970	hypothetical protein; N-acetylneuraminate lyase	5	12.82	0	0	0.02
group_5239	hypothetical protein	5	12.82	0	0	0.02
group_5415	Helix-turn-helix; hypothetical protein	5	12.82	0	0	0.02
group_6205	Glycerol facilitator-aquaporin gla	5	12.82	0	0	0.02
group_8516	hypothetical protein	5	12.82	0	0	0.02
group_8562	hypothetical protein	5	12.82	0	0	0.02
rmlD_1	dTDP-4-dehydrorhamnose reductase	5	12.82	0	0	0.02
group_9955	hypothetical protein	5	12.82	0	0	0.02

Appendix G. Plagiarism report

NATIONAL INSTITUTE FOR COMMUNICABLE DISEASES
Division of the National Health Laboratory Service

Centre for Respiratory Diseases and Meningitis
1 Modderfontein Road, Sandringham, 2131
Tel: +27 (0)11 555 0352 Fax: +27 (0)11 555 0437
Email: nicolew@nicd.ac.za

University of the Witwatersrand
Faculty of Health Sciences, Postgraduate Office
Parktown, 2193
Johannesburg, South Africa

To whom it may concern

RE: Turnitin report for Thabo Mohale (1006116) MSc dissertation

Mr Thabo Mohale is a part-time student doing his MSc Med at University of the Witwatersrand. He published a manuscript titled "Genomic analysis of nontypeable pneumococci causing invasive pneumococcal disease in South Africa, 2003-2013. *BMC Genomics* 2016, 17:470. doi: 10.1186/s12864-016-2808-x.:470-2808". The manuscript is included as a chapter in his dissertation. The overall high similarity index score (27%) in the Turnitin plagiarism report (see below) was due to the published manuscript.

Kind regards,

Dr Nicole Wolter
Lecturer, Clinical Microbiology and Infectious Diseases
Faculty of Health Sciences, University of the Witwatersrand
and
Senior Medical Scientist, Centre for Respiratory Diseases and Meningitis, National Institute for Communicable Disease

1006116:thabomohale.docx

ORIGINALITY REPORT

27%	24%	21%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	bmcgenomics.biomedcentral.com Internet Source	15%
2	Mohale, Thabo, Nicole Wolter, Mushal Allam, Kedibone Ndlangisa, Penny Crowther-Gibson, Mignon du Plessis, and Anne von Gottberg. "Genomic analysis of nontypeable pneumococci causing invasive pneumococcal disease in South Africa, 2003–2013", BMC Genomics, 2016. Publication	1%
3	www.ingim.co.in Internet Source	1%
4	usir.salford.ac.uk Internet Source	1%
5	Submitted to London School of Hygiene and Tropical Medicine Student Paper	1%
6	www.slideshare.net Internet Source	1%

7	Jessica L. Bradshaw, Haley R. Pipkins, Lance E. Keller, James K. Pendarvis, Larry S. McDaniel. " Mucosal Infections and Invasive Potential of Nonencapsulated Are Enhanced by Oligopeptide Binding Proteins AliC and AliD ", mBio, 2018 Publication	<1%
8	bmcmicrobiol.biomedcentral.com Internet Source	<1%
9	core.ac.uk Internet Source	<1%
10	www.biomedcentral.com Internet Source	<1%
11	Submitted to University of Witwatersrand Student Paper	<1%
12	boris.unibe.ch Internet Source	<1%
13	wwwnc.cdc.gov Internet Source	<1%
14	www.rns-pdf.londonstockexchange.com Internet Source	<1%
15	Submitted to University of Bristol Student Paper	<1%
16	wiredspace.wits.ac.za Internet Source	<1%
17	Submitted to University of Salford Student Paper	<1%
18	www.austincc.edu Internet Source	<1%
19	geneprimp.jgi-psf.org Internet Source	<1%
20	du Plessis, Mignon, Mushal Allam, Stefano Tempia, Nicole Wolter, Linda de Gouveia, Claire von Mollendorf, Keith A. Jolley, Nontombi Mbelle, Jeannette Wadula, Jennifer E. Cornick, Dean B. Everett, Lesley McGee, Robert F. Breiman, Rebecca A. Gladstone, Stephen D. Bentley, Keith P. Klugman, and Anne von Gottberg. "Phylogenetic Analysis of Invasive Serotype 1 Pneumococcus in South Africa, 1989 to 2013", Journal of Clinical Microbiology, 2016. Publication	<1%
21	Stephen D. Bentley. "Genetic Analysis of the Capsular Biosynthetic Locus from All 90 Pneumococcal Serotypes", PLoS Genetics, 2006 Publication	<1%
22	www.premiummedical.travel Internet Source	<1%

23	download.mapsforge.org Internet Source	<1%
24	aac.asm.org Internet Source	<1%
25	cid.oxfordjournals.org Internet Source	<1%
26	Submitted to University of Arizona Student Paper	<1%
27	eurosurveillance.org Internet Source	<1%
28	Scott, J. R., J. Hinds, K. A. Gould, E. V. Millar, R. Reid, M. Santosham, K. L. O'Brien, and W. P. Hanage. "Nontypeable Pneumococcal Isolates Among Navajo and White Mountain Apache Communities: Are These Really a Cause of Invasive Disease?", The Journal of Infectious Diseases, 2012. Publication	<1%
29	open.uct.ac.za Internet Source	<1%
30	Submitted to University of St Andrews Student Paper	<1%
31	Submitted to University of Central Oklahoma Student Paper	<1%
32	herkules.oulu.fi Internet Source	<1%
33	www.circumpolarhealthjournal.net Internet Source	<1%
34	www2.kenes.com Internet Source	<1%
35	www.cdc.gov Internet Source	<1%
36	"Investigators at University of London Imperial College Detail Findings in Pneumococcal Disease (Effe", Biotech Week, Jan 20 2016 Issue Publication	<1%
37	Submitted to King's College Student Paper	<1%
38	www.ncbi.nlm.nih.gov Internet Source	<1%
39	dare.ubvu.vu.nl Internet Source	<1%
40	Agarwal, Vaibhav. "Role of PspC interaction with human polymeric immunoglobulin receptor and Factor H in Streptococcus pneumoniae infections and host cell induced signalling", Universität Würzburg, 2009. Publication	<1%

41	216.226.177.68 Internet Source	<1%
42	Wippel, Carolin. "Alterations of brain dendrite and synapse structure by the Streptococcus pneumoniae neurotoxin pneumolysin - Insights and pharmacological modulation", Universität Würzburg, 2012. Publication	<1%
43	pneumocystis.uc.edu Internet Source	<1%
44	Zhao, Jiangqin Ragupathy, Viswanath Liu, Jikun Wang, Xue Vemula, Sai Vikram Mubarak, Haja Sittana El. "Nanomicroarray and multiplex next-generation sequencing for simultaneous identification and characte", Emerging Infectious Diseases, March 2015 Issue Publication	<1%
45	Crowther-Gibson, P., C. Cohen, K. P. Klugman, L. de Gouveia, and A. von Gottberg. "Risk factors for multidrug-resistant invasive pneumococcal disease in South Africa, 2003-2008: the pre-vaccine era in a high HIV prevalence setting", Antimicrobial Agents and Chemotherapy, 2012. Publication	<1%
	www.hiv.lanl.gov	
46	Internet Source	<1%
47	barcdb.se Internet Source	<1%
48	Lim, Efrem S. Zhou, Yanjiao Zhao, Guoyan. "Early life dynamics of the human gut virome and bacterial microbiome in infants. (RESOURCE)(Report)", Nature Medicine, Oct 2015 Issue Publication	<1%
49	in.gov Internet Source	<1%
50	S. Melchiorre. "Point mutations in wchA are responsible for non-typeability of two invasive Streptococcus pneumoniae isolates", Microbiology, 10/27/2011 Publication	<1%
51	Submitted to University of Liverpool Student Paper	<1%
52	Submitted to University of Limpopo Student Paper	<1%
53	www.wits.ac.za Internet Source	<1%
54	Ndlangisa, Kedibone M., Mignon du Plessis, Mushal Allam, Nicole Wolter, Thabo Mohale,	<1%

	Linda de Gouveia, Monica Birkhead, Keith P. Klugman, and Anne von Gottberg. "Two cases of serotypeable and non-serotypeable variants of Streptococcus pneumoniae detected simultaneously during invasive disease", BMC Microbiology, 2016. Publication	
55	Submitted to University of Birmingham Student Paper	<1%
56	Submitted to University College London Student Paper	<1%
57	von Gottberg, Anne, Linda de Gouveia, Stefano Tempia, Vanessa Quan, Susan Meiring, Claire von Mollendorf, Shabir A. Madhi, Elizabeth R. Zell, Jennifer R. Verani, Katherine L. O'Brien, Cynthia G. Whitney, Keith P. Klugman, and Cheryl Cohen. "Effects of Vaccination on Invasive Pneumococcal Disease in South Africa", New England Journal of Medicine, 2014. Publication	<1%
58	jmm.sgmjournals.org Internet Source	<1%
59	www.pharmiweb.com Internet Source	<1%
60	www.aafp.org Internet Source	

		<1%
61	eb5oy.rr.nu Internet Source	<1%
62	adt.waikato.ac.nz Internet Source	<1%
63	medlibrary.org Internet Source	<1%
64	Mastro, Timothy D. Ghafoor, Abdul Nomani. "Antimicrobial resistance of pneumococci in children with acute lower respiratory tract infection in ", The Lancet, Jan 19 1991 Issue Publication	<1%
65	www.pnas.org Internet Source	<1%
66	www.scielo.org.mx Internet Source	<1%
67	www.hivma.org Internet Source	<1%
68	www.tufts.edu Internet Source	<1%
69	www.thl.fi Internet Source	<1%
70	espace.curtin.edu.au	

	Internet Source	<1%
71	bmcinfectdis.biomedcentral.com Internet Source	<1%
72	www.optitex.com Internet Source	<1%
73	www.acsu.buffalo.edu Internet Source	<1%
74	www.isoforlab.com Internet Source	<1%
75	von Mollendorf, Claire Cohen, Cheryl Tem. "Epidemiology of serotype 1 invasive pneumococcal disease, South Africa, 2003-2013.(RESEARCH)(Report)", Emerging Infectious Diseases, Feb 2016 Issue Publication	<1%
76	dspace.ewha.ac.kr Internet Source	<1%
77	Submitted to University of Aberdeen Student Paper	<1%
78	otitismediasociety.org Internet Source	<1%
79	tampub.uta.fi Internet Source	<1%
80	Submitted to 7996 Student Paper	<1%
81	minerva-access.unimelb.edu.au Internet Source	<1%
82	www.nejm.org Internet Source	<1%
83	www.tjbiochip.com Internet Source	<1%
84	jb.asm.org Internet Source	<1%
85	emgm.eu Internet Source	<1%
86	www.nesi.be Internet Source	<1%
87	repub.eur.nl Internet Source	<1%
88	www.cps.ca Internet Source	<1%
89	www.recentmedicalfindings.com Internet Source	<1%
90	www.vics.org Internet Source	<1%
91	theses.ucalgary.ca Internet Source	

		<1%
92	www.nicd.ac.za Internet Source	<1%
93	Kattan, Randa Rayyan, Amal Abu Zheiman, . "Serotype distribution and drug resistance in Streptococcus pneumoniae, palestinian territories.(DISP", Emerging Infectious Diseases, Jan 2011 Issue Publication	<1%
94	MANSOUR, Hashem Mohammed, MOBAYED, Shaban Al, MANASRA, Ahmed El and ABUDALFA, Rola S.. "Comparison of Cefuroxim versus Cefotaxim in the Treatment of Pneumonia in Children Aged 3 Months- 5 Years an Imaging and Clinical Examination Evidence: Randomized Controlled Trial", Medicine Science Publisher, 2014. Publication	<1%
95	Schähle, Yvonne. "Genomische Diversität und Evolution von Virulenzdeterminanten in Streptococcus spp.", KLUEDO, 2011. Publication	<1%
96	"New Findings Reported from Guangxi Medical University Describe Advances in Cholesterol (Association ", Health & Medicine Week, Feb 19 2016 Issue	<1%
	Publication	
97	Whitney, Cynthia G. Pilishvili, Tamor Fa. "Effectiveness of seven-valent pneumococcal conjugate vaccine against invasive pneumococcal disease: ", The Lancet, Oct 28 2006 Issue Publication	<1%
98	Zimmerman, Richard Kent Middleton, Donal. "Routine vaccines across the life span, 2007.", Journal of Family Practice, Feb 2007 Issue Publication	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		